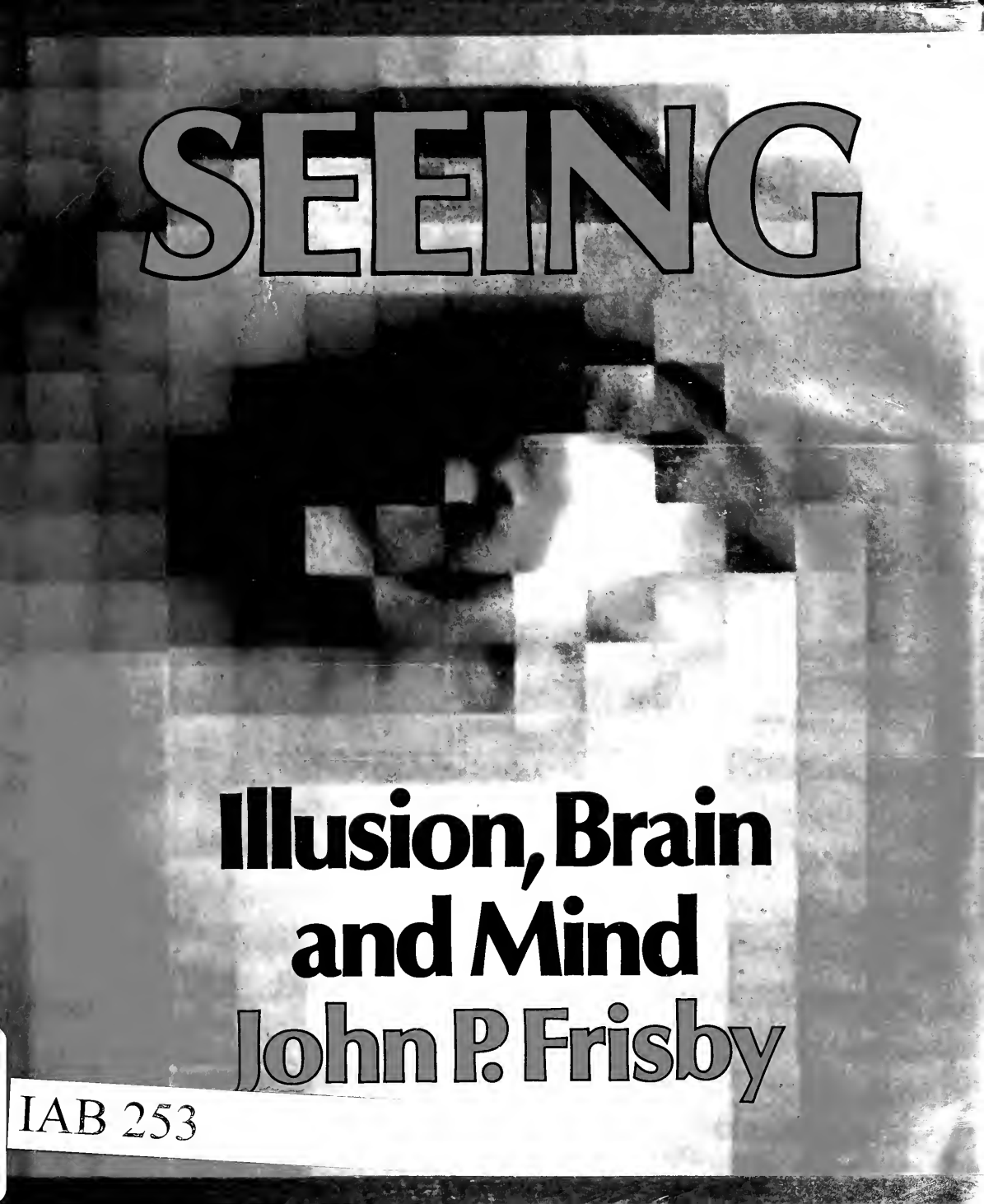


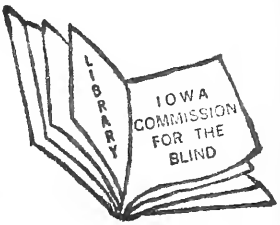
What happens inside our heads when we see? Many people believe that our brain contains a sort of cinema screen, on to which our eyes project images of the outside world, and that it is these images of which we are conscious. But John Frisby demonstrates that seeing is really a very complex process which even specialists are barely beginning to understand, and not remotely like projection on to a screen. With the aid of more than 200 illustrations, most of them in colour, he conveys in clear, non-technical language the excitement and importance of work now in progress on human vision. In particular he explains how psychologists, physiologists and computer scientists are reinforcing each other's conclusions from their three very different standpoints.

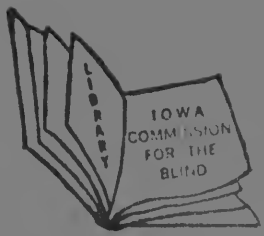
Visual illusions are a recurrent theme of the book, and John Frisby shows their special value in unravelling the mechanisms of seeing: by catching the brain making a mistake, we come to understand how it functions when it gets things right. Chief among the illusions is a large selection of 'random-dot stereograms', never presented to the general reader before, which spring dramatically into three dimensions when viewed through the special red/green spectacles provided with the book.

Seeing not only provides a perfect introduction to the psychology of perception, but raises wider questions about the way we interpret our environment, leaving us with a profound respect for a crucial faculty which we tend to take for granted.

$16.95

FRISBY, John P.
Seeing





SEEING


modigliani

SEEING

Illusion, Brain and Mind

John P. Frisby

Oxford New York Toronto Melbourne
OXFORD UNIVERSITY PRESS
1980

Oxford University Press, Walton Street, Oxford OX2 6DP
OXFORD LONDON GLASGOW
NEW YORK TORONTO MELBOURNE WELLINGTON
KUALA LUMPUR SINGAPORE JAKARTA HONG KONG TOKYO
DELHI BOMBAY CALCUTTA MADRAS KARACHI
NAIROBI DAR ES SALAAM CAPE TOWN

First published 1979

British Library Cataloguing in Publication Data
Frisby, John
Seeing.
1. Visual perception
I. Title
152.1'4 BF241 78-40515
ISBN 0-19-217672-2

Printed in Spain by TONSA S.A., San Sebastian

CONTENTS

PREFACE

Novelists sometimes say that their characters take over even the best laid plot and determine an ending different from the one originally envisaged. Something of the sort happened in the writing of this book, even though it is not a work of fiction, or at least is not meant to be. I started out to write The General Reader's Fun Book of Visual Illusions, with extended captions to give some explanation of why the illusions were as they were. But it proved impossible to say much of interest about them until the scene had been set with some introductory chapters on the fundamental nature of seeing and the basic neural machinery of the visual system. When I had done this, the natural format for the book turned out to be not just a series of illusory figures, but instead a series of chapters dealing with selected topics from the standpoint of how the visual system performs certain jobs. To be sure, pictures of illusions abound – they provide valuable clues about visual mechanisms – but now they complement and illustrate a 'story', rather than performing the central guiding role as originally planned. This change makes the book suitable not only for a general readership, but also for an introductory course on vision, especially one in which the teacher wishes to emphasise a combined psychological, physiological and computational approach. But no attempt has been made to write a conventional textbook. The coverage is anything but encyclopedic, and a course tutor will find many holes that need plugging. But that seems to me no bad thing: why else have a course tutor? More important in my view is to have a text which communicates some fundamental ideas in visual science, as it is being practised today, so laying a foundation upon which the teacher can build.

But despite this change from my expectations and plans, I hope the book is still 'fun'. It is certainly still intended for the general reader, in that no special knowledge is demanded and all terms are defined as the book proceeds. Even so, understanding visual mechanisms is intrinsically tricky, and although I have made every attempt to reduce each problem area to its simplest essentials, the general reader will sometimes need a fairly large dose of motivation to grapple with all the book has to say. But whenever the text becomes a bit too intricate for him to cope with, he is encouraged to flick over the pages and treat the book as it was originally meant to be written – as a pleasure book of visual illusions. There are plenty of them and newcomers to the area will enjoy peering at the illustrations, often in disbelief that their visual systems can possibly get it all so wrong. Visual illusions are, quite simply, fascinating.

Acknowledgements

I owe a special debt to David Marr, whose writings on seeing are for me the work of genius, and whose influence can be felt throughout this book. I am particularly grateful to John Mayhew for our frequent discussions about vision, which have helped illuminate for me so many different aspects of the subject. He read most of the manuscript in draft, made useful comments, and provided all the computer-drawn figures. Kevin Connolly, my Head of Department, provided me with excellent facilities, good advice and encouragement throughout. Without his willingness to arrange my teaching and administrative commitments in a way which still left some time over for peace and quiet, the book would not be written yet. I am very appreciative of these things. Henry Hardy – of the Roxby Press when the project began, now of the Oxford University Press – my editor, and the person who invited me to write the book in the first place, has provided patient and careful guidance at all stages, always helping towards greater clarity of presentation and also offering much useful advice from his own considerable knowledge of perception. Hugh Elwes of the Roxby Press is to be thanked, now the book is finally written, for bullying me into continuing with it when my energies were flagging and I was near to quitting. David Warner and his team of artists have made a first-rate job of designing the book and contributed much inventiveness and skill in transforming my figure sketches into proper illustrations. Numerous other people have contributed in many other ways, some perhaps not realising just how helpful they have been, even if indirectly, and I would particularly like to mention Bela Julesz, Richard Gregory, Chris Brown and Paul Dean. Finally, my family deserves and gets my warm thanks for having put up so gamely with so much while the book was being written. They have amply earned the book's dedication.

John P. Frisby

Sheffield
May 1978

Figure Acknowledgements

1–3,4b Amedeo Modigliani's *Nude* by courtesy of Courtald Institute Galleries, University of London. **15** After Whistler, R. (1947) *OHO*, Bodley Head. **16** Reproduced with permission from *Brain Sciences Information Project* (1973), Ferranti Ltd. **17a** After Tinbergen, N. (1951) *The Study of Instinct*, Clarendon Press. **17b** After Rubin, E. (1915) *Synsoplevede Figurer*, Glydendalske. **17c** Reproduced with permission from Boring, E.G. (1930) 'A new ambiguous figure.' *American Journal of Psychology*, Vol. 42, pp.444–5. **17d** After Fisher, G.H. (1966) 'Materials for experimental studies of ambiguous and embedded figures.' *Research Bulletin of the Department of Psychology, University of Newcastle Upon Tyne*, No.4. **19** Photograph by R.C. James taken from Thurston, J. and Carraher, R.G. (1966) *Optical illusions and the visual arts*. Litton Educational Publishing Inc. Reproduced by permission of Van Nostrand Reinhold Company. **20** Photograph by R.J.C. Blewitt supplied by Ardea London. **21** Photograph supplied by Frank Spooner Agency, London. **22, 25** By courtesy of Escher Foundation, Haags Gemeentemuseum, The Hague. **23, 26** After Penrose, L.S., and Penrose, R. (1958) 'Impossible objects; a special type of illusion.' *British Journal of Psychology*, Vol. 49, pp.31–3. **24, 28** Model based on a design by Gregory, R.L. (1971) *The Intelligent Eye*. Weidenfeld and Nicolson. **29** Reproduced with permission from Ames, A., Jr. (1952) 'Perceptions not disclosures.' In Kilpatrick, F.P. (1952) *Human Behaviour from the Transactional Point of view*. Hanover, N.H. **38, 42** Photomicrograph supplied by Institute of Ophthalmology, London. **47–8** After Hilgard, E.R., Atkinson, R.C., and Atkinson, R.L. (1975) *Introduction to Psychology*, 6th Edition. Harcourt Brace Jovanovich Inc. **49** Histology by S.P.A. Kariyawasam, photography by K. Fitzpatrick. **53–4, 66** Reproduced by permission from LeVay, S., Hubel, D.H., and Wiesel, T.N. (1975) 'The pattern of ocular dominance columns in macaque visual cortex revealed by reduced silver stain.' *Journal of Comparative Neurology*, Vol. 159, pp.559–76. **62, 76–8** After Blakemore, C. (1973) 'The baffled brain.' In Gregory, R.L., and Gombrich, E.H. (1973) *Illusion in Nature and Art*. Duckworth. **63** After Hubel, D.H. and Wiesel, T.N. (1974) 'Sequence regularity and geometry of orientation columns in monkey striate cortex.' *Journal of Comparative Neurology*, Vol. 158, pp.267–93. **64, 65** Reproduced with permission from Hubel, D.H., Wiesel, T.N., and Stryker, M.P. (1977) 'Orientation columns in macaque monkey visual cortex demonstrated by the 2-deoxyglucose autoradiographic technique.' *Nature*, Vol. 269, pp.328–30. **67** After LeVay, Hubel and Wiesel (1975) – see **53–4**. **75** Photograph by Joe Bangay. **79** After Brodman, K. (1914) 'Physiologie des Gehirns.' In *Die Allgemeine Chirurgie der Gehirnkrankheiten, Neue Deutsche Chirurgie*, Vol. 11. Enke. **88–9** After Blakemore, C. and Cooper, G.F. (1970) 'Development of the brain depends on the visual environment.' *Nature*, Vol. 228, pp.477–8. **92–3** After Blakemore, C. and Campbell, F.W. (1969) 'On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images.' *Journal of Physiology*, Vol. 203, pp.237–60. **97** After Mollon, J.D. (1974) 'After-effects and the brain.' *New Scientist*, 21 February, pp.479–82. **100** Reproduced with permission from Blakemore, C. (1973) – see **62**. **103a** After Mackay, D.M. (1973) 'Lateral interaction between neural channels sensitive to texture density.' *Nature*, Vol. 245, pp.159–61. **103b,c** After Klein, S., Stromeyer III, C.F., and Ganz, L. (1974) 'The simultaneous spatial frequency shift: a dissociation between the detection and perception of gratings.' *Vision Research*, Vol. 14, pp.1421–32. **104** After Robinson, J.O. (1972) *The Psychology of Visual Illusion*. Hutchinson. **105b** By courtesy of Gerald Scarfe. **105c** By courtesy of the Mansell Collection. **105d** Cartoon by Strube in *Daily Express*, 8 June 1940. **108** Reproduced with permission from Ullman, J.R., and Rosenfeld, A. (1977) 'Picture recognition and analysis.' *The Radio and Electronic Engineer*, Vol. 47, January/February, pp.33–48. **110, 115, 126** After Marr, D. (1976) 'Early processing of visual information.' *Philosophical Transactions of the Royal Society of London*, Series B, Vol. 275, pp.483–524. **111** By courtesy of Tate Gallery, London. **112** Private collection, England. By courtesy of the artist. **119–20** Reproduced with permission from Julesz, B., Gilbert, E.N., and Shepp, L.A. (1973) 'Inability of humans to discriminate between visual textures that agree in second-order statistics – revisited.' *Perception*, Vol. 2, pp.391–405. **123** By courtesy of Imperial War Museum, London. **130** After Harmon, L.D. (1973) 'The recognition of faces.' *Scientific American*, Vol. 229, November, pp.71–82. **132** Original photographs (shown on p.122) for a and e, b and f, c and d supplied by Popperfoto (London), Mansell Collection, and John Kobal Collection respectively. **134a, b, c** By courtesy of Haags Gemeentemuseum, The Hague. Arrangement after Locher, J.L. (1971) *The World of M.C. Escher*. Harry N. Abrams, Inc. **135** By courtesy of Metropolitan Museum of Art, New York. **137** Reproduced with permission from Gross, C.G. (1973) 'Inferotemporal cortex and vision.' *Progress in Physiological Psychology*, Vol. 5, pp.77–123. Academic Press. **141** By courtesy of the Percival David Foundation of Chinese Art. **151** After Greef, Z. (1900) *Graefe-Saemisch Hb. ges. augenheilk.*, II, Vol. 1, Kap. 5. **152–3** After Pirenne, M.H. (1967) *Vision and the Eye*, 2nd Edition. Associated Book Publishers. **154** After Cornsweet, T.N. (1970) *Visual Perception*. Academic Press. **159** From Tyndall (1867) *Sound: a Course of Eight Lectures Delivered at the Royal Institution of Great Britain*. Longmans, Green. **163–4** After Frisby, J.P., and Clatworthy, J.L. (1975) 'Illusory contours: curious cases of simultaneous brightness contrast?' *Perception*, Vol. 4, pp.349–57. **167** By courtesy of Victoria and Albert Museum, London. **171** From Wheatstone, C. (1838) 'Contributions to the physiology of version. On some remarkable, and hitherto unobserved, phenomena of binocular vision, I.' *Philosophical Transactions*, Part I, pp.371–94.
174 Stereoscope by courtesy of Henry Hardy. **177** Reproduced with permission from Mayhew, J.E.W., and Frisby, J.P. (1976) 'Rivalrous texture stereograms.' *Nature*, Vol. 264, pp.53–6. **186–7** Reproduced with permission from Julesz, B. (1971) *Foundations of Cyclopean Perception*. University of Chicago Press. **188** Reproduced with permission from Frisby, J.P., and Clatworthy, J.L. (1975) 'Learning to see complex random-dot stereograms.' *Perception*, Vol. 4, pp.173–8. **198** After Blakemore, C. (1969) 'Binocular depth discrimination and the nasotemporal division.' *Journal of Physiology*, Vol. 205, pp.471–97. **199a** After Julesz, B. (1971) – see **186**. **204–10** After Julesz, B., and Miller, J. (1975) 'Independent spatial frequency tuned channels in binocular vision and rivalry.' *Perception*, Vol. 4, pp.125–43. **211–12** After Mayhew, J.E.W., and Frisby, J.P. (1976) – see **177**. **213** After Frisby, J.P., and Mayhew, J.E.W. (1978) 'The relationship between apparent depth and disparity in rivalrous-texture stereograms.' *Perception*, Vol. 7, pp.661–78. **216** After Mayhew, J.E.W., and Frisby, J.P. (1979) 'Perceived lightness not predicted by the co-planar ratio theory.' *Science* (in press).

Artwork and illustration acknowledgements

Terry Allan Designs Ltd: 2, 31–37, 39–41, 43–45, 47, 48, 50–52, 57, 58, 68, 69, 71, 72, 76, 78, 81, 82, 86, 90, 91, 97, 142–149, 151–156, 160, 161, 195, 197, 221. Augustine Studios: 46, 75, 139, 182, 183, 185. David Bryant: 1, 3, 12, 30, 56, 59–61, 87–89, 138, 169, 178. Eugene Fleury: 55, 59–63, 70, 73, 74, 95, 96. Clive Spong: 7–9, 22, 26, 27, 98, 102b, 104, 114, 116, 125, 162, 165, 170, 175, 176, 179, 203, 215, 219, 220. Brian Wright/Augustine Studios: 4.
Special photography by Bryce Attwell: 5, 6, 24, 28, 109, 166, 168.

1 PICTURES IN OUR HEADS

What goes on inside our heads when we see? Most people take seeing so much for granted that few will ever have considered this question seriously. But if pressed to speculate, the ordinary person who is not an expert on the subject usually says that there must be an 'inner screen' of some sort in our heads, perhaps like a cinema screen, except that it is made out of brain tissue. The eyes transmit an image of the outside world on to this screen, and this is the image of which we are conscious. Elaborating these suggestions about visual mechanisms somewhat, we arrive at the following theory of seeing, illustrated in **1**.

> Each eye works like a camera. Both camera and eye have a lens, and where the camera has light-sensitive film, the eye has a light-sensitive *retina*, a network of tiny receptive units that form the back surface of the eyeball (Latin *rete*: 'net'). The lens's job is to focus an image of the outside world on to the retina – the *retinal image* – which stimulates the retina so that it sends messages about the image along *optic nerve fibres* to the *brain*. The brain is composed of millions of tiny components called *cells*. Certain cells specialise in vision and are arranged in the form of a sheet – the 'inner screen'. Each cell in the screen can at any moment be either active or inactive. If a cell is very active, it is signalling the presence of a bright spot at that particular point on the 'inner screen' – and hence at the associated point in the outside world. Equally, if a cell is only moderately active, it is signalling an intermediate shade of grey. Completely inactive cells signal black spots. Cells in the 'inner screen' as a whole take on a pattern of activity whose overall shape directly mirrors the shape of the retinal image received by the eye. For example, if a painting is being observed, as in **1**, then the pattern of activity on the 'inner screen' directly resembles the painting. As soon as this pattern is set up on the screen of cells, the observer has the experience of seeing the painting.

This 'inner screen' theory of seeing is easy to understand and is intuitively very appealing. After all, our visual experiences do in some sense seem to 'match' the outside world: so it is natural to suppose that there are mechanisms for vision in the brain which provide the simplest possible type of match – a physically similar or 'photographic' one. Indeed, the 'inner screen' theory of seeing can be likened to television, an image-transmission system which is also photographic in this sense. The eyes are equivalent to TV cameras, and the image finally appearing on a TV screen connected to the cameras is roughly equivalent to the proposed image on the 'inner screen' of which we are conscious. The only important difference is that whereas the TV-screen image is composed of more or less brightly glowing dots, our visual image is composed of more or less actively working brain cells.

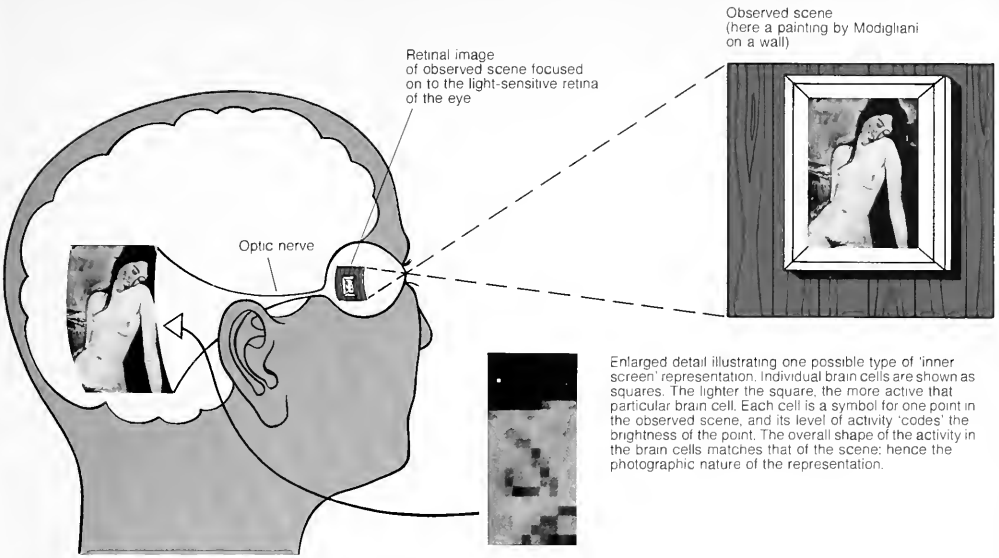
Symbols and Scene Descriptions

The first thing to be said about the 'inner screen' theory of seeing is that it proposes a *symbolic description* as the basis of seeing. In this respect it is like almost all other theories of seeing, but to describe it in this way requires some explanation.

A *symbol* is anything which stands for something other than itself. Words are symbols. For example, the word 'chair' stands for a particular kind of sitting support – the word is not the support itself. Many other kinds of symbols exist, of course, apart from words. A red traffic light stands for the command 'Stop!', the Stars and Stripes stand for the United States of America, and so on. A moment's reflection shows that there must be symbols inside our heads for the things we see, symbols which themselves are unlike the things they represent. Upon opening up a patient's head for a brain operation, the surgeon does not find there a miniature stage-set of the world! All he finds is a pink blancmange-like mass of brain cells. So it is an inescapable conclusion that there must be a symbolic description in the brain of the outside world, a description cast in symbols which stand for the various aspects of the world of which sight makes us aware. In fact, when we began by asking 'What goes on inside our heads when we see?' we could as well have put this question as 'When we see, what are the symbols inside our heads that stand for things in the outside world?' This may seem an odd rephrasing of the question at first, but it must be remembered that neither objects nor properties of objects (such as lightness) exist inside our brains as such. Only brain cells are to be found there, and it is these cells which must be serving a symbolic descriptive function.

For example, the 'inner screen' theory proposes that certain brain cells serve as symbols for *points* in the outside scene which is being viewed. The brain cells are not the selfsame thing as these outside points, but they do the job of representing them inside our heads. These brain cells are symbols standing for parts of the outside world.

The idea of visual experience as a symbolic process may seem a strange one. The likely reason for this is that the world



1 The 'inner screen' theory of seeing. The physical shape of the pattern of brain activity matches that of the observed scene

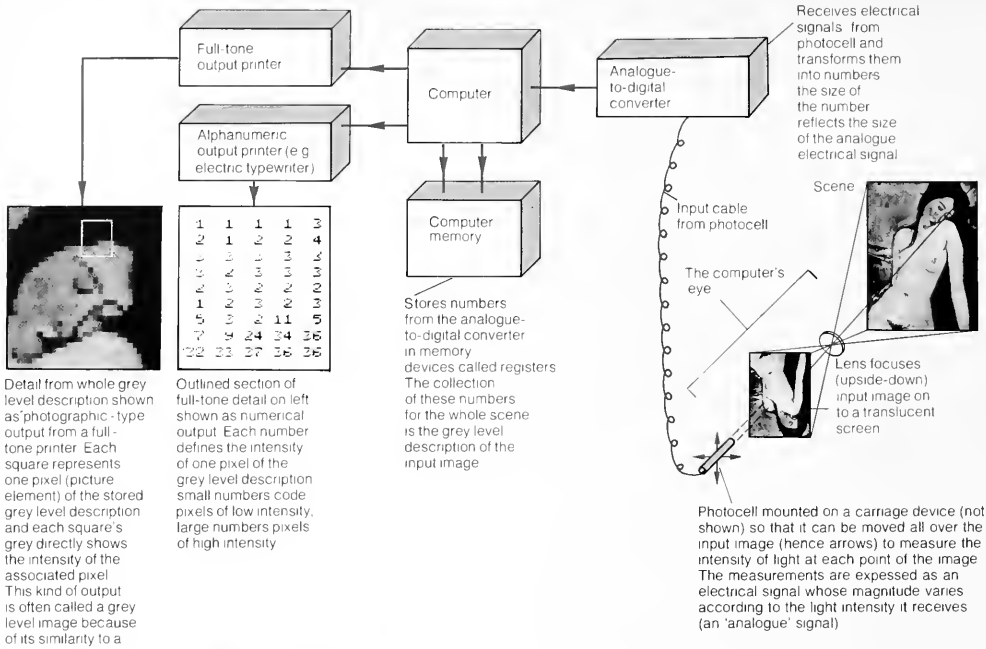
we see – the visual world – is so very clearly 'out there' that it can come as something of a shock to realise that somehow the whole of this world is tucked away in our skulls as an inner representation which stands for the real outside world. It is difficult and unnatural to disentangle the 'perception of a scene' from the 'scene itself', but they must be clearly distinguished if seeing is to be understood. When the difference between a perception and the thing perceived is fully grasped, the conclusion that seeing must involve a symbolic description sitting somewhere inside our heads becomes easier to accept. Moreover, the problem of seeing becomes easier to state clearly. This problem is: what is the nature of the brain's symbolic description of the visual world, and how is it obtained? It is this problem which provides the subject of this book.

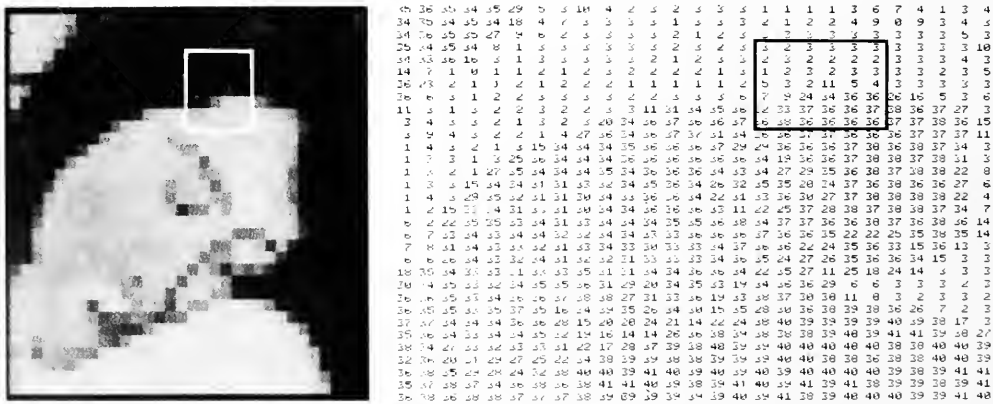
Perception, Consciousness and Brain Cells

Another reason why it might feel strange to regard visual experience as a symbolic process is that the obvious candidates for the symbols, brain components of one kind or another, may seem quite insufficient for the task. The 'inner screen' theory posits a direct relationship between conscious visual experience on the one hand and activity in certain brain cells on the other. That is, activity in certain cells is somehow accompanied by conscious experience. Proposing this kind of parallelism between brain-cell activity and visual experience is characteristic of many theories of perceptual brain mechanisms, as we will see in due course. But is there more to it than this? Can the richness of visual experience really be identified with activity in a few million, or even a few hundred million, brain cells? Are brain cells the right kind of entities to provide, somehow, conscious perceptual experience?

In the face of this kind of question, the 'ordinary person' who was pressed into speculating about perceptual mechanisms, and who came up with the 'inner screen' theory in reply, might wish to withdraw to what seems safer ground at this point, and say that activity in brain cells cannot provide a completely satisfactory account of perceptual experience. If he did so he would be joining in a long tradition of philosophical thought which claims that consciousness cannot be wholly and completely identified with the activity of matter, even brain matter, and that it must be carried by a different 'substance' of some kind. This viewpoint, whose origins lie deep in antiquity, but which is usually associated with the great French philosopher René Descartes (1596–1650), is called *dualism* because it proposes two sorts of substance in the universe, mind and matter. These substances are quite distinct, it is claimed, although somehow they can relate to one another.

If dualism is true, then it is obviously wrong to equate activity in the brain cells of 1 simply and straightforwardly with the conscious experience of the scene observed. The dualist might hold that activity in such cells is necessary for the conscious experience, a precursor if you like, but that consciousness itself must be carried by a substance quite different from brain matter, even though this substance must somehow co-exist with brain cells. Descartes recognised the need for mind and matter to 'speak' to one another and proposed that the site of interaction was a small structure lying deep in the brain, called the pineal gland. He chose this structure for its important role because it is singular, as he thought consciousness to be: most brain structures are replicated in the brain's two halves, so that just as we have left and right legs, so too we have left and right cerebral hemispheres and so on (see





2 [top] A simple image-processing system for obtaining and storing a grey level description

[bottom] Large sections of the grey level description with inset boxes giving the location of the details shown in the top figure

p.39). No one now takes this pineal gland theory seriously, but dualism itself survives as a widely held doctrine espoused by many philosophers, scientists and laymen alike.

Curiously, although perceptual experience is the stock-in-trade of every psychologist who studies perception (except perhaps the behaviourists, who would insist that they do not study conscious experience at all, only perceptual behaviour), psychologists have little to say about the conscious aspect of it. Consciousness remains a great mystery, despite considerable advances in our knowledge of perceptual mechanisms, and little can be said about it sensibly at present in terms of scientific theories. None the less, it is probably true that most psychologists and neurophysiologists, with some notable exceptions, reject dualism and either explicitly or implicitly adopt some version or other of the 'identity theory' as a working hypothesis. That is, they believe that a given conscious experience, such as the perception of a bright point in the painting of **1**, is quite simply an attribute of activity in one or other brain cell, and there is no need to invoke a special and independent substance to explain mind.

But even if it is correct to assert an identity of some kind between experience and brain-cell activity, this view itself leaves many questions unanswered. Is consciousness associated with *all* brain cell activity, or just some? The latter seems most likely, but then the follow-up question arises: why is experience related only to some brain cells and not others? What is special about the ones identified with experience? We have no answer to such questions. Indeed, the psychologist and the neurophysiologist have no sensible way of studying them at present, and so simply put them to one side for the time being while they get on with the business of tackling problems which look soluble.

Machines for Seeing

One such problem, albeit an enormously difficult one, is building a machine which can recognise objects – in the sense that, given an image of a scene as an input, it can print out on a typewriter a statement about what objects the scene contains, and in what relationships: in short, a machine which can do the job of deriving a scene description from an input image. Whether one should call such a device a 'perceiver', a 'seeing machine', or more humbly an 'image processor' or 'pattern recogniser', is a moot point which probably hinges upon whether the user of these terms believes that consciousness can ever be associated with non-biological brains. In any event, most computer scientists who work on the problem of devising automatic image-processing machines would call the activity appearing on the 'inner screen' of **1** a *grey level description* of the painting. This is because the 'inner screen' is signalling the various shades of grey all over the picture (I ignore colour for present purposes, and also many intricacies in the perception of grey: see chapter 6). Each individual brain cell in the screen is describing the grey level at one particular point of the picture in terms of an *activity code*. The code is a simple one: the more active the cell, the brighter the point in the painting. We consider later the details of this type of code, and what is meant by a cell being 'active'.

One way in which a grey level description can be obtained in a computer image-processing system is shown in **2**. The details of how the system operates are given in the figure's labels.

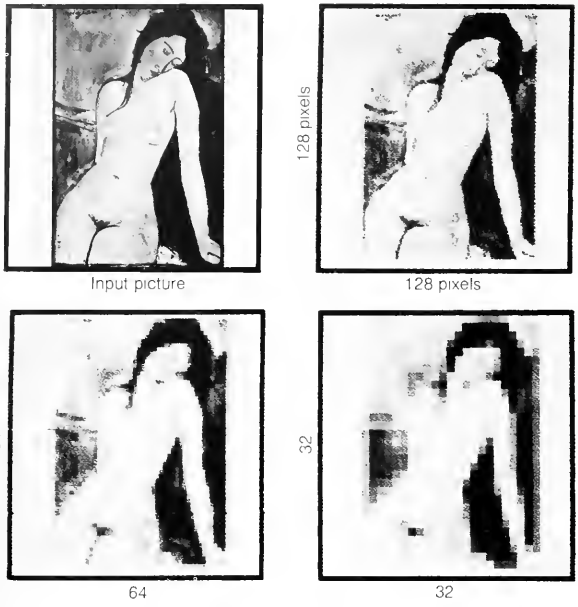
To call, as in **2**, a collection of numbers in a computer's memory registers a 'grey level description' may seem curious, but this is exactly what the numbers are providing. The term 'grey level' arises from the black-and-white nature of the system, with black being regarded as a very dark grey (and recorded with a small number) and white as a very light grey (and recorded with a large number). And the numbers are a description in the sense defined earlier, in that they make *explicit* the grey levels in the input image. That is, they make these grey levels usable by subsequent stages of image processing, as will become apparent in later chapters.

The input image in **2** is upside-down on the translucent screen, because of the laws of optics, but this creates no difficulties in practice, subsequent compensation being easily arranged. It is also the case that the retinal image in the eye is upside-down, and many people are worried by this. 'Why doesn't the world appear upside down?' they ask. The answer is quite simple: as long as there is a constant correlation between the outside scene and the retinal image, then the processes of image interpretation can rely on this correlation and build up the required scene description accordingly. Upside-down is simply interpreted as right-way-up, and that's all there is to it. If an observer is equipped with special spectacles which optically invert the retinal image so that it becomes itself right-way-up, then the world appears upside-down until the observer learns to cope with the new correlation between image and scene, an adjustment process which takes weeks, but is possible. Exactly what the nature of the adjustment process is – does the upside-down world really begin to 'look' right-way-up again, or is it simply that the observer learns new patterns of adjusted movement to cope with the strange world he finds himself in? – is an active research topic at the present time.

The number of *pixels* in a computer's grey level description will vary according to the capabilities of the computer (e.g. the size of its memory) and the needs of the user. For example, the use of a dense array of pixels will require a large memory store and produce a grey level description that picks up very fine detail, so that when output as a full-tone printout, the resulting grey level image will be discriminable from the input image only with difficulty, if at all. On the other hand, it may be necessary to use large pixels, each of which will represent quite a large area of the input image, in which case a full-tone printout produced to the same scale as before takes on a block-like appearance. These possibilities are illustrated in **3**, where the same input image is represented by three different grey level images, with pixel arrays of 128×128, 64×64 and 32×32. Note in this connection that an ordinary domestic TV set produces an image in pixel form with an array size of 625×625. The individual pixels are so tiny that they cannot be readily distinguished (unless a large TV screen is observed from quite close to). We will have more to say later about the interesting effects of reproducing pictures in block-like form with large pixels (see page 118).

Representations and Descriptions

It is easy to see why the computer's grey level description illustrated in **2** is similar in principle to the hypothetical 'inner screen' shown in **1**. In the latter, brain cells adopt different levels of activity to represent (or code) different pixel brightnesses. In the former, the computer holds different numbers in its memory to do exactly the same job. So both systems provide a grey level description of their input image, even though the detailed nature of the *representation of the description* is different in the two cases. This distinction



3 Grey level images of varying pixel size from a computer image-processor

between the status of an image description (the job it performs) and the way the description is actually represented (its physical embodiment in man or machine) is an important one which deserves further elaboration.

Consider, for example, the physical layout of the hypothetical 'inner screen' of brain cells. This is an anatomically neat one, with the various pixel cells arranged in a format which physically matches the arrangement of the corresponding points in the image. In complete contrast to this, the computer registers which perform the same job as the brain cells would not be arranged in the computer in a way which physically matches the input image. That is, the 'anatomical' locations of the registers in the computer memory would not necessarily be arranged in a grid-like form, as the brain cells are. Instead, the registers might be arranged in a variety of different ways, depending on many different factors, most of them trivial ones to do with how the memory was manufactured. None the less, the computer can keep track of each pixel measurement in a very precise manner by using a system of labels for each of its registers, to show which part of the image each refers to. The details of how this is done need not concern us here: it is sufficient to say that the labels ensure that each pixel value can be retrieved for later processing as and when required. Consequently, it is true to say that the brain cells of **1** and the computer memory registers of **2** are serving the same descriptive function – recording the grey level of each pixel – even though the nature of the representation in each case differs radically. It differs both in the nature of the pixel code (level of cell activity versus size of stored number) and in the anatomical arrangement of the units that represent the pixels.

Describing Objects

The 'inner screen' of **1**, then, can be described as a kind of symbolic scene description. The activity of the cells which compose the 'screen' describes in a symbolic form the corresponding points in the scene being viewed. But when we put it this way it is easier to see why this is so inadequate as a theory of seeing: it gives us such an impoverished scene description. The scene description which exists inside our heads is *not* confined simply to the brightnesses of individual points in the scene before us, but tells us an enormous amount more than this. Leaving aside the obvious limitation of not having anything to say about colour vision, the 'inner screen' description does not help us understand how we can know what *objects* we are looking at, or how we are able to describe their various features – shape, texture, movement, size – or their spatial relationships one to another. Such abilities are basic to seeing – they are presumably what we have a visual system for – and yet the 'inner screen' theory leaves them out altogether.

You might feel tempted to reply at this point: 'I don't really understand the need to propose anything more than a photograph-in-the-head on an "inner screen" in order to explain seeing. Surely, once this kind of symbolic description has been built up, isn't that enough? Are not all the other things you mention – recognising objects and so forth – an "immediately given" consequence of having the photographic type of representation provided by the "inner screen"?'

The answer to this question is to point out that the visual system is so good at telling us what is in the world around us that we are understandably misled into taking its effortless scene descriptions for granted, so much so that these do indeed seem to be something 'immediately given' by a photographic representation. But the truth is the exact opposite! Arriving at a scene description as good as that provided by the visual system is an immensely complicated process requiring a great deal of further interpretation of the meagre information contained in a grey level image, as will become clear as we proceed through the book. Achieving a grey level description is only the first and most trivial task.

This point is so important it is worth reiterating. The intuitive appeal of the 'inner screen' theory lies in its proposal that the visual system builds up a photographic-type brain picture of the observed scene, and its suggestion that this brain picture is the basis of our conscious visual experience. The trouble with the theory is that although brain cells are offered as symbols for points in the scene, everything else in the scene is left unanalysed. It simply is not much good having the visual system build a photographic-type copy of the scene if when the task is done the system is no nearer to using the information which is present in the retinal image to decide what is present in the outside scene. The theory fails to explain how we can recognise the various objects and properties of objects represented by the 'brain pictures'. And the ability to achieve such recognition is anything but an immediate consequence of having a photographic representation. Providing a more or less accurate replica of reality (at least of two-dimensional scenes) is easy – witness television. But it is exactly because a television set cannot do more than this that we would not regard it as a 'seeing machine'. Of course, it does produce a (limited) replica of the world as an image on its screen. But because it cannot decide what is 'in' this image, and therefore what is in the scene before the camera,

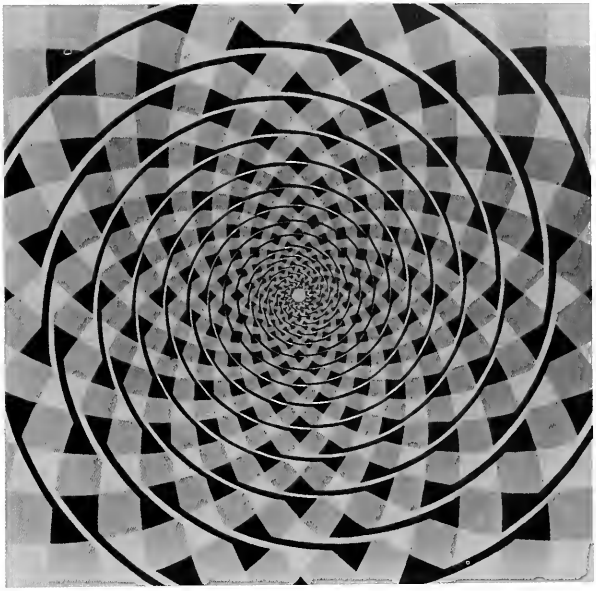
we would not call it a 'visual perceiver'. Devising a seeing machine which can receive a light image of a scene and use it to describe what is in the scene is much more complicated, a problem which is as yet largely unsolved. But whatever the true theory of seeing is, it must include processes quite different from the simple mirroring of the outside world by brain pictures. Mere physical resemblance is not an adequate basis for the brain's known powers of symbolic scene description.

Note in this connection that the computer's grey level image in **2** is both static and colourless: the computer's memory holds a frozen 'snapshot' cast simply in blacks, whites and greys. These limitations are imposed deliberately, because the added complications of considering a changing input picture, or a coloured one, are too great for today's machine-vision systems to handle. At least, the complications are too great if the grey level images are to be analysed to discover what objects they contain and in what relationships. It is easy enough to have a computer system made sufficiently sophisticated to 'observe' a changing scene and update its grey level image accordingly (so that it behaves like a television system). But the processes of image analysis required to discover what lies 'in' the grey level image at any one time are so complicated, poorly understood, and time-consuming on present-day computers that the analysis could never keep up with the changing view. This brings us back to the principal weakness of the 'inner screen' theory, namely that a description of a scene in terms of the brightnesses at every point is an extremely limited form of description.

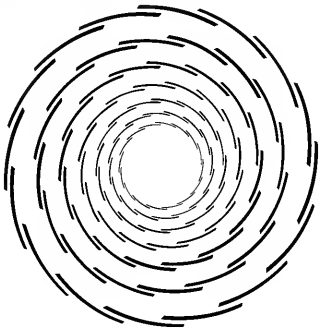
Visual Illusions and Seeing

The idea that visual experience is somehow akin to photography is so widespread and so deeply rooted that many readers will probably not be convinced by the above arguments against the 'inner screen' theory. They know that the eye does indeed operate as a kind of camera, in that it focuses an image of the world upon its light-sensitive retina. One way to break down confidence in continuing with this analogy past the eye and into the visual processes of the brain is to draw attention to the fact that what we see often differs dramatically from what is actually before our eyes. In short, the non-photographic quality of visual experience is borne out by the large number and variety of visual illusions. Many of these are illustrated in this book and almost all of them offer valuable clues about the existence of perceptual mechanisms devoted to building up an explicit scene description. These mechanisms operate well enough in most circumstances, but occasionally they are misled by an unusual stimulus, or one which falls outside their 'design specification', and a visual illusion results.

Look, for example, at **4**, which shows an illusion called *Fraser's spiral*. The amazing truth is that there is no spiral there at all! Convince yourself of this by tracing the path of the apparent spiral with your finger. You will find that you return to your starting point. At least, you will if you are careful: the illusion is so powerful that it can even induce incorrect finger-tracing! But careful tracing shows that the picture is really made up of concentric circles. The spiral exists only in your head. Somehow the input picture fools the visual system, which mistakenly provides a scene description incorporating a spiral – even though no spiral is present. A process which takes concentric circles as input and produces a spiral as output can hardly be thought of as 'photographic'.
It is worth while digressing at this point to consider a ques-



4 Fraser's spiral The fact is that there is no spiral here at all. Try to trace it with your finger. In fact Fraser's figure is based on concentric circles composed of segments angled towards the centre (see below).



5 The teacup illusion Imagine the spoon was stood upright in the cup Which mark on the spoon handle would then be level with the cup's rim? When you have decided, turn the page and look at 6

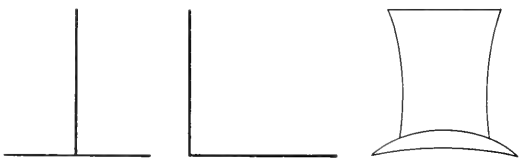
tion which may have occurred to you: are dramatic illusions such as Fraser's spiral genuinely representative of our everyday perceptions, or are they just unusual trick figures dreamt up by psychologists or artists? These illusions may surprise and delight us but are they really helpful in telling us what normally goes on inside our heads when we see the world?

Psychologists interested in perception would answer this question with a definite 'yes'. On the one hand, they regard illusions as important clues in trying to understand perceptual processes, both when they produce reasonably accurate perceptions, and when they are fooled into generating illusions. We will see how this strategy works out as we proceed through this book. And on the other hand, illusions abound in ordinary scenes, but most often go unnoticed by the casual observer. The teacup illusion shown in **5** is a good case in point. The photograph is of a perfectly normal teacup, together with saucer and spoon. Judge which mark on the spoon would be level with the rim of the teacup if the spoon was stood upright in the cup. Now turn the page and look at **6**! The illusory difference in the apparent lengths of the two spoons, one lying horizontally in the saucer and one standing vertically in the cup, is remarkable. Convince yourself that this perceptual effect is not a trick dependent on some subtle photography by investigating it in a real-life setting with a real teacup and spoon. It works just as well there as in the photograph. The conclusion from effects like this is that illusions are far more commonplace than is often realised. Artists and craftsmen know this fact well, and learn in their apprenticeships, often the hard way by trial and error, that the eye is by no means always to be trusted. Seeing is *not* always believing – or shouldn't be.

7 The vertical-horizontal illusion The vertical line appears much longer than the horizontal one.

8 The vertical line still appears longer than the horizontal one, but less so than in 7, showing that a part of the vertical-horizontal illusion is due to the bisection of the horizontal line by the vertical one.

9 The top hat illusion, a clear-cut case of the vertical-horizontal illusion. The hat is as tall as the brim is wide, but appears much taller.



The teacup illusion demonstrates that we tend to overestimate vertical extents at the expense of horizontal ones, particularly if the vertical element bisects the horizontal one. **7** illustrates the simplest version of this effect, and is commonly known as the vertical-horizontal illusion. The effect is weaker if the vertical line does not bisect the horizontal [**8**] but is still present. It is easy to draw many realistic pictures containing the basic effect. The top hat of **9** is a good example: the brim is as wide as the hat is tall! The perceptual mechanisms responsible for the vertical-horizontal illusion are not well understood, though various theories have been proposed since its first published report in 1851 by A. Fick.

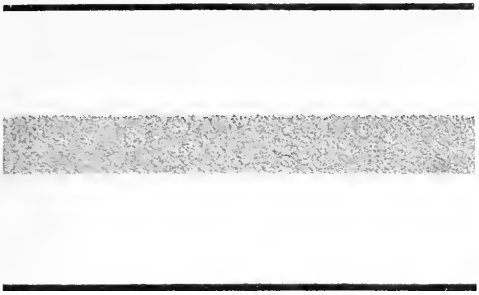
The illusions just considered are instances of *spatial* distortions: concentric circles can become a spiral, vertical extents can be stretched, horizontal ones shortened, and so on. A different type of illusion, a *brightness* illusion, is shown in **10**. The grey strip shown on its own has the same grey level along its entire length. This is true both if the strip is measured objectively with a physical instrument (such as a photocell) or if the strip is measured subjectively by our visual system (i.e. it appears of even brightness). Now look at the strip flanked by greys of varying intensity. Unbelievably, this grey strip is physically identical to the one shown on its own! Measured with a photocell, *both* strips would give identical light-intensity measurements at all points along their length. And yet our visual system's output is very different in the two cases. When the line is on its own, it gets the answer 'right' – we see a line of even brightness. When the line is flanked by the other greys, the system produces a huge perceptual error, with the perceived brightness at any point being strongly influenced by the intensity of the flanking greys. If these greys are dark, then the central strip appears relatively bright; if they are light, then the strip appears relatively dim. A further example of this *brightness contrast effect* is shown in **11**. The effect is eloquent testimony to the fact that perceptions cannot be thought of as simple 'photographic copies' of the world, even when it comes to a visual experience as apparently simple as that of brightness.

Scene Descriptions Must be Explicit

Explanations of some of these illusions will be offered in due course. For the present, we will return to the theme of seeing as a process of symbolic scene description, and attempt to articulate in a little more detail what this means.

The essential property of a scene description is that it makes some property of the scene *explicit*. In the 'inner screen' theory of **1**, the various brain cells make explicit the various shades of grey at all points in the scene. To say that the cells 'make explicit' the various greys is to say that they signal the intensity of these greys in a way which is sufficiently clear for subsequent processes to be able actually to *use* them for some purpose or other.

A scene description, then, is the result of processing an



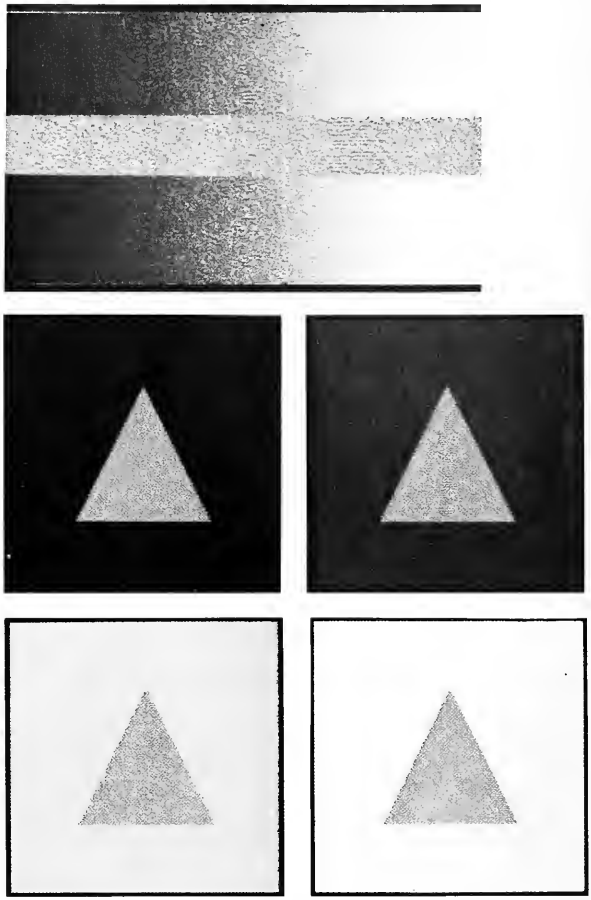
10 A brightness contrast illusion

image of the scene in order to make an attribute of the scene explicit. A simple example which illustrates this kind of system in action is shown in **12**, perhaps the most primitive 'seeing system' conceivable – a burglar alarm operated by a photocell. The corridor is permanently illuminated, and when the intruder's shadow falls over the photocell detector hidden in the floor an alarm bell is set ringing. Viewed in our terms, what the photocell-triggered alarm system is doing is:

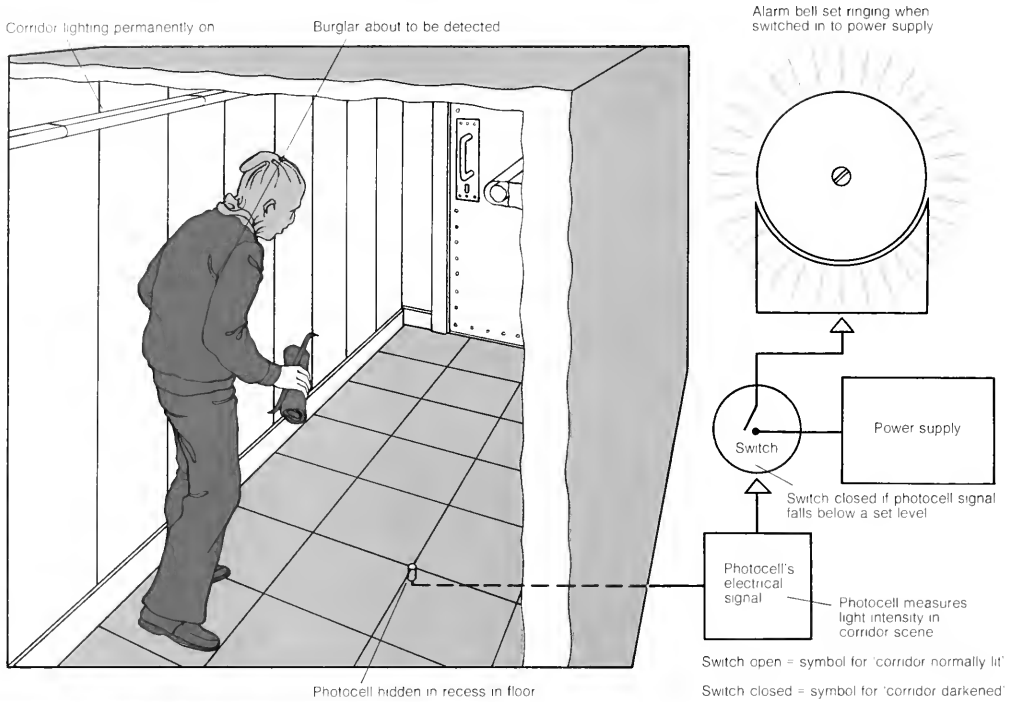
1] Inspecting a scene – a section of the corridor;
2] Collecting light from all over the scene – the equivalent in this primitive visual system of forming an image in a complex one;
3] Measuring the intensity of the light collected – the job of the photocell;
4] Using the measurement to build up a symbolic scene description of the corridor illumination – 'switch open' symbolises 'corridor normally lit' and 'switch closed' symbolises 'corridor darkened';
5] Using the symbolic scene description as a basis for action – ringing the alarm bell or leaving it quiet.

The effortless fluency with which our visual system delivers its explicit scene description of normal scenes is so beguiling that the sceptical reader might still doubt that this process is what seeing is all about. To help persuade him, it can be helpful to show certain 'trick' figures which catch the visual system out in some way, and reveal something of the scene description process at work. Consider, for example, the picture shown in **13**. A large bird-creature (called a Roc in the caption) is holding a girl in its beak. Now turn the book upside-down for a moment and look at the picture again. What a transformation! The bird-picture turns into a fish-picture! The bird's head becomes the fish, the bird's body becomes an island, the girl becomes a man sitting in a boat which was once the bird's beak, and so forth. The point is that only one image or 'stimulus' is physically present: but it induces the visual system to provide radically different scene descriptions depending on which way up it is presented.

The full cartoon strip from which the picture is taken is shown in **14**. It is one of a series of two-way-up cartoon strips drawn by Gustave Verbeek and published in the *Sunday New York Herald* in the 1900s. Although restricted to the normal 6-panel comic-strip format of his time, Verbeek was not content with this and decided to have 12 panels with no increase in space. The first 6 panels he used to set up some desperate plight facing his heroine, Little Lady Lovekins. The next 6



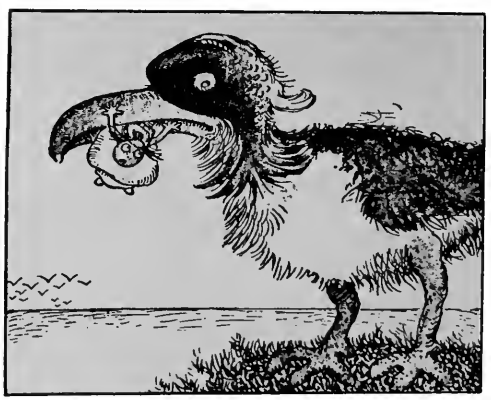
11 Another brightness contrast illusion The small inset grey triangles all have the same physical intensity, but their apparent brightnesses are very different. Triangles on a dark ground appear brighter than triangles on a light ground.



12 A simple burglar alarm system operated by a photocell

6 [left] The teacup illusion again: the vertical spoon seems much longer than the horizontal one

13 Drawing from a cartoon strip by Gustave Verbeek



14 A complete upside-down cartoon strip drawn by Gustave Verbeek
Panel 5 is 13 upside-down.

panels, which were the first 6 viewed upside-down, showed how his hero, Old Man Muffaroo, came to the rescue. Having each panel serve twice meant that each had to make sense whichever way up one looked at it. Verbeek thus made his characters interchangeable, and Little Lady Lovekins usually becomes Old Man Muffaroo on inversion. The skill and ingenuity shown by Verbeek in doing this was remarkable, particularly when it is remembered that he had to keep the story-line going between panels.

Verbeek's upside-down comic strips demonstrate the visual system at work building up scene descriptions which best fit the available evidence. Inversion subtly changes the nature of the evidence in the retinal image about what is present in the scene, and the visual system reports accordingly. Notice too that, for most observers at least, the two alternative readings of the picture actually *look* different. It is not just that we attach different verbal labels to the picture upon inversion – 'fish' versus 'bird', 'man' versus 'woman', 'leg' versus 'tree trunks'. Rather, we actually *see* different pictures in the two cases. The pattern of ink on the page stays the same, but the experience it gives rise to in us as we look at it is made radically different simply by turning the picture upside-down. (At least, this happens in the best of Verbeek's pictures, such as **13**; some others in **14** are not quite so good in this respect, but this does not alter the basic point.) Of course, certain small features remain similar upon inversion; but the overall scene description changes radically and qualitatively, and produces correspondingly different perceptions.

The 'inner screen' theory has a hard time trying to account for the different perceptions produced by inverting **13**. The 'inner screen' theorist wishes to reserve for his screen the job of representing the contents of visual experience. Fundamentally different experiences emerge upon inversion; therefore fundamentally different contents must be recorded on the screen in each case. But it is not at all clear how this could be done. The 'inner screen' way of thinking would predict that inversion should simply have produced a perception of the same picture, but upside-down. This is not what happens.

Other examples of the way inversion of a picture can show the visual system producing radically different scene descriptions of the same image are given in **15** and **16**. The former is a portrait which changes extraordinarily when turned upside-down. And the latter shows a scene with steep cliffs and craters alongside one with gently rounded hills: it is difficult to

15 A portrait adapted from a drawing by Rex Whistler

16 The two pictures are identical, as can be proved by inverting the book. The brain assumes that light comes from above, and interprets the shadows accordingly to build up radically different scene descriptions in each case

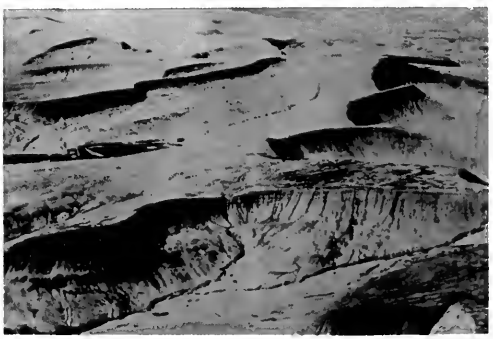
believe that they are one and the same picture, but turning the book upside-down proves the point.

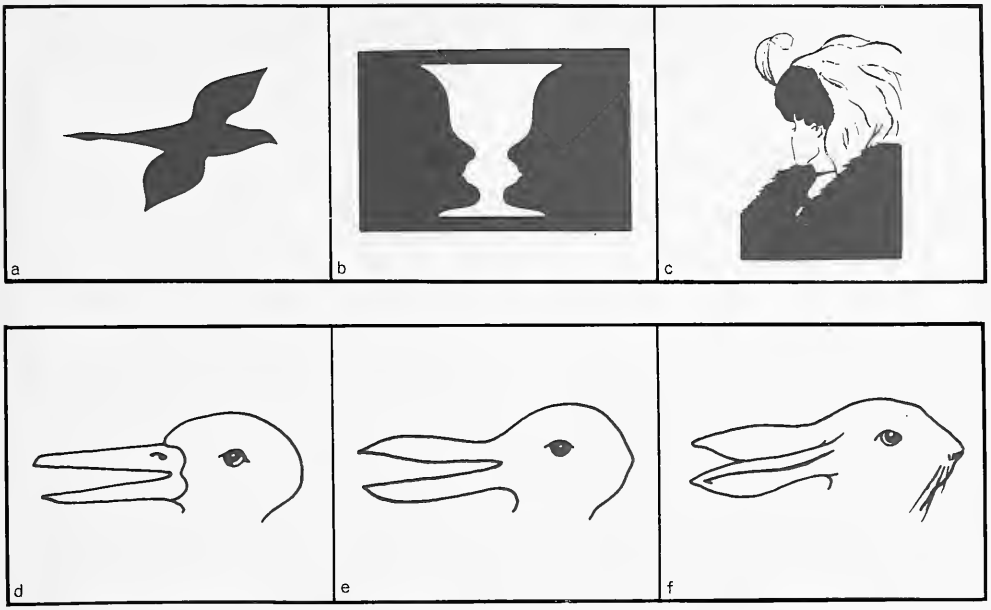
Another trick for displaying the scene-description abilities of the visual system is to provide it with an ambiguous input which enables it to arrive at different descriptions alternately. Figure **17** shows a number of ambiguous figures, one of which forms part of a graded set that makes it easier to see the ambiguity. The significance of these various ambiguous figures is that they show different scene descriptions in force at different times. Once again, the image remains constant, but the way we experience it changes radically. Some aspects of the scene description do remain constant throughout – certain small features for instance – but the overall look of the picture changes subtly as each possibility comes into being. The scene description adopted thus determines the way we see the pictures. Just as with the upside-down figures, it is not just a case of different verbal labels being attached at different times. Indeed, the total scene description, including both the features described and the overall interpretation, quite simply *is* the visual experience each time.

One last trick technique for demonstrating the talent of our visual apparatus for scene description is to slow down the process by making it more difficult. Consider **18** for example. What do you see there? At first, you will probably see little more than a mass of black blobs on a white ground. No particular object is evident. But now turn the book clockwise 90 and have another look. It may be that you will still see nothing special – but for many readers this rotation will have brought out the hidden object. For the unfortunate ones who need yet more help, what the blobs portray is printed on page 25.

Once the hidden scene has been found (described) and made explicit, the whole appearance of the pattern changes. Here the visual system's normally fluent performance has been slowed down, and this gives us an opportunity to observe the difference between the 'photographic' representation postulated by the 'inner screen' theory, and the scene description that really occurs when we see things. The latter requires active interpretation of the available data. It is not 'immediately given' and it is not a passive process.

A further opportunity to observe your visual system battling with a difficult task of scene description is given in **19**. Few readers will be able to see the hidden object this time, even though the picture is the right way up. For the correct answer, look on page 25.





17 Ambiguous figures (a) Hawk/goose; (b) vase/faces; (c) wife/mother-in-law; (d) – (f) duck/rabbit series.

One interesting property of **18** and **19** is that once the correct scene description has been arrived at, it is difficult (and perhaps even impossible) to lose it. One cannot return easily to the previous naïve state, and experience the pictures as one did to start with. This says something about the visual system's reluctance to give up hard-won victories in battles of interpretation. A final pair of hidden-object figures is shown in **20** and **21**. These are not degraded images like **18** and **19** but naturalistic ones containing a camouflaged object. Again, most readers will need the benefit of being told what is in the scene before they can find the hidden object (see page 25 for correct answers). The use of prior knowledge about a scene as an aid to its interpretation will be considered in much greater detail later.

Three-Dimensional Scene Descriptions

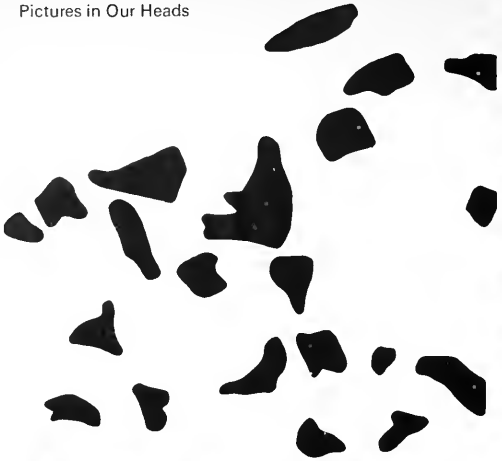
So far we have confined our discussion of explicit scene description to the problems of extracting information about objects from flat (two-dimensional) pictures. The visual system, however, is usually confronted with a scene in three dimensions. It deals with this challenge magnificently and provides an explicit description of where the various objects in the scene, and the parts of these objects, lie in space.

The 'inner screen' theory cannot cope with the three-dimensional character of visual perception: its representation is inherently flat. An attempt might be made to extend the theory in a logically consistent manner by proposing that the 'inner screen' is really a three-dimensional structure, a solid mass of brain cells, which represent the brightness of individual points in the scene at all distances.

It is doubtful whether complex three-dimensional scenes could be re-created in brain tissue in a direct physical way. But even if this was physically feasible, what happens when we see M. C. Escher's drawing entitled *Ascending and Descending* [**22**, **23**]? The monks walk up or down a never-ending staircase. (Escher was a Dutch artist and to appreciate why he chose monks to ascend and descend endlessly one must know that the Dutch term for useless labour is 'monk's work'!) The staircase is part of an impossible building. It could not be physically created in the way we see it (although we will return to the question of building trick models of 'impossible' objects in a moment). But if the building is physically impossible, how then could we ever build in our brains a three-dimensional physical model of it? We must look elsewhere for a possible basis for the brain's symbolic representation of depth ('depth' is the term usually used by psychologists to refer to the distance from the observer of items in the scene being viewed).

Another impossible object is shown in **24** [plate 1], this time an impossible triangle. Escher has again provided a superb narrative drawing as an elaboration of the basic effect, using it to portray an impossible waterfall [**25**]. The water in his drawing would circulate for ever – a realisation of the dream of a perpetual motion machine, and of course a solution to the energy crisis.

What do impossible figures tell us about the brain's symbolic representation of depth? Essentially, that small details are used to build up an explicit depth description for *local*



parts of the entire scene, and that the overall consistency of the representation is not treated as so important. Just how the local parts of the impossible triangle make sense on their own is shown in **26**, which gives an exploded view of the figure. The brain interprets the information about depth in each local part, but loses track of the overall description it is building up. Of course, it does not entirely lose track of this global aspect, otherwise we would never notice that impossible figures are indeed impossible. (Though many people are completely taken in by Escher's drawings [**23** and **25**], and often have to seek help in finding out what is wrong with them. They do not notice anything impossible about them and wonder what all the fuss is about.) But the overall impossibility is a rather 'cognitive' effect – a realisation in thought rather than in experience that the figures do not 'make sense'. If the visual system took the global aspect very seriously, then it would have dealt with the figures rather differently. For example, it could in principle have 'broken up' one corner of the impossible triangle and led us to see part of it as coming out towards us and part of it as receding. This is illustrated in **27**. But the visual system emphatically does not do this, not from a line drawing [**26**, **27**], nor from a physical embodiment of **27** (devised by Richard Gregory) [**24**, **28**, plate 1]. That is, if a three-dimensional model of **27** is made and viewed from just the right position, so that it presents the same retinal image as the line-drawing, then our visual apparatus still gets it wrong, and delivers a scene description which is impossible globally, albeit sensible locally [**28**]. Viewing of this 'real' impossible triangle has to be one-eyed, otherwise other clues to depth come into play and produce a proper global perception. (Two-eyed depth processing will be discussed in detail in chapter 7.)

One interesting game that can be played with the trick model of the impossible triangle is to pass another object, such as one's arm, through the gap while an observer is viewing the model correctly aligned, and so seeing the impossible arrangement. As the arm passes through the gap, it seems to the observer that it slices through a solid object!

An important point illustrated by **28** is the inherent ambiguity of flat illustrations of three-dimensional scenes. The real object has two limbs at very different depths: but viewing with one eye from the correct position can make this real object cast just the same image on the retina as one in which the two limbs meet in space at the same point. This inherent ambiguity, so difficult to comprehend fully because we are so accustomed to interpreting the flat retinal image in just one way, is brought out clearly in a famous demonstration by Ames, shown in **29**. The observer peers with one eye through a peep-hole into three separate rooms and sees a chair in each one (**29**a). But when he looks at each room from above he realises that only one room has a chair in it (the one shown in **29**b). The other two have an odd assemblage of luminous lines, etc., suspended in space by invisible wires. None the less, the collection of lines is cunningly arranged in each case to produce a retinal image which mimics that produced by the real chair. In the middle room, the lines are not even formed into a single object, but lie in space in quite different locations – and the 'chair seat' is painted on the rear wall! In the right-hand room, the lines-plus-seat *are* a coherent object, but one which is very distorted. The point is that all three rooms have things within them which result in a chair-like retinal image being cast in the eye. The fact that we see them all the same – as chairs – is a tribute to the visual system's ability to interpret the retinal information in just one way in each case, the way which yields the most 'conventional' perception. It is 'blind' to all other possibilities, but that should not deceive us into thinking that these possibilities do not in fact exist. How the visual system arrives at the same scene description in each case is still a great mystery, as is the question of whether it is important that we 'know' the shape of normal chairs, and use this knowledge to guide the interpretation of the retinal image (see p. 116).

Normal scenes are usually interpreted in one way and one way only, despite the ambiguity just referred to of the information in the retinal image. But it is possible to catch the visual system arriving at different descriptions of an ambigu-

18 [facing, top left] What figure do the blobs portray?

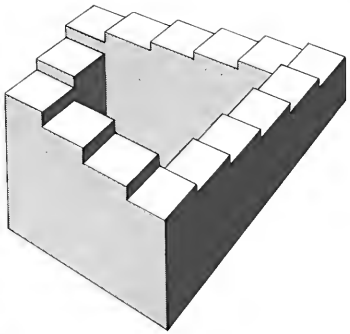
19 [facing, bottom left] What object is hidden amongst the blobs?

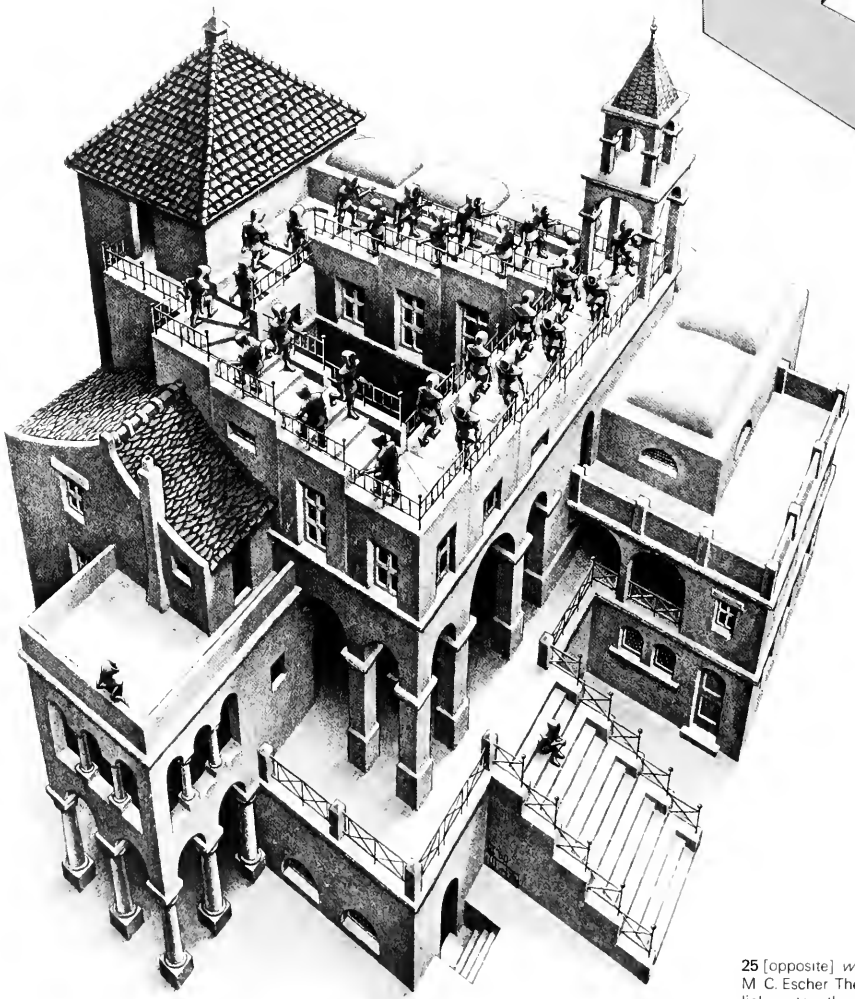
20 [facing, top right] A case of natural camouflage

21 [bottom right] A case of military camouflage

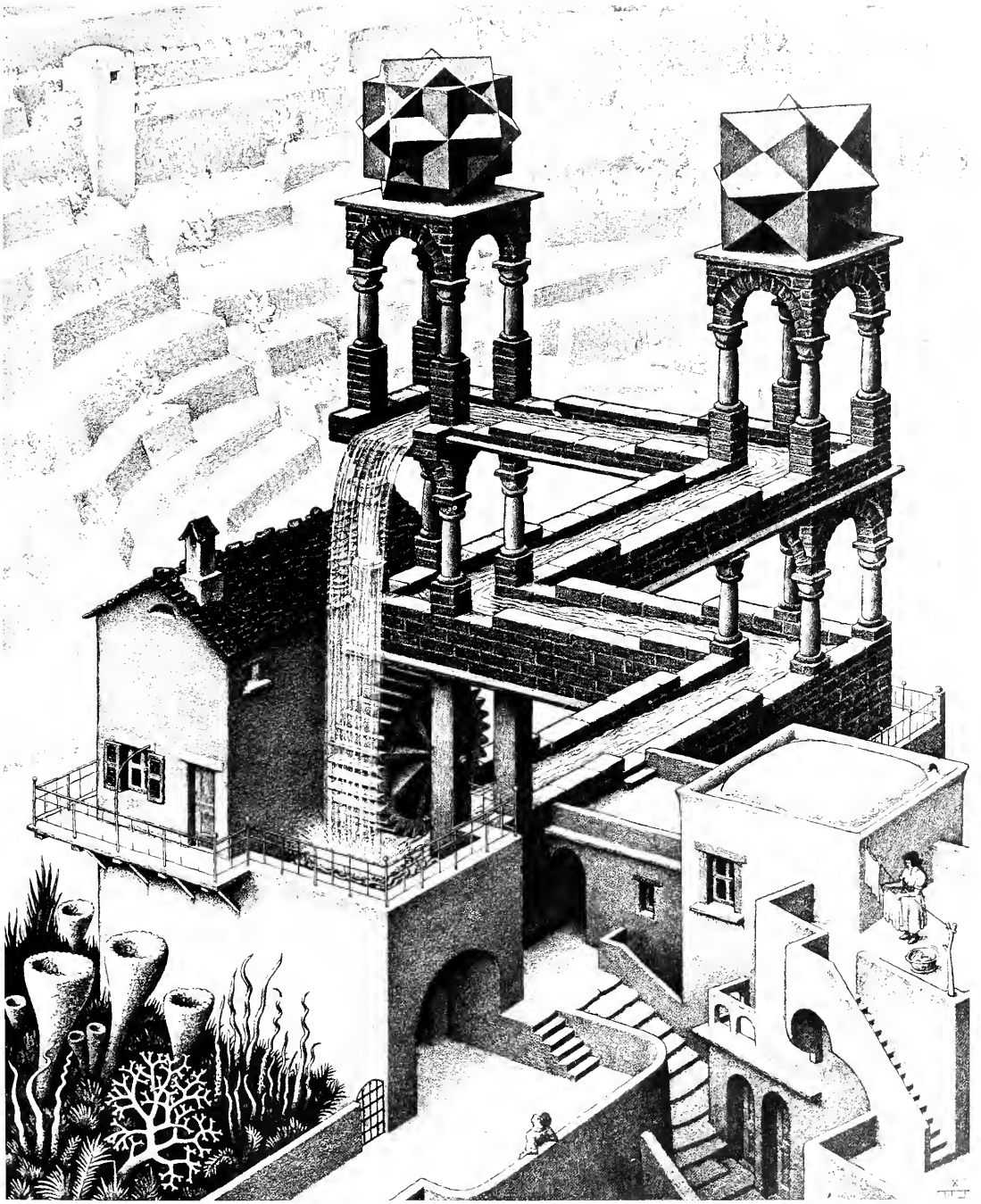
22 *Ascending and Descending*, a lithograph by the well-known illusionist artist M. C. Escher, showing movement up and down a never-ending staircase.

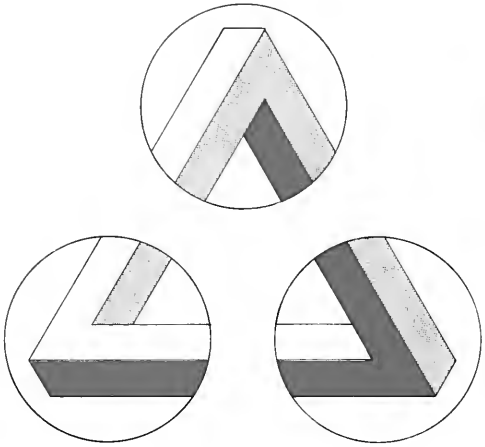
23 Drawing after Penrose on which Escher based his lithograph. It shows the essential structure of the impossible building.

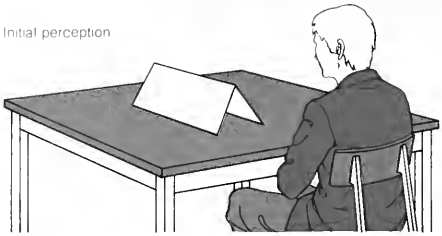


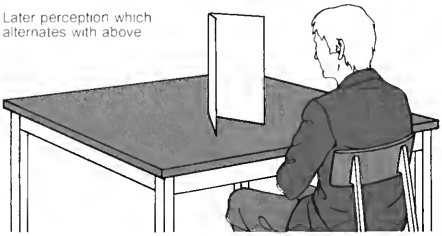


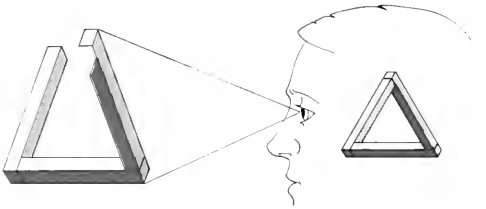
25 [opposite] *waterfall*, lithograph by M. C. Escher. The waterfall is based on a linking-together of two impossible triangles. The falling water forms one limb of each of the triangles.









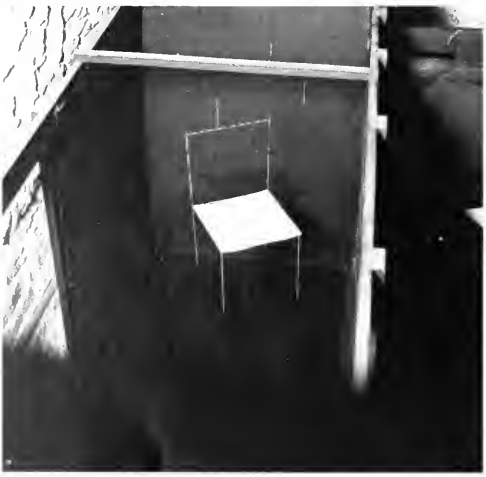


30 Try staring at a piece of folded paper resting on a table. After a while it suddenly appears not as a tent but as a raised corner.

26 [top left] The local parts of the impossible triangle make sense: it is the overall organisation which is nonsense.

27 A real object (left) which when viewed from the appropriate vantage point gives the same retinal image as the drawing of the impossible triangle shown inside the observer's head.

29 [below] **The Ames chair demonstrations.** (a) What the observer sees when he looks into the rooms shown in b, c and d through their respective peepholes; (b) normal chair, (c) scattered parts of a chair; (d) distorted chair.



a

b

ous three-dimensional scene in the following way. Fold a piece of paper along its mid-line and lay it on a table [**30**]. Stare at a point about mid-way along its length, using just one eye. Keep looking and you will suddenly find that the paper ceases to look like a tent as it 'should' do, and instead looks like a corner viewed from the inside. The effect is remarkable and well worth trying to obtain. The point is that both 'tent' and 'corner' cast identical images on the retina, and the visual system sometimes choses one interpretation, sometimes another. It could have chosen many more of course, and the fact that it confines itself to these two alternatives is itself interesting.

Conclusions

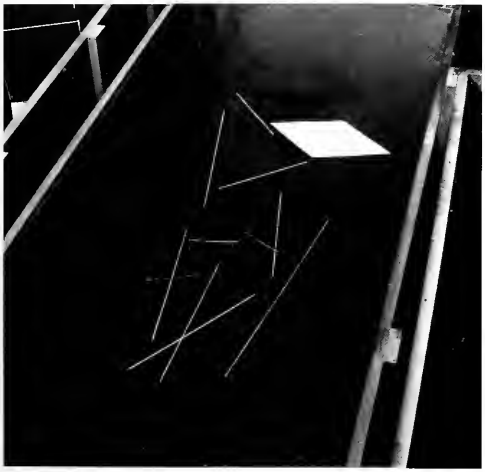
Perhaps enough has been said by now to convince even the most committed 'inner screen' theorist that his photographic conception of seeing is quite inadequate. Granted then that seeing is the business of arriving at explicit scene descriptions, the next problem becomes: how is this done?

It turns out that understanding how to extract explicit descriptions of scenes from retinal images is an extraordinarily baffling problem, despite being a superbly interesting one. The problem is at the forefront of much scientific and technological research at the present time, but still remains largely intractable. Seeing has puzzled philosophers and scientists for centuries, and continues to do so. To be sure, some notable advances have been made in recent years on several fronts within psychology, neurophysiology and machine image-processing, and this progress will be reviewed in this book. But we are still a long way from being able to build a machine which can match the human ability to read handwriting, let alone one capable of analysing and describing complex natural scenes. And this is so despite multi-million dollar investment in the problem, because of the immense industrial potential for good processing systems – think of all the handwritten forms, letters etc. which still have to be read by humans even though their contents are routine and mundane. An understanding of seeing so fundamental that it enabled a highly competent visual robot to be built would be a remarkable achievement, and this objective is at the core of much scientific effort at the present time. Whether we will witness a successful outcome to the quest this century, or even in the next, is debatable, as is the question of whether a solution would impress the ordinary person. It is a curious fact, but also a significant one in that it highlights both the difficulties inherent in understanding seeing and the way we take seeing so much for granted, that computers can already be made which are sufficiently clever to play competent, even good, chess, and yet they cannot be made to match the visual capacities even of some quite primitive animals. The computer cannot be made to 'see' the chessboard (the moves are fed into it in non-visual ways) even though it can be made to 'think' about it! Even so, most people would probably be more impressed with a good chess-playing computer than with a good image-processor, despite the fact that the former is nearer realisation than the latter. It is one of the prime objectives of this book to bring home to the non-specialist reader why the problem of seeing remains so baffling. Perhaps by the end of the book the reader will have a greater respect for his magnificent visual apparatus.

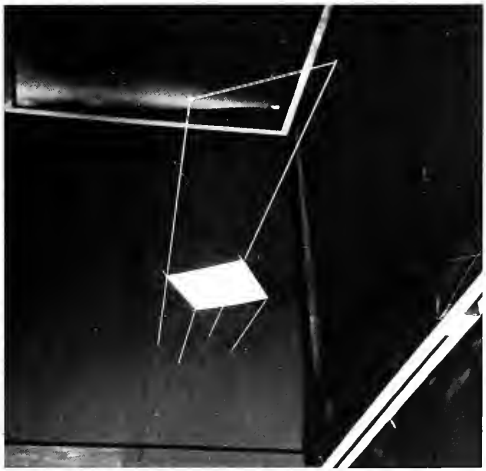
Meanwhile we have said enough in this opening chapter to make abundantly clear that any attempts to explain seeing by building representations which simply mirror the outside world by some sort of physical equivalence are bound to be insufficient. So we can finally dispatch the 'inner screen' theory to its grave and concentrate henceforth on theories which make *explicit description* their objective.

Solutions to problem pictures on pp. 20–1

18 shows a horseman.
19 shows a dalmatian dog sniffing at fallen leaves.
20 shows a woodcock in thick ground cover.
21 shows a Vietnamese soldier.



c



d

2 SEEING FEATURES

The problem of seeing is the problem of building up a symbolic description of a scene using information contained in an input visual image. That much was made clear in the last chapter. The central questions of interest become therefore: first, how can a symbolic scene description be arrived at; and second, how is the description represented in our brains? The first question is a matter of the *principles* involved in the analysis of visual scenes – what is the general nature of the problem, and what general methods can be used in its solution? The second question concerns the particular visual machinery with which human beings are equipped – how do the eye and brain actually carry out the job of visual scene description? (Other visual systems, such as those in other animals, or in computers, might be doing exactly the same job as far as the principles of operation are concerned, yet with a different set of components – e.g. human brain cells versus insect brain cells versus transistors.) This distinction between the *status* of a description (the job it performs and the principles involved in obtaining it) and the *representation* of the description (the physical embodiment of the description in man, other animals, or machines) was referred to earlier (p. 11), and it is an important distinction to grasp. In this chapter and the next, we will be concerned with both of these aspects of the problem of obtaining feature descriptions.

There are many aspects of natural scenes which human observers find it easy to describe on the basis of sight (i.e. 'see'), but there are various advantages in beginning with the problem of *seeing features*. A look around the visual world before you enables you to describe it effortlessly in terms of small elements, such as the edges of objects, corners, blobs, lines, etc. Could this ready ability for seeing the features within a scene reflect the fact that there is a 'feature representation' inside our heads – a collection of symbols which stand for features in the outside world, symbols which contribute in an important way to the explicit description of the visual scene which constitutes seeing?

Features and Objects

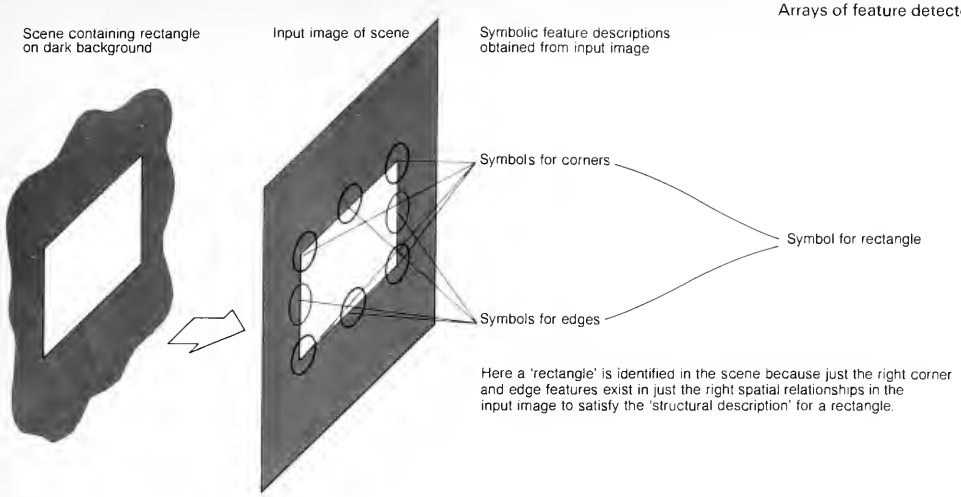
Proposing brain symbols for features cannot be the whole answer to seeing, of course, because we are able to describe scene characteristics much more complicated than simple features. Even so, a feature representation might serve not only as a basis for our ability to see features in themselves, but also as an important first step on the way to the more complex perceptual jobs of object recognition, depth perception, etc. Consider, for example, how it is that we are able to recognise a simple shape such as a rectangle [**31**]. A rectangle is built up of various features, usually four corners and connecting edges or lines. Perhaps when we 'see' a rectangle, what has happened is that the visual system has identified certain corner/edge/line features as being present and then discovered that these are so arranged with respect to one another that they satisfy the scene description 'rectangle present'. If the features had been arranged differently, then a different scene description would have been arrived at (such as 'square present', for example). And if different features and a different arrangement of them had been discovered, then a quite different shape would of course be described. The point is that perhaps a phase of feature analysis precedes a phase of object description in seeing, rather than an object being identified 'all at one go', as it were.

Tackling the problem of object recognition by looking for structures within a collection of sub-parts is called, straightforwardly enough, the *structural description* approach to object recognition. It is by no means the only way to go about the job, but there are good grounds for supposing that the human visual system works on this basis (perhaps in addition to others), and much more will be said about it later. For the present, **31** simply illustrates a possible advantage for subsequent processes of object recognition of having an early phase of feature analysis.

To summarise what has been said so far: one principle of early (sometimes called 'low-level') visual processing is *to encode symbolically all the useful information contained in the input image in terms of a vocabulary of feature symbols*. The word 'vocabulary' is appropriate because feature symbols are rather like 'visual words' which stand for features in the scene. Just what form these 'words' might take as an actual representation in the human brain we will leave until the next chapter. For the moment, we will concern ourselves solely with some general principles of extracting information about features from an input image.

Arrays of Feature Detectors

First of all, what symbols can we use for features in this general discussion? Earlier, when discussing the problem of building a burglar detector (p. 15), the symbol used for 'burglar present' was a closed switch. The same idea will be applied here, as illustrated in **32** [plate 2]. The input image is shown being inspected by an array of detectors. Each detector is portrayed as a square in a grid, and each square is to be thought of as a switch (see detail). If any switch is closed,



Here a 'rectangle' is identified in the scene because just the right corner and edge features exist in just the right spatial relationships in the input image to satisfy the 'structural description' for a rectangle.

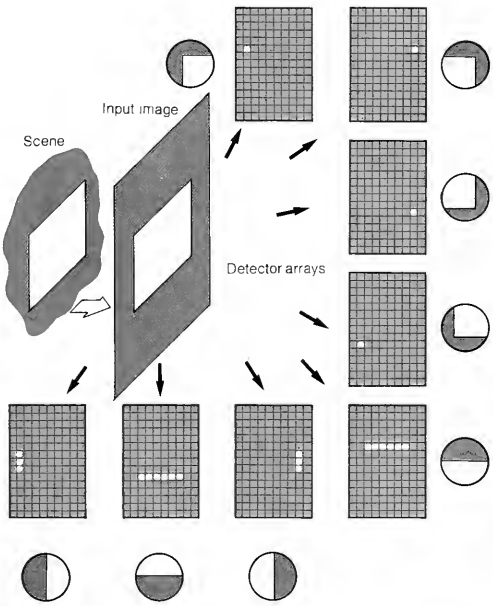
then this is the signal that a certain feature exists in the input image at the location dealt with by the detector in question. Note that all the detectors in the array are searching for the same *trigger feature*: the only difference between them is that they are 'looking' in different parts of the input image. The patch of input image which each detector deals with is called the *receptive field* of the detector. Adjacent detectors in the array have somewhat overlapping receptive fields, as shown in the enlarged detail, even though the centres of adjacent receptive fields are shifted with respect to one another.

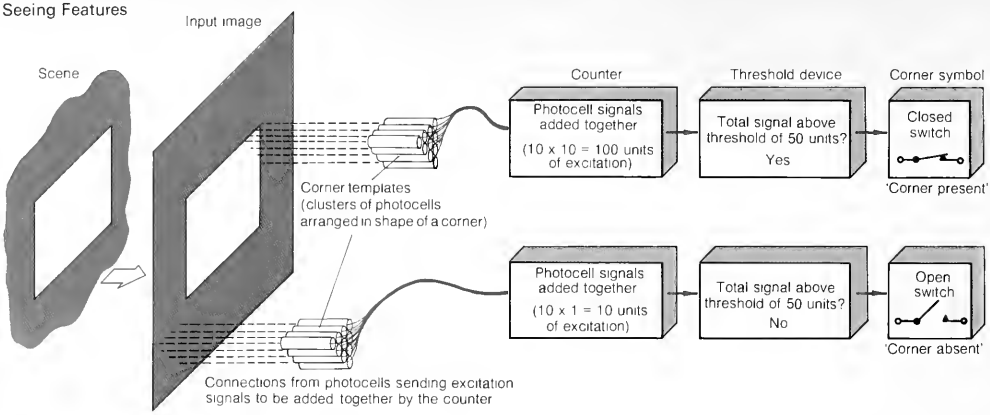
In **32**, all the detectors in the array are hunting for the same trigger feature. That is, they are all looking for a 'corner-pointing-upwards-to-the-right'. Other features are dealt with by other detector arrays [**33**]. Thus we have to imagine myriad different detector arrays, one array for each possible type of feature to be identified. This might seem a very expensive way to go about feature analysis: so many different types of feature exist that a huge number of switches would be required! However, if we were to consider each switch as equivalent to a brain cell (an idea which will be discussed in full later), we might take comfort from the fact that there are also a huge number of brain cells potentially available. Besides, the human visual system is not equally good at feature analysis over the entire field of view. We cannot identify small detailed features in peripheral vision, but by constant adjustment of the direction in which our eyes are pointing we can use the particularly good machinery available for straightahead (central) vision successively over the whole field of view.

For the case of a white rectangle on a dark ground, at least eight different detector arrays will be required, because the rectangle has eight main types of features [**33**]. Each detector array is said to be *tuned* to its own particular feature in the input image. Closed switches in each array are shown as light squares; open switches as brown squares. Thus detectors which have found their trigger features in the input are readily visible. Moreover, subsequent stages of scene description (e.g. object recognition) would find the closed-switch symbols an explicit and usable form of feature description.

31 Object recognition via structural description

33 Multiple feature-detector arrays Each array is composed of many detectors, all tuned to the same type of feature. Feature types (corners and edges) are shown as insets. Active detectors in each array are shown as light squares.





34 A simple corner detector.

The task of these subsequent ('higher-level') processes would be, in the case of object recognition, to decide what any given cluster of closed switches in the various arrays actually represented, i.e., in the present case, to 'discover' the rectangle.

In order to understand fully what follows, it is very important to realise that the light squares in **33** do *not* signify 'bright spots' in the input image. They are not part of a grey level description of the scene. Rather, each light square signifies the presence of a feature – a corner or an edge as the case may be. Thus, a light square does not have the same status as a pixel. A pixel is a single spot of the grey level description; an activated feature detector represents a cluster of light/dark pixels arranged in the shape of the relevant feature.

The reason the edge-detecting arrays possess several closed-switch detectors arranged in a line is that the edges of the rectangle are long enough to trigger several detectors That is, several adjacent receptive fields will be presented with different sections of the edge (overlapping sections in fact), and so the associated adjacent detectors will all be triggered. What the various edge detectors are therefore in effect saying is: an edge is present over such-and-such a length of the input image. Of course, subsequent stages of analysis are required to make explicit from this edge-detecting information just how long the input-image edges are: this information is only implicit in the multiple responses to the separate sections of the edge. But the essential principle is that the lines of light squares in the various arrays do *not* represent lines of bright spots in the input image. This point cannot be emphasised enough. Rather, they represent a set of edge-detections.

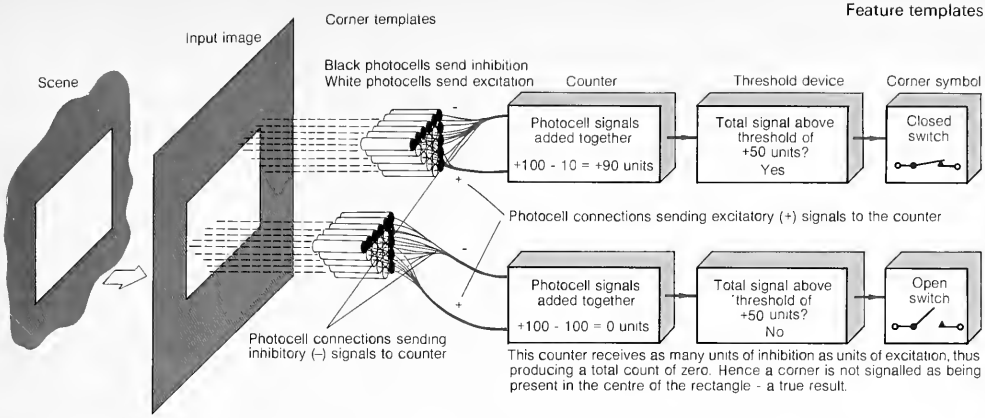
The scheme shown in **33** illustrates the objective to be attained, as far as feature analysis is concerned. The next problem is: how is each detector to be so wired up that it becomes tuned to its own particular feature and no other? This is by no means a trivial problem, and its successful solution is still an active research issue at the present time. The rest of this chapter is devoted to illustrating the difficulties and some of the possible solutions.

But first, one fundamental attribute of the symbolic feature descriptions shown in **33** is worth mentioning. This is that features of all types are represented with physically identical switches. The corner switches have just the same physical structure as the edge switches. Subsequent processes know what features the switches represent simply because these processes know the array in which the switches reside, and not because the switches themselves 'close differently' in any way once they are triggered. This form of representation is known as *place coding*: it is the place where the switch is to be found that signifies what the closure of the switch means. That is, a switch closure in one array means 'corner' and in another it means 'vertical edge'. Place coding of this type is a common form of coding used in the brain, as we will see in due course.

Feature Templates

Suppose we wished to build a detector for 'corner-pointing-upwards-to-the-right', i.e. a device which could close a switch when it found a corner feature of this type in its receptive field. How could this be done? One way of going about the problem is to apply to the receptive field a *feature template* for the corner.

The idea of a feature template is best understood by imagining the receptive field being inspected by a cluster of light-sensitive photocells grouped together into the shape of the feature to be detected. We are considering the case of the feature 'corner-pointing-upwards-to-the-right' (corner for short in the discussion that follows). Accordingly, imagine 10 photocells fixed together in the shape of a corner, as shown in **34**. Each of the photocells in this *corner template* receives light only from the point of the image at which it is directed, and each photocell generates a photoelectric signal whose size varies according to the strength of the light stimulation received. For example, let us suppose that very intense light produces a maximum photoelectric signal from each photocell of 10 units and that no light at all gives zero units. A cable connects each photoelectric cell in the template to a *counter* which adds up all the units of *excitation* received to give a measurement of the total excitation generated by the template as a whole. The excitation count is then checked by a *threshold* device whose job it is to decide whether the count



35 An improved corner detector based on excitation and inhibition

is large enough to justify closing the switch which is the symbol for 'corner present'. If the count is above-threshold (e.g. 50 units of excitation or greater), then the switch is closed and a corner has been described as being present in the input image at the location covered by the receptive field. If the count is below-threshold (e.g. less than 50 units of excitation), then the switch is left open and a corner description does not occur.

These two possibilities are shown in **34**. One corner template is inspecting a corner in the input image, and so each one of its photocells receives strong stimulation. Assuming that the rectangle is very brightly lit, each photocell is giving a maximum response of 10 units and so the counter adds these up to arrive at the total excitation tally of 100 units. This is above the set threshold of 50 units and so the corner is symbolised as being present by the switch being closed. The other corner template is inspecting a portion of the dark surround and each photocell in this template is only weakly stimulated – let us say it sends 1 unit of excitation to the counter. The total count of 10 units is well below threshold and so the switch remains open: no corner has been detected.

Feature Templates with both Excitation and Inhibition

As you may have realised already, the attractively simple scheme of **34** for wiring up a corner detector has a crucial weakness. It would lead the switch to close not only when the receptive field was positioned over an appropriate corner but also when it was positioned in the middle of the rectangle! This light region would be sufficiently large to cover the corner template, so that each photocell would be stimulated fully, and an above-threshold excitation count would result. Consequently, a corner would be detected even though no corner is present. Any system relying on this arrangement would experience an *illusory* corner in this region of the input image! Obviously, this is hopeless. Any seeing machine so vulnerable to an illusion as this would be a very poor device indeed – it could never be sure when a corner was or was not present. Some way has to be found to make the corner detector much more selective, so that it is triggered only by real corners.

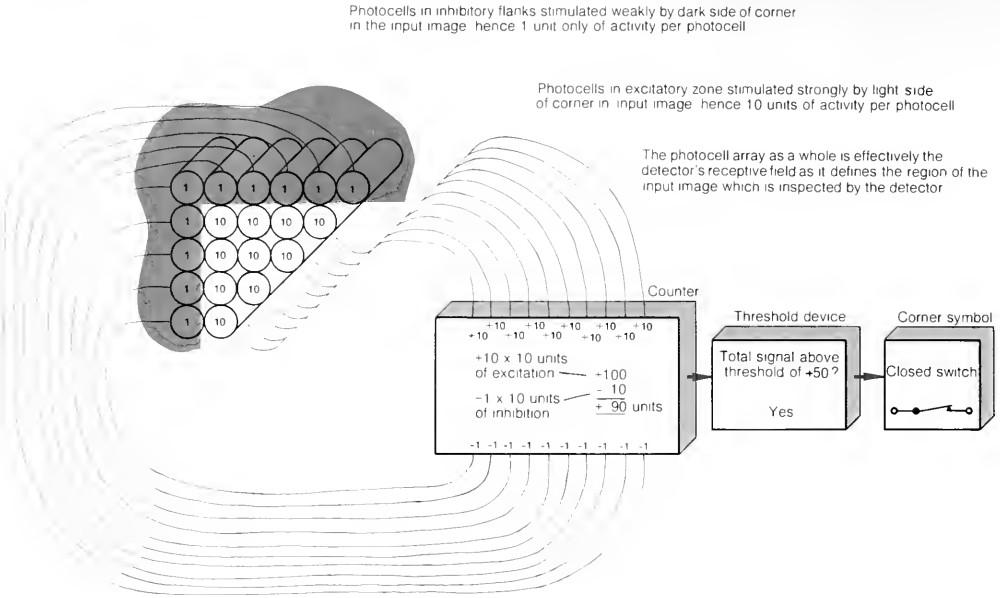
One way round this type of inefficiency is shown in **35**. Here there is a corner-shaped cluster of photocells which send *positive* signals to the measurement device (as before), together with flanks of further photocells which send *negative* signals. Both sets of photocells work only if stimulated by light: they differ simply in the nature of the signals they send to the counter. The counter thus receives both excitation and *inhibition*, the latter being the term for the negative signals.

The job of the inhibitory flanks is to prevent the corner detector being illusorily sensitive to a large patch of light that covers the whole of its template. When this happens, the excitatory signals would add up to 100 units just as before. But now the inhibitory signals would also equal 100 units (there are 10 photocells in the flanks and each would be maximally stimulated to signal 10 units each, just like the other photocells). Consequently the counter would be faced with $+100$ units and -100 units, giving a total of zero. This is obviously below threshold and so a 'corner illusion' is prevented. Adding the inhibitory zones to the template has clearly given us an improved corner detector. These zones do not substantially interfere with the detection of a real corner because in this case the inhibitory flanks receive only weak stimulation (e.g. 10 photocells each signalling 1 unit each $=10$ units of inhibition in all: see **35**). Consequently, the small amount of inhibition does not cancel the strong excitatory signal and the total count remains above threshold, so that the switch can be closed.

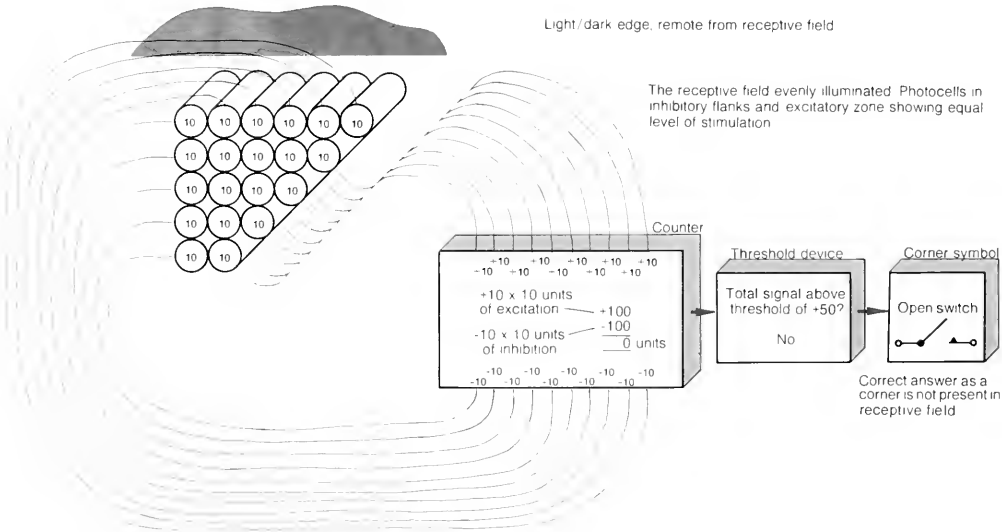
The improved detector of **35** is based on an interplay between excitation and inhibition which is found in many sensory mechanisms, both biological and man-made. Because it is commonplace and because it embodies a key principle used in visual image processing, it is important to understand it. Let us look at it in more detail.

An enlarged view is shown in **36** of the detector's template of photocells inspecting a suitable corner in the input image (see shading: note that we are now considering a corner-pointing-upwards-to-the-left). Each photocell is shown as a circle and the number inside each circle represents the stimulation level of the photocell caused by the light it receives from the input image point at which it is directed.

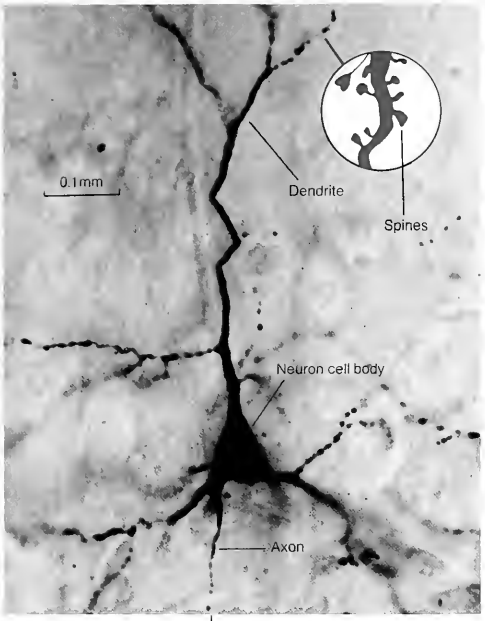
The counting part of the detector must be thought of as a



36 The corner detector as a counting device: enlargement to show details of receptive field



37 The corner detector faced with even illumination



38 Microscopic view of a single brain neuron The enlarged detail of the dendrite shows in diagrammatic form spiny protuberances which receive inputs from axons of other neurons (the latter are not shown either in the enlarged detail or in the original picture).

device which receives and counts signals from its set of sensors – the photocells comprising the template. This set of sensors inspect a region of the input image called the receptive field of the detector. However, as the sensors themselves define what region is to be the receptive field, the collection of sensors itself is also sometimes called the receptive field – the template of sensors and the region of the input they inspect are two sides of the same coin, as it were.

Each sensor sends either excitatory (+) or inhibitory (−) signals to the counter. These signals are obviously opposite in their effects, one type tending to make the count larger and the other type tending to decrease it. The count at any one moment will reflect the prevailing balance between excitation and inhibition. The counter part of the detector can thus be thought of as a kind of ballot box which receives 'votes' cast for the parties of Excitation and Inhibition. It adds up the votes cast of each kind, subtracts one tally from the other and sends the result to the threshold device. If the resulting number is positive, and large enough, then the switch is duly closed and a corner has been detected.

Details of how the improved corner detector fails to respond to an area of even illumination, such as that found in the centre of a rectangle, are shown in **37**. Now the counter finds that excitatory and inhibitory signals cancel each other, so producing a zero total which falls well below threshold. The same kind of result would occur regardless of the absolute level of illumination. For example, the detector would be inactive throughout the dark surround to the rectangle because the weak excitatory signals would be exactly nullified by weak inhibitory ones.

Excitation and Inhibition in Nerve Cells

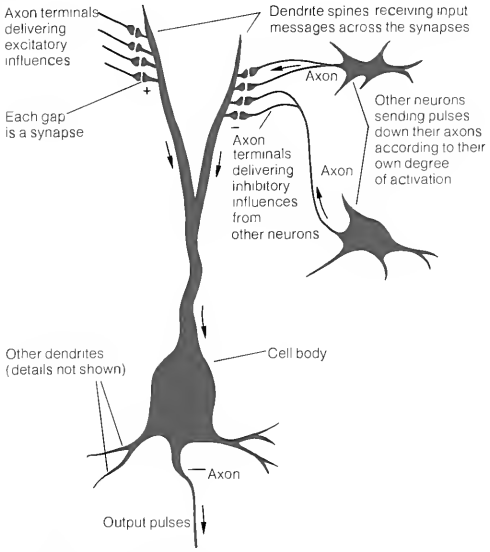
Attention has already been drawn to the fact that the kind of interplay between excitation and inhibition just described for the corner template counter is commonplace in biological systems. We will digress briefly from our discussion of how to build up feature descriptions to consider excitation and inhibition as processes governing the activity of nerve cells.

Nerve cells, technically known as *neurons*, are far too small to be seen with the naked eye, so the help of the microscope is required if we are to study their physical structure. The usual technique is to examine microscopically-thin slices of nerve tissue, suitably stained with special dyes which enhance the visibility of the tissue's various components.

A microscopic enlargement of a single neuron, stained to appear completely black, is shown in **38**. It has a *cell body* about 0.05mm in diameter (neurons differ considerably in size), to which various *nerve fibres* are connected.

One fibre is called the *axon*. It is relatively smooth and its job is to carry messages to other neurons, or to body structures such as muscles and glands. Brain axons are typically short, but neurons in the spinal cord may have axons over a metre long (e.g. those running from the tip of the spine to a toe). The axon is the *output* fibre of the neuron and we can think of it as roughly equivalent to the cable leading from the counter of **36** to the threshold device.

All the other fibres of the neuron are called *dendrites*, and



39 Diagram of typical neuron, representing it as a device which adds up excitatory and inhibitory influences and responds accordingly with a firing rate of so many pulses per second (compare with **38**).

these are its *input* fibres. They collect messages from the axon terminals of other neurons or, in the case of certain neurons in the retina, from *light-sensitive receptors* (roughly equivalent to the photocells we have been discussing). Dendrites are often covered with spiny protrusions, in which case it is these spines that collect incoming messages. However, incoming axons can terminate directly on the cell body as well. Usually, a neuron receives inputs from a great many other cells, often numbered in the thousands.

The junction between the axon terminal of one neuron and the dendrite or cell body of another is called the *synapse* (Greek: 'joint'). This junction is not a direct connection: a small physical gap (see **39**) exists at the synapse between axon and dendrite, or axon and cell body. Incoming signals are transmitted across this gap by chemical intermediaries, called transmitters, but the details of how these work need not concern us.

The dendrites of **38** and **39** have no direct equivalents in the counters of **36** and **37**, but the synapses would be where the incoming cables from the photocells terminate on the surface of the counter.

Nerve messages carried by axons take the form of low voltage (about 70 millivolt) electrochemical *pulses*, often also called impulses. Each pulse is a brief local change in the electrical potential which exists between the inside and the outside of the axon, caused by an equally brief change in the permeability of the axon's skin. At any one point along the axon, this change in electrical potential lasts only about 1 millisecond, but it spreads along the fibre comparatively slowly, at velocities of between 1 and 100 metres per second depending on the nature of the axon. Eventually each pulse arrives at a synapse with another neuron, whereupon it proceeds either to *excite* this other neuron or to *inhibit* it (via chemical intermediaries – see above). Which type of influence occurs depends on the type of axon delivering the pulse. Excitatory inputs to a neuron tend to make it 'fire', i.e. send pulses down its own axon. Inhibitory inputs tend to do the reverse, i.e. stop the neuron sending pulses down its own axon. Exactly how these two types of input achieve these effects is immaterial here. All we need to know is that some inputs are excitatory and others inhibitory.

A neuron receiving both types of input is drawn schematically in **39**. Some dendritic spines are shown receiving excitatory influences from axon terminals, while other spines are receiving inhibition. In reality, the two types of input would not be clustered together according to type as in **39**, but would be intermingled. The point of showing them separately is to emphasise that the neuron is rather like the counters described previously (compare **36** with **39**, for example). Just like those counters, the neuron is a kind of ballot box. It adds up Excitatory and Inhibitory votes and expresses the resulting balance as an output which varies in size accordingly.

Of course, the neurons are not literally counters: the way they combine excitatory and inhibitory influences is not literally to use arithmetic as though they were electronic calculators. But they show an interplay between excitation and inhibition which is similar in principle to that described for the counters. They weigh up their excitatory and inhibitory inputs and respond accordingly. Very few other details about neuron mechanisms need be known for our purposes. Simply to understand this interplay between excitation and inhibition will be enough to take us a long way along the road to understanding our current knowledge about how the brain seems to work when building up a representation of the visual world.

Mention has just been made of the fact that each neuron, like the counters, expresses its excitation-minus-inhibition tally as a varying output. The neuron's output is not a number, however, as was the output of the counters. Nor does the variable nature of the output take the form of a variable size of impulse. Nerve impulses either happen or they do not happen: they are said to follow the *all-or-none law*; each individual pulse is remarkably similar in size and duration. The variation in the output of a neuron is expressed not in terms of the size of its impulse but in terms of its *rate of firing*, i.e. the number of impulses it generates every second. A very active neuron can fire as often as 1000 times a second, but 50–200 impulses per second is more usual for an activated cell. Essentially, rapid rates of firing are found if excitatory inputs greatly exceed inhibitory ones, and the exact rate of firing depends upon the extent to which the excitation dominates the inhibition. A neuron whose excitatory input is balanced by its inhibitory one is either completely silent, or else fires with a *low resting discharge rate*. If a cell is firing at this resting rate, then inhibition has to exceed excitation by a certain amount before the cell becomes completely inactive, and sends no impulses at all down its axon.

As a point of detail, it is perhaps worth noting at this juncture that some neurons (e.g. the bipolar cells of the retina: p. 133) respond to the tally of excitation and inhibition which they receive in terms of a *graded potential change* across their cell membranes. But this does not alter the basic point just made, that most neurons transmit their outputs along their axons in all-or-none pulses which vary only in their rate of firing, not in their size. Receptors too respond in a graded potential fashion, rather than in terms of impulses, with the size of the voltage potential proportional to the intensity of light in the input image which falls upon them. This graded voltage level is thus wholly equivalent to the numbers in a computer's grey level description (see later in this chapter, p. 35, and also p. 134).

Because neurons express their output as a rate of firing and because their resting rate of firing is usually quite low, neurons have a difficulty compared with the counters in registering strong inhibition. The counters could send forth a negative number quite happily (think of how easy negative numbers are for an electronic calculator). But neurons cannot produce a negative firing rate – they cannot produce fewer than zero pulses per second! The best they can do is to reduce their resting rate to zero, and many neurons do not show even a low resting rate which can be reduced in this way. It might be wondered at this point what is the advantage, anyway, of being able to register strong inhibition. If the inhibition cancels the excitation, is that not enough? The answer to this query is that being able to detect strong inhibition does make it possible to recognise a feature with reversed brightness relationships. In **36**, a white corner on a dark surround gives a strong excitatory signal. If the corner had been a dark one on a light ground, then an equally strong, but inhibitory, signal would result [**40**]. This is because the light falling on the inhibitory flanks would stimulate a great deal of inhibition, and this would not be cancelled by the limited excitation deriving from the excitatory zone. The strong inhibitory signal is perfectly good evidence in favour of a corner feature, albeit one with reversed contrast relation-

ships. And if a threshold of *either* greater than + 50 *or* less than − 50 was set in the threshold device, then each condition would be a satisfactory one for closing the switch symbolising 'corner present' [**40**].

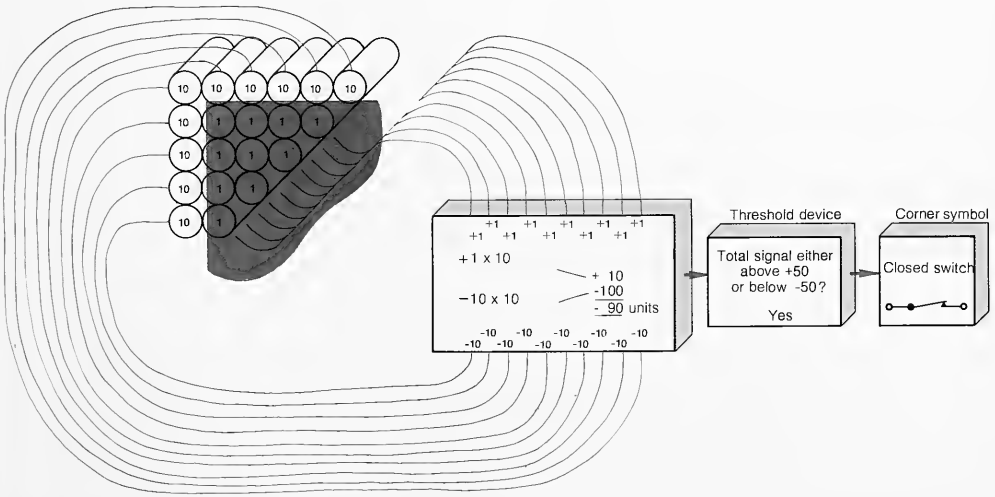
The use of a strong inhibitory signal to symbolise dark-on-light corners is not open to the brain. Its neurons receive messages from the light-sensitive receptors in the retina [**41**], which are roughly equivalent to the photocells referred to throughout the chapter. If the excitation-minus-inhibition tally for any sensory neuron favours inhibition, the best this neuron can do is remain silent. It cannot express a minus quantity in its output, as explained above. So what the brain does is to have two sets of neurons, one for dealing with light-on-dark features, the other for dealing with dark-on-light. The way each set is wired up is shown schematically for a hypothetical corner template neuron in **41**. The brain's trick is simply to reverse the nature of the effects of any given region on each particular cell, to suit the cell's needs. In the top (light-on-dark) neuron of **41**, the corner itself stimulates excitation, the flanks inhibition, just as for the corner template counter of **36**. But in the bottom (dark-on-light) neuron, it is the corner which stimulates inhibition and the flanks which stimulate excitation – the reverse of the arrangement for the neuron above it. Now each neuron can send a positive signal to the threshold unit, and the problem of not being able to send negative firing rates down axons is removed. A cunning trick indeed!

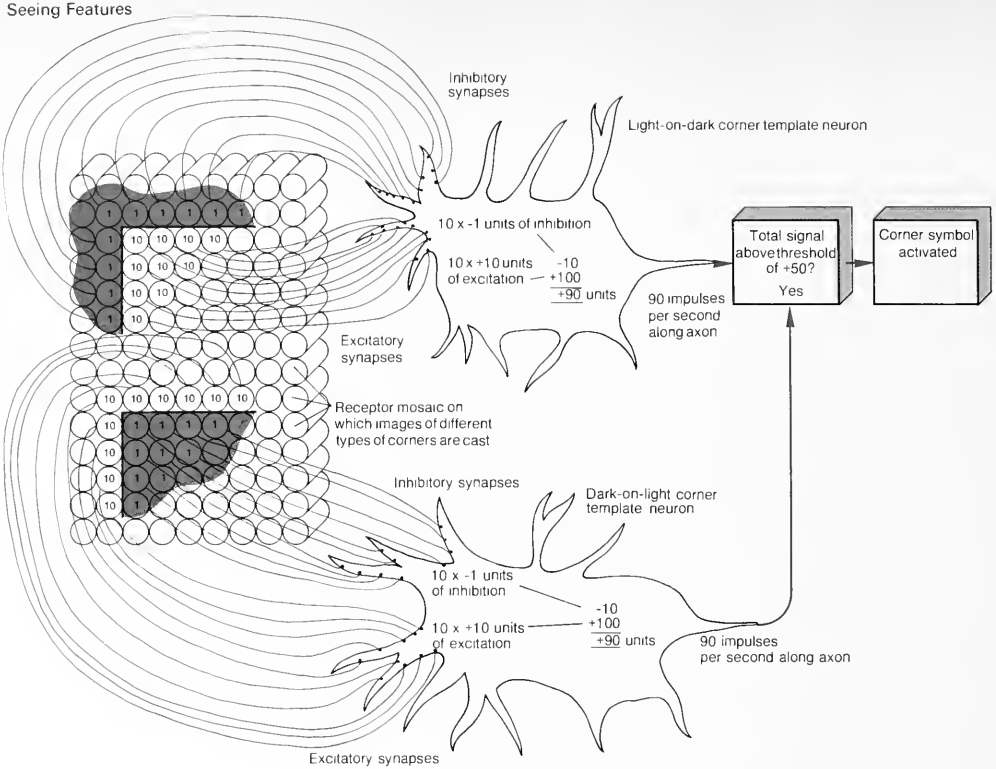
The Receptor Mosaic of the Eye

One aspect of **41** which makes it different from the equivalent figures **36**, **37** and **40** is that a *receptor mosaic* (in the retina) replaces the photocell templates. These templates were clusters of photocells, each with a particular arrangement of excitatory and inhibitory wiring connections to their associated counters. But consider the problem which would exist in using a given photocell cluster all over an image. Remember the objective that was set up in **33**, to have a switch close in each detector array in each position where the feature dealt with by the array is found. If each switch had its own photocell cluster, there would be a real difficulty in giving them all a good view of the input image at once. In **35**, two separate photocell clusters are shown inspecting the input image at the same time, but they are doing so in quite different locations. What would happen if the regions of the image they needed to inspect overlapped? (Refer back to the detail of **32** [plate 2], which shows overlapping receptive fields, for a reminder on this point.) Clearly, the clusters would get in each other's way. One cluster would block the view of the other. For example, the cluster of photocells inspecting a given region of the image for a corner-pointing-upwards-to-the-left might block the view of another cluster trying to inspect the same region of the image for an edge. And this problem would be magnified enormously if photocell clusters from *all* the detector arrays shown in **33** were to try and inspect the input image at one and the same time.

Computer systems often get around this kind of difficulty by *serial* scanning of the imput image. In **2**, an arrangement was illustrated whereby a single photocell could be moved all over the input image successively for the purpose of building up a grey level description. If this single photocell was replaced by a suitable photocell cluster, then a similar scanning operation could be conducted for feature detection, with the results at any given location being used to close the switch symbolising 'feature present' for the equivalent location in the associated detector array. And once the input image had been scanned for one type of feature with one type of photocell cluster, the cluster could then be replaced by another one, suitable for recognising the next type of feature, and new

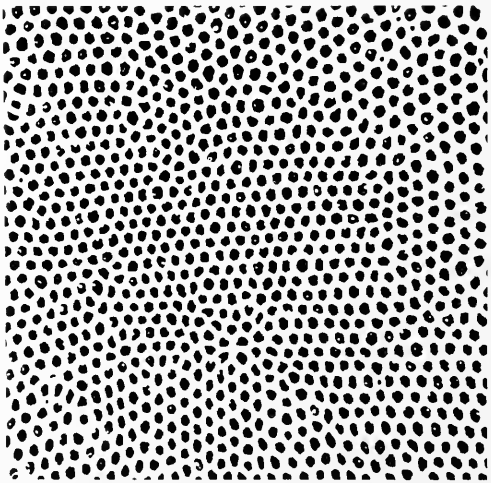
40 The corner detector faced with a dark corner on a light background





41 The brain's solution to the problem of dealing with both light-on-dark and dark-on-light features. (Sensory neurons in the brain do not in fact receive inputs from the eyes directly, but via intermediate nerve cells: this diagram is abbreviated and schematic.)

42 Photograph of cells in the part of the human retina that deals with straightahead vision (the fovea), seen head-on under a microscope (×1200)



scans performed till all the detector arrays of **33** had been catered for.

Although the serial scanning just described, in which a photocell cluster is physically moved all over an input image, is theoretically feasible, most computer systems find it more convenient to do the serial scan in a different way. It will be remembered from **2** that a grey level description of an input image can be stored in a computer's memory as a set of numbers. This description is indeed usually built up by some kind of scanning system, perhaps as described in **2**, but more often using the electronic scanning system of a television camera. But once the grey level description has been obtained by this scanning, no further direct scans of the input image are required. Instead, it is an easy matter to instruct the computer (technically, to 'program' it) to perform the required scans on the grey level description itself. The computer is told about the corner template (or other feature template) design, and it applies this template successively all over the collection of numbers in its memory. That is, it looks at the region of the grey level description covered by the template in a given receptive field position, notes which grey level numbers are to be deemed excitatory and which inhibitory for that position of the template, and arrives at a count accordingly. It does this counting operation for each receptive field sub-region of the grey level description in turn, and closes switches in the detector arrays wherever above-threshold counts are found. This scanning is still serial, one location being dealt with after another in a long series, and it is wholly equivalent to scanning the input image directly with a photocell cluster. The benefit of doing the scan this way, however, is that it takes advantage of the tremendously fast circuits which present-day computers possess. Thus the serial counts can be done on the grey level description held in the computer's memory much more quickly than an equivalent count performed by a photocell cluster moved all over the image in a mechanical system of some sort.

The human visual system, in common with most other animal visual systems, solves the problem of multiple feature-template scans differently. Its components, the neurons, are in many ways dreadfully slow devices compared with computer components. If the brain had to rely on serial scanning, it would hardly ever finish the job of analysing a scene before the scene had changed into another one requiring a fresh analysis! Indeed, even fast, modern computers find the serial scanning requirements of image processing so demanding that their ability to deal satisfactorily with input images is very limited. It is not uncommon in the present state of the art to find a computer locked in thought for several hours while dealing with a single image. Such a rate of performance would be quite hopeless for an animal trying to use visual information to enhance its survival prospects in a realistic environment, and biological visual systems have solved the problem by developing richly endowed *parallel* processing capabilities. That is, they have chosen the method of replicating similar components, so that all parts of an input image, and all types of features, can be dealt with in parallel, i.e. simultaneously. The human visual system does show some signs of serial processing, in that central (straightahead) vision is, as we have seen, particularly well-endowed with visual processing capability, and is used successively over the field of view. This serial mode of operation is made possible by eye and head movements which point the line of central vision wherever it is most needed. However, the basic design of the visual system is parallel, and the first sign we see of this is in the receptor mosaic of the retina.

An enlarged view of a small region of the human receptor mosaic is shown in **42**. This is the light-sensitive surface of the eye, and it is the surface on which the input image is focused by the eye's lens. Each receptor is a specialised cell which is capable of responding to light and passing on a message about the intensity of this light to the next layer of cells in the retina. Chapter 6 will describe retinal mechanisms in some detail. For the present, it is sufficient to regard each receptor as equivalent to one of the photocells so often referred to in the present chapter.

The most important thing to note about the receptor mosaic is that it provides a grey level description of the input image. But here, instead of the grey level description being built up by successive scans of a single photocell as in **2**, a multitude of receptors simultaneously record grey levels for each point of the image. There are about 160 million receptors in the human eye, and the output of the whole receptor mosaic can be thought of as equivalent to the grey level description shown in **2**, with each receptor providing one pixel.

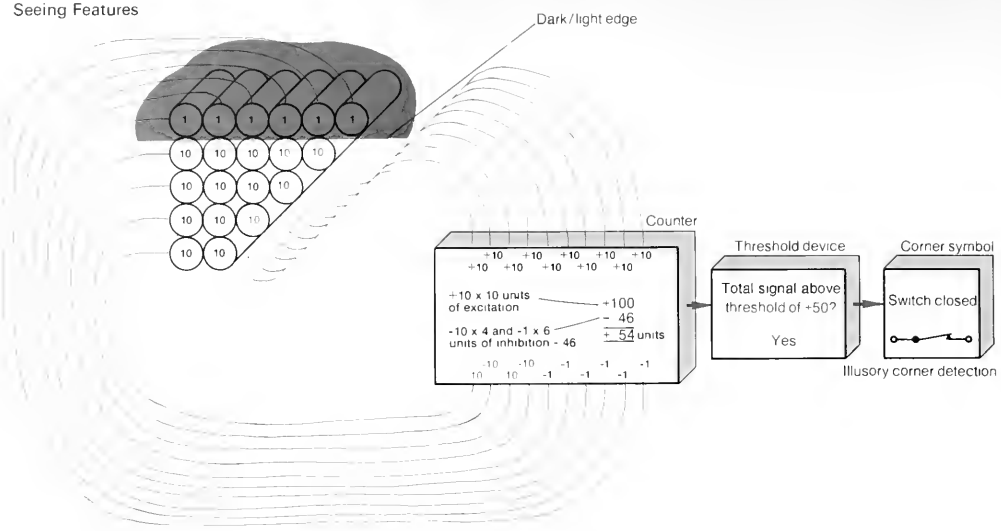
In **41**, each receptor is shown with a cable (fibre) taking its grey level intensity signal to just one of the corner template neurons. This unique, direct type of connection does not occur in practice, and is shown in **41** only for simplicity and convenience. In reality receptors feed their signals to other cells in the retina (see chapter 6), and these cells then feed brain neurons via the visual pathways (see chapter 3). Moreover, and much more importantly for our present theme, each receptor in practice influences not just one brain neuron, as in **41**, but hundreds of thousands of neurons, perhaps even millions. In short, *the grey level measurement provided by each receptor is used over and over again by many, many brain cells, all working in parallel.* Many analyses of the image, represented in **41** only by corner-feature analyses – but a great many more are in fact conducted – are all run at the same time for all parts of the field of view. It is this tremendous power of the visual system to compute in parallel that makes it able to deal so quickly with the job of scene description. The parallel operation imposes tremendous burdens on the wiring diagram of the visual system, because each receptor has to be connected ultimately to so many brain cells, but this is a burden well carried by the biological tissue of the brain, as we will see in chapter 3. At the present time, it looks as though computers too must be designed to operate in a parallel fashion if they are ever to be able to deal effectively and speedily with natural scenes viewed in realistic circumstances.

The Problem of Feature-Template Ambiguity

So much for our digression into the mechanisms the visual system uses for counting excitation and inhibition. We will take up the topic of the visual machinery of the brain again in the next chapter, but for the present we will return to the major topic of this chapter, namely the principles involved in building up a feature description.

So far we have discovered that the improved corner detector would work well for a genuine corner of its required kind, and would not be fooled by large, even areas of illumination. But what of other areas of the input image, such as the edges joining the corners together? Might there be problems to be overcome there?

In **43**, the result of placing a dark/light edge on the receptive



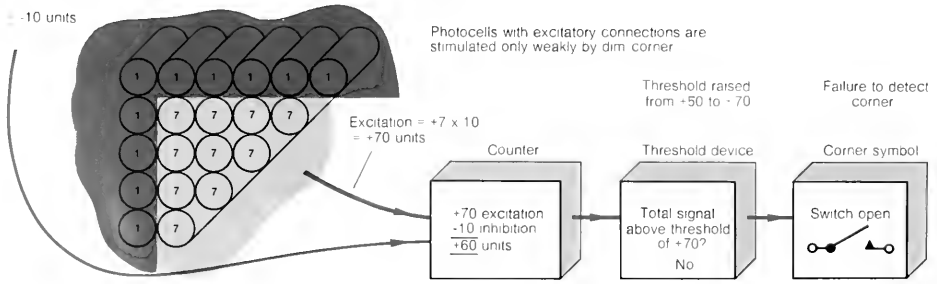
43 Corner detector inspecting an edge

field is shown. The excitation count is just as before: +100 units. But the inhibition tally is much less than the −100 units produced in the case of even illumination over the entire field (see **37**). Now the inhibition is only −46 units. The inhibitory flank which is well illuminated provides the bulk of this: −40 units. The top flank is only dimly lit and so provides only −6 units of inhibition. The resulting overall count is +54 units – which is above-threshold. Consequently, a corner is signalled when only an edge exists! Here is another problem which must be dealt with: the corner template is producing an ambiguous signal. It is impossible to tell from the closure of the switch whether we are confronted with a corner or an edge.

The obvious solution to the problem is to raise the threshold required for the corner switch to be closed, say from +50 to +70. But this leads immediately to another difficulty: dim corners could then be missed, as demonstrated by **44**. In fact, the problem of detecting dim corners was always implicit in the detector designs of **34** and **35**. The threshold level was set arbitrarily at +50 for the purposes of those illustrations, but even that level would inevitably exclude very dim corners from being detected. Thus we have arrived at a tricky problem: if the threshold is set low we get illusory corner detections at edges, and if the threshold is set high we fail to detect dim corners. How can the dilemma be resolved?

The fundamental difficulty with the corner template approach to detecting corners is simply stated: *the template's signals are intrinsically ambiguous*. This is neatly illustrated by **45** which shows, using our old threshold setting of +50, how a Y-shape of bright spots on a dark surround can be registered as a corner! This is ridiculous, of course, and emphasises just how ambiguous the template signals are. In other words, it is not at all clear whether a large response by the template to some feature in the input image genuinely reflects the presence of a corner or not. Although many tricks have been devised to improve the performance of feature templates of this general

44 Corner detector inspecting a relatively dim corner



type (see Suggestions for Further Reading, p. 159), the next approach to feature detection described here will be radically different and based on:

1] abandoning the attempt to build templates for features as complex as a corner; instead, we will concentrate solely on recognising edges and then build up a feature description of a corner from two edge descriptions found to lie at right angles to one another in a suitable corner-like arrangement;
2] taking *several measurements* using different templates, followed by
3] *interpretation of these measurements* as a whole to give a unique and certain answer about the type of edge feature present.

This threefold approach tackles the fundamental problem of template-signal ambiguity directly and, moreover, it seems to be the approach adopted by the visual systems of many animals, including man. The details of this approach are, however, best left until after the visual machinery of the brain has been described in the next chapter.

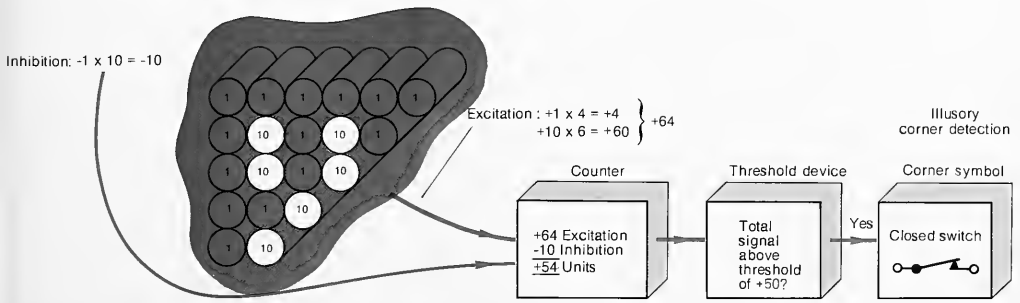
Convolution

Before proceeding to that chapter, I want to emphasise fully the inherent ambiguity of the corner template by showing it at work on a richly detailed grey level image. This is a particularly good way of demonstrating that the template is sensitive in an unwanted way to many other features besides corners of its specified type. The image I have chosen for this purpose [**46**] is a framed version of the Modigliani painting with which I began this book [**1**]. The upper left corner of the painting's mount (enclosed within a receptive field) provides an ideal candidate for identification by our light-corner-pointing-upwards-to-the-left-on-a-dark-ground detector. But the image also contains many other features to which the detector should remain 'blind' if it were a really good corner detector. How in practice, then, does the detector fare when faced with such a richly detailed grey level image?

Now it would be a formidable task to apply by hand, as it were, our corner template to every sub-region of the grey level image of the Modigliani and to calculate all the excitation/inhibition arithmetic accordingly. Fortunately, this laborious exercise is not required: we can instruct a computer to do the job for us. It will be remembered from **2** that it is possible to store in a computer's memory a grey level description of an input image. These grey levels are stored as numbers and are wholly equivalent to the various photocell readings shown in so many of the illustrations of this chapter (see also p. 134 on the retinal mosaic). It is an easy matter to program the computer to perform the required excitation/inhibition counts defined by the corner template design. It simply notes from its memory the photocell readings for the pixels covered by the receptive field of the template, identifies which readings are to be excitatory and which inhibitory according to the template design, and adds them up to give the required excitation-minus-inhibition count. It does this counting operation for each receptive field sub-region of the image in turn (refer back to **32** [plate 2]) and stores each tally in another part of its memory. The problem then is: how can the counts for the whole image be 'made visible', so that we can see how the computer has got on with its task? The answer is: print out a full-tone image in which each small square has a grey value assigned to it, according to the size of the tally [**46**]. If the tally is large, then a bright square is printed. If the tally is zero, then a mid-grey square is printed. And if the tally is negative, because the inhibition exceeds the excitation, then the square is printed towards the black end of the range. The result is a full-tone *convolution image*, so called because the process of applying a template all over an image and seeing how well it fits at every point is called *convolution*. Each square in the convolution image of **46** is in fact too small to be visible as a separate entity because this image contains 16,384 (128×128) squares in all (cf. the fine-resolution image of **3**). Even so, it is important to *think* of this convolution image as composed of many small elements, with the brightness of each one representing a corner template count.

Convolving a detector with an input image is a widely used technique for processing images, in both man-made and biological visual systems, and it is essential to understand it. The key point to realise is that the convolution image is *not* a grey level description: its various elements vary in the density of their greys but *not* for the purpose of representing the light-intensity of points in the input. Rather, the elements vary in grey to represent the goodness of fit of the convolution template everywhere in the input image. Each element in the convolution image has a position which signifies the point in

45 The corner detector misled by a Y-shape of bright dots



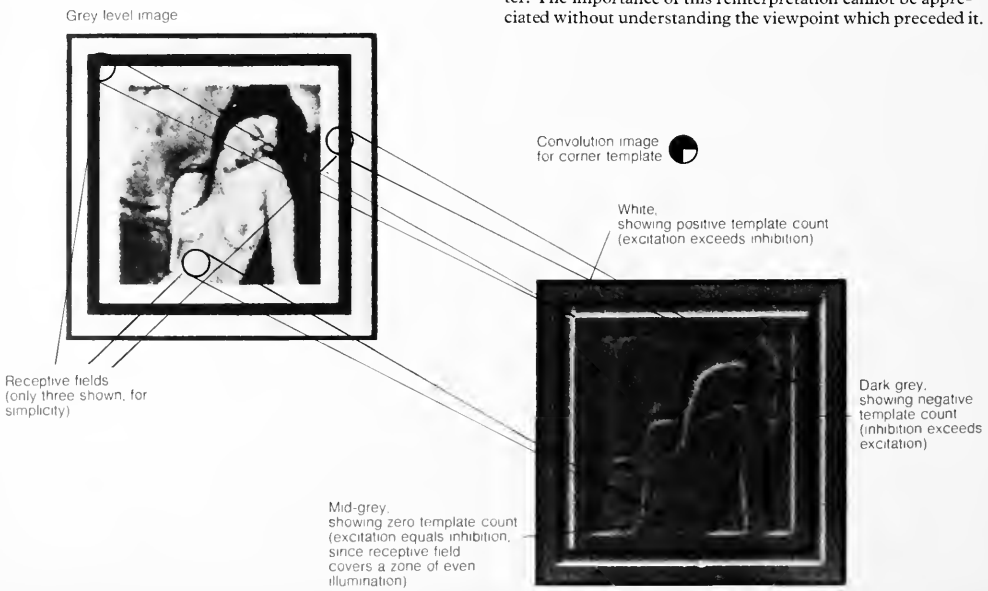
the input image on which the template was centred when the template's goodness of fit with the relevant point of the image was calculated. This is illustrated in **46** by the elongated cones connecting certain elements in the convolution image to their receptive fields in the grey level image. For simplicity, only three such cones are shown in **46**, but one should imagine every element in the convolution image having its own receptive field, with neighbouring fields overlapping as illustrated in the very similar diagram of **32** [plate 2].

The most obvious, and at the same time the most surprising, aspect of the convolution image in **46** is that it has the shape of the painting faithfully reflected in it! What should have happened, if our corner template was really doing what we wanted of it, is quite different. Just one element in the convolution image should have been very white – the element whose receptive field contained the upper left corner of the painting's mount – and all the other elements should have been mid-grey (refer back to **33** to be reminded of this type of objective). Actually, we would have been satisfied if there had been some 'very near mid-grey' elements as well (representing very low level but non-zero counts) because these would fall below any reasonable threshold level we might set for closing switches symbolising 'corner present'. Also, the presence of very black elements would not be worrying because these would be usefully signalling the presence of a dark corner on a light ground (i.e. a corner of the right type but one whose brightness relationships are reversed: see **40**). In short, the trouble with the convolution image of **46** is the presence of many elements whose brightness falls neither close to the mid-grey level, nor at the white or black extremes of the range. These elements illustrate forcibly *the inherent ambiguity of the corner template*, and it is they which pose the major problem. How are they to be interpreted? Do they represent faint, low-contrast corners of the right type, which certainly could have produced them? Or are they spurious signals deriving from edges, corners of the wrong type, blobs, etc., as in fact we know them to be in the present instance? Later processes trying to use the template counts represented by the convolution image of **46** for corner identification have no way of knowing, with things as they stand. Obviously, therefore, the corner template counts cannot be used straightforwardly as a basis for closing switches representing corners (or for the activation of any other corner symbols, including those used by the brain). The objective shown in **33** cannot be achieved on the basis of such *ambiguous data*.

Conclusions

It might seem odd to have spent so much time examining carefully the behaviour of a corner template, only to reject it finally as an inadequate basis for corner detection! In fact, there are good reasons for having gone through this careful examination. The main one is that it has served to introduce certain fundamental concepts of visual image processing, such as excitation, inhibition, templates, convolution, threshold, receptive fields, place coding etc. These concepts will be necessary for understanding the material in the next chapter. A secondary reason is that it is only relatively recently that ambiguity difficulties inherent in the idea of a feature template have become really clear, and this realisation has led to a radical reinterpretation of some classical findings on the visual machinery of the brain, as we will see in the next chapter. The importance of this reinterpretation cannot be appreciated without understanding the viewpoint which preceded it.

46 Convolution (see text for explanation)



3 THE VISUAL MACHINERY OF THE BRAIN

The human brain is an immensely complicated structure which sits tightly packed inside its protective casing, the skull. The brain's overall appearance is similar to that of a greyish pink blancmange with a wrinkled skin, and its softness and consistency is about that of a blancmange also. Somehow this superficially unimpressive structure mediates or controls such complex processes as thinking, remembering, feeling, talking, walking, seeing, etc. The brain has about 100,000,000,000 components called *neurons* (see chapter 7, p. 31) for carrying out these jobs, and an even greater number of *glia cells* which surround the neurons. The role of the glia cells is uncertain at the present time. They have usually been thought of as serving simply nutritive and supportive functions for the neurons, so enabling the latter to get on with the 'information-processing' jobs listed above. It may well turn out, however, that the glia cells play a direct role in information processing themselves. The mysteries of brain function far outweigh the items of certain knowledge at the present time, even in the case of something so seemingly basic as what different brain components are engaged in doing.

Only a simplified account can be given here of how the brain is built and seems to work, and this only for the parts most directly concerned with vision. The aim will be to give a glimpse of what recent neurophysiological findings have to say about how the visual world might be represented in brain tissue, and to discuss some of the clues these findings have provided about how to solve image-processing problems in general.

The Cerebrum

The most prominent part of the brain, and the part you would see if you lifted off the top of the skull as though it were a cap, is the *cerebrum* (Latin: 'brain'). This structure contains the bulk of the brain's machinery for vision, although there are other important sites, as we shall see. The cerebrum sits on top of many other brain structures, as shown in the exploded brain diagram of **47** [plate 3], the cross-sectional diagram of **48** [plate 3] (which also indicates some of the functions of these structures), and in the section of a real human brain given in **49** [plate 3].

The cerebrum is divided down the middle, from front to back, into two halves, the left and right *cerebral hemispheres*. Communication between the cerebral hemispheres is possible via a structure called the *corpus callosum*, which connects them.

The surface layer of each hemisphere is called the *cerebral cortex* (Latin *cortex*: 'bark'). The cortex is about 3–4mm thick in man and its natural colour is a greyish pink, but in the brain section of **49** [plate 3] the cortex has been specially stained so that it shows up as a wrinkled blue ribbon. To visualise the structure of the cortex, imagine a football perhaps twice the size of the head which has been deflated and crumpled up to fit inside the skull. If you now think of the crumpled skin of the football as the cortex then you will have quite a good idea of the latter's structure. Obviously, given this folded packaging of the cortex inside the skull, the cortex appears in cross-section as a multilayered structure [**49**, plate 3], but in fact it is best thought of as a sheet which has been wrinkled up.

The cerebral tissue lying just beneath the cortex is the *white matter*, so named because its natural colour is white. It is composed of billions and billions of *nerve fibres*. Each fibre can be thought of as a tiny telephone cable which carries messages either from place to place within the brain or between the brain and the rest of the body. The cortex has fibres within it too, but its principal constituent is huge numbers of *cell bodies*. Cell bodies and nerve fibres are the two main parts of the type of brain cells called neurons (chapter 2, p. 31).

Visual Pathways to the Brain

The major visual pathway carrying the messages from the eyes to the brain is shown in broad outline in **50** [plate 5], and in fuller detail in **51** [plate 4], in which the eyes are shown inspecting a person, and the locations of the various parts of this scene 'in' the visual system are shown with the help of numbers.

The first thing to notice is that the eyes do not receive perfectly equivalent images. The left eye sees rather more of the scene to the left of the central line of sight (region 1 to 2), vice versa for the right eye (region 8 to 9). There are also other differences between the left and right eyes' images in the case of three-dimensional scenes: these will be described fully in chapter 7, which deals with how differences between the images in the two eyes are used for the perception of the third dimension.

Next, notice that the *optic nerves* join at the *optic chiasma*. Some fibres within each optic nerve cross over at this point and therefore send their messages to the cerebral hemisphere on the side of the brain *opposite* to where they originated. Other fibres stay on the same side of the brain throughout. The net result of the partial crossing-over of fibres is that

messages dealing with any given region of the field of view arrive at a common destination in the cortex, regardless of which eye they come from. In other words, left- and right-eye views of any given feature of a scene are analysed in the same physical location in the *striate cortex*. The nature of this *binocular analysis* will be described at some length in due course (again in chapter 7).

The left and right *lateral geniculate nuclei* are the first 'relay station' of the fibres from the eyes on their way to the brain. That is, axons from the retina terminate here on the dendrites and cell bodies of new neurons, and it is the axons of these latter cells that then proceed to the cortex. We will not discuss in any detail the possible functions of the lateral geniculate nuclei, but simply note that a good deal of mystery still surrounds the question of what these structures do. Some think that because they receive inputs not only from the eyes but also from other sense organs they might be involved in filtering messages from the eyes according to what is happening in other senses. We are able to attend selectively to one sense or another, and it may be that the geniculate nuclei are involved in shutting out visual inputs to the brain when attention is being devoted to some other source of information.

Before we go on to discuss the way fibre terminations are laid out in the striate cortex, note that the optic nerves provide visual information to two other structures shown in **51** – the left and right halves of the *superior colliculus*. This is a brain structure which lies underneath the cerebrum and it seems to serve a function different from that performed by the regions of the cortex devoted to vision. The weight of evidence at present suggests that the superior collicus is concerned with what might be termed guiding visual attention. For example, if an object suddenly appears in the field of view, mechanisms within the superior colliculus detect its presence, work out its location, and then guide eye movements so that the novel object can be observed directly with the full visual processing power of central (straightahead) vision. This detailed examination for features, object recognition, etc., seems to be the special role of the visual machinery of the cortex.

It is important to realise that other visual pathways exist apart from the two main ones shown in **51**. In fact, in monkeys and most probably also in man, optic nerve fibres directly feed no less than six different brain sites. This is testimony to the enormously important role of vision for ourselves and similar species.

But the most important thing of all to grasp from **51** is the orderly, albeit somewhat curious, layout of fibre terminations in the striate cortex. First, note that a face is shown 'mapped out' on the cortical surface ('cortical' = 'of the cortex'). This is the face which the eyes are inspecting. Second, the representation is upside-down (as indeed are the retinal images too, but this is not shown in **51**; refer back to **2** for a reminder on this point). Third, the mapping is such that the representation of the scene is cut right down the middle, and each cerebral hemisphere deals with just one half. Fourth, and perhaps most oddly, the cut in the representation places the regions of the scene on either side of the cut farthest apart in the brain! Fifth, the mapping is spatially distorted in that a greater area of cortex is devoted to central vision than to peripheral: hence the relatively swollen face and the diminutive arm and hand. (This doesn't of course mean that we actually *see* people in this distorted way – obviously we don't. It is just a graphic way of showing that a disproportionately large area in our brain is assigned to the inspection of what lies straight ahead.) This dedication of most cortical tissue to the analysis of the central region of the scene reflects the fact that we are much better at seeing details in the region of the scene which we are directly looking at than we are at seeing details which fall out towards the edge of our field of view.

All in all, the cortical mapping of incoming visual fibres is curious, but none the less orderly. That is, adjacent regions of cortex deal with adjacent regions of the scene, except for the strange split down the mid-line. The orderliness of the mapping is reminiscent of the 'inner screen' proposed in **1**. But the oddities of the mapping might give any die-hard 'inner screen' theorist pause for thought. The first 'screen' we meet in the brain is a very strange one indeed!

The striate cortex is not the only region of cortex to be concerned with vision – far from it. Fibres travel from the striate cortex to adjacent regions, called the *prestriate cortex* because they lie just in front of the striate region. These fibres preserve the orderliness of the mapping found in the striate region, and in fact several more neatly-arranged scene representations are to be found in the pre-striate zone. The details of these further mappings are beyond the scope of this book. But each one seems to be specialised for a particular kind of visual analysis, such as colour vision, three-dimensional vision, movement vision. Little is known for certain at present, though, about the exact function of these extra mappings, the bulk of research having been concentrated on the striate cortex. One big mystery is how the visual world can appear to us as such a well-integrated whole if its analysis is actually conducted at very many different sites, each one serving a different analytic function. Yet other brain sites must be involved in this process of integrating the results of the striate and prestriate analyses, sites probably in the *visual association cortex* [**51**].

The Hypercolumn Theory

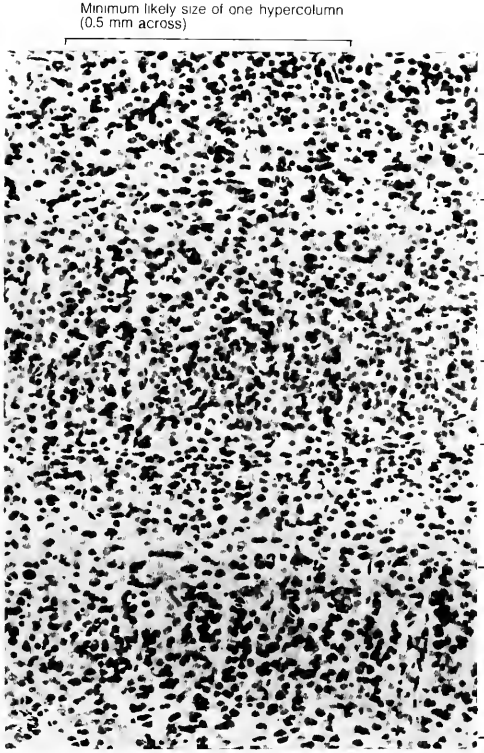
Let us now consider in a little more detail the nature of the striate mapping. Some remarkable recent research by David Hubel and Torsten Wiesel on the striate cortex of monkeys suggests that this mapping may be rather like the detector array mapping described in the previous chapter (see **32**), although much more complicated in many ways. The equivalent picture for the striate mapping is shown in **52** [plate 5]. The striate region of the right cerebral cortex is shown divided up into squares, just like the detector array of **32**. Each square on the striate surface represents a processing sub-unit, called a *hypercolumn* by Hubel and Wiesel. The reason for this name will become fully apparent in due course. For the present, think of each hypercolumn as a cluster of many constituent *columns* of cells, each one extending from the surface of the cortex down through to the white matter (fibres) below. Thus each hypercolumn is a small block of cells, organised into column sub-units, with a total area about 0.5–1mm square on the cortical surface, and about 3–4mm deep (remember that the cortex is about 3–4mm thick). Each hypercolumn contains tens of thousands of cells, perhaps up to a quarter of a million. The job of all these cells is to inspect jointly a particular region of retina, a region we will call the *hyperfield*. (Think of the hyperfield as the receptive field of the hypercolumn, but also remember that the cells making up the hypercolumn will have their own receptive fields scattered over the region covered by the larger area of the hyperfield.)

The hypercolumn is a processing machine which has to decide what image feature is currently present in its hyperfield. This makes the hypercolumn considerably more complicated than the units in the detector array of **32**, which are just concerned with a single type of feature. In contrast, the hypercolumn is concerned with many different types of features. None the less, the basic idea illustrated in **32** for the corner detector is found also in the striate cortex: the analysis of the input image by myriad similar visual processors, each concerned with just one part of the image-to-be-analysed. Hyperfields overlap to some degree, just like the receptive fields of the corner detectors (see enlarged detail of **32**), but essentially each hypercolumn is concerned with just one region of the input image. The hypercolumns thus all chatter away simultaneously about what features they are 'seeing' in their own restricted domain, and it is the job of later processes to sort out from this feature description what objects are present in the scene.

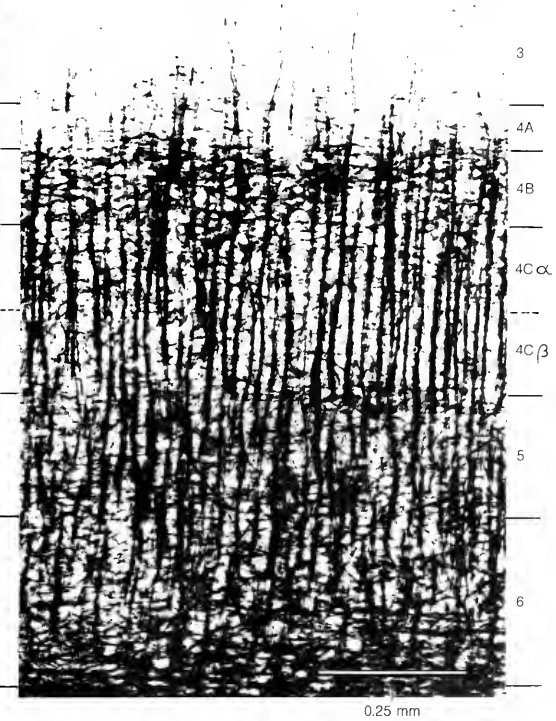
Although the hypercolumns are distributed evenly over the cortical surface, hypercolumns concerned with the central retina differ from those dealing with the periphery in having smaller hyperfields. This is shown in **52**, where the small circle close to the nose (around location number 5 in the retinal image) represents the hyperfield of a central hypercolumn, and the much larger circle dealing with the elbow represents the hyperfield of a peripheral hypercolumn. The approximate area of *cortex* in each hypercolumn remains roughly constant right over the cortical surface: it is just that peripheral hypercolumns have to deal with a much larger area of retina and hence can be concerned only with a relatively crude feature analysis of this larger area. Central hypercolumns on the other hand, with their smaller hyperfields, can engage in a much finer analysis. This accounts for the difference between the finely-detailed vision of objects inspected directly (central vision) and the crude appreciation of objects seen off-centre (peripheral vision). Of course, because central hypercolumns have smaller fields, there have to be more of them to cover a given area of the retinal surface. This fact gives the cortical map its curious spatial distortion, illustrated by the large nose, smaller ears and tiny hands.

The squares in **52** cannot be seen on the striate cortex itself. The squares simply represent a functional organisation, whose details will shortly be described, imposed upon a structure which itself is remarkably uniform in appearance.

53 Monkey striate cortex: microscopic enlargement of a section stained to emphasise cell bodies



54 Monkey striate cortex: as in **53** but now stained to emphasise nerve fibres.



A Cautionary Note

The research to be described on hypercolumns and related topics in this chapter is very new and much of it not yet fully confirmed. Moreover, the research is based principally on one species, the macaque monkey, and it is always possible that the human visual system (to say nothing of other visual systems) works quite differently, despite the great similarities in many other ways between the visual systems of man and the macaque monkey. None the less, the findings of Hubel and Wiesel on hypercolumns fit in remarkably well with certain recent advances in understanding the *principles* of how feature information can be extracted from visual images. Consequently, the hypercolumn story is included in this book as the current 'best bet' about how feature descriptions are arrived at in our own visual systems, not as 'received fact'. It could well be that further research over the next few years will substantially alter present ideas about hypercolumns. Indeed, such revision is more likely than not, because our ignorance of visual mechanisms far outweighs our knowledge. But revision of theories in the light of new data is the everyday business of science, and it is best not to shield the non-specialist reader, for whom this book is intended, from this fact. It is preferable to give him an account of where the latest research seems to be heading rather than to keep him too cautiously within the confines of well-established knowledge. In this way he can perhaps share in some of the excitement of current research into vision, an excitement bred of a fascinating blend of psychological, physiological and computational discoveries. But at the same time, the non-specialist must recognise what the full-time scientist takes for granted: the provisional nature of scientific theories. There is a constant need to seek out evidence relating to them, and they must be relinquished when the data show that they are in fact unsustainable, despite their achievements to date in helping us understand some aspect of nature. Present evidence favours hypercolumns, but future evidence might not. Even so, the hypercolumn idea has been chosen as a worthwhile topic for this book because whatever its ultimate longevity or otherwise, it provides a marvellous framework for discussing the fundamental problems to be faced in building up a feature description of a visual scene. In using it in this way, I will be speculatively elaborating the hypercolumn concept considerably beyond the account given by Hubel and Wiesel.

Microscopic Neuroanatomy of the Striate Cortex

The time has now come to have a close look at the fine structure of a hypercolumn. This entails microscopic examination of thin slices of striate cortex. Figures **53** and **54** show suitably enlarged views of slices of monkey striate cortex. Both sections come from the same region of the cortex, but they are stained differently to emphasise different features. Thousands of small blobs can be seen in **53**. Each one is an individual cell body of a neuron – the magnification is much less than in **38**, which shows just one cell.

A very different impression of striate cortex is given in **54**, which shows a section stained to emphasise fibres rather than cell bodies. In fact, the dark stripes running from top to bottom are bundles of several fibres. The predominance of fibres running vertically emphasises the *columnar organisation* of the striate cortex. Look back and consider **53** once again. Previously you probably saw **53** as quite undifferentiated, perhaps rather like coarse sandpaper. But now look more carefully and you will be able to distinguish *columns* of cells, somewhat like strings of onions. It can be helpful in finding them to hold the book almost flat in front of you so that you are looking 'up' the columns, as it were, a trick which helps make them discernible. The bundles of fibres shown in **54** deal with individual columns of cells, taking messages up and down the column from cell to cell. Fibres running horizontally also exist, connecting cells in different columns, but these fibres are less striking in **54** than the vertical ones.

Incidentally, your perception of **55** is now probably quite different from what it was initially. You have 'learnt' how to see it, helped partly by the information given in the text and partly by your continued re-examination of it. This is a fine example of perceptual interpretation to add to those described in chapter 1.

The striate cortex is organised not only into columns but also into *layers*. These are relatively thick bands of tissue which have subtly different anatomical properties. The numbers shown between **53** and **54** are the technical labels for the various layers. We will not be much concerned with such details here, although it is interesting to note that fibres from the lateral geniculate nuclei (carrying messages from the eyes) terminate in layers 4A, 4Cα and 4Cβ. That is, they arrive right in the middle of the cortex (see also **47**). Layer 4B is special in that it does not receive incoming fibres, but instead contains a dense network of horizontally-running fibres. This network is visible with the naked eye in brain sections, where it appears as a white stripe. This stripe, called the Stripe of Gennari after its discoverer, is the structure that gives the striate cortex its name (Latin *strea*: 'a fine streak').

The overall anatomical picture of a hypercolumn is as shown in **55**. Many of the details of this structure are as yet unclear, but it seems roughly square in cross-section, with sides of about 0.5–1mm. Each side is composed of about 20–40 columns of cells, so that there must be somewhere between 400 ($20 \cdot 20$) and 1600 (40×40) columns in all. Each column probably has something like 120 cells in it, so that the total number of cells in the hypercolumn as a whole could be nearly a quarter of a million ($1600 \times 120 = 192{,}000$). All these figures are very approximate, but they serve to indicate that the hypercolumn is a processing sub-unit which is richly endowed with components and immensely complex – certainly much more so than our simple detector arrays of **33**. None the less, the complexity is somewhat relieved by the great orderliness of function which apparently exists within each hypercolumn's constituent parts. To understand something of the functioning of the hypercolumn, however, requires moving from the study of the structures formed by neurons to the study of their behaviour – from neuroanatomy to neurophysiology.

Recording from Single Cells

The neurophysiological technique which has advanced our understanding of sensory systems more than any other is called *single cell recording*, and is illustrated in **56**. An experimental animal (in this case a monkey, but many other creatures have been studied including crabs, fish, frogs, birds, rats and cats – and even human patients being operated on for brain abnormalities) is shown looking at a screen on which a variety of stimuli can be displayed. The monkey looks alert in the picture, and indeed fully conscious animals are sometimes used (the animal shows no sign of pain with careful procedures), but most often the animal is anaesthetised to achieve

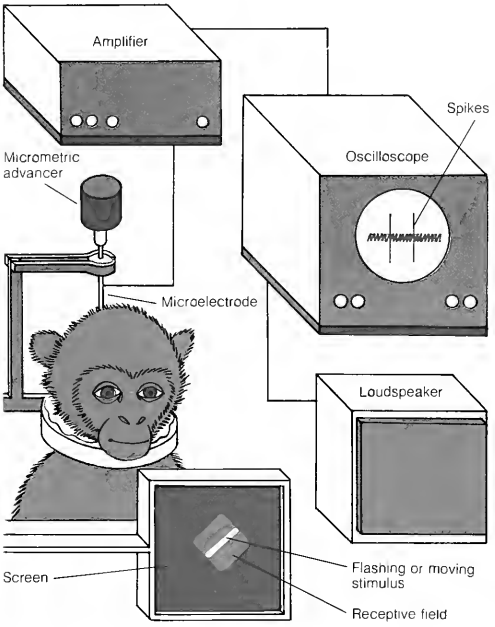
complete immobilisation, which in turn makes it possible to control accurately where the eyes are looking. A delicate probe, called a *microelectrode*, is inserted with great precision through a hole drilled in the skull into a chosen part of the visual system. This electrode, usually a thin wire insulated except at its tip, picks up the tiny electrical changes which take place around active nerve cells. These minute voltage changes (of about 70 millivolts) are amplified and displayed on the screen of an oscilloscope (an electronic device whose display screen is like a television). As explained previously, an active neuron emits a series of pulses: these appear on the oscilloscope screen as vertical 'spikes'. If a cell is responding strongly (i.e. excitation outweighs inhibition), then a fast and prolonged series of spikes is seen on the screen. If the cell is receiving little excitation, or if its excitatory inputs are balanced by inhibitory ones, then just a few spikes are seen – the resting discharge rate. And if the cell is strongly inhibited, then no spikes whatever appear on the screen.

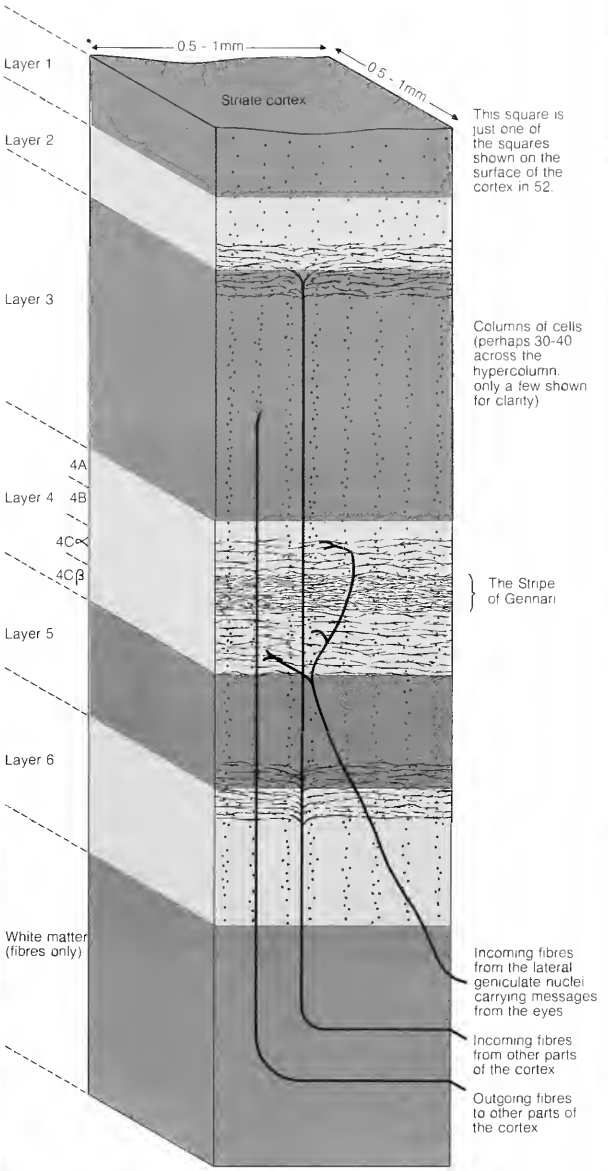
The pulses from the amplifier are also sent to a loudspeaker so that the neurophysiologist can literally listen in to the brain's chattering language of pulses. What he hears from an active neuron is a series of clicks, each click being an individual pulse.

The general strategy underlying the technique of single cell recording is to advance the microelectrode very gradually into the part of the visual system chosen for study until some

55 [right] Schematic diagram showing organisation of a hypercolumn into columns and layers

56 [below] Single cell recording in the striate cortex The microelectrode picks up nerve impulses from cells sensitive to patterns on the screen.





activity from a neuron is detected. Having pulses made audible via the loudspeaker is very useful in this process because it frees the experimenter from having constantly to observe the oscilloscope screen. Once having picked up a neuron's activity with his microelectrode probe, the experimenter then proceeds to try out a range of visual stimuli on the stimulus screen in an attempt to discover what visual inputs excite the cell and what ones inhibit it.

We will first of all describe the results obtained by using this technique to explore the properties of cells in the striate cortex. However, the technique has been used with success in many other locations within the visual system, including the retina, and we will turn to some of these other results in due course.

The first important property of striate cells which was discovered by single cell recording is that each cell is concerned just with a limited patch of retina – its receptive field. This is entirely as expected, given all that has been said so far about the neuroanatomical mapping of the retina on to the striate cortex – the way neurons from each area of the retina connect up to a corresponding area in the cortex in an orderly way. In the experiment shown in **56** the monkey's eyes are immobilised, which enables the receptive field of the cell being recorded from to be drawn on the screen. Obviously, if the animal's eyes were allowed to move about (as is permitted in some experiments) then the receptive field of the cell would be swept successively over different regions of the screen.

Also in keeping with the neuroanatomy are results obtained when the microelectrode is driven into the cortex perpendicular to its surface, so that it passes from one cell to another within a given column. If recordings are taken from each cell encountered in turn, then it is found that they all have their receptive fields in the same general region of the retina. The whole column of cells is therefore 'looking' at more or less the same part of the retinal image (there is some scatter in the locations of receptive fields). What is each cell looking for?

It turns out that there are many different types of cells in each column. Some are called *simple cells*, others *complex cells*, and yet others *hypercomplex cells*. These titles were invented by the discoverers of the cells' properties, Hubel and Wiesel, and they reflect the relative complexity of the analysis performed by each neuron. But although there are many different types of cells within every column – types we will shortly describe – all the cells within any one column share a very important property: they are all maximally excited by *line stimuli with the same orientation*. This common property of the cells in a column will become clear as we proceed further. The only exception to the rule that all cells in a column share the same orientation preference occurs in layer 4, in which the cells are not orientationally tuned at all. Some parts of layer 4, it will be remembered (p. 42), receive input fibres from the lateral geniculate nucleus, and at this early stage of cortical processing the orientational analysis has not yet got under way. Note that lateral geniculate cells, like those of the retina (pp. 129 and 133), are not orientationally tuned either: orientation selectivity is a property exclusively of cortical cells.

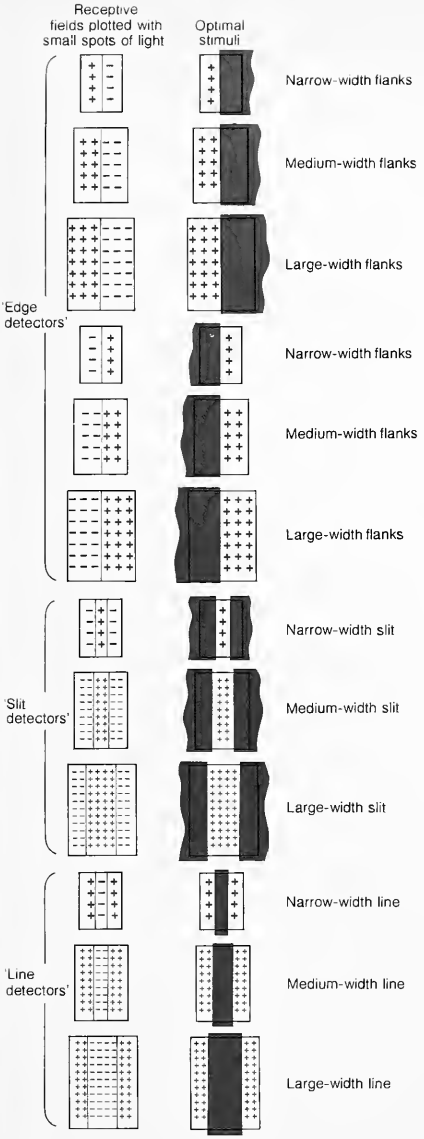
Simple Cells

Let us begin with the simple cells, which as their name implies are the easiest ones to understand. Their distinguishing characteristic is that their receptive fields can be divided into excitatory and inhibitory sub-regions using *stationary* stimuli. That is, if a stationary stimulus such as a spot of bright light is flashed in certain regions of the field, then it is found that the cell becomes excited and emits a burst of impulses. Equally, if a stationary spot is flashed in other regions of the field, the cell becomes inhibited, i.e. it stops emitting impulses, or its firing rate at least falls below its normal resting discharge rate. (Actually, many simple cells have very low, even zero, resting discharge rates. Thus, if we wish to demonstrate the presence of an inhibitory region in the receptive field, we may first have to excite the cell by flashing a stimulus on the excitatory zone of its field, and then flash a stimulus in the inhibitory zone. The latter flash reveals its inhibitory effect by reducing the ongoing rate of firing caused by the excitatory stimulus.)

The next question becomes: what are the *shapes* of the excitatory and inhibitory regions of the receptive fields of simple cells? This question can be answered by exploring the effects of flashing small spots of light all over the field, noting each time the effect of the flash on the cell. If a record is kept of the excitatory and inhibitory effects by marking plus-signs and minus-signs respectively on a sheet of paper which itself represents the general region of retina covered by the receptive field, then some typical field maps of simple cells are as shown in the left-hand column of **57**. At first sight the variety might seem rather bewildering, but on careful examination it is possible to group the different cells into certain categories, called *edge detectors*, *slit detectors*, and *line detectors*. Unfortunately, it turns out that these labels (the customary ones to date) are probably misnomers in that the cells possessing these fields are almost certainly not *directly* signalling edge/slit/line feature detections at all – but more of that anon.

Consider first the receptive fields of the so-called edge-detecting simple cells. Note that each field is divided into two sub-regions – one excitatory and one inhibitory. This is shown by the + s and − s, the former indicating that when a spot of light was flashed in their location the cell became active, the latter that in their position the spot of light caused inhibition. The boundary between the excitatory and inhibitory regions has an orientation which defines the *orientation tuning* of the cells in question. In **57**, all the edge detectors are vertically tuned: this means, as we shall see, that they respond best to vertical edges. Indeed, all the field maps shown in **57** are vertically tuned – edge, slit and line ones. This is because all the field maps come from cells within a single column. Remember that all cells in any one column share the same orientation tuning, with different columns differing in the orientation to which they are tuned.

The appropriateness of calling the cells of **57** vertically tuned will perhaps be more apparent when the *optimal stimuli* for exciting these cells are considered. The optimal stimulus for each cell is shown alongside its +/− mapping. One has to imagine these optimal stimuli lying over the receptive fields, with the light region of the stimulus falling on the + s, the dark region on the − s. Because light falls on the + zone, shown by the + s drawn in the appropriate region of the optimal stimuli of **57**, the cell would obviously be receiving a lot of excitatory stimulation from a stimulus of the kind illustrated. That is, each + represents a spot on the receptive field which is receiving light stimulation (compare the equivalent photocell inputs of the last chapter), and as *all* the + s are simultaneously being excited by the light they receive, the cell is obtaining the most excitation possible – hence the



57 Receptive field maps and optimal stimuli of simple cells from a vertically tuned column

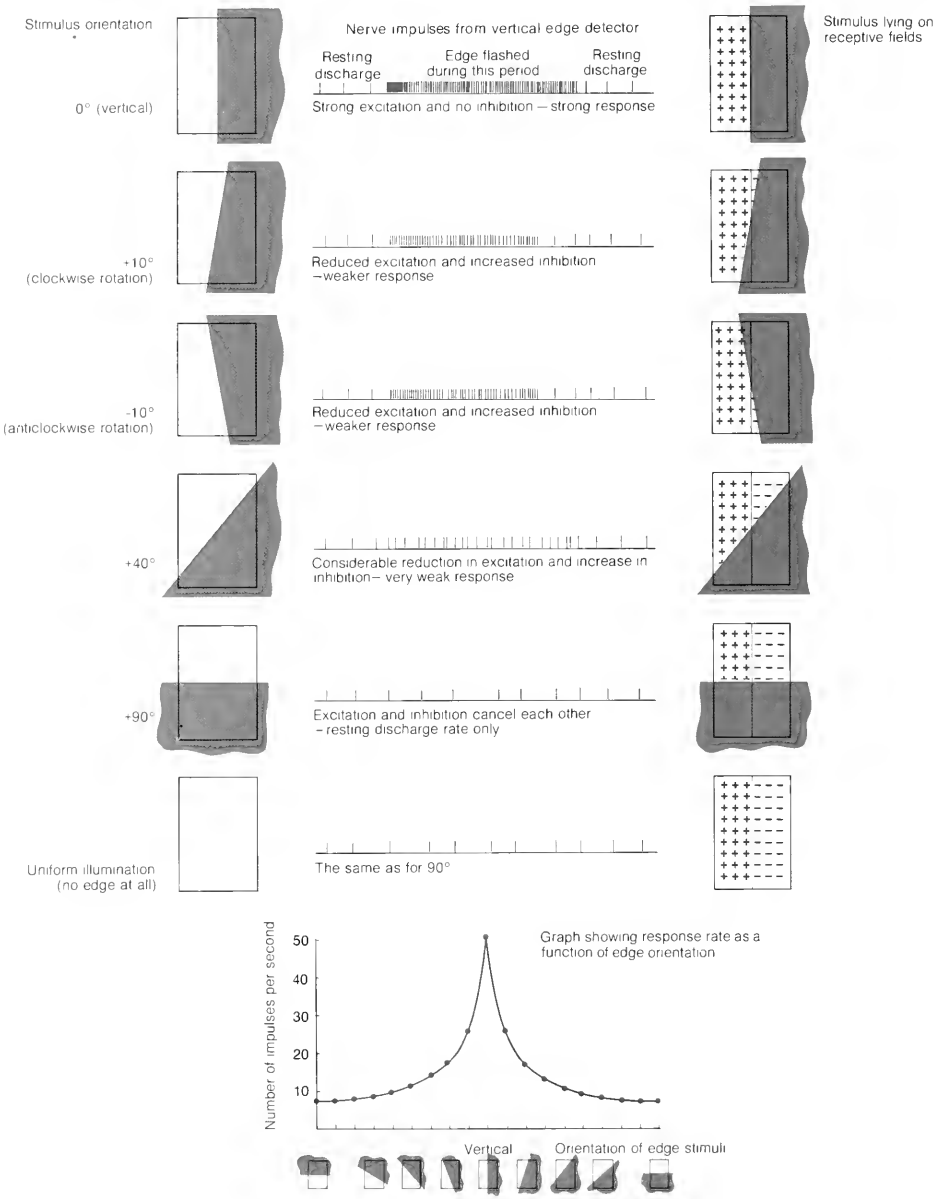
use of the word 'optimal'. The word 'optimal' is doubly fitting because not only is each cell receiving maximal excitation from the stimuli shown in **57**; it is also receiving no inhibition. This is because each inhibitory spot is receiving only darkness, as it were. Thus the inhibitory zone is getting no stimulation and therefore sends up to the cell no inhibition. In short, the input of excitation is very strong and is not reduced by inhibition, so that the firing rate of each cell is at its maximum: hence the use of the term 'optimal' to describe the stimulus in each case. Note that the absence of inhibition from each optimal stimulus is depicted in **57** by the absence of − s on the dark areas of each stimulus.

What has just been said about the optimal stimuli for edge detecting cells applies equally well to all the other types of cell in **57**. In every case, the optimal stimulus is one which provides maximum excitation and minimum inhibition.

The term 'slit' to refer to an optimal stimulus of a white line on a dark surround may seem odd, but it has become a conventional term ever since Hubel and Wiesel created these stimuli for use in their experiments by shining light through a slit – literally. The term 'line' has also become a customary one for a dark line on a light surround, an unfortunate usage in many ways because all the optimal stimuli shown in **57** are line stimuli of a sort, and not just the 'line' ones. Sometimes, however, the slit and line stimuli are called light and dark 'bars' respectively.

If you look down the column of optimal stimuli for the different kinds of simple cell, you can easily see why these cells have so often been dubbed feature detectors. That is, it has commonly been assumed that because each cell has as an optimal stimulus one or other line feature, each cell must be a signalling device for saying whether this feature is present on the patch of the retina inspected by the column of cells as a whole. But recent work, particularly that by David Marr, has suggested that this is far too oversimplified a view of their function. He has discovered, by attempting to build an artificial (computerised) feature-detecting system, that the output of a cell of this kind cannot be taken straightforwardly as a signal for a particular feature. Activity in any one cell considered on its own is far too ambiguous to be usable directly as a feature description, ambiguous in just the same way as was the corner detector of the last chapter. We will return to this very important issue later on in this chapter. But at this stage it is worth mentioning Marr's view that the outputs of simple cells are best thought of as *measurements* of the input image which function as a first step towards working out what feature has produced any given set of measurements. The key insight that Marr's work has produced is that a further stage of *interpretation* of the simple cells' measurements is necessary before a feature description can be obtained: the simple cells are not themselves feature detectors. I will expand considerably on Marr's views in due course. I mention them in outline at this stage in order to emphasise that the labels 'edge detector', 'slit detector' and 'line detector' should not be taken too literally. They are convenient terms in that they describe the kinds of input features which optimally excite the simple cells: but they must not be taken any further than that.

I have already mentioned that all the receptive fields shown in **58** are vertically oriented, and that the cells associated with them are therefore said to be vertically tuned. Let us explore this in a little more detail. First, the axis of orientation is given by the boundary between the excitatory and inhibitory



58 Orientation tuning of a vertical edge detector

areas. In **57**, of course, this boundary is always vertical, and the optimal stimuli are therefore vertical. What happens if non-vertical stimuli fall on these receptive fields? The answer is that the response of the cells is reduced according to the amount by which the non-vertical stimulus diverges from the vertical. This is illustrated in **58** for an edge detector optimally responsive to a left-side-light/right-side-dark vertical edge. Nerve impulses are shown as they would appear on the oscilloscope screen of **56** before, during and after an edge stimulus had been briefly flashed on the screen. Various types of edge, differing in their orientations, are shown on the left of the figure. The impulse records show the response waning as the orientation is changed from vertical, reaching in the end the resting discharge rate (the rate found before and after the stimulus presentation) when the edge is horizontal. The overall effect of stimulus orientation, for many more edge stimuli than are shown on the left of the figure, is illustrated in the graph below the records. This graph shows a strong peak at vertical, reflecting the vertical optimal stimulus required for the cell's maximal response, with a sharp falling off in response rate as the stimulus edge is rotated away from the vertical, either clockwise or anti-clockwise.

The right-hand side of **58** shows why the edge detector exhibits this vertical tuning effect. As the stimulus edge is rotated from vertical, so darkness falls on some of the excitatory zone (so reducing the excitatory input to the cell), and at the same time some light falls on the inhibitory zone (so causing inhibition to be set going, which then partially offsets the excitation). In the end, when the edge is horizontal (shown on the graph as 90° orientation difference from vertical) the cell receives as much inhibition as excitation and so its firing rate is reduced to its resting discharge level. The balance between excitation and inhibition in this case is shown by the equal numbers of +s and −s triggered by the horizontal edge. Smaller shifts of orientation from vertical, either clockwise or anticlockwise, produce smaller reductions in firing rate than a full 90° twist because they do not take so much light off the excitatory zone and do not place so much on the inhibitory zone. Thus the vertical edge detector of **59** is vertically tuned because with a stimulus at any other orientation the total excitation-minus-inhibition count is not so great as with a stimulus at vertical itself.

It is best to think of the excitatory and inhibitory zones of each cell as carrying equal weight overall. That is, if uniform illumination is covering the receptive field (bottom stimulus of **58**), then it is generally true that the cell does not 'notice' it at all because all the +s are cancelled by the −s: there is an exactly equal balance between them.

The basic point about excitation/inhibition counting illustrated for the vertical edge detector of **58** would apply to all the types of simple cells illustrated in **57**. As the optimal stimulus for any one of them is rotated, so the fall-off in excitation and growth in inhibition would be noted by the cell and its firing rate affected accordingly.

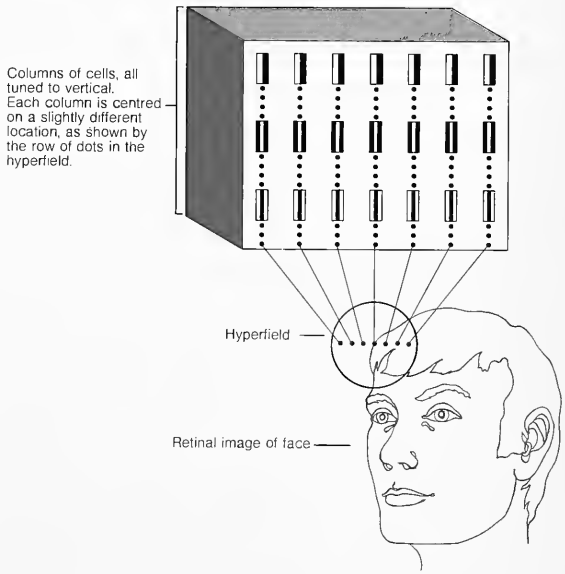
Notice that the column of cell types shown in **57** shows subdivisions of cells within the basic sub-classes of edge, slit and line detectors. Thus some cells have 'narrow-width slit' optimal stimuli, others 'medium-width slit' optimal stimuli, and so on. The cell fields shown are in fact just a sample of the whole population of cell types in the column, and many more types of field exist covering a wide range of slit-widths, line-widths, and widths-of-flanks on either side of an edge. Just why so many different widths within each sub-class should be required will be discussed later; for the present, think of the cells within a column taking a broad range of different measurements from the image feature in their receptive fields. The usefulness of these measurements will become clear in due course.

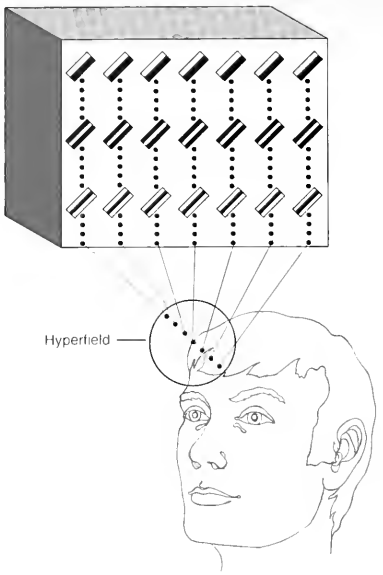
Orientation Columns and Orientation Slabs

So far we have described just one vertically tuned column of cells. But in fact each hypercolumn is equipped with many such columns, all similarly tuned but each dealing with a slightly different area of the hyperfield **[59]**. That is, each vertically tuned column is 'centred' on a different spot in the hyperfield so that there is a *slab of columns* covering the whole width of the hyperfield. Use of the term 'slab' is wholly appropriate because this is how the columns are physically arranged in the cortex: the columns within the slab are literally side by side.

Of course, although each column is centred on a slightly different patch of the hyperfield, the receptive fields of cells in neighbouring columns within the slab will overlap. The manner of overlap is as shown in **32** [plate 2] for the (hypothetical) corner-detector array described earlier. But despite this overlap, the centres of the column's receptive fields will be shifted slightly with respect to one another so that the whole hyperfield is covered, as indicated in **59**. Thus the

59 A slab of vertically tuned columns within a hypercolumn There are many cells in each column. Only a few are shown, as dots. Just three cells in each column are enlarged to show their receptive field types (from top down, edge, slit and line fields).





60 A slab of obliquely tuned columns within the same hypercolumn as that referred to in 59

measurements provided by the slab of cells are spread over the whole area of the field, so that the vertically tuned slab of columns is giving a broader spread of measurements than would just one column on its own.

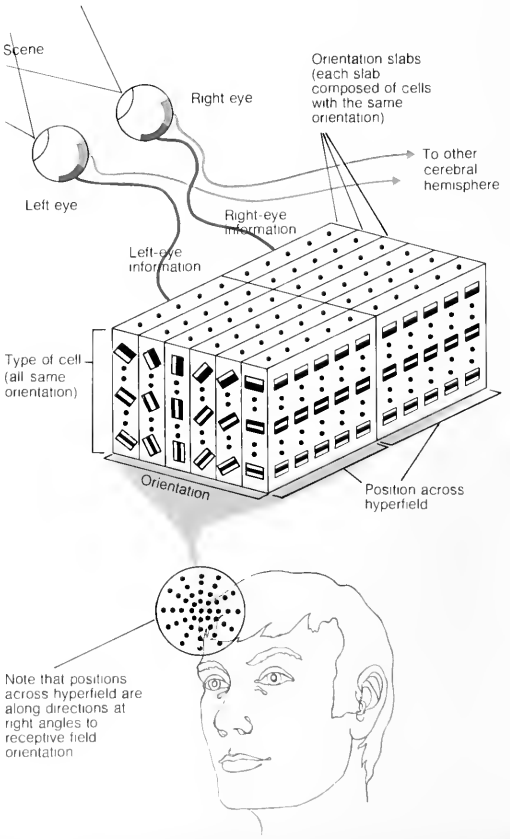
So much for vertically tuned columns. What about other orientations? There are in fact columns for orientations all around the clock, with the tuning of each column differing by about 10° from its nearest 'orientation neighbour'. A set of obliquely tuned columns, again arranged in a slab, is illustrated in **60**. Note that each column within the slab has a receptive field centred on a slightly differing point within the hyperfield. But be careful to note also that these points straddle the hyperfield at a different angle from that shown in **59** for the vertically tuned slab. The rule for the way these points straddle the hyperfield is that it is always at right angles to the orientation of the tuning of the slab. Thus the vertically tuned slab of **59** has its columns inspecting points spread out horizontally across the hyperfield, and the obliquely tuned slab of **60** has columns inspecting points spread out along the oblique opposite to its own orientation tuning. So the key point to grasp is that each slab is 'scanning' the hyperfield in a direction perpendicular to the orientation tuning of its own cells. Again, the significance of this attribute of hypercolumn organisation is best left until after an account has been given of how feature descriptions are arrived at from the whole mass of measurements taken by the hypercolumn.

Enough has now been said to make clear that each hypercolumn is really a set of slabs, each slab composed of cells tuned to the same orientation. How then is the whole hypercolumn structure built up? As far as present knowledge indicates, this structure is as shown in **61**. This is a highly abbreviated, speculative and schematic figure, of course. There are in reality many more orientation slabs than shown – perhaps 18–20 in all. This number fits in with the size of the shifts in orientation-tuning between slabs. If each slab is tuned to an orientation about 10° different from its immediate neighbours, then 18–20 slabs would be needed to go right around the clock. (Remember that if a line is twisted, then after 180° of twist orientations begin to repeat themselves.)

A further simplifying property of **61** which must also be borne in mind is that there are in fact myriad different cell types within each column. Only a few are shown in **61**.

One extra feature displayed in **61** which I have not yet mentioned in connection with the hypercolumn is the fact that it really has two halves – a left one and a right one. I pointed out early on in this chapter that the crossing-over of some fibres from the eyes at the optic chiasma brought left- and right-eye views of the scene together to a common location in the striate cortex. This common location is the hypercolumn. Again for simplicity, the left- and right-eye halves of

61 The hypercolumn (highly schematic and speculative).



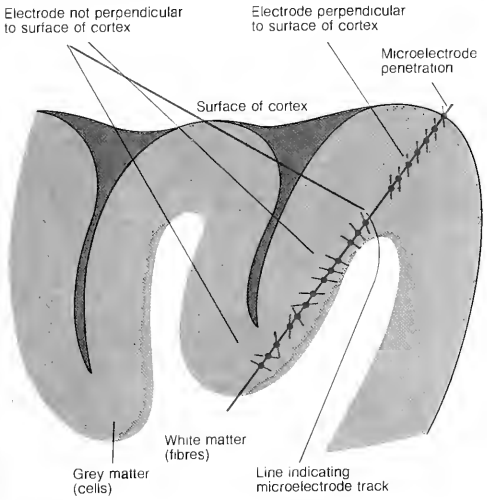
the hypercolumn are shown as quite separate entities in **61**. In fact, some cells in the striate cortex are found to be *binocularly driven*. That is, they respond actively to optimal stimuli in either eye. Others are preferentially driven from just one eye. Therefore the division of the hypercolumn into two strictly defined halves, one left- and one right-eyed, is almost certainly an over-simplification. None the less, the *monocular dominance* (i.e., predominant sensitivity to stimuli in one particular eye) of certain regions of the hypercolumn has been amply documented by the work of Hubel and Wiesel, as we will soon see. Note in passing that the left and right halves of the hypercolumn can also be thought of as slabs, albeit larger ones than the orientation ones.

Evidence Favouring the Hypercolumn

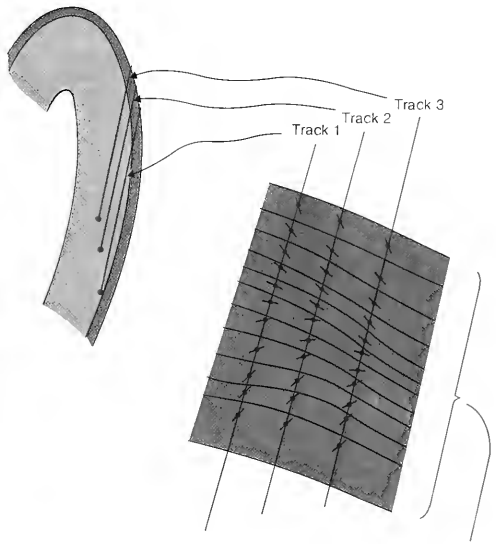
The hypercolumn structure outlined in **61** is hypothetical, but it does fit a great deal of neurophysiological data. Consider, for example, **62**, which shows the results obtained from a microelectrode penetration which traversed quite a large region of striate cortex. Each dot on the microelectrode track represents a cell from which recordings were taken. Each cell was dealt with in turn beginning with the one closest to the surface, and each cell's optimal stimulus was discovered as described earlier. In **62**, just the orientation tuning of each cell is illustrated (and not whether it was an edge, slit or line detector) by the lines drawn through each dot. Thus the vertical lines indicate vertically tuned cells, the horizontal lines horizontally tuned cells, and so on. It can be seen that where the path of the electrode remained perpendicular to the surface of the cortex (as in the initial part of the penetration), all the cells shared the same orientation tuning, which at this point happened to be vertical. But where the electrode's path is at some other angle to the surface of the cortex (as in the later part of the penetration), the orientation tuning of successive cells changes in a fairly regular progression. Note that the electrode path is itself always straight, but its angle to the cortical surface changes as it penetrates the brain because the cortex is folded (see p. 39).

Results such as those shown in **62** are obviously consistent with the idea that individual columns of cells share the same orientation tuning. When the electrode stays within a single column (or at least within a single slab), cells are found to have the same orientation preference. But when the electrode starts to pass from one column to the next (or from one slab to the next), cells are found to vary systematically in their

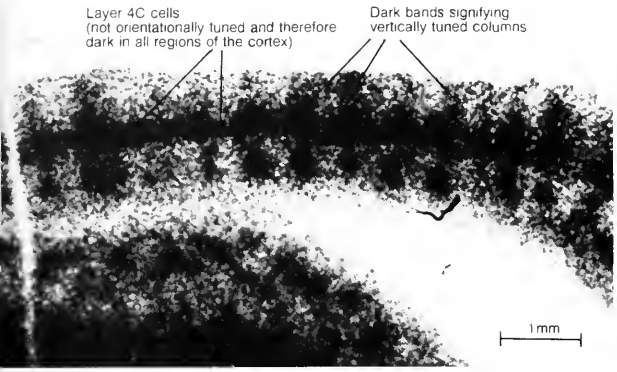
62 Results from a microelectrode penetration which initially stayed perpendicular to the surface of the cortex, and so found cells all sharing a common orientation preference (in this case a preference for vertical). Thereafter, the electrode crossed many different columns as it traversed a fold of cortex non-perpendicularly, and so it found many different orientation preferences. Each dot on the track represents a single cell, and the line drawn through the dot represents that cell's orientation tuning.



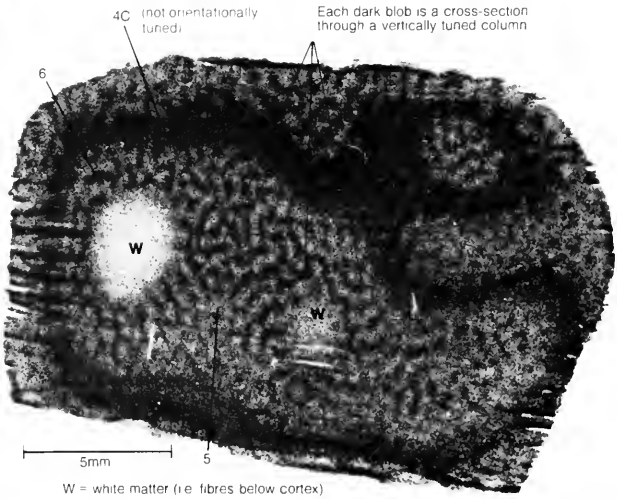
63 (a) Section through striate cortex cut perpendicular to its surface (as in **62**), showing three electrode tracks running approximately parallel to the cortical surface. The tracks lie slightly above one another, as shown, but more importantly they are also displaced laterally across the surface of the cortex. This latter arrangement cannot be shown easily on the flat page, but imagine the tracks lying at different depths above the page so that they would fall in slightly different sections. Now imagine yourself looking down on the surface of the cortex (or looking at the page from its left-hand side) so that the three electrode tracks are seen in plan view (b), which brings out the lateral displacement. Also shown in (b) are the preferred orientations of cells found by the electrodes as they penetrated through the cortex. The existence of orientation slabs is thus supported by this technique (see text).



64 Perpendicular section through the striate cortex of a monkey just after it had been shown a pattern of moving vertical stripes of irregular width filling the whole field of view of each eye. A special technique (see text) enables vertically tuned columns which are active during exposure to the stimulus to be seen as dark bands



65 Section through the same region of striate cortex as that shown in 64, but now cut approximately parallel to the cortical surface. The section actually cuts through several layers of cortex, labelled according to number (4C, 5 and 6)



W = white matter (i.e. fibres below cortex)

orientation preference, in the general fashion shown in **62**.

Some parts of the non-perpendicular track record of **62** show successive cells with the same orientation preference: this is because for some short distance the electrode happened to stay within a particular orientation slab. But the slabs are very narrow, perhaps only one cell thick, and the electrode rarely stays in one slab for long. So a better way of showing the slab structure of the cortex is to record successively from three or so parallel non-perpendicular penetrations, each one separated from the next by a fraction of a millimetre. Typical results obtained by Hubel and Wiesel using this technique on monkey cortex are shown in **63**. Three electrode tracks are shown. Track 1 was recorded first, then track 2, and finally track 3. It can be seen that the regular progression in orientation tuning found in track 1 was matched by those found in tracks 2 and 3. The inference is that the electrode tracks happened to cross almost at right angles a set of orientation slabs, and these are sketched in by means of a set of parallel lines marking out the slab boundaries. These lines would not, of course, be visible in the cortex: they are an inference from the data and show the functional organisation of the striate cortex which Hubel and Wiesel propose as an explanation of their results.

Figures **62** and **63** illustrate, in greatly abbreviated and simplified form, the kind of neurophysiological evidence favouring the hypercolumn theory of striate cortex organisation depicted schematically in **61**. Corroborative evidence of a neuroanatomical kind also exists, however, which strengthens considerably the hypercolumn story. One particularly remarkable experiment recently performed by Hubel and Wiesel, working with a colleague called Michael Stryker, involved an anaesthetised macaque monkey whose open eyes were exposed to a pattern of vertical stripes continuously for a period of 45 minutes. The stripes were of irregular width, filled the whole field of view of each eye, and were moved about to make sure that all the vertical line detectors of the monkey's striate cortex would be very active throughout the period of observation. The monkey was injected just prior to this period with a special chemical which has the property of being much more readily absorbed by active brain cells than by inactive ones. Consequently, during the 45-minute exposure period the active vertical line detectors excited by the vertical stripes would be absorbing this chemical from the monkey's bloodstream in much greater quantities than other striate cells tuned to non-vertical orientations, and so not so active during exposure to the stimulus. Following the stimulus presentation, the monkey was instantly killed and a microscopic examination made of where in the striate cortex the special chemical had collected. Exactly as expected, given the neurophysiological data already described, columns of brain tissue containing the chemical could be identified in slices of striate cortex. Thus in **64** the labelled dark bands show the regions of high chemical uptake, and they are arranged in columns perpendicular to the cortical surface. These bands therefore signify columns of vertically tuned cells which were constantly active during the period of exposure to vertical stripes. The bands are quite broad, and certainly broader than the width of a single column, because each striate cell responds not only to its optimally oriented line stimulus but also to orientations on either side of this optimum (refer back to **58** for a reminder on this point). Consequently, cells with optimal orientations to one or other side of vertical would have been activated to some extent by

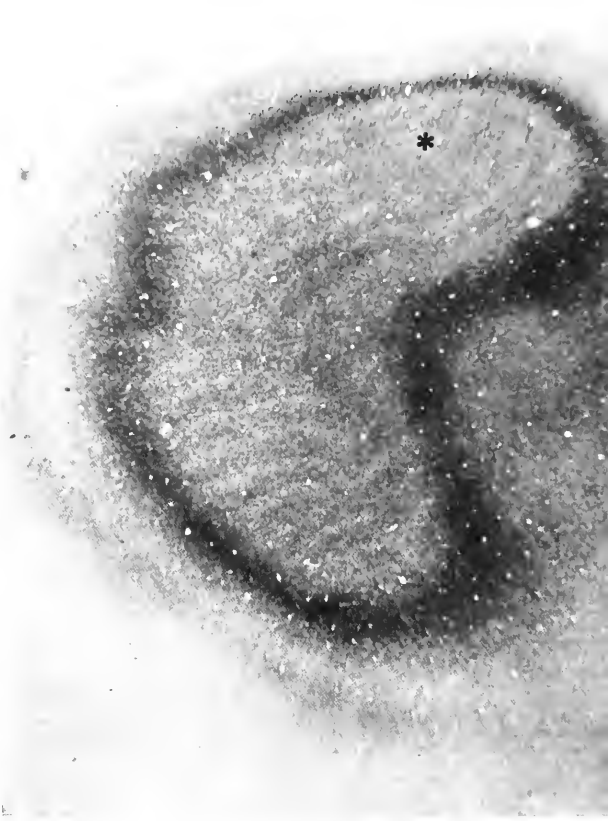
the vertical stripes and would therefore show up in **64** as part of the dark band of active cells. The bright bands between the dark ones signify columns tuned to orientations around horizontal (say ± 40–60°), columns which would not have been activated to any appreciable extent by the vertical stripes (see **58**), and so would not have absorbed much of the special chemical.

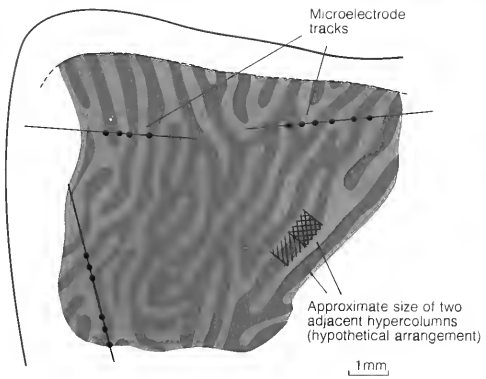
The dark band cutting through the dark columns and running parallel to the cortical surface in **64** shows active cells in layer 4C. These cells are not orientationally tuned (see p. 44) and so all of them would have been triggered to some degree by the vertical stripes. Hence all the cells in this layer appear dark.

Figure **64** presents a perpendicular section of the experimental monkey's striate cortex. Also of interest is the view obtained by the non-perpendicular section of **65**, obtained by slicing almost parallel to the surface of the cortex. Here the active columns do not figure as dark bands as they did in **64**, but rather as dark blobs. This is because the columns are now seen in cross-section. In fact, each dark blob represents a slab of vertically tuned columns (plus orientation neighbours, of course, also activated to some degree by the vertical stripes, as we noticed earlier in relation to **64**). Equally, light blobs show inactive slabs tuned to within ± 40–60° or so of horizontal and therefore not excited by the vertical stripes. The whole section presented by **65** is thus strong evidence in favour of a distribution of orientation slabs of all types over the cortical surface. Each hypercolumn would, of course, contain a full set of slabs of all orientations from each eye. It is instructive to compare **65**, which shows the experimental evidence from neuroanatomy, with **52** and **61**, which show in schematic form present beliefs about striate cortex organisation based upon this evidence, and upon the evidence from neurophysiology already described.

The kind of technique for labelling active tissue which I have just explained has another use of interest to us in this context, besides the neuroanatomical plotting of orientation slabs. It can be used with equal effect for discovering the layout of the slabs which are predominantly sensitive to stimulation from just one eye – the 'monocular dominance slabs'. Indeed, Sokoloff, who developed this procedure for labelling active brain tissue, first used it for this very purpose. In his experiment the anaesthetised monkey views a richly patterned stimulus which would stimulate almost all line detectors in the cortex, but he views it with only one eye. The microscopic appearance of his striate cortex upon subsequent examination is found to be rather like that shown in **64**, with alternating dark and light bars now signifying alternating slabs of cells, each preferentially tuned to one eye.

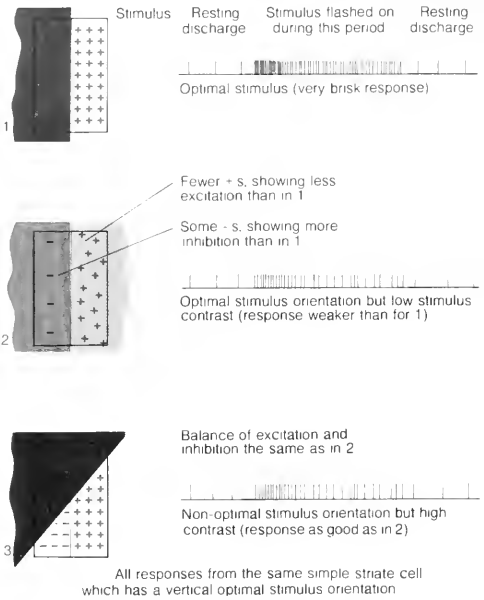
It has recently been discovered, however, by Hubel and Wiesel and a colleague, Simon LeVay, that the monocular dominance slabs can be seen directly under the microscope using ordinary staining techniques, without need of the enhancement provided by Sokoloff's procedure for labelling





66 [top] Section through striate cortex at an angle to its surface, stained to show bands of left- and right-eye-dominant cells. The asterisk marks a region where the section grazes layer 5, where the bands are not visible. The dark staining ring is the Stripe of Gennari (layer 4B), and within this is a region of layer 4C in which the bands can be seen.

67 [bottom] Schematic diagram of left- and right-eye-dominant bands of cells from various sections of the type shown in **66**.



68 The ambiguity of simple cell responses

active brain tissue. It turns out that fibres parallel to the surface of the cortex in layer 4C are denser *within* the monocular slabs than *between* adjacent slabs. This means that a suitably-stained section of the striate cortex which cuts through layer 4C parallel to the cortical surface reveals alternating dark bands ('fibre-dense': monocular slabs) and light bands ('fibre-sparse': boundaries between monocular slabs). These bands can be seen in the section shown in **66**, and the general arrangement is made clear in the schematic diagram **67**, which shows a reconstruction, from many sections like **66**, of how the monocular bands appear over a fairly wide region of cortex It is important to realise in **67** that the pale bands are depicted by the boundaries demarcating the left/right monocular slabs (each monocular type indicated by shading). Sketched in on **67** are a pair of (hypothetical) adjacent hypercolumns to illustrate how the bands are composed of many adjacent monocular slabs from different hypercolumns. The swirling pattern of the bands makes it clear that hypercolumns cannot be as neatly laid out over the striate surface as was shown in the schematic picture **52**, but this is a complicating detail we can ignore. The principle of organisation shown in **52** fits the data. There is a distribution of processing sub-units – hypercolumns – all over the striate cortex, each pretty much the same in its design. In particular, given the evidence of **67**, each region of the striate cortex (at least in layer 4C) contains areas preferentially tuned to inputs from both the left and the right eye.

One particularly fascinating aspect of **67** concerns the microelectrode tracks. Hubel and Wiesel collected single cell recordings along these tracks before conducting the microscopic examinations which were used to build up the reconstructed pattern, and then superimposed the tracks upon the reconstruction afterwards. The tracks have dots marked on them and these dots show the points at which cells changed from being preferentially driven by one eye to being preferentially driven by the other. The fascinating finding is that the dots lie exactly on the pale bands, i.e. on the boundaries between the microscopically ascertained monocular bands. This matching of neurophysiological and neuroanatomical data is a tremendous achievement. When evidence from two such different approaches blends as neatly as this, the confidence to be had in the overall picture presented by both is greatly increased.

How the Hypercolumns Work

The evidence for the existence of hypercolumns distributed over the striate cortex is quite good, as the last section has made clear. The next obvious question is: how does each hypercolumn perform the job we have ascribed to it, namely, examining its own patch of retina (its hyperfield) and arriving at a feature description of the image falling on this patch? Here we must take the clues offered by neurophysiological findings as to how striate cells are selectively responsive to particular line stimuli, and marry them to advances in the area of image processing by computers. As far as the latter area is concerned, the treatment given here will be heavily dependent on the work of David Marr.

The first point to emphasise is that we cannot take each cell so far described in the hypercolumn as a straightforward neural code for a particular feature, namely the optimal stimulus for the cell. (If things *were* as simple as this, then activity in a cell whose optimal stimulus was, say, a vertical edge in a given retinal position would mean 'vertical edge present in such-and-such a position', and so forth.) I mentioned this before when I pointed out that the terms 'line detector', 'edge detector' and 'slit detector' were misleading in this regard. The reason why each cell cannot be regarded directly and simply as a feature detector is that the output of any given cell is ambiguous, just like the output of the corner detector discussed in chapter 2. For example, consider a simple cell whose optimal stimulus is a vertical edge. Such a cell would of course respond most strongly when a high-contrast black white edge falls in the appropriate position in its receptive field (**68**, top). If the contrast of the edge was reduced, by making the black zone a dark grey and the white zone a light grey, then the cell would still respond, but less strongly (**68**, middle). The response is lowered because less excitation and more inhibition is being fed to the cell than before. Now consider the cell's response to a high-contrast edge rotated a few degrees (say $\pm 10^{\circ}$ or so) from the optimal vertical orientation (**68**, bottom). The cell would respond just as well to this non-optimally orientated stimulus as it would to the optimally oriented one of lower contrast! If activity in this cell was to be taken simply and directly as the neural representation of a vertical edge, we would be susceptible to some very awkward illusions! We would confuse faint vertical edges with high-contrast just-off-vertical ones, a quite unsatisfactory state of affairs which fortunately does not arise.

The general answer to the ambiguity problem posed in **68** is to consider each cell's output not on its own but in the *context* of the activities of other cells in the hypercolumn. For example, suppose that mechanisms were present within the hypercolumn which were designed to take notice of the *most*

active neuron when assessing the orientation of an input feature. In this case, if the faint vertical edge of **68** was the input feature, the most active cell would still be found in a vertically tuned orientation column [**69**], even though this cell would not be firing as briskly as it would have been if the edge were a high-contrast vertical one. That is, a vertically tuned cell would 'come out on top', as it were, against all the opposition from other, even less active cells. Hence the hypothesised mechanism within the hypercolumn (a mechanism as yet quite unspecified, of course) would note that this cell was the most active one, and register the input feature as having a vertical orientation. Now when the high-contrast but just-off-vertical edge stimulus appears, the vertically-tuned neuron fires off at just the same rate as it did in response to the faint vertical edge (lower half of **69**). But in this case the adjacent cell fires off at an even greater rate because the just-off-vertical edge is at its optimal orientation. Consequently, the mechanism which detects the most active neuron would register the input feature not as a vertical one but at its true orientation of just-off-vertical. In this way, by taking advantage of the context in which any given cell's activity occurs, the hypercolumn need not be fooled by the intrinsically ambiguous responses of individual cells.

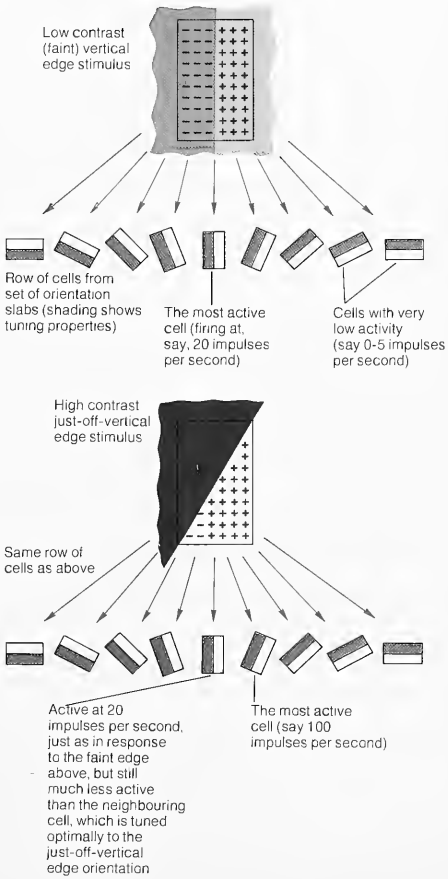
This orientation example illustrates very nicely the general approach of regarding simple cell responses as *measurements for interpretation*, and not as the brain's feature descriptions on their own. In fact, the need for interpretation is even greater than this example suggests. Consider for instance the fact noted earlier, that there seem to be only about 18–20 orientation slabs per hypercolumn, to cover all orientations in the relevant hyperfield. If cell responses were interpreted as suggested in **69**, and just the most active taken as signalling the orientation of the input feature, then we would be limited to discriminating between line stimuli as different in orientation only if their real orientations differed by about 10° or more. This, of course, is a gross underestimate of our perceptual capabilities in this regard: we can easily manage discriminations of around 2°. Clearly, there is a need for some method of *interpolation* between pairs of orientation measurements. That is, if one cell with an optimal orientation of vertical is firing at, say, 40 impulses per second, and its orientation neighbour, optimally tuned to 10° off vertical, is firing at, say, 35 impulses per second, then these two measurements need to be interpreted as coming from an input feature with an orientation somewhere in between the two optimal orientations – perhaps 4° or so off vertical in this case. This is a much more sophisticated approach to interpretation of simple cell measurements than that of simply finding the most active neuron and assuming its optimal orientation to be the actual orientation of the input feature.

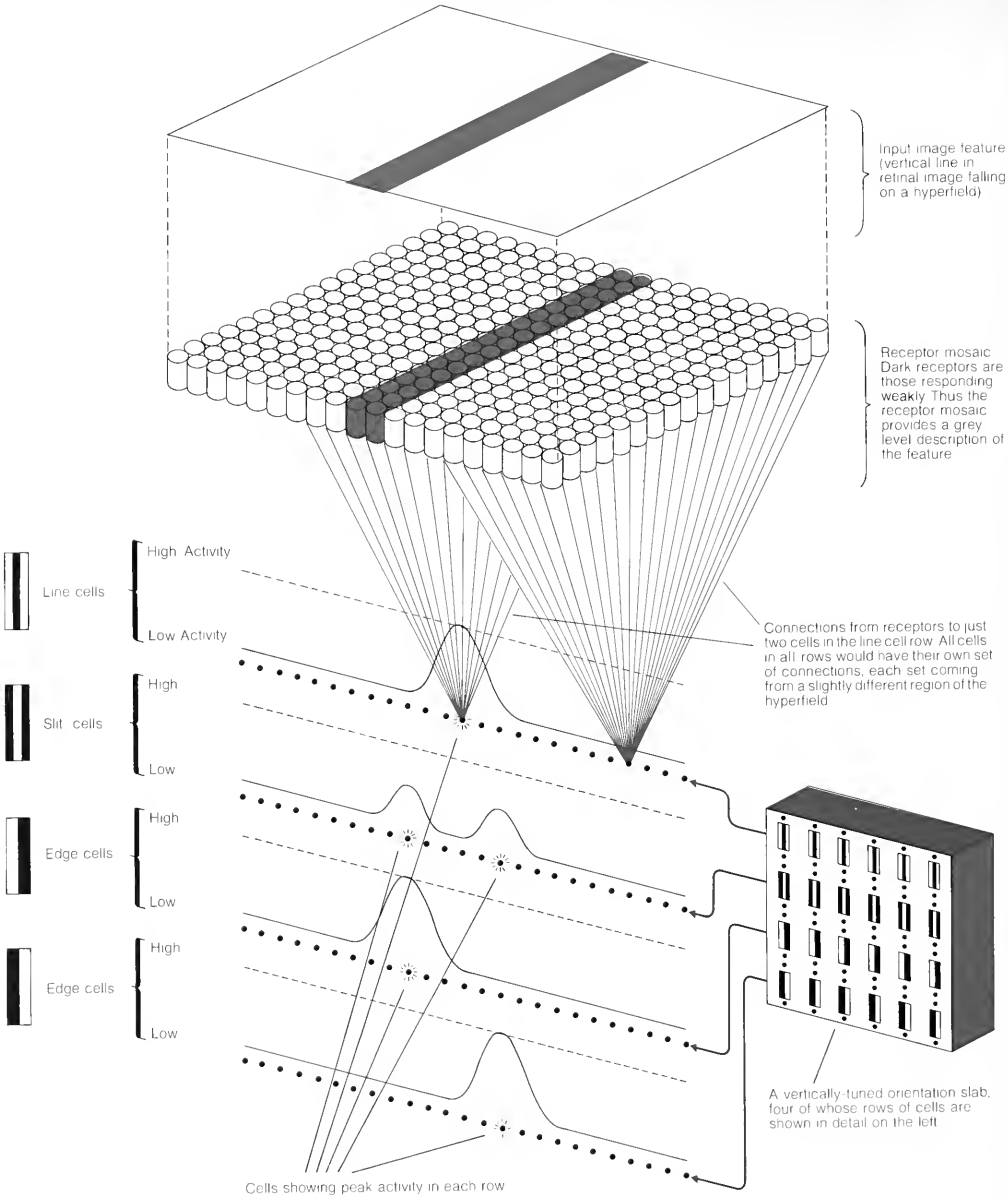
This demonstration of the need to interpolate between the orientation methods provided by the relatively few orientation slabs in each hypercolumn forcefully illustrates the need to break away entirely from the idea that simple cells are feature detectors, pure and simple. It is this realisation which lay behind my earlier remarks about the inappropriateness of the usual labels for these cells – 'slit detectors', 'line detectors', and so on. The conclusion, arrived at by Marr from his efforts to incorporate units like simple cells in a computerised feature-description machine, is that simple cells must be regarded simply as image-measuring devices which provide data about features of the input image, data which must be further interpreted before a proper feature description can be arrived at.

But if the need for interpretation of activity in simple cells is now clear, what is the code in which the outcomes of the interpretative process are written? What is the feature description lodged in the brain which is the neural correlate of our conscious awareness of features? If we see a 'sharp edge oriented at 23° clockwise from the vertical and of medium contrast', what is the neural representation of this feature description, this percept? We *see* the feature well enough; so what is its code in the head?

The extraordinary answer is that we have almost no idea! Despite all the tremendous advances made over the last twenty years or so in our knowledge of sensory neurophysiology, advances which have culminated in the hypercolumn concept and all the intricate neurophysiology and neuroanatomy described in this chapter (and much more besides, of course), we still have only tentative hypotheses, and rather

69 Interpreting simple cell responses in context





70 Activity profiles in one slab of a hypercolumn when dealing with a vertical line

poorly worked out ones at that, about what the feature code might be. A Nobel prize surely awaits the discoverer of this elusive language.

Possible candidates for the code seem to be fourfold. First, it might be that features are indeed coded as *active single cells*, with different cells representing different types of features, more or less as originally thought when the line detectors were dubbed just that, but with cells other than those so far described being the feature code. Second, it could be that a cluster of *many active cells* is the code for a feature, rather than just one cell. Third, features may be coded in terms of patterns of pulses in nerve fibres, the so-called *pulse coding* possibility. This idea goes beyond just asking whether a cell is or is not active and asks whether or not the pattern of pulses it emits over time varies in a way which could represent features. Fourth, it could be that the code is wholly *biochemical* and that features are represented not in terms of cell activity at the level of impulses, but at the much finer level of what biochemical constituents certain cells contain and in what states.

We will return later on to discussing these possibilities in more detail. But it is worth saying that we are at present not only unclear about the nature of the feature code, we are also uncertain about where it is lodged. The hypercolumn has been described throughout as the device which sends forth a message about what feature is presently falling on its hyperfield. Actually, this is a hypothesis which extends beyond present evidence. Certainly the mass of simple cells in the hypercolumn seem to provide the essential image measurements which are necessary for building up a feature description. But exactly where these measurements are interpreted is not yet known. Still, a location within the hypercolumn seems the present best bet. After all, each hypercolumn may have around a quarter of a million cells, and not all of them are simple ones. Moreover, the arrangement of the hypercolumn's components suits remarkably well the processes of interpretation of image measurements which seem to be necessary, judging from work in the computerised image-processing field (see next section). But caution is obviously desirable when talking about a structure like the brain whose workings still remain so mysterious. It could be, for example, that the hypercolumns take the business of interpretation only so far, leaving other brain sites to complete the work. Future findings on this as on so many other aspects of brain function should be fascinating.

Descriptions of Feature Types

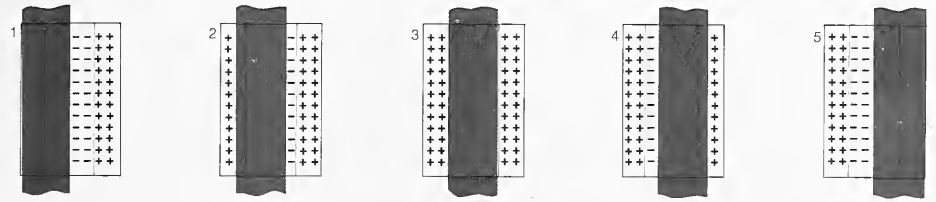
The previous section discussed how in principle the responses of simple cells could be used to assess the orientation of features. But what of feature type (slit, line, edge, etc.)? Is it a matter of equating slit detector responses with light lines on a dark ground, line detector responses with dark lines on a light ground, and edge detector responses with light/dark edges? Once again, Marr's work suggests that direct equations of this kind cannot be made, and that these basic responses are bound to be subject to a process of further interpretation.

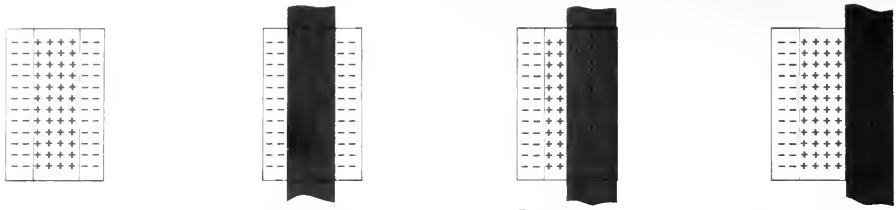
Consider **70**. It shows an image of a vertical line focused on a region of the retina's receptor mosaic which forms the hyperfield of a particular hypercolumn. The first thing to note is that the line feature is recorded as a grey level description in the receptor mosaic itself. Next, note the cables leading from the receptors to two cells in a vertically tuned slab of the hypercolumn. (Of course, these cables would not in reality be direct connections as shown; there are intervening stages both in the retina and in the lateral geniculate nuclei, but these are details which we can ignore for present purposes.) Cables (i.e. nerve fibres) are shown for two cells only, in just one row of cells in the slab. This is solely for simplicity: similar collections of cables must be imagined for all the cells in the slab, so that each individual cell can 'look' at its own region of the hyperfield, with its own type of receptive field organisation, slit, line or edge.

Note that each cell in each row 'looks' at a slightly different part of the grey level description in the receptor mosaic. That is, it receives input connections from slightly different regions of the hyperfield, as shown in **70** for just two cells in the upper row. Adjacent cells in each row would have overlapping cable connections, but each cell would have a receptive field centred on a different spot. The connections to the cells are either excitatory or inhibitory, and the arrangement of these two types of input determines the type of feature to which each cell is optimally sensitive – slit, line or edge.

Above each row of cells is shown an *activity profile*. The lines forming these profiles do not exist in the slab itself, of course. They are merely a graphic aid for showing the activity of each cell. Thus the height of the line above each cell represents how active that cell is. In the top row, for example, the most active cell is the one centred directly on the line feature. Cells to either side are less well activated, because their excitation-minus-inhibition counts are smaller – the line feature is not quite perfectly positioned for their optimal stimulus requirements. The fussiness which simple cells (i.e. all the

71 The effect of stimulus position on the response of a line-detecting simple cell





72 Responses of a slit detector to a dark line imaged on different locations of the receptive field

line slit edge detectors talked about so far in this chapter) show about the position of their optimal stimulus has not been emphasised so far. But as **71** shows, if a line falls even slightly to either side of the optimal position for a line detector, then that detector's output is reduced. The maximum excitation-minus-inhibition count is no longer maintained.

Different cell types inspecting a given feature in the hyperfield show peaks in their activity profiles at different points depending on the nature of their receptive fields. As already noted, the peak for the top row of cells in **70** is centred over the input line feature. Moreover, this is the only peak which occurs in this particular row for this particular stimulus. But consider the next row down, which shows the responses of slit cells across the hyperfield. Remember that here we have slit detectors, so called, responding to a dark line stimulus on a light ground. Even though there are no actual slits in the stimulus, some cells in certain positions respond quite well! Why this happens is illustrated in **72**, which shows that if the dark line happens to fall over an inhibitory flank of the slit detector, then this flank no longer receives light stimulation, and so does not send inhibition to the cell, with the result that excitation from the central zone goes uncancelled by inhibition from this flank. The other inhibitory flank still sends inhibition, and so the response of the cell is to that extent reduced from what it would be if the optimal stimulus – an appropriately positioned slit – was on the receptive field. But the point is that a quite respectable firing rate can be obtained from a slit detector, so called, even when a line stimulus is being used. Here is another form of the ambiguity problem: slit detectors can 'see' line stimuli if they are suitably positioned on the hyperfield. So much for 'slit detecting' cells being any good as slit detectors, if their outputs are considered on their own!

The reason two peaks occur in the slit detector activity profile is that one peak occurs where one inhibitory flank is covered, the other peak where the other inhibitory flank is covered.

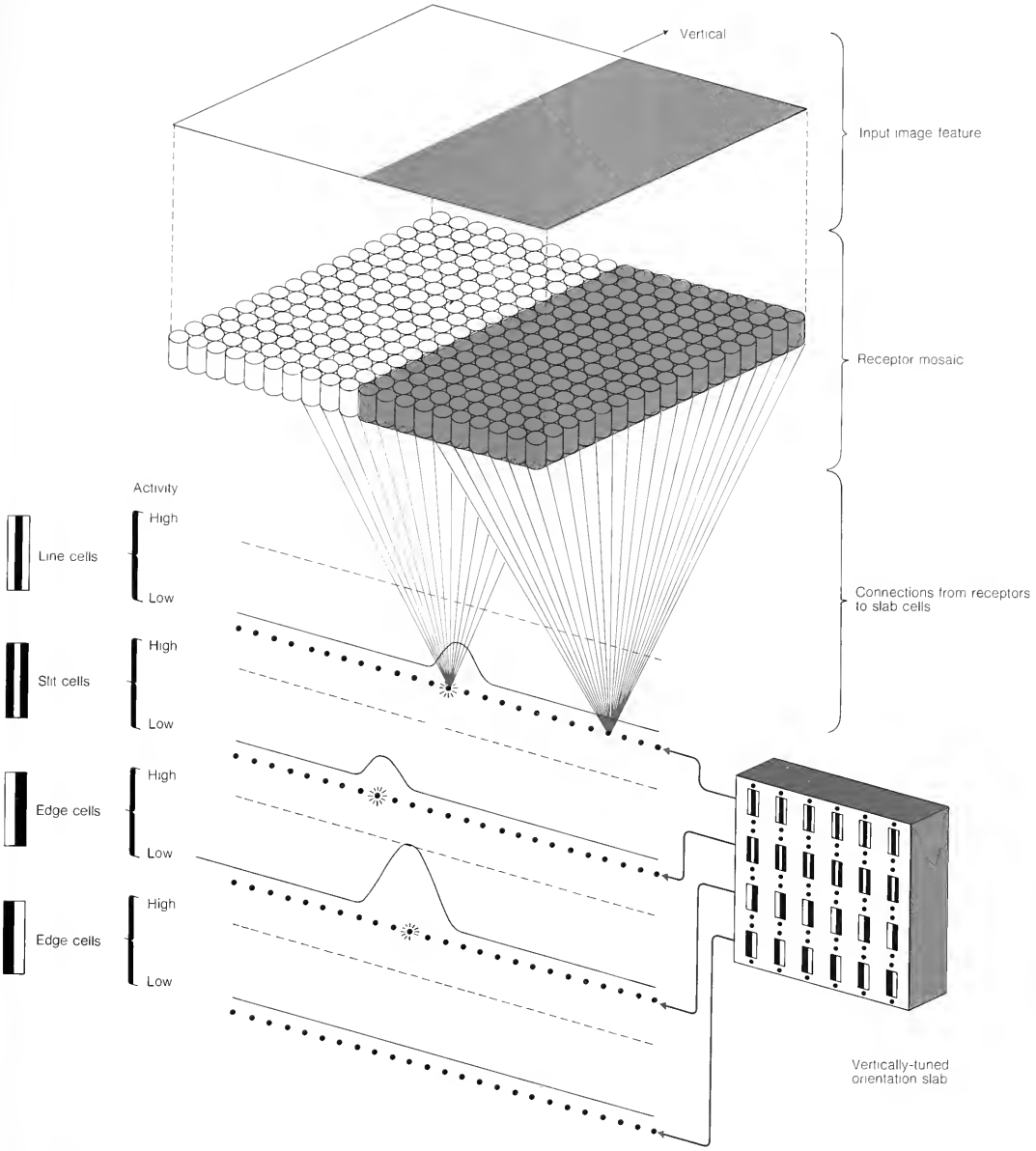
The edge detector profiles also show peaks of activity to one side of the hyperfield centre. Each peak occurs where an edge of the line stimulus happens to fall exactly on the excitation-inhibition boundary within the edge detecting receptive field. This happens, obviously, for cells whose fields fall to one side of the centre of the line stimulus.

The activity profiles shown in **70** are only approximate, but they give the correct impression, namely that a line feature will create activity in various parts of the vertically tuned slab shown. The line creates, if you like, a 'signature tune' of activity. It is the job of other mechanisms to recognise this signature tune as belonging to a particular type of feature, in this case to a dark-line-on-a-light-ground feature. This is another way of saying that the simple cell responses depicted in **70** must be *interpreted*, and not taken simply and directly as a feature code. So here we find, just as for feature orientation, that arriving at a description of feature type is not a straightforward business.

Different types of feature give different patterns of peaks in the activity profiles. Profiles for an edge are shown in **73**. If you like, you can follow through these profiles in detail for yourself as we have just done for the case of a line. But it is enough to recognise the basic point, namely that the pattern of response to an edge is different from the pattern for a line. It is this pattern which must be recognised in each case by the visual system if it is to decide correctly what feature type is present.

Just how the patterns of peak activity are recognised by brain mechanisms as belonging to a particular type of feature is a research question right at the forefront of present knowledge. It is interesting to note in passing that Marr uses the presence of peaks in the pair of edge-detector profiles shown in **70** as the hallmark of a line feature. It is rather ironic that this should prove convenient for his computer system: the 'edge detectors' are used as 'line detectors'! This underlines all that has been said so far about simple cells not being themselves a feature code, but only measurements taken on the way to obtaining one. But there are probably many ways of going about the task of recognising activity profile patterns, and which one is used by the brain is a moot point.

Figures **70** and **73** show activity profiles in a slab tuned to vertical. But remember that many other slabs exist, each tuned to its own optimal stimulus orientation. Each slab would have a pattern of activity within it which required interpretation, just as described for the vertical case. Note, however, that different slabs deal with different orientations around the clock in about 10° steps, so that different slabs will become active depending upon what orientation the input feature has. Only those slabs whose orientational tuning is close to the orientation of the feature being dealt with will become active at any one time. The final assessment of feature orientation must be arrived at by suitable interpolation



73 Activity profiles in one slab of a hypercolumn when dealing with a vertical edge

between the most active slabs, as explained earlier when the general question of feature orientation was being described (p. 53).

Reference back to **59** and **60** will remind you that each slab's cells are set to measure the image on the hyperfield in different locations spread out at right angles to the slab's preferred orientation. In point of fact, this attribute of hypercolumn organisation is pure hypothesis on my and John Mayhew's part, and has not yet been confirmed to our knowledge by neurophysiological evidence. It is a hypothesis which is sensible in that scans across the input image in this way are exactly what Marr uses in his computerised feature-description system. It turns out that this mode of processing considerably aids the business of interpreting activity patterns to determine what feature is present. It will be interesting to see if future neurophysiological findings confirm or falsify the speculation. If they confirm it, it will be an interesting and unusual case of advances in computerised image-processing guiding very specifically, and with profit, biological investigations of brain organisation.

A final point before leaving the question of descriptions of feature type. You may be wondering why on earth we should bother with all the complex machinery of the hypercolumn if the measurements of all its cells are so ambiguous that they need interpretation at every stage. The answer is that patterns of activity are set up in the hypercolumn *as a whole* which are unambiguously related to feature type. Characteristic peak profiles will occur for the different sorts of line feature, despite variation in feature contrast. Also, these patterns are fairly insensitive to input image degradation. So the hypercolumn is a solution to the problems discovered in chapter 2 when a more direct approach to feature description was tried and found sadly wanting. Of course, we have not yet reached the point of describing more complex features such as corners, or even blobs. These matters will be taken up later on.

Description of Feature Fuzziness

So far we have only considered line features which have sharp light dark boundaries. This is, of course, quite unrealistic. Many line features in most natural scenes will have more or less gradual transitions from light to dark. That is, they will be more or less *fuzzy*. A sharp edge can be regarded as of 'zero fuzziness'; a very blurred or extended edge is one with 'high fuzziness'. The problem is: how can one assess this important and obvious characteristic of features from the kind of image measurements provided by simple cells?

The general answer to this question is: inspect patterns of activity in cells with differently sized receptive fields and make interpretations accordingly. This approach is illustrated in **74**. At the top of this figure are shown two edges, a sharp one and an extended or fuzzy one. The luminance profile of each one (i.e. the intensity of light across each) is shown graphically right at the top, and it can readily be seen from these profiles that one edge has a rapid transition from light to dark, whereas the other proceeds along this path slowly. Underneath each type of edge are shown rows of cells taken from a vertically tuned orientation slab. The three rows of cells shown underneath each edge are the same in each case: on the left of the figure they are shown responding when the sharp edge falls on the hyperfield dealt with by the slab, on the right they are shown responding when the sharp edge is replaced with the extended one.

Note that the top row is made up of cells with small receptive fields, the middle row has medium-sized receptive fields, and the bottom row has large fields. As before, the cells within each row have their fields centred on different points in the hyperfield, with neighbouring cells having fields that overlap, but are also slightly shifted with respect to one another.

Consider responses to the sharp edge first. It can be seen that the peak of activity is of the same size in all three cell rows. This fact is a sign that a sharp edge is present, and so the mechanisms of interpretation can use it for registering a zero degree of fuzziness.

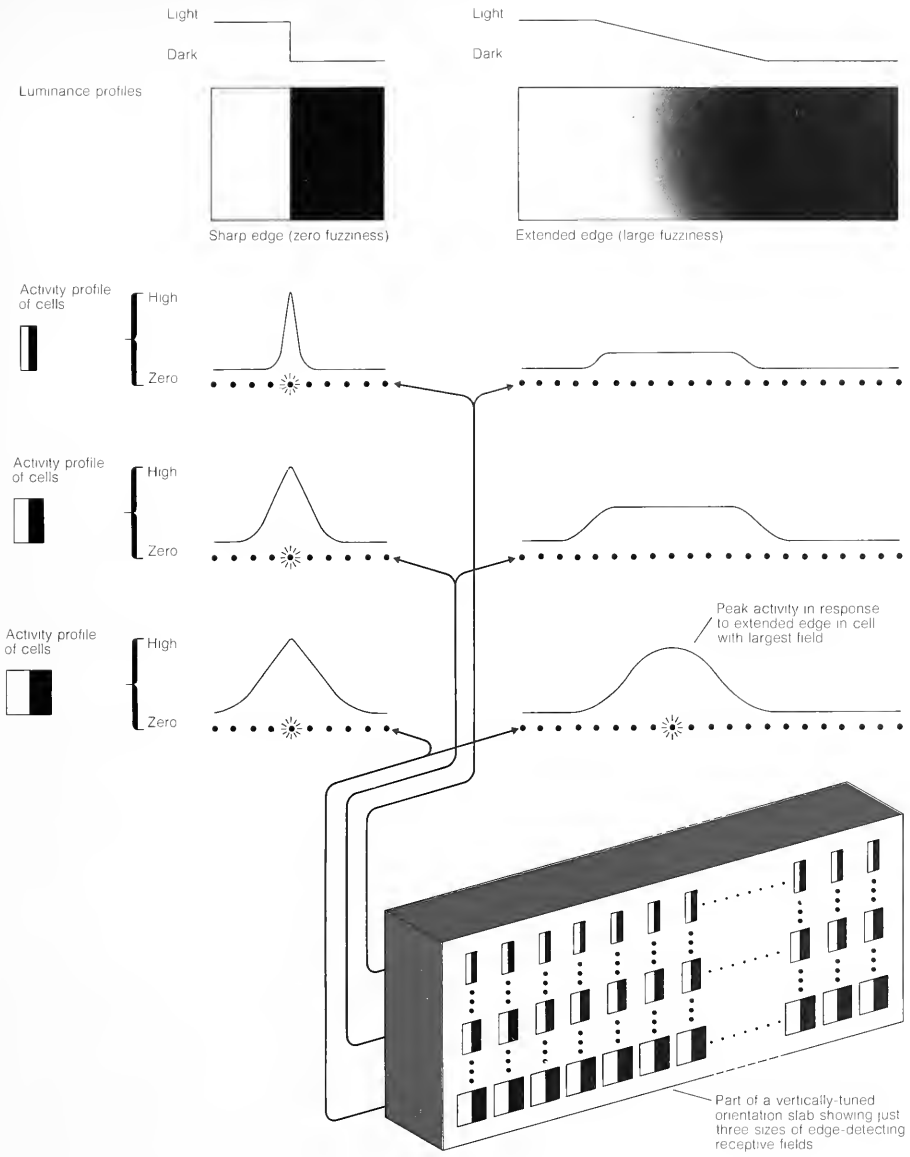
On the other hand, as soon as the edge becomes at all fuzzy the rows show peaks of different sizes. This is shown very clearly in **74** because the input image feature is a very extended edge. When receptive fields of different sizes produce peaks of different sizes, the rule is that the fuzziness of the edge is given by the smallest field size which attains the maximum response of any cell. In other words, if bigger fields do not produce a bigger response, then the rule is to accept the field size of the smallest field which produces the biggest response as a measure of fuzziness. This is the rule which Marr has found useful in interpreting cell responses in order to achieve estimates of fuzziness, but of course it might be that the brain does it rather differently. A description of other possibilities would be out of place here, but enough has been said to convey the general idea that the so-called edge detecting cells are really best thought of as devices for measuring light/dark luminance gradients.

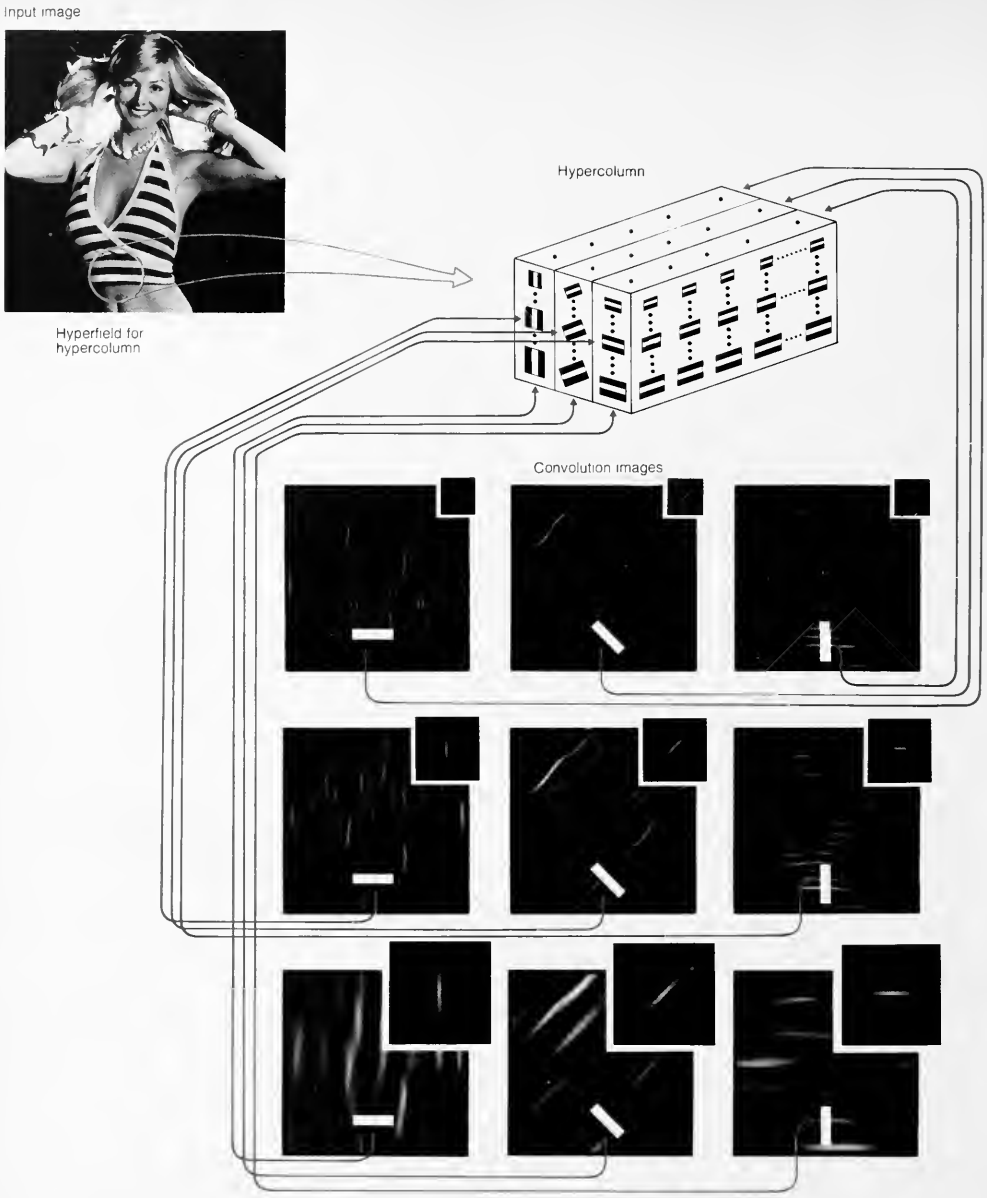
Figure **74** as usual presents a hypothetical and highly schematic account of how the hypercolumn might work while doing a particular job. As already noted, it must be realised that probably many more receptive field widths actually occur than are shown in **74**, and in addition one must imagine many field widths for all types of cells – slit, line and edge detectors. But enough has been said to give a fair idea of the advantage in having a variety of widths, and further details will not be presented here.

One important point to note about **74**, as well as many other figures in this book showing slab organisation, is that the neatly laid out rows of cells sharing a common field type and width are highly speculative. Such a principle of organisation has been indicated in the cat's striate cortex by neurophysiological data obtained by Lamberto Maffei and Adriana Fiorentini, but there have been no reports so far of studies showing this in the monkey. I have taken the liberty of displaying this layout in this book for simplicity. It might be that the activity profiles found so beneficial by Marr in deciding upon feature fuzziness could be arrived at without a neat and orderly layout of cells, but there do seem to be clear advantages in economy of 'wiring up' the striate cortex if similar cells are arranged adjacently. Once again, we must wait for future research to decide this point, as so many others.

Also required in addition to descriptions of feature type, orientation and fuzziness is a description of feature contrast (see chapter 4), but we will not consider this in detail here.

74 Edge detectors with fields of different sizes dealing with sharp and extended edges


Light
Dark
Luminance profiles
Sharp edge (zero fuzziness)
Light
Dark
Extended edge (large fuzziness)
Activity profile of cells
High
Zero
Activity profile of cells
High
Zero
Activity profile of cells
High
Zero
Peak activity in response to extended edge in cell with largest field
Part of a vertically-tuned orientation slab showing just three sizes of edge-detecting receptive fields



75 Hypercolumns and a whole image

Hypercolumns at Work on a Whole Image

The time has come to illustrate hypercolumns at work on a complete image, in order to round off our detailed consideration of how each individual hypercolumn is built and how it seems to operate.

Consider first the convolution images of **75**. These images show the outputs of line detectors of the kind drawn as insets in each image, when convolved with the input image above them. That is, the receptive field types shown in the insets have taken measurements everywhere over the input image, and an excitation-minus-inhibition count has been taken for every location. The size of each count is expressed in the convolution images as a density of grey. Thus each point in the convolution images shows how active a cell with the inset receptive field is when this field is inspecting the equivalent point in the input image. (If you have any trouble understanding this, go back to chapter 2, p. 37, for a reminder on the corner detection convolution of **46**. As in that figure, zero counts are represented by mid-grey and positive counts are shown with the mid-grey-to-white region of the tonal scale, with white representing the highest positive count possible. But here, for simplicity, negative counts – shown in **46** with the mid-grey-to-black region of the scale – have been omitted and all negative points portrayed simply with mid-grey. Thus the convolution images here, by showing only positive counts, give an idea of activity within cells with the receptive fields illustrated when dealing with the particular input image shown.)

How do the convolution images of **75** relate to the hypercolumn cells? The answer is that each hypercolumn would have inside it a pattern of activity corresponding to just one area of each convolution image. For example, the hypercolumn shown in **75** has a hyperfield centred low down in the input image. That is, this is the region of the input 'looked at' by this particular hypercolumn's cells. The hypercolumn shown is in fact a much abbreviated one, for simplicity. Only one half of it is shown (remember each hypercolumn has a left- and right-eye section); just three different orientations are illustrated; only line detecting cells are present, and just three widths of these. But this is a sufficiently rich hypercolumn for our present purposes. And the prime purpose is to point out that each row of cells in this hypercolumn would be showing an activity profile which was just a slice of the overall activity profiles given by each convolution. These slices are drawn in on the convolution images as small rectangles. Note that the slices are at right angles to the receptive field orientations. Think of the slices as in principle one point thick, so that each tiny point in each slice represents just one cell, with the activity of this cell being shown by how white-grey the point is.

Given that the particular hypercolumn illustrated in **75** 'sees' just the activity profiles shown as slices of the convolutions, it can now be understood from the other areas of the convolutions what other hypercolumns would be 'seeing'. Simply move mentally the slice drawn on each convolution to any location you care to choose (being sure always to maintain its orientation perpendicular to the relevant receptive field orientation). The result of doing this mental walk may surprise you. It is easy enough to understand, at least intuitively, that the girl's striped swimsuit would show up strongly in appropriately oriented cells belonging to hypercolumns dealing with the relevant input image areas. But it is not so easy to grasp why all the other activity should be present, such as the stripe of activity appearing in the middle right convolution round about where the girl's eyes are in the equivalent location in the input image. At least, it is not so easy to grasp if one comes to these simple cells with a 'line detector' understanding of their function. But if one recognises them to be simply measurement devices, which sample the input image so that features can be worked out from their outputs, then the surprise is diminished. After all, the initial measurements must be sufficiently rich to allow all sorts of features to be interpreted, both large and small, fuzzy and sharp, etc. (In the example mentioned, the shape of the girl's eyes is such that the receptive fields of the appropriate 'line detectors' receive much more excitation than inhibition, thus giving a light output in the corresponding convolution image.)

The convolution images of **75**, then, give us a good idea of the activity profiles set up all over the striate cortex within the mass of hypercolumns. Remember that adjacent hypercolumns will have overlapping hyperfields (i.e. that the slices shown on the convolution images will in fact overlap with other slices, so that there is a certain duplication within the striate cortex of activity profile data). At least, this is how we currently think hyperfields are arranged, but the question of degree of overlap is a research question at present, about which one cannot be definite just yet. In any event, **75** shows the kind of data obtained by the hypercolumns, data which are then interpreted to arrive at a feature description for each sub-region of the input.

Hypercolumns Summarised

By now you should have an idea of the hypercolumn as a processing sub-unit, everywhere replicated over the striate surface so that each part of the input image can be analysed. Each hypercolumn has within it a set of simple cells (as well as others, soon to be described), and these cells provide a large number of measurements of the feature of the input image which is cast upon the hypercolumn's hyperfield. From these measurements, various feature descriptions are built up, and by the time these are complete, the type, orientation, fuzziness and contrast of the feature on the hyperfield is known. The interpretation of these measurements is made much easier by the neat layout of simple cells in the hypercolumn, which means that peaks of activity can readily be identified, that cells dealing with the same orientation are arranged next to each other, and so on. It seems highly probable, but by no means certain, that the processes of interpretation take place within each hypercolumn, so that the striate cortex as a whole sends to other brain sites statements about what features are present in each sub-region of the field of view. The code in which these feature statements are written is not yet known. Other processes must then exist which receive the feature statements arrived at by the hypercolumns, and group them together into descriptions of larger perceptual entities, such as long lines or edges, straddling many hyperfields; and these larger entities must in turn be grouped into still larger ones, until the materials are to hand for recognising objects (see chapter 5). After all, we are aware of much more than tiny local details of the kind dealt with by hypercolumns considered individually. There must be, as it were, further stages of processing which simultaneously scan all the evidence represented in all the convolution images of **75**, and 'read off' from this mass of data a full description of the whole image – a description, presumably, that surfaces in our consciousness as our awareness of the scene being viewed.

Complex Cells

So far we have described only one type of cell found in the striate cortex, the simple cell. The hallmark of the simple cell is that its receptive field can be plotted into excitatory and inhibitory sub-regions, the effects of which add up comparatively straightforwardly. This means that the simple cell is very demanding about the position in which a stimulus falls on its receptive field, as **71** and **72** made clear. But other types of cell without this property are to be found in the hypercolumn, and it is to their description that we now turn.

The distinguishing feature of a *complex cell* is that it responds equally well to its optimal stimulus wherever this stimulus falls in its receptive field. The optimal stimulus must still be either a line, a slit or an edge, just as for the simple cells, and it must still suit the orientation tuning of the complex cell in question (as for simple cells, so complex cells are found with different optimal orientations, and they are found in the same orientation slabs as the similarly preferential simple cells). But it matters little where the optimal stimulus falls in the receptive field of a complex cell, as **76** makes clear. In that figure, a moving line is shown sweeping across the receptive fields of both a complex and a simple cell, to illustrate the difference between them. The moving line creates a very vigorous response in the complex cell as soon as it appears on the field, and this response is maintained until it leaves the field, having moved right over it. The simple cell, on the other hand, responds only when the moving line happens to 'fit' the excitatory and inhibitory regions, and so its response is crucially dependent on the exact stimulus position. Note that excitatory and inhibitory regions are not shown in **76** for the complex cell because its field cannot be plotted out in this way: hence its label 'complex'.

The fact that complex cells are orientationally tuned, just like simple cells, is illustrated in **77**, which shows how a complex cell tuned to horizontal fails to respond if the stimulus is vertical.

Other typical characteristics of complex cells are that they tend to have larger fields than simple cells, they respond weakly if at all to stationary stimuli illuminated continuously (briefly flashed or, even better, moving stimuli are what really excite them), and they often respond to one direction of movement only (e.g. in **76**, the complex cell might respond well to movement from left to right but not from right to left). These latter two properties have led to suggestions that complex cells are involved in movement perception. This proposal, while interesting, must be regarded with caution at present. Much of this chapter has been devoted to explaining that it is dangerous to equate 'optimal stimulus type' with 'stimulus type described by the cell'. This consideration applies as much to complex cells as it does to simple ones. Unfortunately, whereas in the case of simple cells a body of image-processing theory has been built up which can give at least a plausible role to simple cells, this does not seem to exist for complex cells. At present we have very little idea about their contribution to the hypercolumn's computations.

One final point to note about complex cells is that they may be fed by retinal cells quite different from those that feed simple cells. Certain retinal cells, termed *Y cells*, have large cell bodies and axons, transmit messages relatively rapidly (about 40 metres per second), have large receptive fields, and tend to be located in the periphery of the retina. They tend to be sensitive to movement and respond primarily to changes of stimulation. Other retinal cells, termed *X cells*, have small cell bodies and axons, send relatively slow signals up to the brain (about 20 metres per second), have small receptive fields, and are more common in the central region of retina. It is thought that Y cells might feed complex cells and tell the brain about changes in the visual field, whereas X cells might feed simple cells and contribute to the latter's detailed analysis of stationary stimuli. We will have more to say about retinal mechanisms in chapter 6.

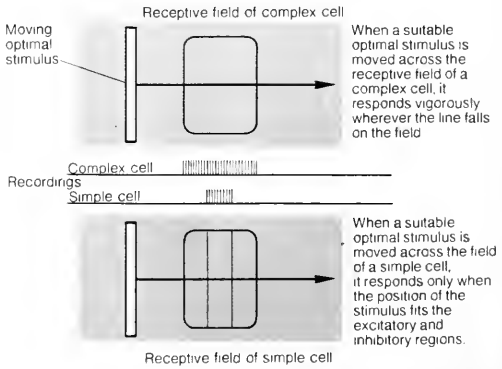
It might be of interest to some readers who learnt the Hubel-and-Wiesel story as students some years ago to note that it is now no longer generally believed that simple cells provide the input to complex cells. Simple cells might make some contribution in this regard, but various lines of evidence have led to a rejection of Hubel and Wiesel's early suggestion that the responses of complex cells could be due to each complex cell being fed by a cluster of simple cells, with each simple cell having its receptive field centred on a slightly different retinal position. This arrangement could, in principle, explain how it is that the complex cell can respond to a line stimulus over a wide region of retina. The truth, however, seems to be otherwise, although the details remain obscure.

Hypercomplex Cells

As their name suggests, hypercomplex cells are even more complicated than complex ones. They are similar to complex cells in having receptive fields which cannot be mapped into excitatory and inhibitory sub-regions, and similar also in preferring moving stimuli. But hypercomplex cells are distinguished by their added selectivity for stimulus *length*. That is, they respond poorly if the line stimulus, be it slit, line or edge, is made too long at one end or both. Thus the best stimulus for them is either a bar of defined length, or a corner, as in **78**.

Again, we have little idea at present about what hypercomplex cells are for. One tempting possibility is to speculate that activity in the hypercomplex cell of **78** is the feature code within the hypercolumn for 'corner present'. In short, is an active hypercomplex cell the brain's symbol for a corner feature, just as switch closures were corner symbols in the appropriate arrays of **33**? We do not know, but it is a hypothesis which is currently entertained seriously by many brain

76 Complex and simple cells compared



scientists. What the proposal amounts to is that each hypercolumn is equipped with a full set of feature cells and each set is equivalent to a sub-region of one array of **33**. Certainly there are many different cells within each hypercolumn (perhaps as many as a quarter of a million), so this possibility cannot be immediately ruled out on the grounds of shortage of components. But other possibilities exist for the feature code, as mentioned earlier on p. 55.

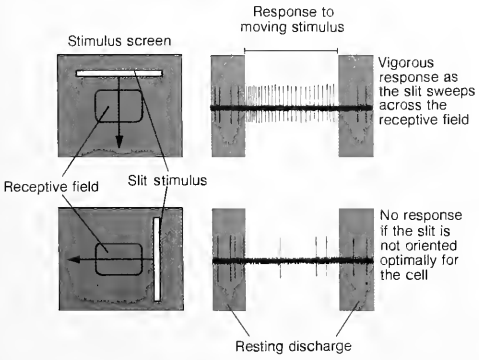
Perhaps the most likely possibility, however, is that the striate cortex is not really the site of a fully elaborated feature description at the level of corners etc. at all. Perhaps its output is just a set of local symbols (i.e. symbols for features in small regions of the visual field) – a rather primitive level of analysis. These symbols could concern the feature type which is discovered to lie in each particular hyperfield, and its orientation, fuzziness and contrast. These symbols might take the form of cell activity, a pulse code, and/or a biochemical change, but whatever they are they might not be the symbols which underlie our conscious awareness of features at all. Further stages of interpretation might well be required. After all, each hypercolumn is only concerned, it seems, with a relatively small patch of retina, and yet the features we perceive are often much larger entities which straddle many hyperfields. Thus it may well be that the hypercolumns are really only a very early stage on the way to a proper, full-fledged feature description.

Evidence from the Effects of Brain Damage

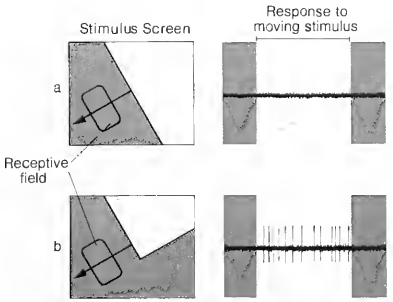
What happens to humans when they are deprived of their striate cortex, either by injury or by disease? Recent studies by Larry Weiskrantz, Elizabeth Warrington and their colleagues suggest that these patients sometimes have a visual sense which these workers call 'blindsight'. That is, when presented with visual stimuli, the patients deny seeing them. But when asked to guess where the stimulus might be, they often accurately locate the stimulus while all the time protesting that they cannot see it! This blindsight might be mediated by other brain mechanisms concerned with visual analysis, such as those in the superior colliculus (p. 40), mechanisms which might remain reasonably operational despite the damage to the striate cortex.

This conception of the different roles of the superior colliculus and striate cortex receives support from many animal experiments using the *lesion technique*. Here, precisely localised brain damage is created in selected sites and the consequences in behaviour are observed. The general picture which is emerging is that the striate cortex seems specially adapted for feature analysis whereas the superior colliculus mediates the detection of novel events in the periphery of the visual scene, perhaps with the associated function of controlling eye movements. Thus the superior colliculus might ensure that novel events are looked at directly and thus examined carefully with the full feature-recognising capabilities of striate mechanisms.

77 Orientational tuning of a complex cell



78 A hypercomplex cell This cell does not respond when a long white edge moves across its receptive field (a), but does respond when the edge is shortened to form a corner (b).



This cell does not respond when a long white edge moves across its receptive field (a), but does respond when the edge is shortened to form a corner (b).

Some Complicating Considerations

This chapter has given just a bare outline of the visual machinery of the brain. On the one hand, it must be remembered that there exist in the striate cortex and superior colliculus many variations on the basic cell types described, and that in all probability other cell types wait to be discovered. And on the other hand, the striate cortex and the superior colliculus are by no means the only two brain sites to be concerned with vision.

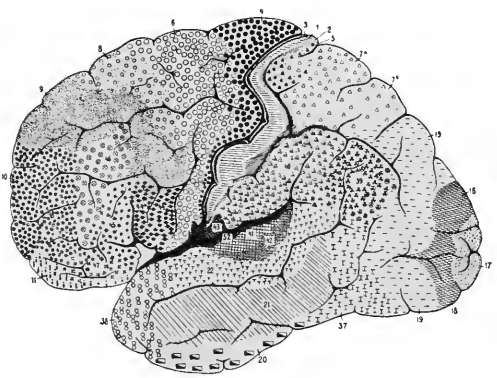
The last point is illustrated by Brodmann's map of the cortex **[79]**, which dates from early this century. Brodmann studied sections of the cortex microscopically and divided it up into different regions according to the size and shape of the nerve cells and fibre bundles. The tiny section at the rear, Area 17, is the striate cortex. Forward of Area 17 are relatively vast areas of cortex also given over to visual analysis. For example, as Semir Zeki has recently shown, Areas 18 and 19 ('prestriate cortex') contain several further maps of the visual scene. Each one of these is similar to the one found in Area 17 **[51]** in that it is laid out in an orderly fashion over the surface of the cortex, but each one differs in the details of its layout and in the function it seems to serve. The exact number of these extra mappings is still doubtful, but maps specialising in the analysis of feature movement, feature colour and feature depth seem to exist. The study of these extra mappings of the visual scene is still in its infancy and I will not attempt to describe the work here (although the binocular perception of the third

dimension will be considered at some length in chapter 7, together with its possible neurophysiological basis in striate and prestriate mechanisms). The essential point to grasp is that whereas mechanisms in Area 17 seem to be devoted to the analysis of feature shape, other feature attributes seem to be analysed elsewhere. So our present understanding of the visual machinery of the brain is that it has many sub-parts each specialising in a particular job. It is almost as though the sense of sight were not a single sense at all but really a collection of many separate senses.

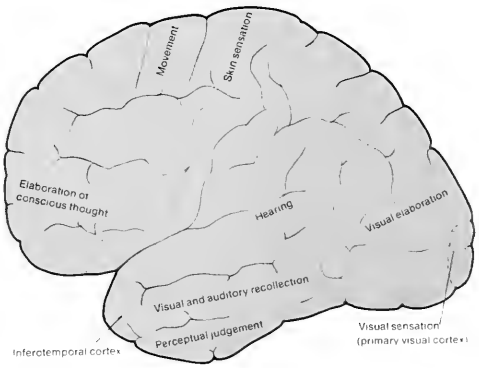
And yet the visual world of which we are conscious is obviously a fully integrated one. Although we can explicitly describe the various qualities attached to each feature in this world, its shape, colour, distance, whether it is moving or stationary, etc., we are at the same time aware of each feature as a single entity, and indeed of sets of features as forming the constituents of larger perceptual wholes, such as objects. Somehow the results of the separate analyses must be brought together, and each feature description for one property related to that for another. But we have almost no idea of how this is done in the brain.

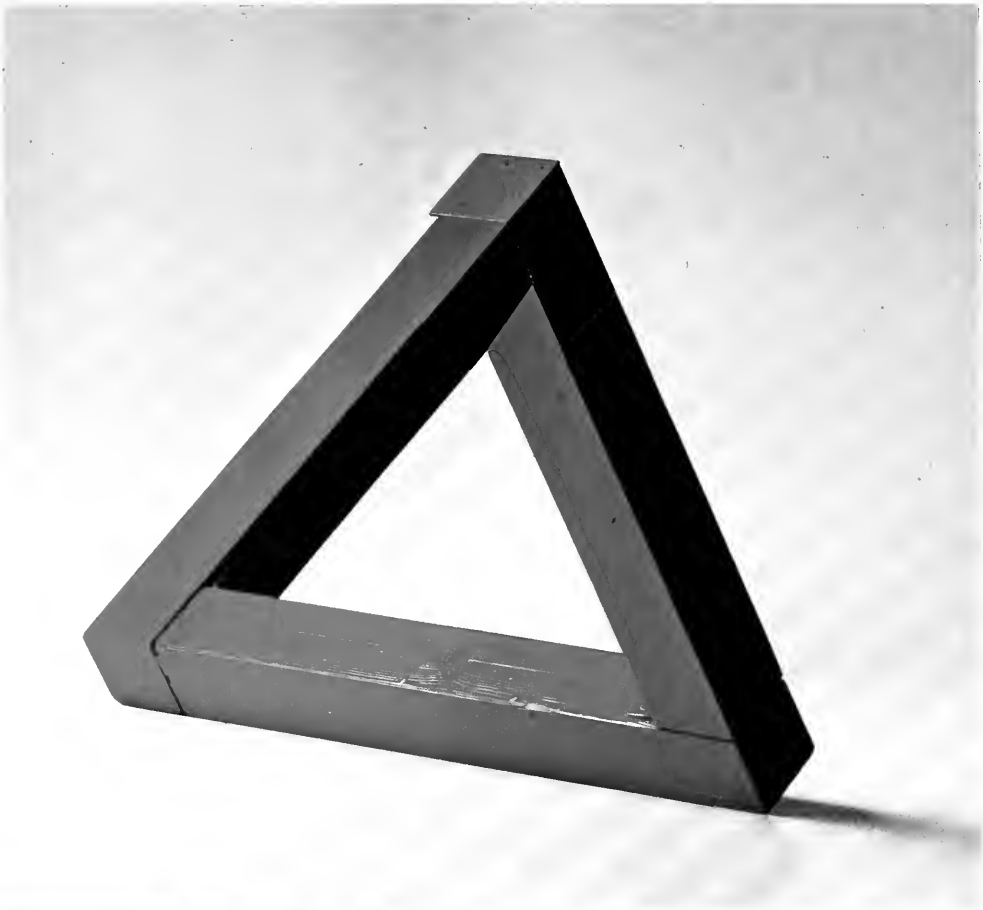
But if we do not yet know how the brain achieves this remarkable feat, we do have some notion of where it does it. Yet another brain map is shown in **80**, this time one based on the work of Wilder Penfield. Penfield drew up his map by observing the results of gentle electrical stimulation of the surface of the cortex, exposed during brain operations on human patients. The surgery was performed under local anaesthetic and so the patient was awake to report on conscious sensations created by the stimulation. Electrical 'tickling' of the striate cortex made the patient see swirling coloured shapes. These shapes were perceived as 'out there', somewhere in front of the patient's eyes, and not lodged inside his skull. Stimulation further forward produced much more complex visual sensations. Indeed, in the inferotemporal cortex (see **80**) stimulation sometimes caused patients to see whole scenes, complete with details of recognisable objects. So it seems possible that sites in the inferotemporal lobe might be those concerned with the 'final' integration of the relatively primitive feature analyses conducted in Areas 17, 18 and 19. But this speculation awaits confirmation. What can be said with some confidence is that a large proportion of brain tissue is given over to seeing, which should not perhaps surprise us, given the dominant role which this sense plays in our lives. What *should* astonish us is the fact that the brain manages to mediate seeing at all! We take its scene description, the visual world of which we are conscious, so much for granted: but how the brain produces it is deeply mysterious.

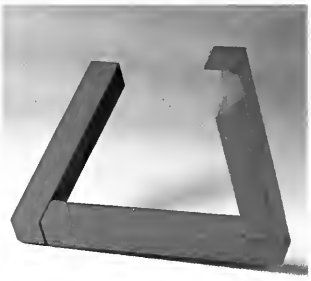
79 Brodmann's map of the cortex Each numbered region differs in the size and type of its nerve cells. Area 17 is the striate cortex. Areas 18 and 19 are prestriate cortex



80 Map of cortical functions based on Penfield's stimulation experiments

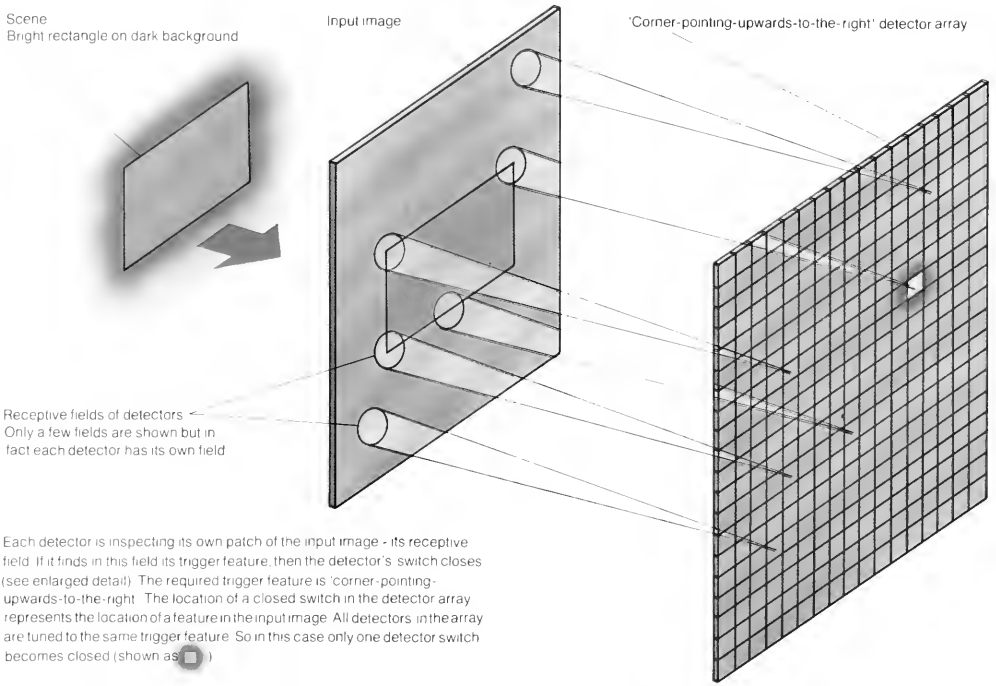




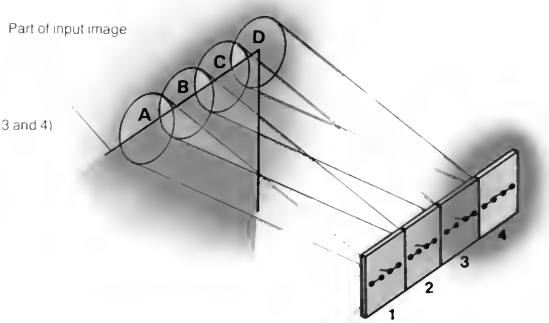


24 [p. 21] **Impossible triangle**: three-dimensional version due to Richard Gregory (see text)

28 [p. 21] If the visual system was determined to make the impossible triangle of **24** sensible globally, it would have 'broken up' one corner and made us see two components at different distances, i.e. the object that is really there. The figures illustrate this eminently possible outcome by showing the object viewed slightly from one side so that the potential perceptual break is made clear. In fact, of course, the visual system does not allow us to see this break.



Each detector is inspecting its own patch of the input image - its receptive field. If it finds in this field its trigger feature, then the detector's switch closes (see enlarged detail). The required trigger feature is 'corner-pointing-upwards-to-the-right'. The location of a closed switch in the detector array represents the location of a feature in the input image. All detectors in the array are tuned to the same trigger feature. So in this case only one detector switch becomes closed (shown as ■).

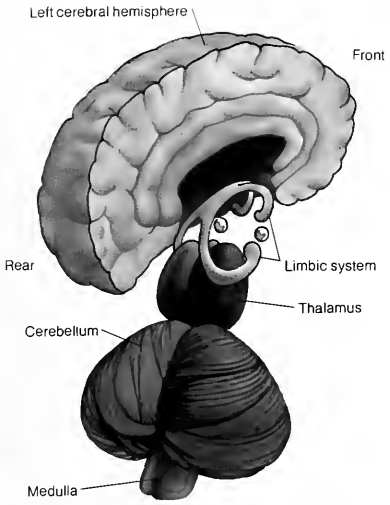


Enlarged detail, showing how adjacent detectors in the array (1, 2, 3 and 4) have overlapping receptive fields (A, B, C and D)

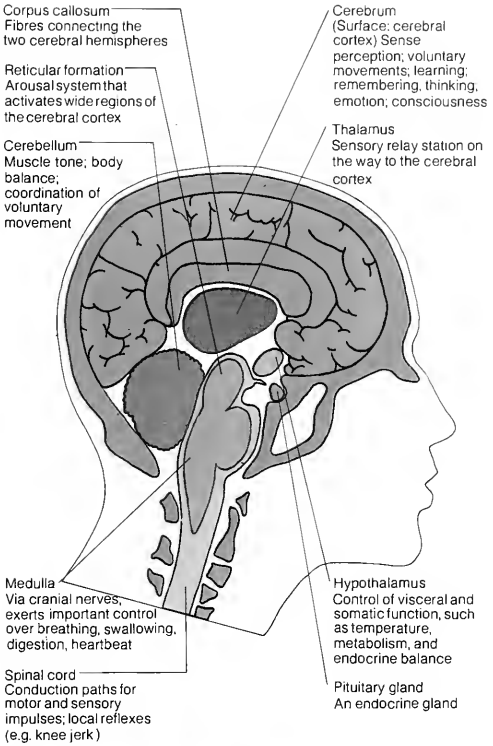
Part of detector array

Each detector is a switch. The closed switch (4) is symbolising 'corner present' and all the other switches (1, 2 and 3) remain open, as they do not find a corner feature in their receptive fields, only an edge

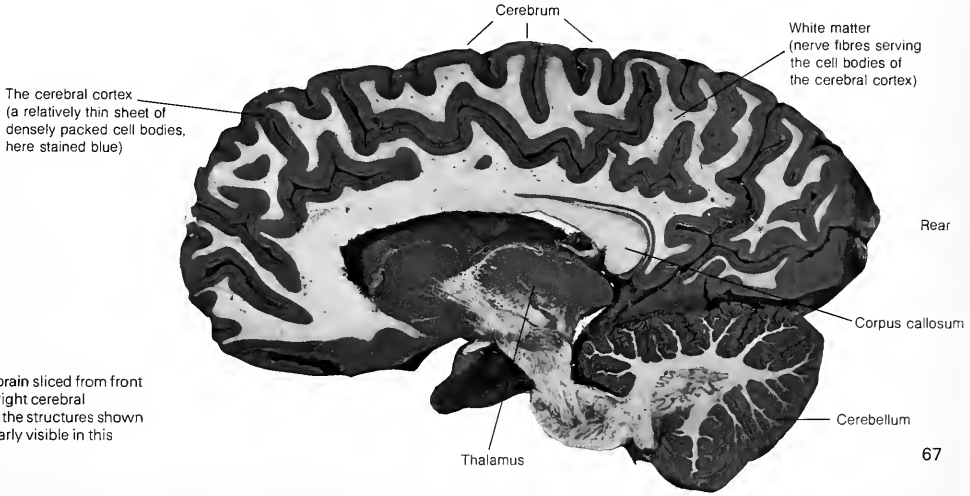
32 [p. 26] A feature detector array



47 [p. 39] **Exploded view of principal brain structures** The limbic system is concerned with emotions: see **48** for the functions of the other labelled parts.

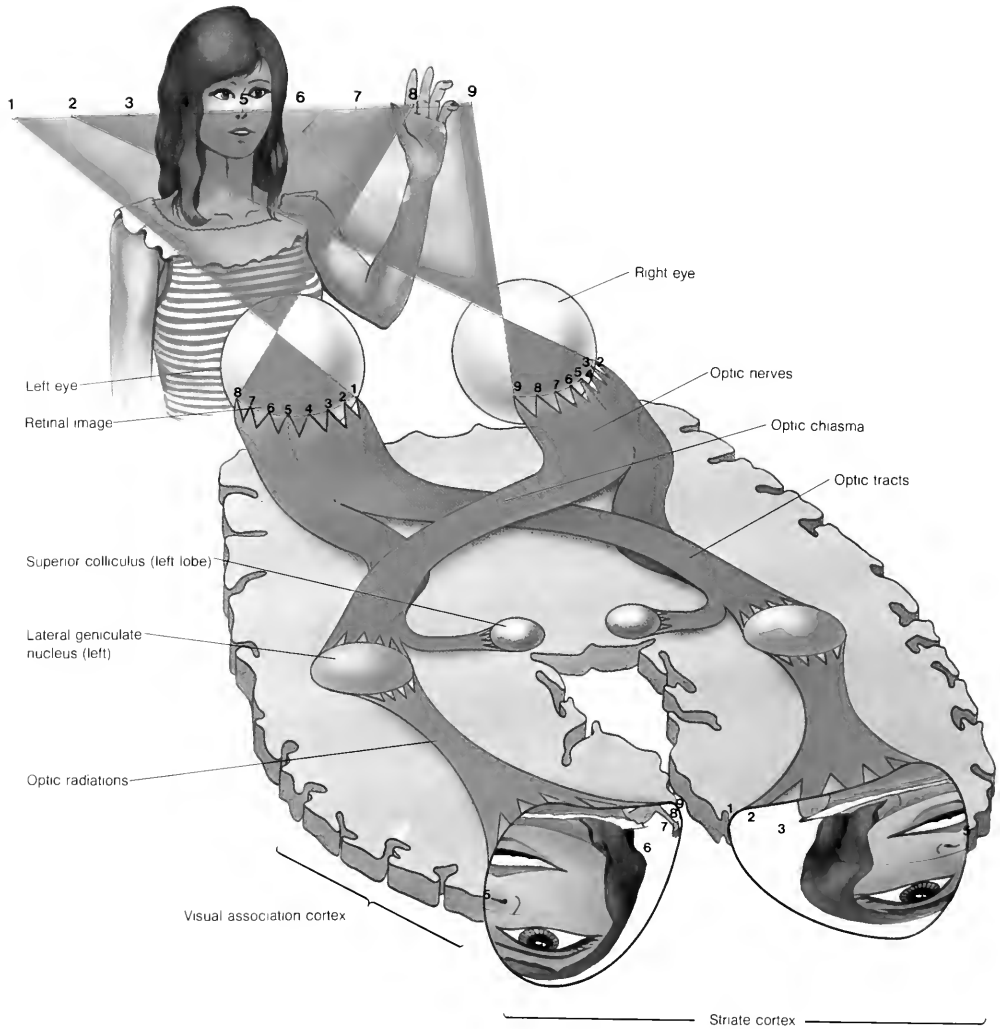


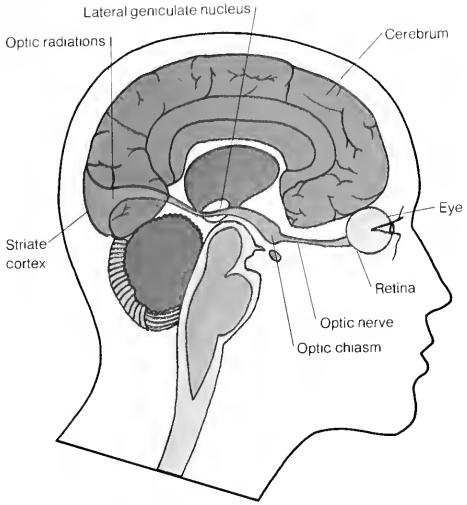
48 [p. 39] **Diagrammatic cross-section of brain**, showing its major parts, and listing their functions



49 [p. 39] **Human brain** sliced from front to rear through the right cerebral hemisphere. Not all the structures shown in **47** and **48** are clearly visible in this particular section.

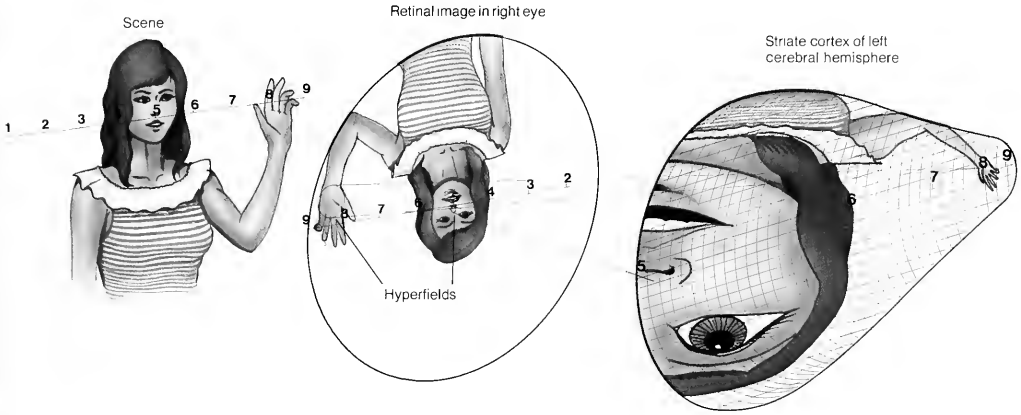
51 [p 39] Schematic illustration of two important visual pathways, one from eyes to striate cortex and one from eyes to superior colliculus. The messages in the first pathway begin in the light-sensitive retina of each eye, travel from each eye in an optic nerve, pass through structures called the optic chiasma and the lateral geniculate nuclei, proceed on their way via the optic radiations, and finally arrive in a region of the cerebrum right at the back of the head called the striate cortex





50 [p. 39] **Diagrammatic section through head** showing principal features of the major visual pathway that links the eyes to the cerebrum

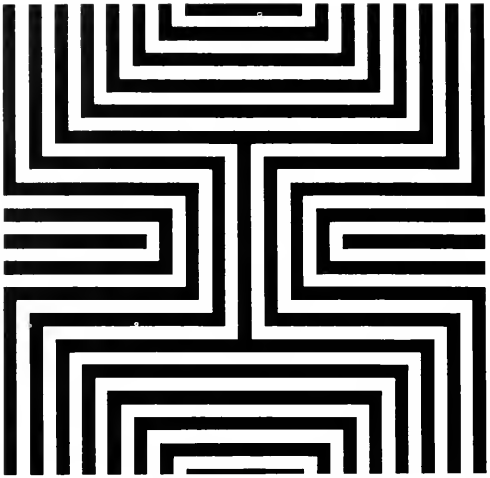
The spatial distortion shown is only an approximation to that really existing. Turn the book upside down for a better appreciation of the layout of the scene in the cortex.



Each 'square' represents one hypercolumn - a block of cortex which performs the job of analysing one section of the retinal images, called here its hyperfield.

52 [p. 40] **Mapping of the retinal image in the striate cortex** Turn the book upside down for a better appreciation of the layout of the scene in the cortex.

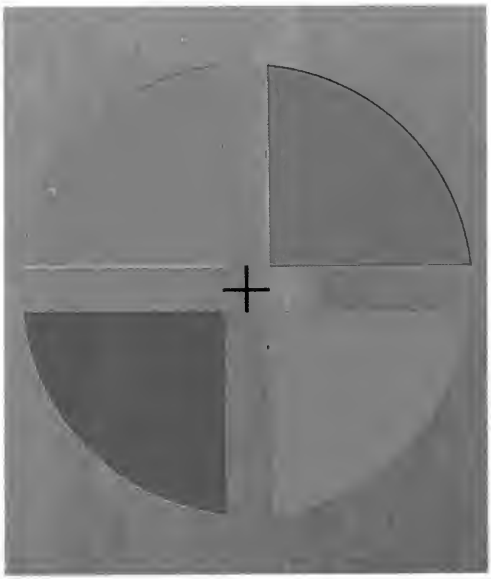
99 [p. 102] The McCollough effect



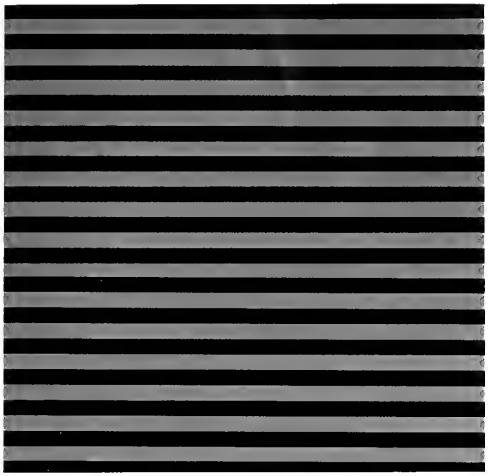
99a Test stimulus

98 [p. 102] Look briefly at the cross in 98a and note that it is set in a uniform grey field. Next, spend 45 seconds or so adapting to the coloured patches in 98b. Be sure to fixate the central cross steadily throughout this period. When the adaptation period is up, quickly transfer your gaze back to the cross in 98a. You will now see, superimposed on the grey field, illusory coloured patches **complementary** in hue to those of 98b. Thus the red patch gives a green after-image, and the green patch a red one. Equally, the blue patch gives a yellow after-image and the yellow patch a blue one. Complementary colours are those which when mixed in appropriate proportions using coloured light sources (not paints) produce the perception of grey. Strictly speaking, the complementary hue for green is a reddish-purple, for a red a blue-green, for yellow a violet-blue, and for blue an orange-yellow. The careful observer with normal colour vision might note these strictly-defined complementary hues as those of his after-images from 98b.

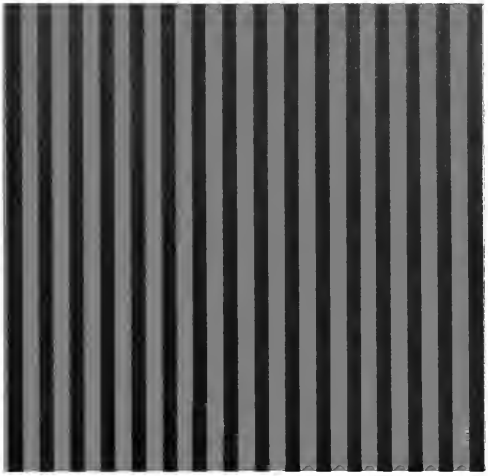
98a Plain grey test field



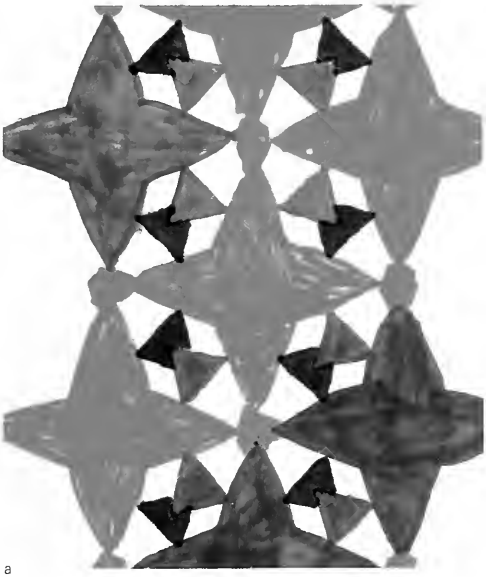
98b Adapting stimulus for negative coloured after-images



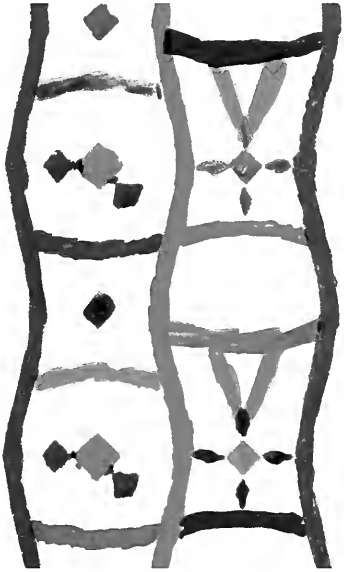
99b Green adapting stimulus



99c Red adapting stimulus



a

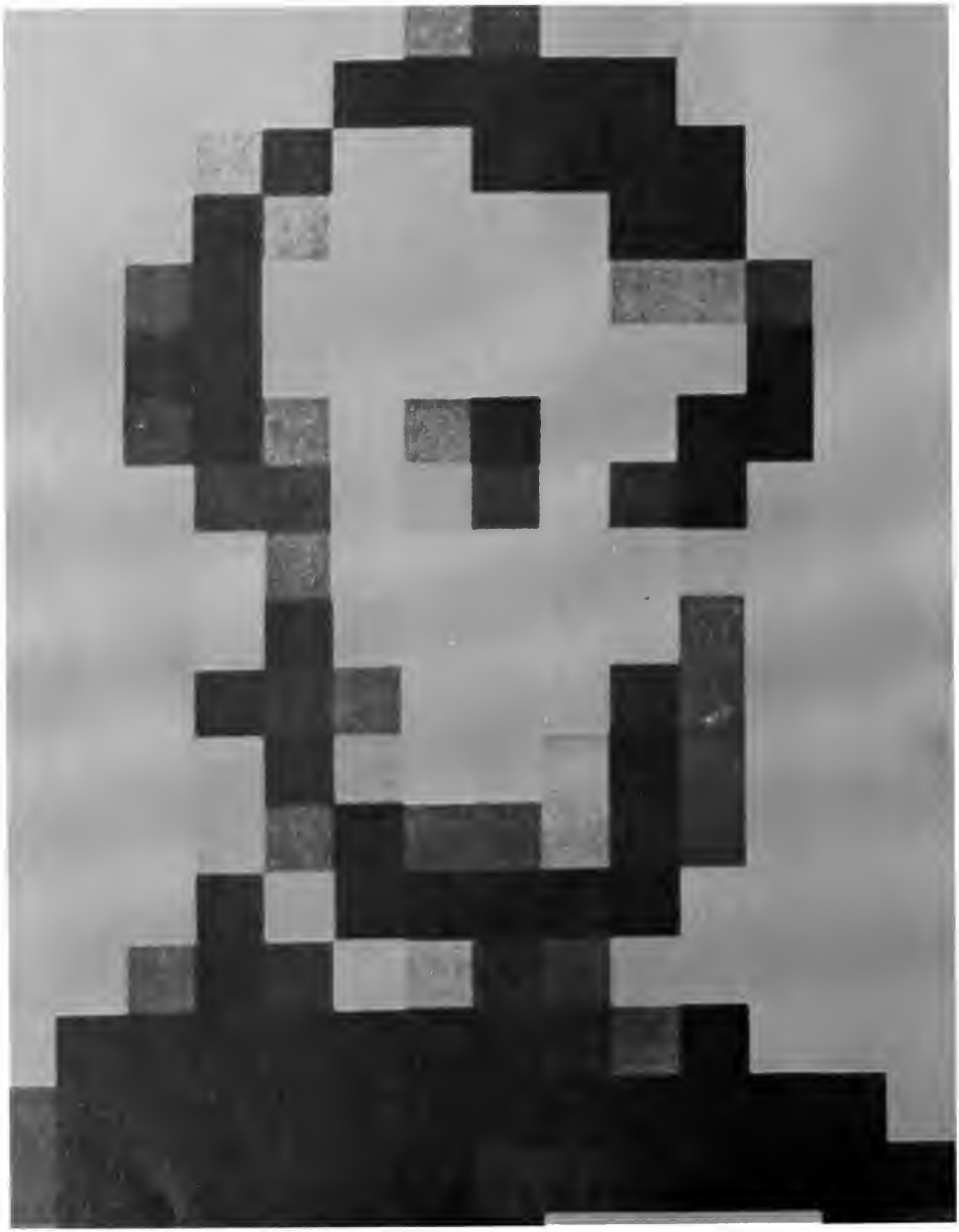


b

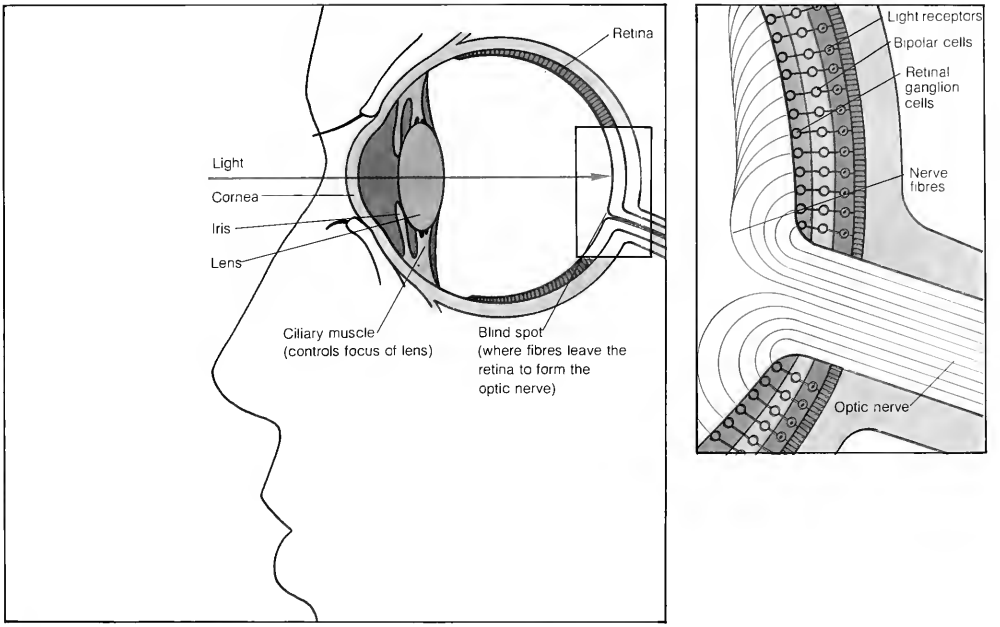
117 [p. 114] **Figure-ground**: my two young daughters were given the same Altair design to doodle with and saw radically different figure-ground structures, as their choice of filled-in areas demonstrates.

(a) [left] Louise's version
(b) [right] Abigail's version

Plate 8

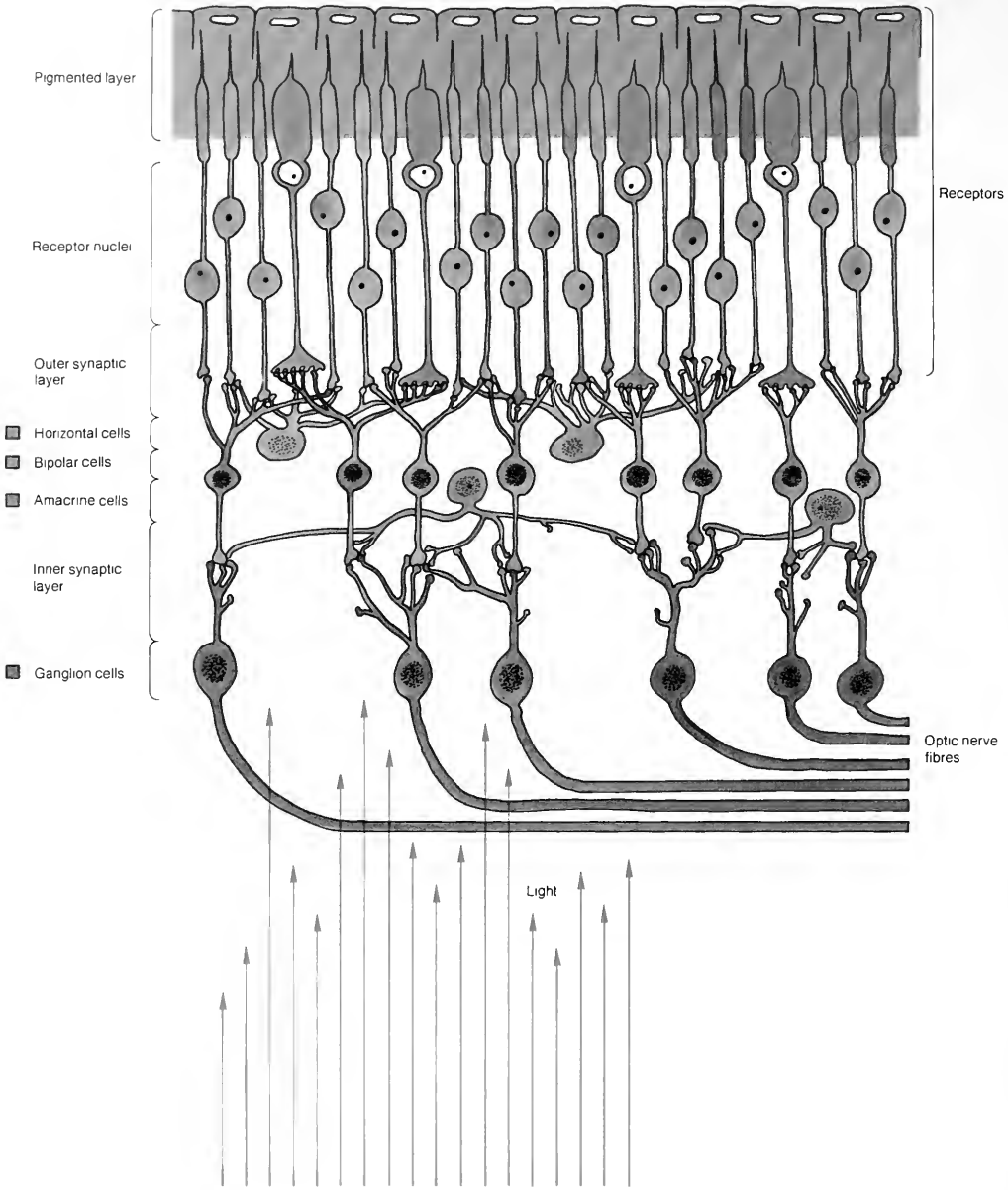


149 [p. 131] The position of receptors, bipolar cells and retinal ganglion cells in the eye Amacrine and horizontal cells not shown



130 [left; see p. 118] A block picture devised by Leon Harmon Can you recognise the hidden object? If not, try screwing up your eyes to blur your vision, and also view the picture from about 2 metres. You might then suddenly see what is present.

Plate 10



154 [p. 133] Schematic wiring diagram of retina

165 [p. 139] Kanizsa's figure drawn in delicate pastel shades The illusory contours are, in the author's opinion, slight or non-existent.

Blue 'grey levels'

Red 'grey levels'

Green 'grey levels'

166 [p. 139] Land's three lightness scales The three black-and-white photographs were produced using a black-and-white film and a coloured filter in front of the camera lens, so that light of just one colour was selected in each case.

Enlarged version of anaglyph to aid viewing with the red/green spectacles

Left half

Printed as red

Anaglyph

Red light

Green light

Green light only gets through green filter, and so into left eye, leaving the red-printed image visible as a dark image on a light green background

Left eye

Red/green spectacles

Right eye

For full wiring diagram, see (plate)

Stereopsis

Stereogram (also called a stereo pair)

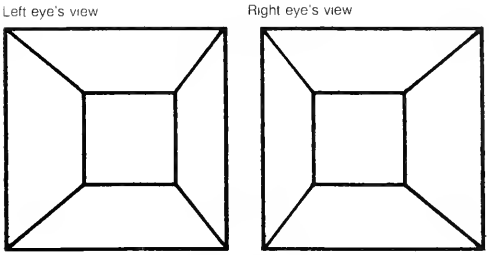
Red light

Green light

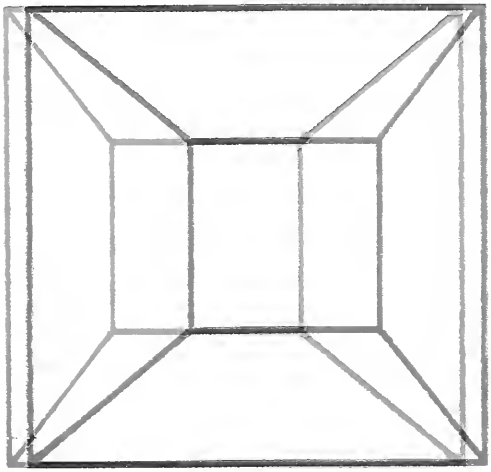
Red light only gets through red filter, and so into right eye, leaving the green-printed image visible as a dark image on a light red background.

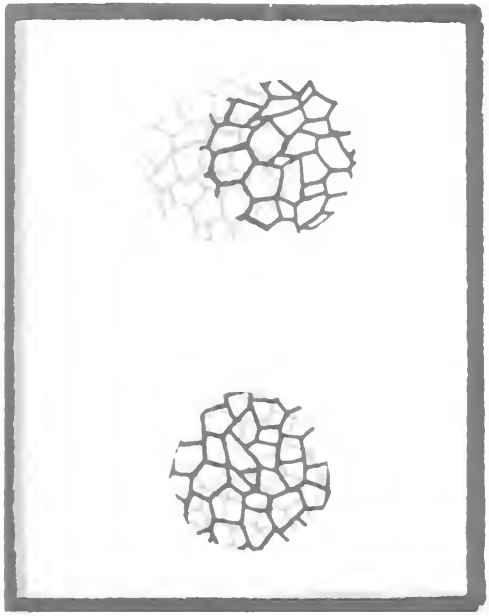
168 [p. 141] A stereogram in anaglyph form Light reflected from the anaglyph strikes the red/green spectacles and the red filter lets only red light through, the green filter only green light. This makes the component picture printed in green look like a dark-printed image to the right eye because the green print reflects little red light and so little gets through the filter from this component. The same thing happens *mutatis mutandis* for the left eye: it sees a dark-printed image of the red component. Look at the anaglyph first with one eye, then the other, using the red/green spectacles. Note that the red and green inks are, sadly, not perfectly matched to the red/green filters, so that each eye receives one image of quite strong contrast together with a second much fainter one. The latter, unwanted, faint image tends to obscure the depth effect only with simple stereograms composed of just a few lines. In most other cases it is quite unnoticeable.

169 [p. 142] The two eyes receive slightly different views of depthful scenes (see text).



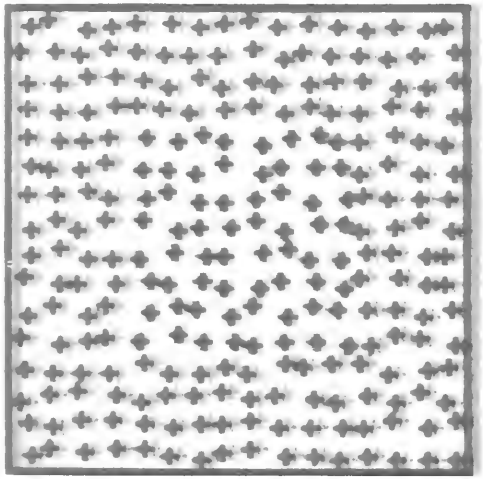
173 [p. 143] Anaglyph producing a 3D view of a sliced-off pyramid (compare with 169 and 170)



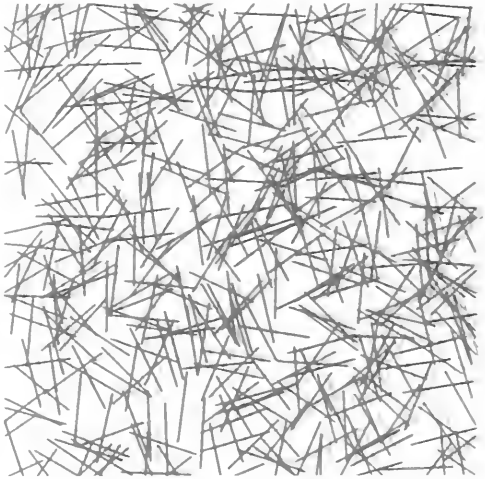


177 [p. 144] The upper blob appears to protrude in front of the lower one despite the many mismatches in the details of left and right stereo halves.

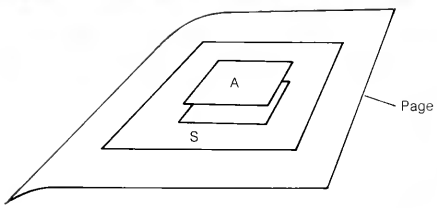
180 [right, p. 145] **A random-dot stereogram** (see text for details) (a) anaglyph; (b) sketch showing appearance of anaglyph; (c) left stereo half; (d) right stereo half



181 [p. 145] Textures for 'random-dot' stereograms can be of varied types



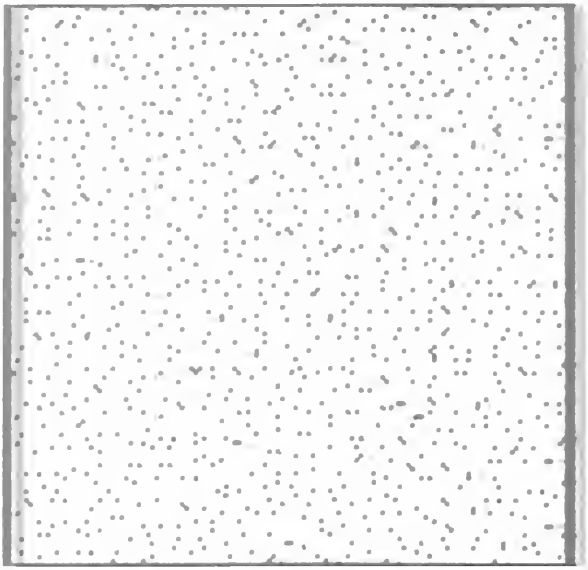
a



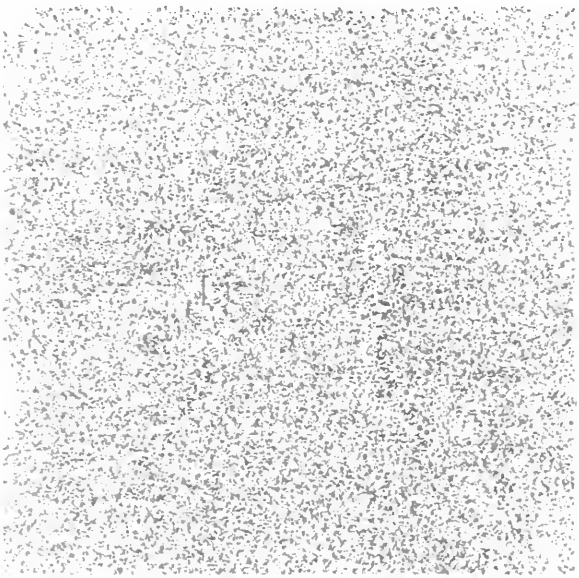
b

c

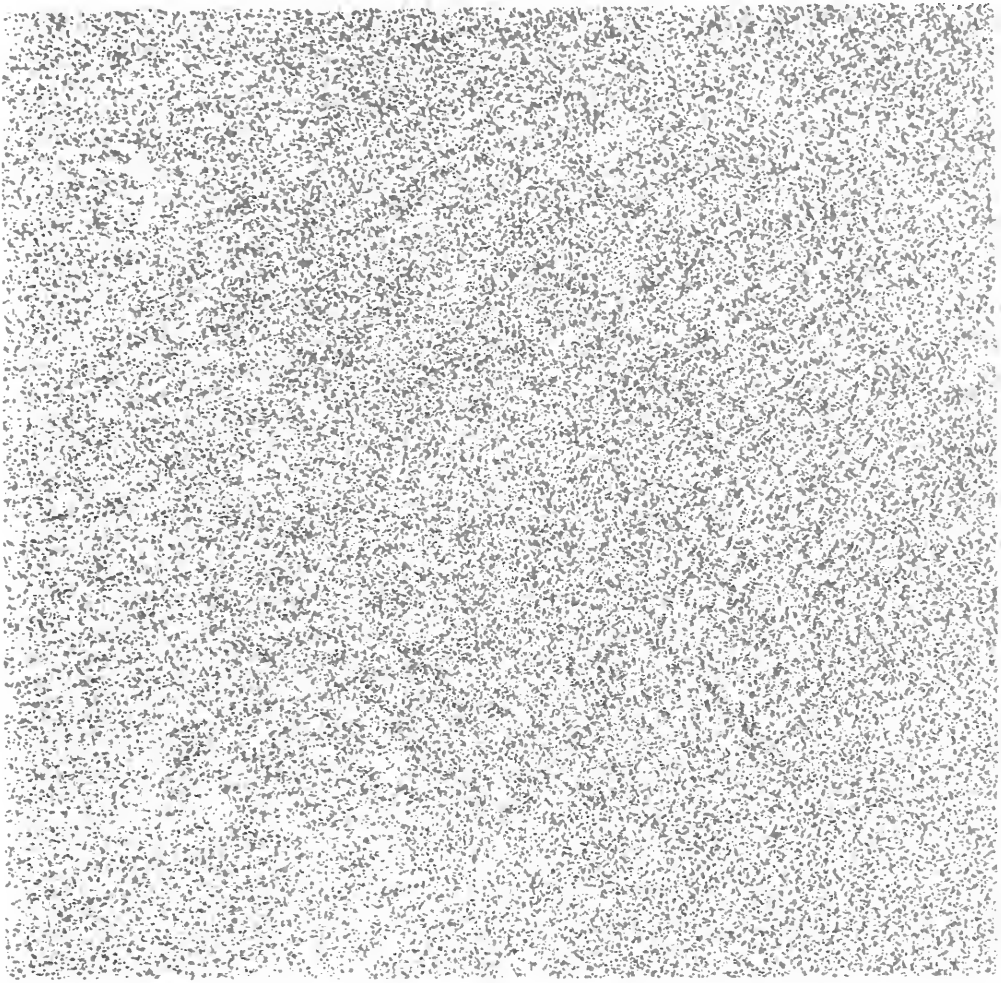
d



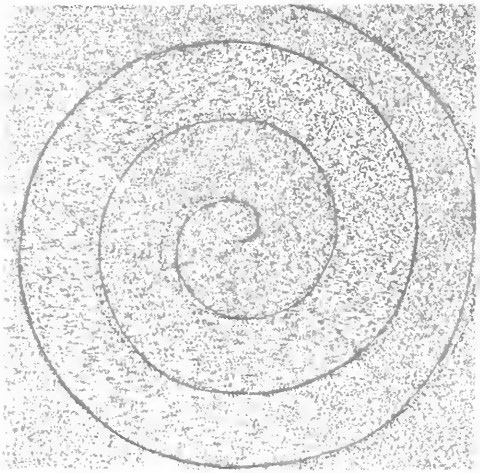
185 [p. 147] **The problem of global stereopsis** Random-dot stereogram whose texture of small dots emphasises the ambiguity problem of finding which left dot matches which right dot



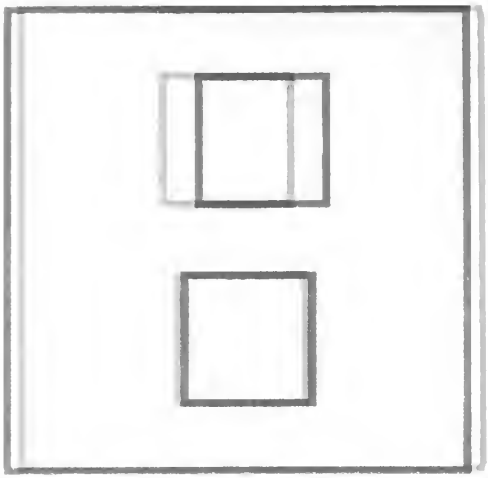
186 [p. 148] **Hyperbolic paraboloid with torus portrayed by a random-dot stereogram**



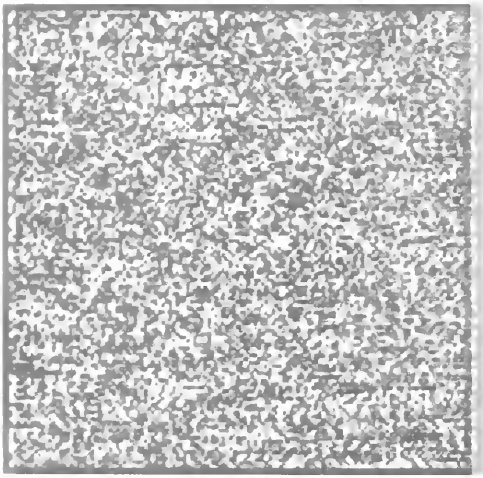
187 [p. 148] Spiral staircase portrayed by a random-dot stereogram



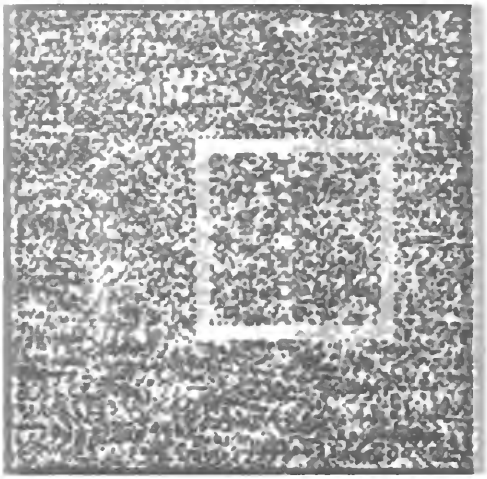
188 [p 148] Anaglyph of spiral staircase (see 187), but with shape of staircase marked in outline in each stereo half to help fusion



189 [p 148] Size of disparity shift determines amount of perceived depth Upper square (large shift) appears to protrude more than lower square (small shift)



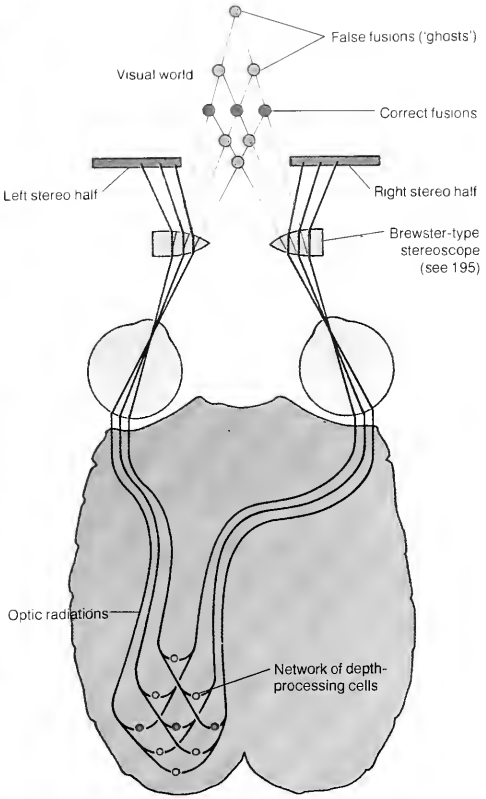
190 [p 148] Large disparity random-dot stereogram Fusion is difficult



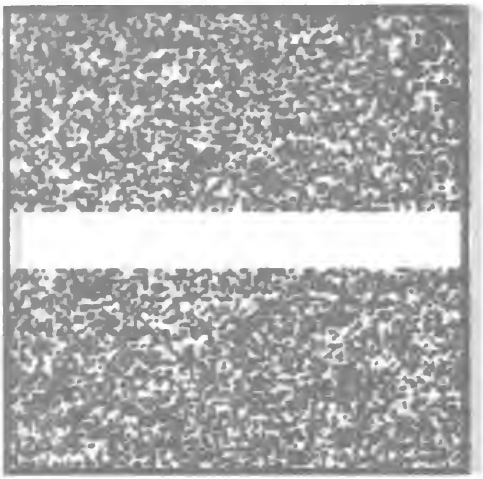
191 [p 148] Large-disparity stereogram helped by monocular contour

192 [p. 148] Small disparity version of 190 Fusion is easier.

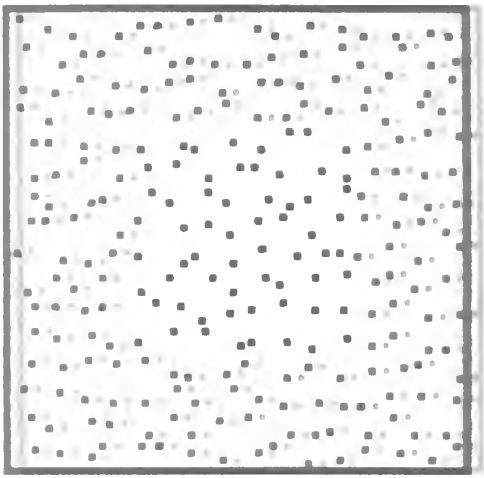
193 [p. 148] Large disparity version of 190, but with monocular marks not giving away shape of disparate object. The triangle helps even though the object is a square.



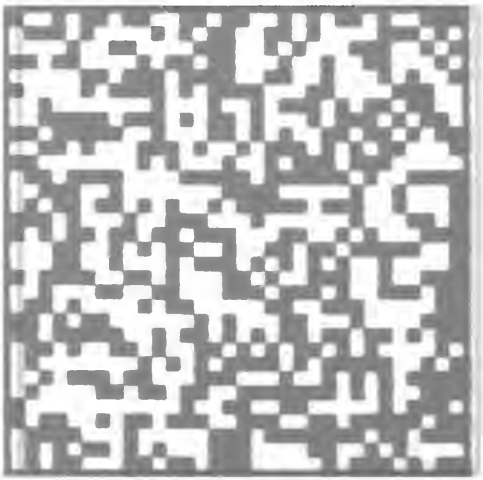
194 [p. 149] The problem of global stereopsis Correct and false local fusions both in the 'perceived visual world' and in a network of depth processing units are illustrated.



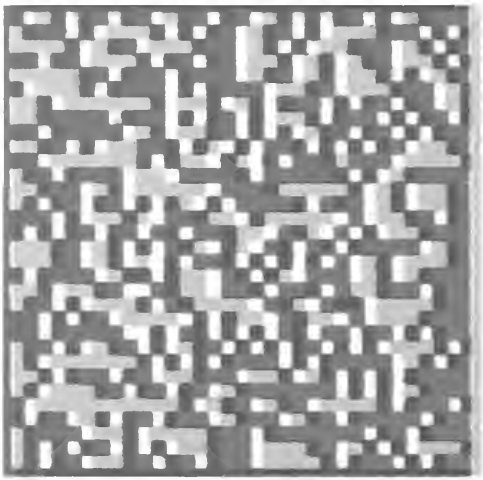
199a [p. 153] Illusory depth contours bridge the gap (see text)



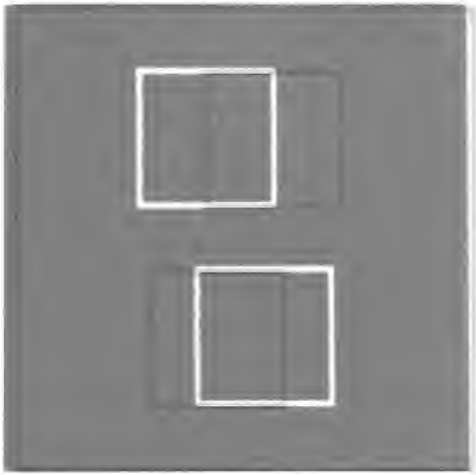
199b [p. 153] The dots of the square-in-depth 'pull up' the white ground on which they lie (see text).



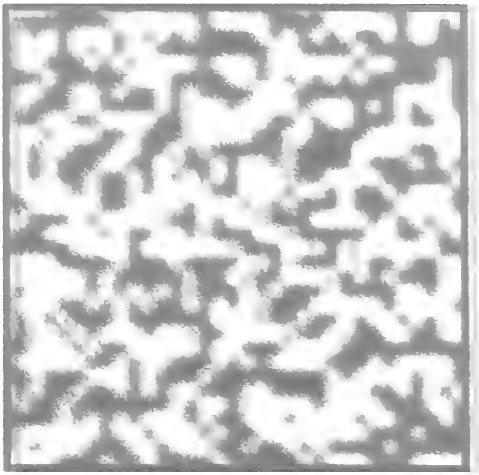
200 [p. 153] Low- and high-contrast stereo halves can be fused.



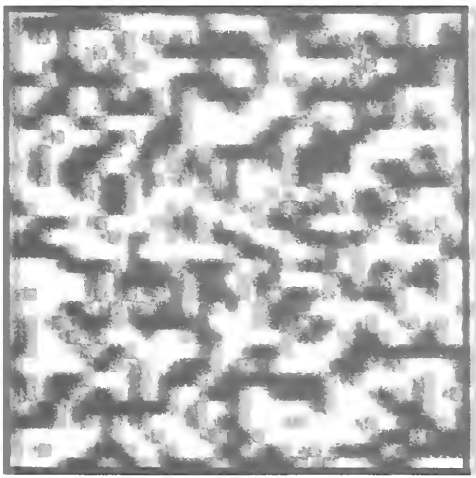
201 [p. 153] Stereopsis is impossible if the right stereo half is a contrast reversal (black-for-white and white-for-black) of the left half.



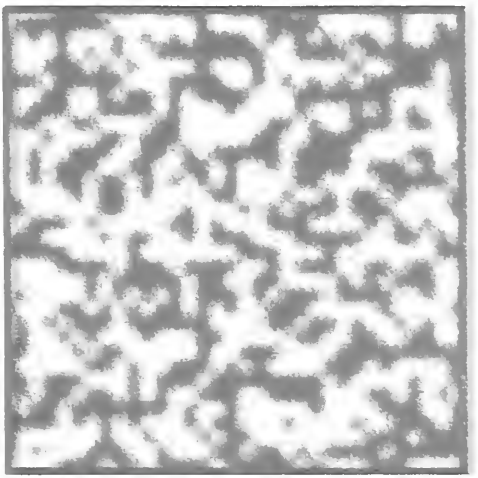
202 [p. 153] Simple stereogram with reversed contrast Stereopsis is possible.



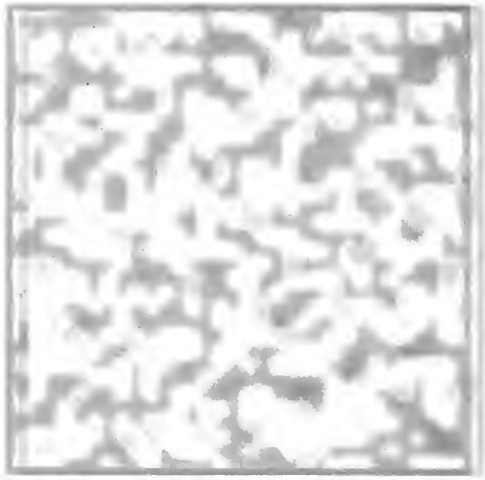
204 Stereopsis survives the blurring of both stereo halves This is a blurred vision of an ordinary random-dot stereogram such as 180. (Technically, the low-spatial-frequency information has been kept, and the high-spatial-frequency detail – sharp edges etc. – filtered out.) Stereopsis survives this degradation of the image quite well.



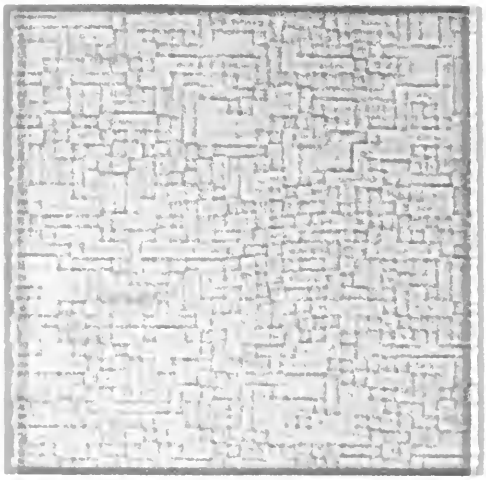
205 Stereopsis survives the blurring of one stereo half This is more surprising than the survival of stereopsis in 204. Blurring one stereo half gives a visual input rather like that received by someone with poor vision in one eye and no spectacles on.



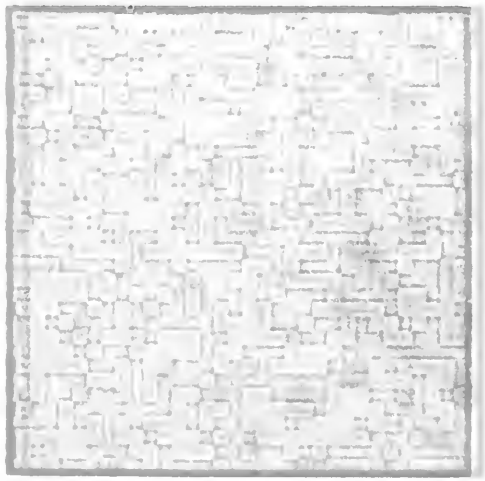
206 Stereopsis carried by blurred blobs survives speckled noise added to one field. One half of this blurred stereo pair has had added to it a high-spatial-frequency noise-speckle ('noise' is the technical term for irrelevant content added to an image). This does not prevent stereopsis taking place. But see 207.



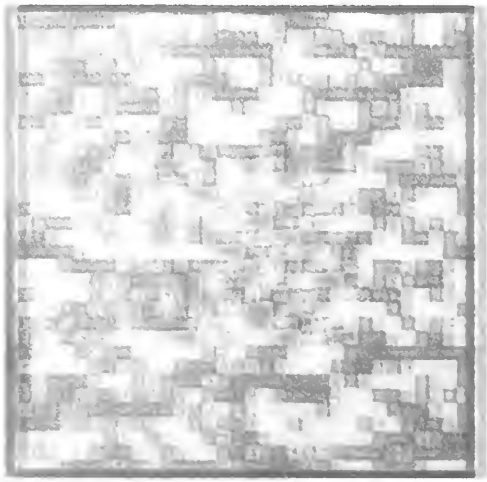
207 Stereopsis carried by blurred blobs is destroyed if similar blurred noise is added to one field. Here the added noise (see 206) is of low spatial frequency, and thus similar to the texture carrying the disparity cue. This does destroy stereopsis.



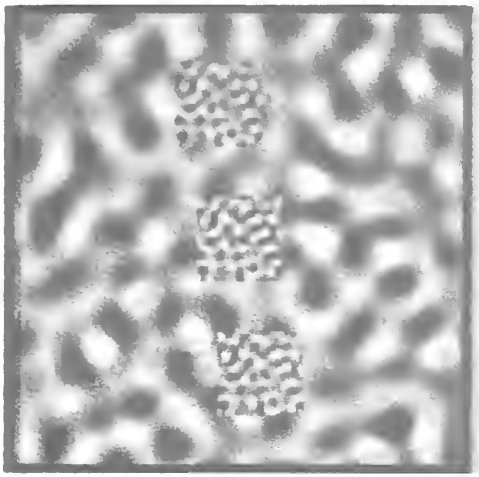
208 Stereopsis is possible if just the edges (high spatial frequencies) are present. This stereogram has been so filtered that it contains only sharp edges, and is technically called a 'high spatial frequency filtered stereogram'.



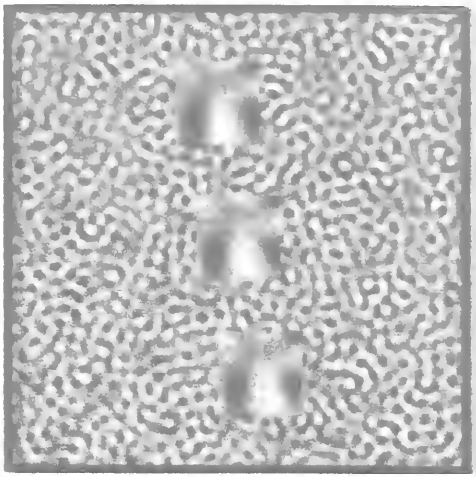
209 Stereopsis carried by edges (see 208) survives (just!) a blurred-blob noise added to one field. Here a low-spatial-frequency noise has been added to one half of 208, and stereopsis can still be obtained. This demonstration is the reverse of that in 206. If the added noise is of high spatial frequency (i.e. similar in frequency to that of the underlying stereogram: compare 207), stereopsis becomes impossible.



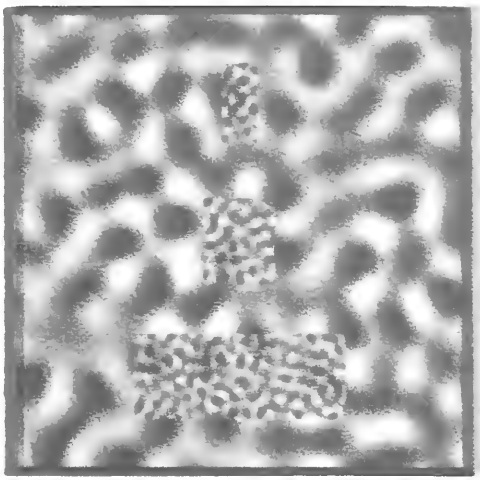
210 A random-dot stereogram with two superimposed surfaces in depth. A central square of low-spatial-frequency blobs recedes behind a protruding lacework of high-spatial-frequency speckle. Compare scenes such as goldfish swimming behind pond weed, or a lamp-post behind lacy curtains, where two depth planes are seen one behind another. This violates the Marr/Poggio rule that any given stimulus point can be placed in only one depth plane.



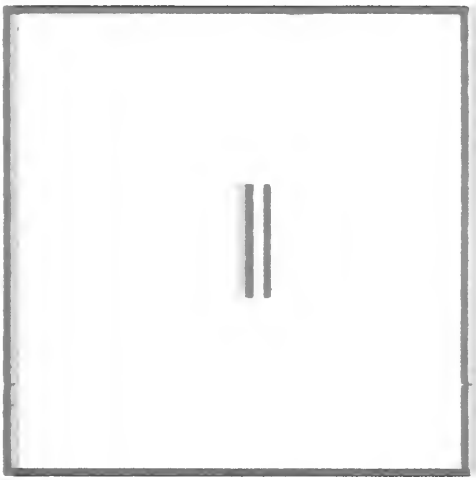
211 [p. 154] Stereogram with rivalrous texture (see text)



212 [p. 154] Rivalrous texture stereogram from which stereopsis is difficult and perhaps impossible to obtain (see text)

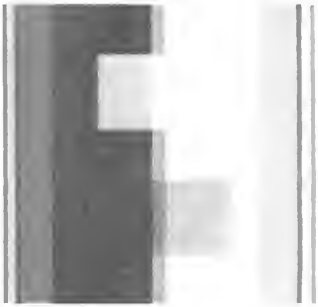


213 [p. 154] Paradoxical stereopsis (see text)

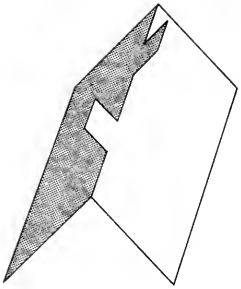


214 [p. 154] Panum's limiting case

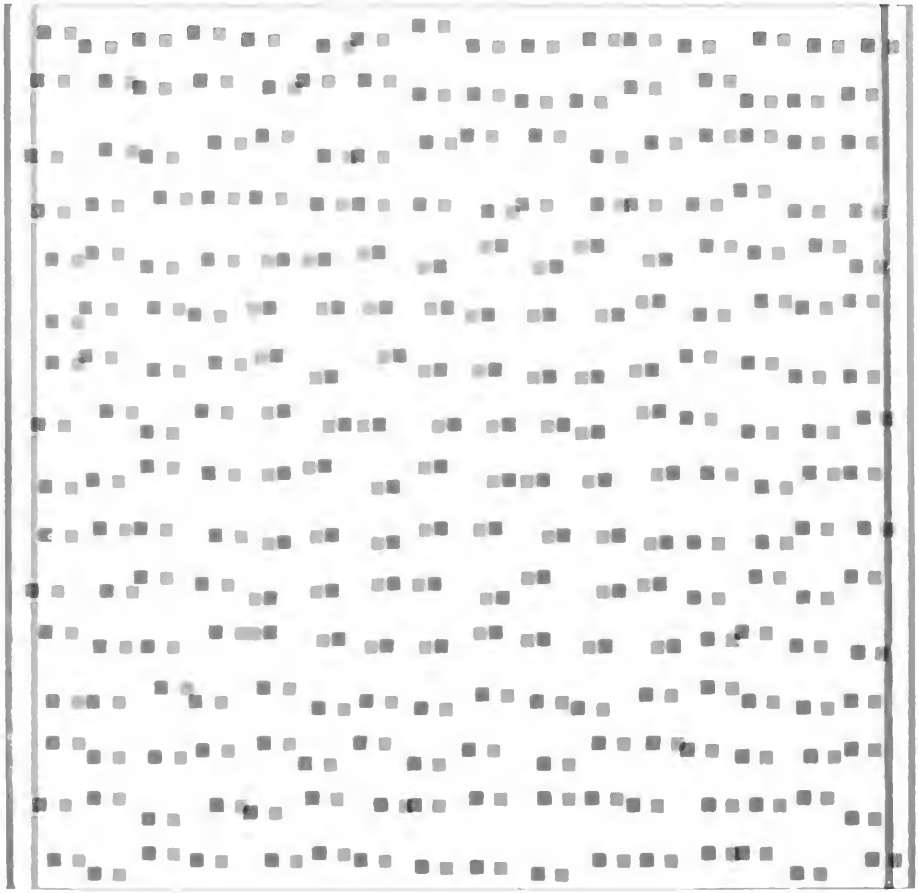
216 [p. 154] The computation of lightness seems to precede stereopsis (see text) Following fusion, a tent-shaped pair of surfaces is seen, with the tabs forming a continuation of the surface on the 'opposite side' (see 217)



217 Sketch showing appearance of 216 when fused binocularly



218 [p. 154] A large-element stereogram suitable for long-range viewing (see text)



4 AFTER-EFFECTS – THE PSYCHOLOGIST'S MICROELECTRODE

Neurophysiology has told us a great deal about the visual machinery of the brain, as chapter 3 made clear. In particular, it has told us that line detectors of various sorts are prominent in the early stages of processing by visual brain mechanisms. Taking up suggestions stemming from work on image processing by computer, we drew the inference that these line detectors provide measurements taken from the retinal image, measurements which are used to build up a symbolic feature description of the scene inspected. In the next chapter we will consider how this early feature description might be elaborated into a richer symbolic description, one which includes explicit descriptions of objects, for example. But in the present chapter we ask: is there any *psychological* evidence relating visual experience to what we know about the neurophysiology of line detectors? Are there any illusions, for instance, which reveal the operation of these detectors? The claim was made in chapter 1 (p. 13) that visual illusions afford interesting clues about the mechanisms of vision. Can this claim be made good in the present instance by finding illusions which demonstrate, or at least strongly suggest, that an analysis of the retinal image by line detectors really does contribute in an important way to seeing?

After-effects

One large group of illusions which has been used to relate neurophysiology to psychology relies upon the fact that certain curious illusory phenomena are experienced after a period of prolonged or intense constant stimulation. Such illusions are called *after-effects* and occur widely in sensory systems.

Perhaps the most commonly experienced after-effect comes from accidental observations of bright lights: the sun, for example, or a naked light bulb. Following such unwanted exposures, it is usual to find that an *after-image* of the light source remains apparent for some time afterwards, superimposed upon whatever scene we next happen to observe.

Bright lights are not essential for obtaining after-images, however. For example, look at the cross in **81** and note that it is set in a plain grey field. Next, stare fixedly for 15 seconds or so at the cross in **82**, which has white discs above it and black ones below. Having completed this period of unvarying stimulation, quickly transfer your gaze back to the cross in **81**. You will now no longer see around this cross a plain grey field, but instead a mass of superimposed after-images: dark discs above the cross and light ones below it. Because the after-images have the opposite brightness to the discs which induced them, they are said to be *negative after-images*. They last for only a few seconds in the present instance, as contrasted with after-images from very bright lights, which can last several minutes. The persistence of the after-images from **82** can, however, be increased by lengthening the time spent staring at the cross. The illusory after-images are almost certainly due to some kind of fatigue, or related processes, probably set up in the retina and caused by the prolonged exposure to the unvarying stimulation. We will have more to say about the possible causes of these after-images later in this chapter, and also in chapter 6.

As stated already, after-images are but a common example of a wide class of perceptual phenomena called after-effects. They are not restricted to vision but occur in other sensory systems as well. For example, if you run a finger to and fro along a curved edge with your eyes closed, a straight edge felt subsequently will seem to be curved in the opposite direction. If you spend a long while (say 90 minutes or so) in an atmosphere containing an excess of carbon dioxide, then fresh air can subsequently give an illusory smell of ammonia. If one listens repeatedly to a tone that increases in intensity, then a subsequently presented tone of constant intensity is likely to sound as though it is decreasing in intensity.

After-effects have long been studied by psychologists, but they have recently become the subject of particularly intensive research. The reason is that they give the psychologist a sensitive tool for probing the workings of sensory mechanisms discovered by the neurophysiologist. Or at any rate, it turns out that many of these effects find a ready explanation in terms of known neurophysiological mechanisms, and so the psychologist has tried to turn the tables and study after-effects as a way of getting at these sensory mechanisms indirectly. And it is because of the neat relationship which often exists between, on the one hand, psychological findings from after-effects, and, on the other, neurophysiological discoveries made using microelectrodes, that after-effects have been dubbed the 'psychologist's microelectrode'. Just as a physiologist can discover by microelectrode recordings from a single cell in the brain that this cell is specially responsive to a given type of stimulus – say, a thin dark vertical bar – so a psychologist, by discovering that we are subject to an after-effect when exposed to prolonged stimulation by a figure full, say, of thin dark vertical bars, concludes that there must be a 'population' of cells in the visual system specially tuned to measure that type of feature. He has made the same kind of discovery as the physiologist, but without having to probe the visual system physically with a microelectrode. In other

words, after-effects give the psychologist his own probe into the early processes of visual analysis, a probe which enters the visual system via the indirect but none the less revealing route of studying certain illusions. And in some ways this indirect form of probing is more powerful than probing with a microelectrode because it reveals facts about whole populations of cells working together at once, rather than about one cell operating on its own.

It will be helpful at this stage to outline the typical procedure used in studies of after-effects, and to introduce some associated terminology, as follows:

First, the observer looks at a *test stimulus* (e.g. the grey field of **81**) and notes some aspect of its appearance, such as its brightness, its colour, and so on. In experiments, as distinct from demonstrations, the perceptual attribute of interest would be measured using *psychophysical* techniques. (*Psychophysics* is the business of measuring the relationship between the perceived and the physical attributes of stimuli: Greek *psyche*: 'mind'; *physic*: 'of nature'.)

Second, the observer stares at an *adapting stimulus* for a prolonged exposure or *adaptation period* to create fatigue or some other adaptation disturbance in the sensory mechanisms of interest.

Third, the observer reverts to the *original test stimulus* and notes its appearance once more. Again, in formal experiments the perceptual attribute of interest is re-measured.

This sequence is called a *test-adaptation-test cycle*. The effect of the adaptation is noted (or measured) as the change in appearance of the test stimulus between its pre- and post-adaptation presentations.

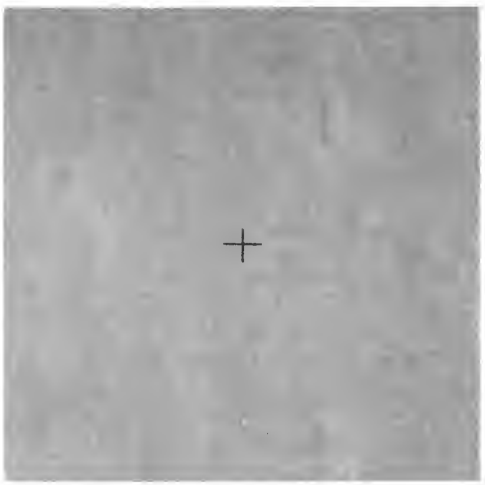
Grating Stimuli

Perhaps the type of stimulus most frequently used during the last 15 years or so of visual psychophysics has been the *grating*. A grating simply consists of a repeating collection of line elements. The line elements can vary in their orientation, their width, the sharpness of their edges, and their contrast.

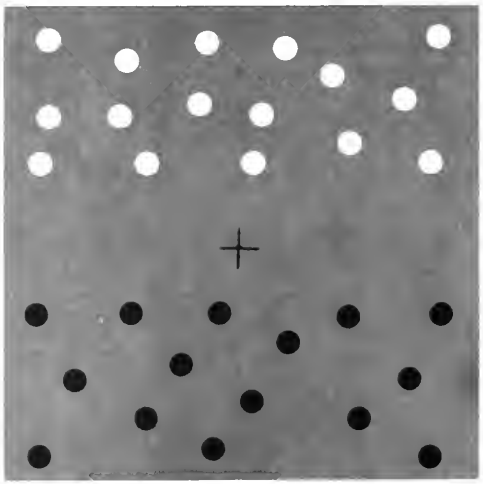
To take *contrast* first, if a grating has very dark black lines and very light white ones (top grating, **83**) then it is said to have a high contrast. Equally, if the 'black' lines are not very dark and reflect light with almost equal intensity to the not-very-light 'white' lines (bottom grating, **83**) then the grating is said to have a low contrast; and, of course, intermediate con-trasts are also possible (middle grating, **83**). Each grating in the series of **83** has a *luminance profile* associated with it, that is, a graph showing how the intensity of reflected light varies across the surface of the grating. These profiles help clarify what is meant by contrast.

Note that although contrast varies from low to high in **83**, the total amount of light reflected from each grating remains the same throughout. That is, if all the light from the top grating was collected somehow and its intensity measured with a photocell, and then this measurement compared with a similar measurement taken from the grating with the lower contrast, the two measurements would be found to be identical. So a higher contrast is not obtained by increasing the total amount of light reflected from the page. Rather, contrast is increased by 'packaging' the same total quantity of light to create 'whiter' regions, at the expense of also creating 'blacker' regions. Thus the *average luminance* of each grating in **83** is identical (within printing error of course), even though the contrast difference between the upper and lower gratings is enormous.

81 A plain grey test field



82 Adapting stimulus for obtaining negative after-images



Contrast Thresholds for Gratings

The gratings shown in **83** vary in their contrast and hence in their visibility. If the series from high to low contrast was extended further at the low end (a difficult thing to do in a printed illustration: hence the limited range given in **83**), then the gratings would become harder and harder to detect, and finally a point would be reached when a very low contrast grating would not be seen at all. That is, the very low contrast grating would appear no different from a field of uniform grey and, technically, it would be described as falling below *contrast threshold.*

The term 'threshold' is a good one because it denotes a step – here the contrast value which marks the step between seeing and not seeing. Contrasts below threshold cause the observer to see nothing except a uniform grey; contrasts above threshold produce the perception of a grating. Contrasts only just above threshold produce a grating of low apparent contrast – a grating of washed-out grey stripes which can only just be discriminated one from another in intensity. Contrasts well above threshold produce clearly visible gratings appearing as a collection of black and white stripes.

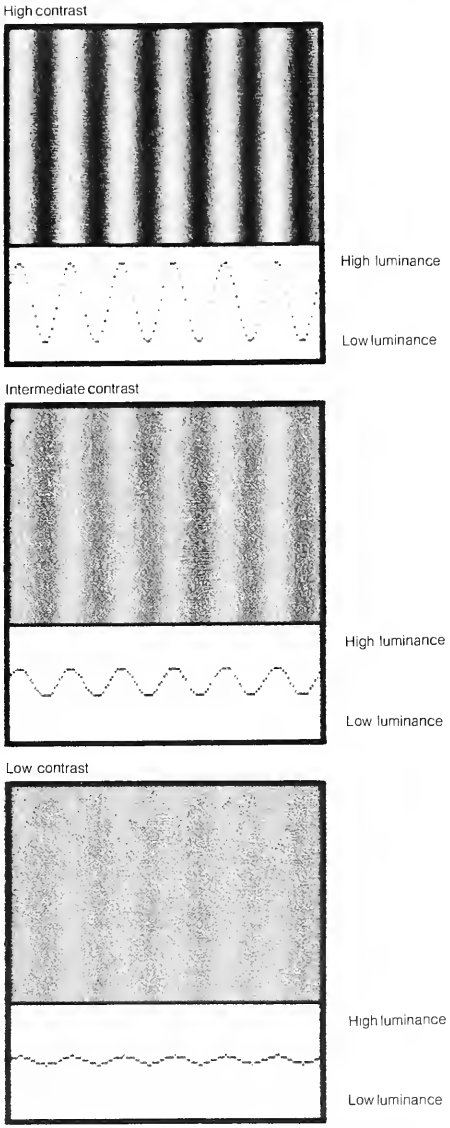
Contrast thresholds, however, like sensory thresholds in general, are not unvarying stimulus values. That is, there is never a particular contrast which strictly defines for all occasions the boundary between seeing and not seeing. Rather, an observer's ability to detect a near-threshold grating varies from moment to moment, even if the contrast of the grating itself is held rigidly constant, so that sometimes the observer will see it, and sometimes he will not.

Contrast Threshold Elevation Caused by Adaptation

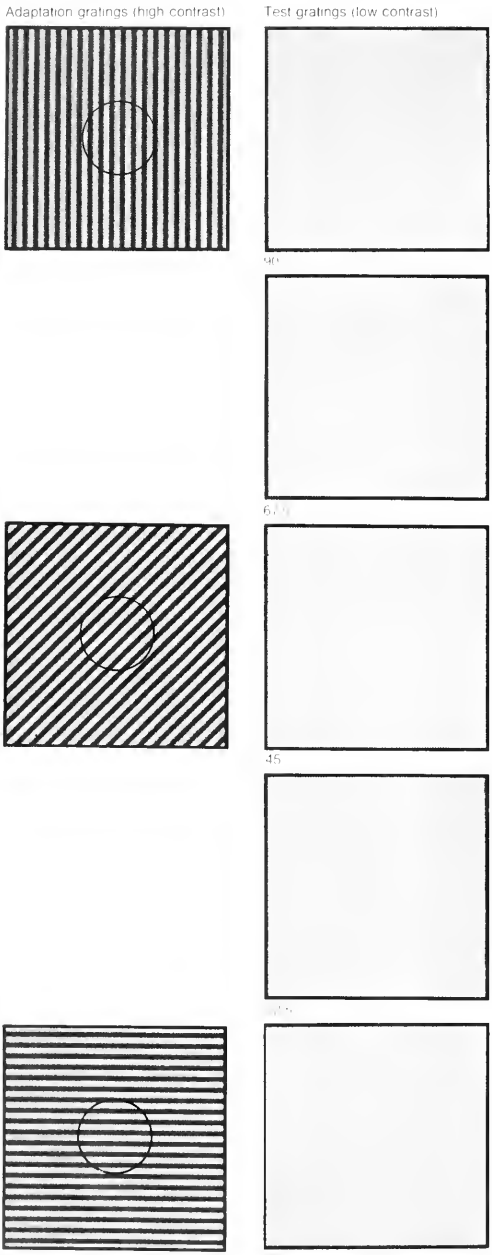
Look at the column of low-contrast test gratings shown in **84** from a distance of about 2 metres. Note that they are only just visible and that their angles vary between vertical and horizontal. If you cannot discriminate the gratings from this viewing distance (i.e. if they fall below your contrast threshold), move the book a little closer until you can just see them. Next, adapt your visual system to vertical lines by gazing at the high-contrast vertical adaptation grating (top left) for about one minute. Use the same viewing distance as you used for inspecting the test gratings. Allow your gaze to wander within the circle drawn on the adaptation grating during your one-minute inspection of it. This is an important precaution which prevents the build-up of an after-image of the kind illustrated in **82**.

Finally, when the full minute of adaptation has expired, quickly transfer your gaze to the column of test gratings. You will observe (if all has gone well!) that the vertical test grating is no longer visible. All that can be seen from that test stimulus is an impression of uniform grey. On the other hand, note that the horizontal test grating (bottom of column) remains visible after the period of adaptation. That is, it is just as detectable post-adaptation as it was pre-adaptation. Look up and down at the test gratings of intermediate orientations and try to decide whether they remain visible post-adaptation or not. Be careful to note that the state of adaptation fades after a few seconds. Therefore, keep returning to the vertical adaptation grating for 'top up' adaptation periods of at least 15 seconds each.

What has this demonstration shown us? The most important point to note is that the period of adaptation to the vertical high-contrast grating raised your contrast threshold for seeing



83 Gratings of varying contrast Technical readers may like to note that, for gratings with such luminance profiles (called **sine wave gratings** from their shape), the contrast is defined by the formula (Maximum Luminance − Minimum Luminance) ÷ (Maximum Luminance + Minimum Luminance), which produces a number varying from 1 (very high contrast) to 0 (zero contrast, i.e. no grating present).



84 The orientational specificity of contrast threshold elevation caused by adaptation See text for viewing details. The equipment used by psychophysicists to study this kind of effect is illustrated in **86**.

vertical stripes. That is, vertical stripes which you could see before adaptation, albeit ones of low contrast which only just exceeded your pre-adaptation contrast threshold, could no longer be seen post-adaptation. It is as though your visual system became 'worn out' during the adaptation period, and when it was later called upon to see a low contrast grating it was just not up to the task. It was too fatigued. Moreover, this fatigue was *limited just to vertical stripes*. Other gratings in the test series of **84** could be seen just as well post-adaptation as they could pre-adaptation (though the visibility of the 67.5° grating might have suffered a little post-adaptation for some readers: see later, especially **90**). This phenomenon is called *orientation-specific* elevation of contrast thresholds.

You should convince yourself that the orientational specificity of this after-effect holds just as well for other adaptation orientations, besides vertical. For instance, adapt to the 45° high contrast grating shown in **84** and note that the 45° low contrast test grating becomes momentarily invisible post-adaptation, whereas the other test gratings are relatively unaffected. Similarly, adapt to the horizontal high contrast grating and observe that now the after-effect is restricted to the horizontal test grating. Allow a few minutes' rest between changes of adaptation stimulus, so that the effects of adaptation to vertical have time to dissipate before you proceed from one adaptation to another.

Width-Specific Elevation of Contrast Thresholds

The demonstration just described for orientation-specific adaptation can be matched by a similar demonstration of width-specific adaptation. Look at the column of test gratings in **85** and note that they all have a low contrast and that they vary in the width of the lines composing each grating. Inspect the gratings from about 2 metres as before and use this viewing distance throughout. Next, adapt your visual system to the grating of thin lines shown at the top of the column of adaptation stimuli. Again, allow your gaze to wander within the circle on this grating to avoid an ordinary after-image. After a minute's adaptation, quickly look at the test grating alongside, whose stripes are of similar width, and note that these stripes are invisible for a few moments. Renew your state of adaptation by re-inspecting the high contrast grating for a further 15 seconds or so and then look at the other test gratings. These gratings are not affected by the period of adaptation to the thin-line grating, their own lines remaining visible post-adaptation. Here again, then, is a specific after-effect, this time one specific to the width of the lines of the grating.

Check that this specificity holds for other adaptation widths by repeating the test-adaptation-test cycle for the other adaptation gratings shown in **85**. Follow the same precautions throughout about viewing distance, length of adaptation, topping up the adaptation, etc. You should always find that the lowered sensitivity caused by the adaptation is restricted to stimuli of the same width as those used for adaptation.

Relating Psychology and Physiology via After-effects

You have probably guessed by now why grating stimuli have figured so prominently in this chapter. Gratings provide a convenient way of presenting line stimuli which differ in their orientation and width. And these two stimulus characteristics

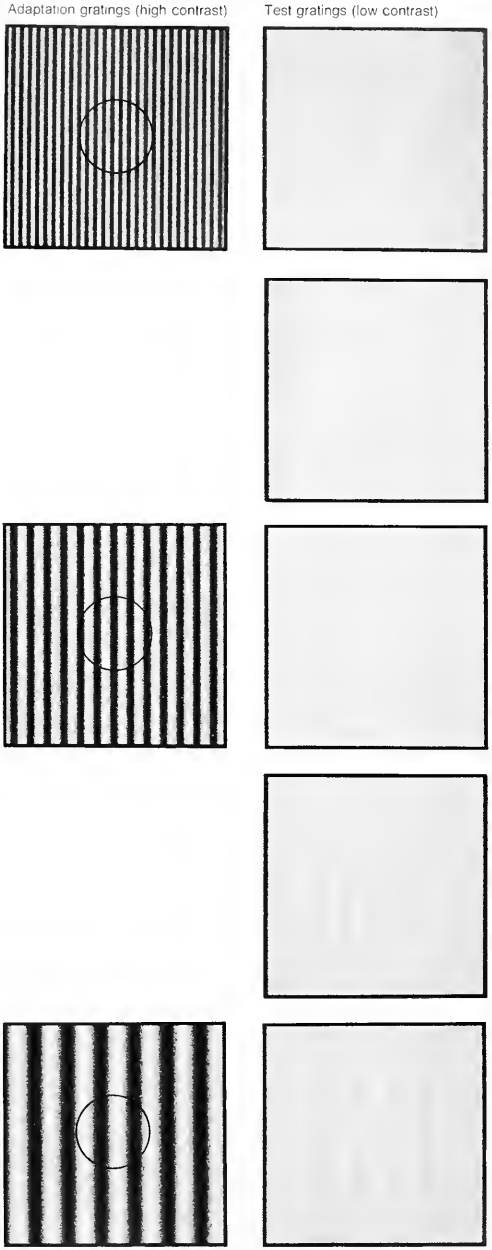
are, of course, amongst those which are very important for determining the optimal stimuli for certain neurons in the visual machinery of the brain, as chapter 3 made clear. You will doubtless remember that some striate neurons were most excited by vertical thin lines, others by horizontal thin lines, yet others by oblique thick lines, and so on, in a large range of different permutations (refer back to **57** for a reminder on this point if you are in any doubt). Thus any one grating stimulus of the kind shown here in **84** and **85** will stimulate optimally a different 'population' of striate neurons. Each grating is large enough to fall on many hyperfields and so cells in many hypercolumns will be stimulated by any one grating. But all cells stimulated optimally will be of the same type. Each grating is, therefore, a powerful probe stimulus for exciting a limited number of cells, cells of a specific type. And a period of adaptation to a high-contrast grating will cause prolonged activity in this stimulated population.

Now it turns out that when brain cells are stimulated constantly for a long period they become 'fatigued'. That is, they respond less well at the end of the period of activity than at the beginning (in fact this loss of sensitivity may be due to a build-up of inhibition). In this respect they are like most other cells in the body: one can imagine a cell constantly faced with its optimal stimulus for a full minute being in a similar position to a muscle cell in an arm which is asked to hold up a heavy weight for a full minute. Very soon, the muscle cell becomes tired, with all sorts of consequences for the overall performance of the arm. This analogy seems to hold quite well for striate neurons. If they are caused to be very active for a long while, they become 'tired'.

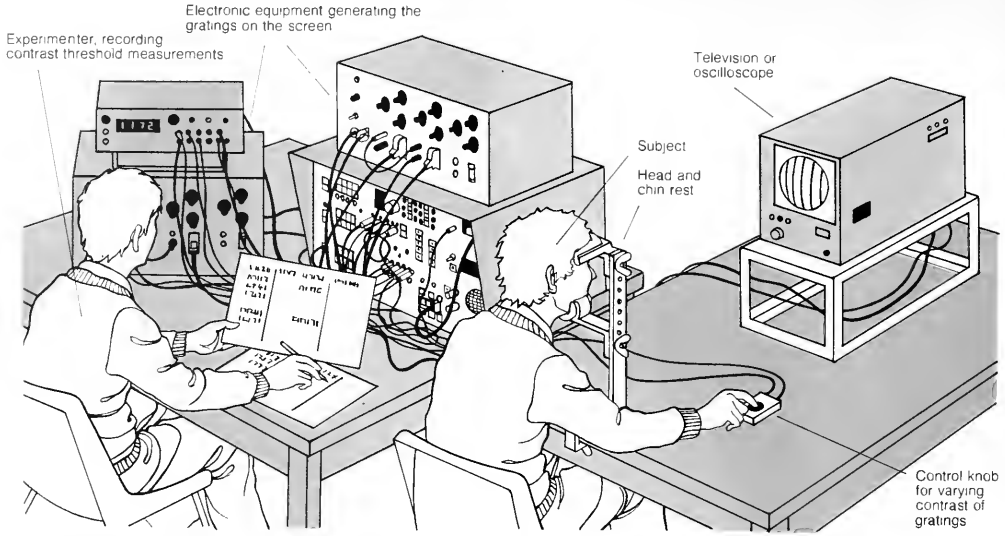
If neurons stimulated optimally by a given adaptation grating become tired by their constant activity over the adaptation period, it is hardly surprising that when they are faced later with a very weak stimulus, such as a low contrast grating of 'their' type, they do not respond to it. And if they fail to 'notice' the weak stimulus, then the observer will fail to see it. The crucial measurements normally provided by the fatigued striate neurons for arriving at the appropriate feature description will simply not be present for a short while post-adaptation. Until the necessary cells recover, a low-contrast grating close to the pre-adaptation threshold will not be detected.

We are now in a position to understand why the contrast threshold elevation caused by the adaptation is *specific to the grating type* used for adaptation. As striate neurons are tuned selectively for line orientation and line width, the period of adaptation will affect some cells but not others. The cells not affected will quite happily detect low contrast gratings post-adaptation, even though the tired cells will not. Hence the observed selectivity of the after-effects.

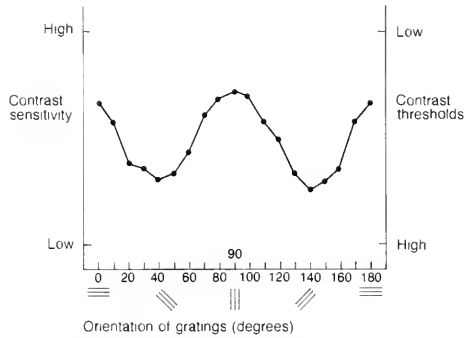
This neat interplay between known neurophysiological discoveries on the one hand and psychophysically measured effects on the other has motivated an enormous number of experiments in psychology and physiology laboratories over the past two decades or so. Literally hundreds of studies have been published, each contributing some new twist to the story, or an added detail on how the tuning of striate cells for orientation, say, relates to the specificity for orientation of the perceived after-effect. The overall body of knowledge provided by this work reflects a remarkably intimate relationship



85 The width-of-line specificity of contrast threshold elevation caused by adaptation See text for viewing details.



86 Simplified experimental arrangement for measuring contrast sensitivity. This equipment allows the experimenter to 'get inside' the head of an observer and measure what the observer sees. The observer turns a knob to adjust the contrast of the grating on the screen without changing its average luminance. He is asked to adjust the grating until it is just visible, and in this way his contrast threshold is measured



87 Graph showing how contrast sensitivity varies according to the orientation of a grating stimulus (imaginary but representative data). Technically, sensitivity is defined as the reciprocal of the threshold

between the two approaches: the dovetailing of psychological and physiological findings is so good that few doubt that both kinds of study are tapping the same mechanisms. Of course, care is necessary in any given instance of interpretation. Finding a similarity between physiological and psychological results is not conclusive proof that the latter stem directly from the mechanisms revealed by the former. But when all is said and done, the use of after-effects as psychophysical probes to match the microelectrode probes of the neurophysiologist is one of the great achievements in psychophysics.

Contrast Sensitivity and Orientation

The graph plotted in **87** presents results from an imaginary experiment conducted with standard psychophysical equipment such as that illustrated in **86**. It shows how our sensitivity to contrast varies according to the orientation of the grating stimuli. Note that high sensitivity (left-hand vertical axis) is shown by low contrast thresholds (right-hand vertical axis), and low sensitivity by high thresholds: if, for instance, the subject had a very low contrast threshold for a grating, this meant he was able to see this grating when it was set to a very weak contrast, so that it is sensible to say that he was very sensitive to it.

The data shown in **87** are based on genuine data, simplified and extended somewhat for our purposes. The graph shows that the subject in question was most sensitive to vertical and horizontal gratings and least sensitive to oblique ones. That is, he required considerably more contrast to see an oblique grating than he required for detection of a vertical or horizontal one.

The special character of vertical and horizontal stimuli is frequently observed in psychophysical work. Various hypotheses have been entertained about why this should be so, ranging from optical factors to do with the lens system of the eye (an improbable basis, in fact, at least for explaining all the effects), to structural neurophysiological causes. This latter

type of hypothesis suggests that the hypercolumn machinery described in the last chapter is somehow particularly specialised for dealing with vertical and horizontal orientations. Just how this specialisation might be built in is uncertain. One possibility is that vertical and horizontal orientation detectors are more finely tuned than those dealing with obliques. Alternatively, or perhaps additionally, there may be slightly more neurons dealing with angles near vertical and horizontal, an increased density which might somehow increase our ability to detect and deal with these orientations: but neurophysiologists have so far found no firm evidence for this hypothesised increased density in their experimental animals.

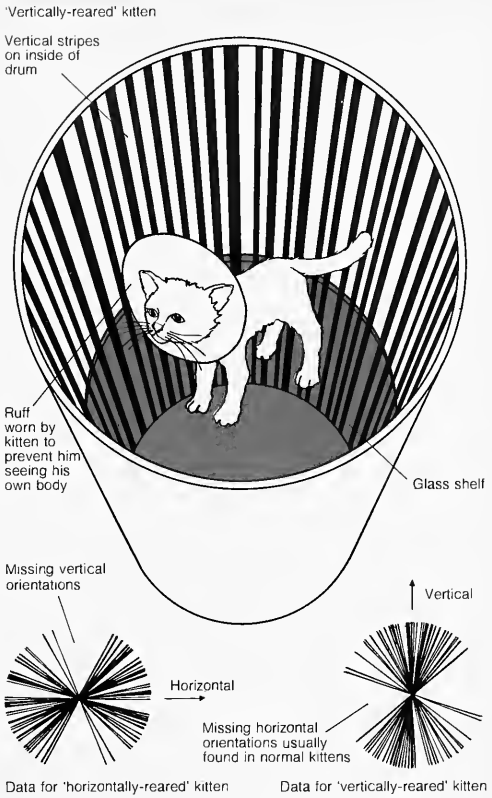
Without undertaking a full evaluation of these and other possibilities, we may just note in passing that it might be that the visual system's preference for vertical and horizontal, however it is embodied in neural machinery, is a result of special tuning to suit our environment. Our modern world is one with a predominance of vertical and horizontal lines, as you can easily confirm by looking around almost any room or typical urban landscape. Could it be that somehow the visual system takes note of this environmental fact and becomes specialised accordingly?

Evidence in favour of this hypothesis has come from an experiment conducted by Blakemore and Cooper, who have shown that in cats, at least, early visual experience can play an important part in the development of the striate cortex. They bred kittens in complete darkness except for certain limited exposures each day to a special visual environment. Some kittens were exposed only to vertical stripes, others only to horizontal ones. This was achieved by placing the kittens in a drum whose internal walls were suitably decorated to give the required stimulation [**88**]. The kittens wore a ruff to ensure that they could not even see their own bodies, so careful were Blakemore and Cooper to restrict visual experience to either vertical or horizontal lines, and nothing else. Subsequent neurophysiological recordings from the kittens' brains showed that the 'vertically-reared' kittens possessed only vertically tuned striate neurons and that the 'horizontally-reared' kittens possessed only horizontally tuned ones [**89**]. In other words, it seemed as though the strange visual environments in which the kittens were reared had had a dramatic effect on the development of their visual cortex.

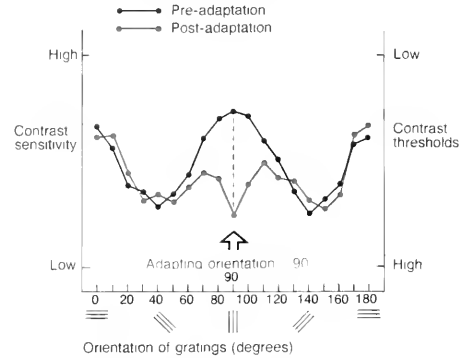
Subsequent work has shown that this environmental tuning effect is not always easy to obtain, and the reasons for this are not yet clear. But the investigation of the role of early visual experience in mediating the development of visual mechanisms is in a particularly active state at the moment, and we may expect that the next few years will both resolve this oddity and add important new discoveries. The key question is: to what extent is the visual system pre-wired by the unfolding of genetically determined growth processes in early life, and to what extent are genetically determined structures 'adjusted' to deal with the particular visual environment they encounter? Different research workers hold different views on this question of *visual plasticity*, as it is called, and the answers will surely be different for different animals. Some creatures are born with their eyes closed (e.g. kittens) and have a relatively gradual visual development which seems to be most active several weeks after birth (at around the 5th week of life for kittens). Other animals (e.g. lambs) are born in a high state of readiness to deal with their environments, both by running away from predators and, it seems, by seeing them in the first place. Where man fits into this range is debatable, but there is no doubt that babies improve in their visual capacities in early life, and that if they suffer certain eye abnormalities this can affect them for life. For example, a congenital squint which is uncorrected for several years after birth is almost certainly followed by an inability throughout life to see binocular (two-eyed) stereoscopic depth (chapter 7). Discoveries about the role of early visual experience in animals are having an important influence on the way ophthalmologists (and others) concerned with the vision of children are thinking about how best to treat visual disorders. The general impetus given by the animal work is to try early treatment, hopefully while the visual system is still in a sufficiently flexible or 'plastic' state of development to take advantage of it.

Before leaving the question of why vertical and horizontal line stimuli are treated preferentially by the visual system, I should point out that not all races seem to show this effect

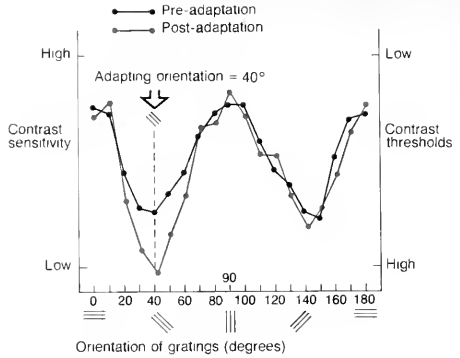
88 Blakemore and Cooper's experiment



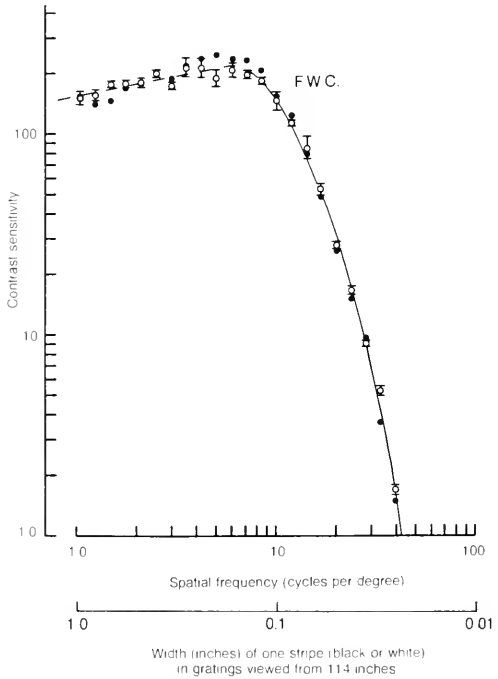
89 Each line represents the optimal orientation of a striate cell found in neurophysiological studies of each type of kitten in Blakemore and Cooper's experiment. The vertically-reared kittens (see 88) had only cells tuned to vertical, and the horizontally-reared kittens had only cells tuned to horizontal.



90 Contrast sensitivity to gratings of varying orientation before and after adaptation to a vertical grating (90°)



91 Contrast sensitivity to gratings of varying orientation before and after adaptation to an oblique grating (40°)



92 Contrast sensitivity for gratings of varying line width (the widths of the lines are expressed – as usual in psychophysics – in terms of spatial frequencies, i.e. the number of black/white cycles per unit angle of visual space)

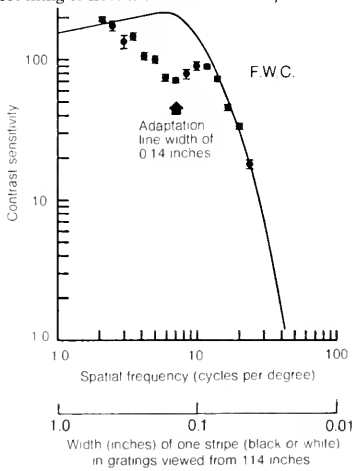
93 [right] Contrast sensitivity for gratings of varying line width after adaptation to a grating of lines 0.14 inches wide (viewed from about 114 inches)

equally. Some studies have suggested, for example, that Asians show little difference between vertical/horizontal and obliques when compared with Caucasians – even Asians brought up in a westernised environment which emphasises vertical and horizontal features. Consequently, one cannot rule out a wholly genetic explanation for the effect.

We are ordinarily more sensitive, then, to vertical and horizontal stimuli than to oblique ones. But what would the graph shown in **87** look like if the data for it had been collected after a period of adaptation to a high contrast grating of, say, vertical orientation?

This experiment could be performed by placing a second oscilloscope, with a high contrast vertical grating on its screen, alongside the one shown in **86**. The observer could adapt to this grating and then immediately transfer his gaze to the other oscilloscope, showing test stimuli, for the purpose of collecting post-adaptation threshold measurements.

The results which would be expected are shown in **90**. The first thing to note is that the data of **87** are included as the



pre-adaptation sensitivity curve. This curve represents the subject's pre-adaptation baseline, against which the effects of adaptation can be judged. The next and most important aspect of **90** is the 'notch' cut out of the post-adaptation curve, as compared with the pre-adaptation curve. This notch is centred on the adaptation orientation of vertical. Sensitivity to gratings more than 20–30° away from vertical is hardly affected by the adaptation exposure: this is shown by the way the pre- and post-adaptation curves intertwine for all orientations other than those around vertical. In other words, what **90** shows is a psychophysical demonstration of orientation-specific adaptation, which matches the informal demonstration which you experienced for yourself when inspecting **84**.

Suppose the subject had been adapted not to vertical but instead to an oblique orientation of, say, 40°. The data would then have looked rather like those in **91**. Again, the pre-adaptation sensitivity curve is shown for comparison, and again there is an adaptation notch cut out from the post-adaptation curve. This time the notch is centred on 40°, as one would expect because this was the adaptation orientation. So once again we have orientation-specific adaptation, but now to a different orientation.

The advantage of graphs like those of **90** and **91** is that the effects of adaptation can be seen in detail and for a wide range of test-grating orientations. Thus a measurement can be made of the *tuning* of the adaptation effect, i.e. the range of test-grating orientations affected. And it turns out that the nature of this psychophysically observed orientation tuning fits in remarkably well with the neurophysiologically observed orientation tuning of striate cells. It is this kind of intimate interplay between data from the two approaches which has caused such excitement and such a vast body of interrelated work over the past few years. When such data as those shown in **90** and **91** are compared with an orientation tuning graph for a simple cell [**58**], it is no wonder that adaptation after-effects have been called the psychologist's microelectrode.

The graphs of **90** and **91** have been explained in terms of lowered contrast sensitivity caused by adaptation. But it would have been equally possible to talk about them in terms of elevated contrast thresholds caused by the adaptation. Try turning the book upside-down and you will see peaks of raised contrast thresholds, again centred on each adaptation orientation. These peaks are identical to the notches cut in the sensitivity curves, of course. As for **87**, one can talk in terms of contrast sensitivity or contrast thresholds, as convenience or preference dictates.

Contrast Sensitivity for Width of Grating Lines

The experiment just described for orientation-specific adaptation can be repeated just as easily for stimuli which vary in the width of the grating lines (their size if you like). Here the experiment would begin by measuring the subject's contrast-sensitivity curve for a wide range of different widths. Then the subject would be adapted to a grating of just one line width and his contrast sensitivity for all widths re-measured. But in this experiment orientation would be kept constant throughout, as the variable stimulus property of interest is simply line width.

This time we can look at actual data taken from a classic paper by Colin Blakemore and Fergus Campbell written in 1969. Thus **92** shows Campbell's pre-adaptation contrast-sensitivity curve (his initials are F.W.C.) for gratings of varying line width. The technical terms used in the axis labelling need not be explained fully for our purposes, the extra added horizontal axis showing clearly the line widths involved (but see caption).

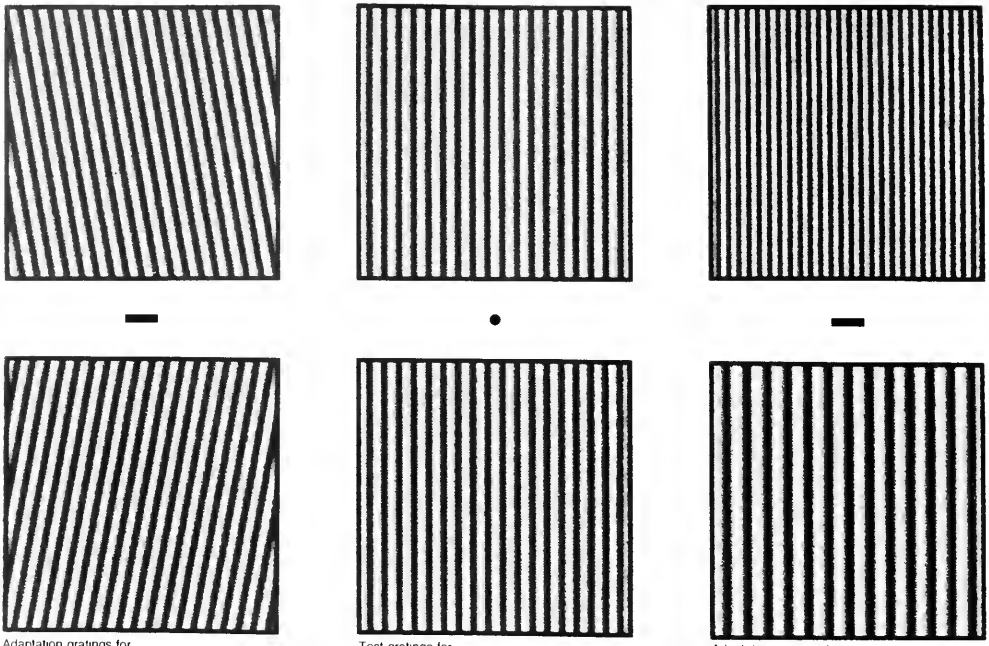
The most prominent point illustrated by Campbell's contrast sensitivity curve in **92** is that line widths smaller than about 0.1 inches viewed from 114 inches away need much more contrast to be seen than do larger widths. In other words, there is a sharp decrease in contrast sensitivity in this region. The limit of this decrease is reached for widths of about 0.05 inches. Beyond this point, Campbell was unable to see a grating no matter how high its contrast, so this point marks the limit of his *visual acuity*. Beyond it, his visual system can no longer pick out the grating lines: they are invisible to him, and he sees just a uniform grey. This is the kind of limit opticians are usually interested in when they assess whether a person needs glasses, although they go about discovering it in other ways (typically with letter charts composed of letters with decreasing sizes: the observer has to say what the letters are and when he fails to get them right then the limit of his visual acuity is reached). The extent to which the limit shown in **92** is caused by optical factors (e.g. limitations in the capacity of the eye's lens to cast a sharp image on the retina), and the extent to which it is due to limitations of a neurophysiological kind, is a question I shall leave on one side.

The visual system also shows a decrease in sensitivity to very large line widths, although the line width range shown in **92** does not bring this fact out very well.

What happens to Campbell's contrast sensitivity curve after adaptation to a grating of, say, lines 0.14 inches wide (again viewed from 114 inches)? The results obtained from an experiment of this kind are shown in **93**. The pre-adaptation curve is plotted as a continuous line: it is in fact the line drawn through the data points of **92**. The notch of decreased sensitivity caused by the adaptation is clearly evident in the post-adaptation curve (imagine a new curve drawn through the new data points), a notch centred on the adaptation line width. There is some spread of the adaptation effect, just as there was in the orientation case, but line widths well away from the one used for adaptation show no decrease in sensitivity. So once again we see the specific nature of the adaptation, this time line-width-specific adaptation.

The Tilt After-effect

So far we have considered only after-effects of contrast threshold elevation caused by adaptation to grating stimuli, but there are in fact other revealing consequences of such adaptation as well. Look, for example, at **94**. Begin by fixating (i.e. fixing your direct gaze on) the dot between the two central test gratings and note that the lines in each grid appear vertical. Next, adapt to the pair of tilted gratings for about 45 seconds or so. Avoid a normal after-image by allowing your gaze to run to and fro along the short horizontal bar between the gratings. Finally, when the adaptation period is up, quickly transfer your gaze back to the central test gratings, again fixating the spot between them. You should now see that the test gratings momentarily appear tilted away from their true orientation of vertical. The illusory tilt for each grating is in a direction away from the corresponding adaptation grating. Thus whereas the adaptation gratings together look like a chevron pointing to the right, the vertical test gratings look, post-adaptation, like a chevron pointing to the left. This illusion, discovered by James Gibson and known as the tilt after-effect for obvious reasons, shows that adaptation to lines



94 Gratings for obtaining the tilt and size after-effects

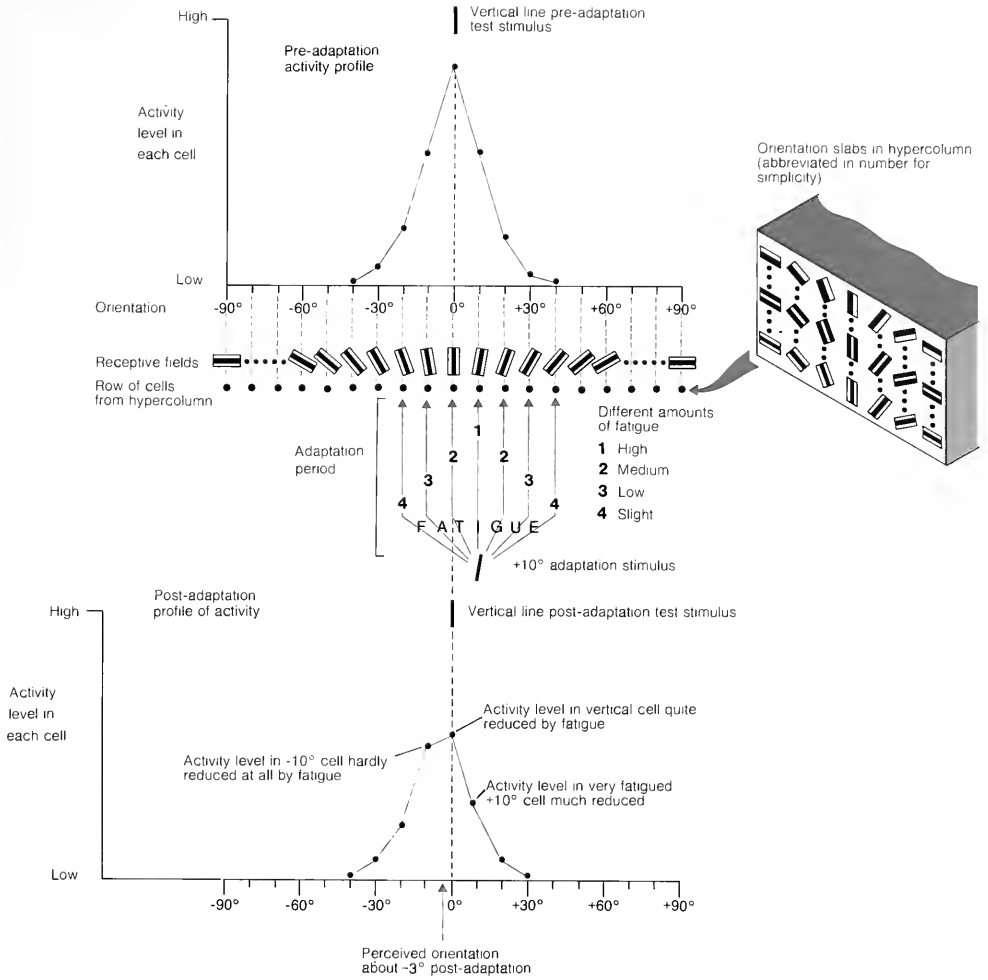
of one orientation can subsequently affect the perceived orientation of lines of another orientation. The nature of the effect is to twist the apparent orientation of test lines *away* from the orientation of the adapting lines. How can the effect be best explained?

A highly plausible neurophysiological explanation, given all we know about the visual machinery of the brain (chapter 3), is to suppose that the period of adaptation causes some kind of fatigue in line detectors which then interferes with the subsequent computation of the perceived orientation of the test lines. Let us consider the details of how this might occur.

Reference back to chapter 3 (**69, 70**) will remind you that it was suggested that feature orientation might be computed by interpreting the *activity profile* in orientationally-tuned simple cells located within the hypercolumn. In the present case, and considering just one line instead of gratings for simplicity, an activity profile set up by pre-adaptation inspection of a single vertical line would be as shown in the top graph of **95**. Activities in a row of cells taken from a hypercolumn are illustrated in this graph, each individual cell coming from a different orientation slab. Activities of line detectors only are shown for simplicity, but of course other cells in the hypercolumn would be chattering away as well and would presumably be contributing to the computation of feature orientation. The rows of cells and their receptive fields are shown beneath the graph's horizontal axis; the plotted points are the activities of these cells caused by a vertical line test stimulus.

The middle section of **95** indicates fatigue created by adapting to a line rotated $+10^{\circ}$ (10° clockwise) from vertical. The idea is that this line stimulus produces fatigue due to its prolonged inspection. The fatigue is highest for the line detector with an optimal orientation of $+10^{\circ}$, because this is the orientation of the adapting stimulus, but adjacent cells would also be stimulated by the adaptation line and so would also suffer some fatigue. A spectrum of fatigue running from high to slight is shown, according to how far away in orientational tuning the cells are from the orientation of the adaptation line ($+10^{\circ}$).

The lower section of **95** shows what would happen when the partly fatigued row of line detectors was again faced with a test stimulus of vertical. The profile of activity would no longer be as it was pre-adaptation. Instead of a symmetrical profile with a peak centred on the vertically-tuned cell, the profile would now be asymmetrical. The vertical cell would still be the most active even though its response would be considerably reduced by the fatigue it suffered during the adaptation period. (Remember, it suffered this fatigue even though the adaptation stimulus was not vertical, because these cells have the property of being sensitive to orientations near to their optimal one as well as to the optimal one itself.) The next most active cell, indeed one nearly as active as the vertically tuned one, is the cell tuned optimally to -10°. This cell was relatively far away, in terms of orientation, from the adaptation stimulus, and so suffered only a low level of fatigue. Cells tuned to orientations even more anti-clockwise from the adaptation stimulus than the -10° cell are hardly affected at



95 A neurophysiological theory of the tilt after-effect See text for explanation.

all by fatigue and respond much as they did pre-adaptation. But note that the cell tuned to $+10°$ suffers most fatigue of all, because it was tuned exactly to the adaptation orientation, and so its post-adaptation response to the vertical line is very much reduced.

The net upshot of the adaptation, then, has been to produce an asymmetrical activity profile in response to the post-adaptation test stimulus of vertical. And this particular asymmetrical profile is exactly the kind of activity profile which would be produced by a line of about $-3°$ in normal circumstances. (Again see p. 53.) So the crucial points to grasp are: that the adaptation caused fatigue, that this fatigue was greater in some cells than others, and that it was such as to skew the post-adaptation profile of activity to a shape like that which would normally be produced not by a test stimulus of vertical but by a line twisted slightly away from vertical in a direction opposite to that of the adaptation orientation. Hence the tilt after-effect. The kind of explanation exhibited in **95** was first devised in 1929 by Georg von Békésy.

One curious feature of the tilt after-effect which a Békésy-type explanation copes with very readily is the so-called *distance paradox*: the paradox is that the adaptation has its effect when we use test stimuli with an orientation *to one or other side* of the adaptation orientation, but not when we use a test stimulus of the same orientation as the adapting stimulus – a sort of remote control effect, as it were. If the adaptation orientation *itself* was used as a test stimulus, then a symmetrical activity profile would be generated, with therefore no illusory shift in perceived tilt. That is, the whole distribution of activity would be depressed by the adaptation, but there would be no skewing in profile shape, and so no change in perceived orientation. It is only when a test stimulus with an angle some 'distance' from the adaptation orientation is presented that the effects of the adaptation become 'one-sided', as it were, and so produce the skewed activity profile which is held to produce the illusion.

A distance paradox is characteristic of feature attributes which are said to be *coded by place*. That is, differences in orientation are detected by neurons in a different 'place' in the brain. Of course, the difference in physical position in the brain is slight: the key point is that the difference in stimulus attribute is represented by activity in a different but physically similar component. (Perhaps a better technical term for this type of coding might be 'coding by connections'.) So far, neurophysiology has shown us that each cell in a row of hypercolumn units is in itself identical. It seems to have the same structure, the same type of nerve impulses, etc. Thus all the components in the hypercolumn row shown in **95** are, it seems, alike – just as were the switches in the detector arrays of **32** [plate 2]. What serves to distinguish them is the connections they receive, so that some are 'turned on' by certain feature orientations, some by other feature orientations, and so on. This is a kind of coding much used in computers: each transistor can be identical in electrical design but each can serve a vastly different function.

The Size After-effect

If the neurophysiological explanation of the tilt after-effect is true, then this effect is based on orientation-specific adaptation in line detectors. Now we know that these detectors are also fussy about line width, as well as line orientation, and we know also that line-width-specific adaptation can be demonstrated using the elevation-of-contrast-thresholds technique. Consequently, it should be possible to generate a width after-effect, similar to the tilt after-effect, but causing an illusory shift in perceived line width following adaptation. This is how Colin Blakemore and Peter Sutton reasoned, and they discovered just such an effect, which you can experience for yourself using **94**.

First, fixate the spot between the two central vertical gratings and check that they are composed of stripes with the same width. Next, adapt your visual system, using the gratings on the right of the figure. Allow your gaze to run along the horizontal bar between the upper (thin) and lower (thick) bars. Continue the adaptation for about 45 seconds or so: even longer can be helpful in seeing this after-effect, which is not quite so noticeable as the tilt after-effect. Following the adaptation period, quickly look back to the spot between the two central test gratings and, if all has gone well, you should notice that these gratings briefly look different in width. The upper grating seems to be composed of thicker lines than the lower grating. Do not expect too big an effect, but the effect is clear enough despite not being a large one.

96 Spiral stimulus for obtaining a strong movement after-effect See text for details.

It is easy to apply exactly the same type of physiological reasoning to explain Blakemore and Sutton's so-called size after-effect as was used for the tilt after-effect. Thus the adaptation causes width-specific adaptation in rows of hypercolumn cells tuned to the widths of the adaptation lines [74]. Later on, when the test gratings are re-inspected, the profile of activity in width-specific cells is disturbed so that the apparent width of the lines in each test grating is 'pushed' away from the width of its adaptation grating. Thus the upper test grating appears thicker and the lower test grating thinner in its line composition. If the labelling of the horizontal axis in **95** for orientation is replaced by labelling for width, then the graphs apply as well to the size after-effect as they do to the tilt after-effect.

The size and tilt after-effects have been much studied in recent years, again operating as a microelectrode for the psychologist. Thus the tuning of these effects – the degree and extent of their influence – can be established using psychophysical techniques, and inferences drawn about the probable tuning of orientation and/or width detectors in man.

The Movement After-effect

The emphasis of this chapter has so far been placed on orientation-specific and width-specific after-effects, because these relate most directly to the hypercolumn machinery described in chapter 3. But the family of after-effects is a large one, and lest a false impression be given to the contrary, we will now turn to a quite different example of the genre – the *movement after-effect*. This is the title given to the fact that after watching movement in one direction for a prolonged period of time, a stationary scene viewed afterwards seems to be moving in the opposite direction. The effect is a strong one and is often experienced in everyday situations. Indeed, it is most commonly known as the *waterfall phenomenon* because it can be obtained by staring for a minute or so at a waterfall, which provides a convenient constantly moving image. When the observer subsequently transfers his gaze to a stationary scene, such as the river bank, this scene appears to move upwards for a few seconds or so.

You can witness a particularly vivid version of the movement after-effect for yourself using the spiral shown in **96**. Copy this out on to a circular piece of paper and then mount this disc on your record-player turn-table. Set the table to rotate at 33⅓ revs per minute and stare at the centre of the disc while it is rotating. Keep looking for about one minute, to get a really good effect. When this adaptation period is up, stop the turn-table and see what happens. What should happen is that the genuine movement of the spiral, perceived as a continuous contraction of the spiral rings towards the disc centre, is replaced by illusory movement in the opposite direction – the spiral seems to be continuously expanding rather than contracting. This movement illusion is very vivid for a few seconds and then it weakens, with quite a long tail-off, perhaps up to 20 seconds or so, during which a slow movement drift can just be detected. Re-adapt to the disc several times, to make sure you perceive the effect clearly: it is well worth the effort. After one of these adaptation periods, try looking not at the stationary spiral as a test stimulus but instead at any convenient stationary surface. You will find that the illusory movement will still be seen quite readily. Notice also that although the after-effect gives a very clear illusion of movement, the apparently moving features nevertheless seem to stay still! That is, we are still aware of features remaining in their 'proper' locations even though they are seen as moving. What we see is logically impossible! This perceptual paradox, like so many others, suggests that the visual system can detect one feature attribute quite independently of others, and that if one part of the feature-description system is 'suffering an illusion' for some reason, this does not mean that other aspects of the total feature description are 'brought into line', and so adjusted to obtain an overall perception which is 'sensible'. The visual system is quite happy to live with paradox, its individual parts in some respects seeming to work like separate modules, almost separate sensory systems. One is reminded here of remarks made in chapter 1 about impossible objects.

The most commonly accepted explanation of the movement after-effect is illustrated in **97**, and is due to Sigmund Exner, writing about this effect as early as *1*894. The general idea is that different directions of stimulus movement are detected by different neurons, and that pairs of neurons dealing with opposite directions are coupled together in some way. A prolonged period of exposure to one stimulus direction fatigues one member of each pair, leading to an imbalance in their activity which produces the movement after-effect. Let us follow through the details of this explanation: it is important because it illustrates an *opponent-process* system at work, and such systems are common in perceptual mechanisms.

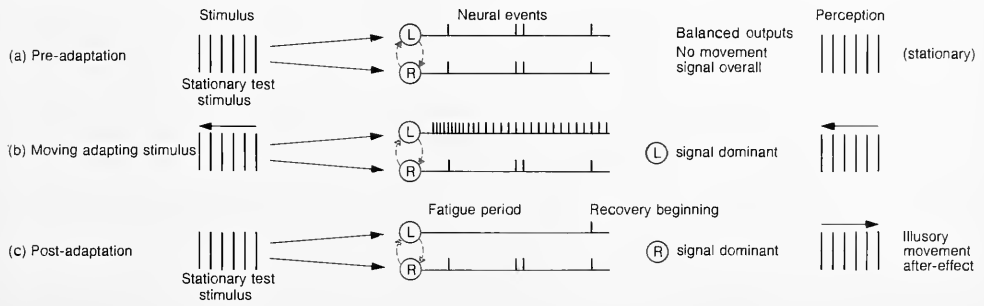
First, mention was made in chapter 3 of certain visual neurons which show *directional selectivity*, i.e. they respond optimally to a particular direction of movement and not at all to movement in the opposite direction (see the section on complex cells: p. 62). A pair of such neurons is shown in **97**a responding to a stationary grating. As would be expected for movement-detecting cells, a stationary stimulus such as this hardly excites them at all and they are shown responding simply with a low resting discharge rate (artificially drawn as identical in the two cells for simplicity). It is assumed that when their outputs are equal in this way, later processes infer that the stimulus is stationary, as indeed it is.

Second, **97**b shows the same two cells responding to an adapting stimulus, a grating moving to the left for a prolonged period. The movement-to-the-left neuron is shown responding vigorously at first, with its response diminishing somewhat as fatigue (or some other adaptation process) sets in. The movement-to-the-right neuron keeps firing with its resting discharge rate, as though it were simply 'seeing' a stationary stimulus. As the leftward-movement neuron shows a much larger output than the rightward-movement neuron, the result is perceived movement to the left – a 'truthful' perception.

Third, **97**c shows what happens when the adapting stimulus is removed and replaced with a stationary grating. The leftward-movement neuron is fatigued and does not even show a resting discharge rate until it recovers. The rightward-movement neuron, on the other hand, keeps on firing with a resting discharge rate as it has done throughout. But now this resting discharge rate is treated by later processes as a movement signal because it is not cancelled as in **97**a or exceeded as in **97**b by activity in the leftward-movement neuron. As a result, an illusory movement to the right is perceived – the movement after-effect – which lasts until the fatigued leftward-movement cell recovers.

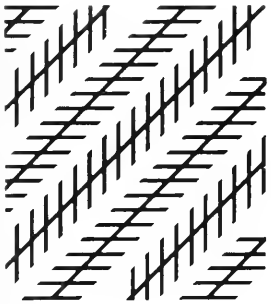
A variation on the above explanation is to suppose that the leftward and rightward neurons are coupled together with inhibitory connections, shown in **97** as the dotted arrows joining the pairs of cells. If these connections transmitted inhibition according to the degree of actvity of each cell, then the most active cell would eventually 'come out on top' as it

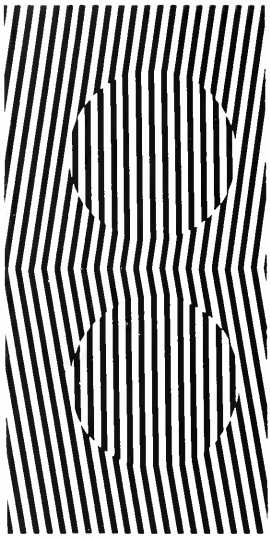
97 Exner's explanation of the visual movement after-effect L and R represent movement-to-the-left and movement-to-the-right respectively.



100 The two discs are composed of vertical lines, but these are made to appear tilted by the surrounding lines.

101 The Zöllner illusion
The long oblique lines are parallel but look slanted with respect to each other.





were, having silenced its 'opponent'. In this way, the later processes of interpretation might be relieved of having to judge the relative activities of the leftward and rightward neurons because only one signal would 'come through'. The after-effect might then be the product of the rightward cell being released from inhibition for a moment or two because of the fatigue built up in the leftward cell during adaptation. I shall not go into the details of this scheme, but I mention it to illustrate the computational power of connections between neurons. A pair of simple inhibitory fibres might go a long way in 'interpreting' or 'balancing out' or 'cancelling' opposing movement signals.

One might ask: why should such cancellation be necessary? The reason is the familiar one: without it, any given neuron signal is ambiguous. For example, does a resting discharge type of output derive from a stationary pattern or from a genuinely moving pattern but one moving in a direction to which the neuron is not tuned? By pairing two opponent processes together (left and right movement detection), the ambiguity is resolved in a simple and economical way: in both **97**a and **97**b the rightward cell's activity is the same, but the eventual perception is different. This trick is open to the nervous system when the stimulus dimension in question has a two-ended character to it, with a 'null point' in the middle when neither stimulus attribute is present. Here we have leftward movement or rightward movement as the two ends of the continuum, with 'stationary' in the middle.

Other two-ended sensory dimensions exist, for example lightness darkness, with mid-grey in the middle as the null point. Consequently, the negative after-images observed in **81** might find an explanation in terms of opponent processes, in this case the mechanisms mediating lightness and darkness perception (see chapter 6). Coloured negative after-images can also be obtained [**98**, plate 6], which similarly suggests opponent processes between different types of colour mechanism (red-versus-green and blue-versus-yellow, the null points on each continuum again being mid-grey).

To return to the question of mechanisms for movement detection, it is of some importance to note that although the *direction* of movement in **97** is coded by *place*, that is, a different nerve cell is used to code activity in different directions (p. 62), the *velocity* of the movement could well be coded by *frequency*. In other words, the faster the movement, the more active the neuron and so the more frequent its nerve impulses.

Finally, one should note that the pair of neurons shown in **97** is only one pair among many. Each pair deals with a different continuum along which movement can occur – left/right, up/down, or some other angle. The pairs might be located in the striate cortex, as the hypercolumns certainly contain directionally sensitive cells, or in the retina, as cells with this property are also found there in many species. For a rotating spiral stimulus such as **96**, cells dealing with all these continua would be stimulated during the adaptation period, but of course only one member of each pair would be activated (hence the contracting movement seen in the case of the spiral, i.e. movement inwards from all angles, followed by the illusory expansion). A further complication is that each cell pair would have a limited receptive field and so one needs to think of a full set of pairs devoted to analysing each part of the field of view.

Contingent After-effects

In 1965 Celeste McCollough discovered a particularly odd kind of after-effect which has since attracted a great deal of research; to date over a hundred published reports have appeared on it, and on related effects. You may see the after-effect for yourself using **99** [plates 6–7]. First, look at the test stimulus **99**a, which is simply a pattern of black and white lines, some vertical and some horizontal. Note that this stimulus is quite colourless. Next, adapt your visual system using **99**b and **99**c. Look for 10 seconds at the red-vertical pattern and then for 10 seconds at the green-horizontal one, then go back to the red-vertical for a further 10 seconds, and so on, alternating between the two adapting stimuli for a period of at least 3 minutes, or, if you can bear it and want to obtain a good effect, for 10 minutes. When this adaptation period is up, transfer your gaze back to **99**a and note what you see. The so-called *McCollough effect* is that after adaptation the black and white test stimulus appears coloured. That is, negative coloured after-images can be seen. However, these colour illusions are not after-images of the usual sort because if you look carefully you will see that the vertical lines in the test stimulus appear faintly greenish (the vertical adaptation lines were red) and the horizontal test lines appear faintly reddish (the horizontal adaptation lines were green). However, as all parts of the retina were equally exposed to red and green during the adaptation phase, the illusory colours cannot be due to the same mechanisms as are reckoned to underlie the normal negative coloured after-images of **98**. Somehow what has happened is that negative colour after-effects have become 'tied to' the orientation of the inducing stimulus. Try the effects of turning **99**a through 90°: you should find that the illusory colours in the test stimulus change place, showing that 'vertical' and 'horizontal' are defined with respect to the retina, not the world.

The McCollough effect was the first of many such after-

effects to be discovered over the past 10 years or so. These after-effects are called as a group *contingent after-effects*, because the appearance of an after-effect of some sort is made dependent, or 'contingent', upon some other stimulus characteristic (in the case of the McCollough effect demonstrated in **99**, it is contingent on the presence of vertical and horizontal lines). Similar colour after-effects can be made contingent on direction of movement (e.g. leftwards versus rightwards movement), on line width (thick versus thin stripes), or on curvature (convex versus concave lines). Moreover, it is possible to do things the other way round and, for example, make a movement after-effect contingent upon the colour of the test display. All sorts of such combinations can be employed: hence the fact that so many research studies on these effects have been published so far – and no doubt there are many more to come yet.

But there have been reasons other than sheer novelty, and a curiosity to try out every likely, or unlikely, combination, behind the flurry of research into contingent after-effects. The hope has been that this new class of after-effects would provide a new tool for probing the workings of the visual system, a new microelectrode if you like. Thus various people have speculated that their existence might reflect the presence in the visual system of 'doubly-tuned' neurons, neurons which require that a stimulus should have two (or more) characteristics before the neuron will fire. For example, there may be neurons which require their optimal stimulus to be not only of the right orientation but also of the right colour. Examples might be cells tuned to vertical-and-red, oblique-and-red, horizontal-and-red, vertical-and-green, oblique-and-green, and so on. Now if such cells were coupled in an opponent-process manner, much as we supposed to be the case when explaining the movement after-effect, then a neural basis obviously exists for explaining the after-effects. The period of adaptation can be held to fatigue, or otherwise adapt, just one half of each opponent pairing, with the after-effect then emerging as the post-adaptation consequence. Some encouragement for this line of explanation comes from neurophysiological reports that some orientation detectors are indeed also colour-specific.

But there are problems for the explanation in terms of doubly-tuned neurons. First, given the large number of contingent after-effects, this explanation implies vast numbers of doubly-tuned cells, enough to cover all the possible combinations of stimulus attributes now known to give contingent after-effects. This seems rather wasteful of neural machinery, but, on the other hand, there are indeed vast numbers of cells in the visual cortex. Second, and more importantly, contingent after-effects can be very durable. If you managed to last out the full 10 minutes of adaptation to **99**b and **99**c, your contingent after-effect is still probably present. So try looking again at the test stimulus **99**a. Did you see a fleeting reappearance of the illusory colours? If so, your experience is in keeping with many reports of the long-lasting nature of these effects. They can even last for weeks if the initial adaptation is long enough (e.g. half an hour or so). This longevity suggests a mechanism quite different from the usual short-term adaptation of neurons posited for normal after-effects. It favours an explanation involving some kind of learning. As John Mollon has succinctly put it, it suggests that if we do not have neurons jointly specific to colour and orientation before we adapt to **99**b and **99**c, perhaps we do when we have finished. In other words, perhaps adaptation can *create* doubly-tuned cells.

Simultaneous versus Successive Illusions

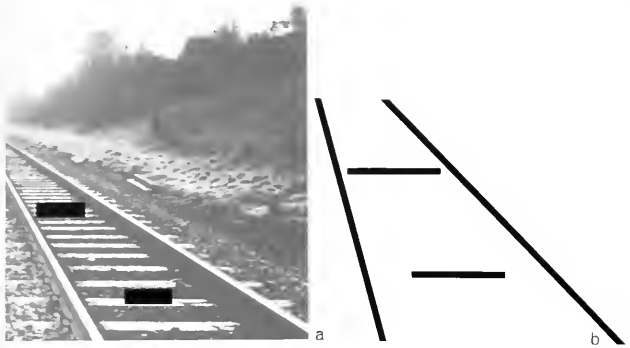
After-effects are examples of *successive* illusions, ones in which stimulation at one moment in time has consequences for the perception of subsequent events. Such illusions stand in contrast to *simultaneous* effects, whereby two or more stimuli presented together interact to cause an illusory effect. It turns out that there are very many illusions which stem from the presentation of several line stimuli at one and the same moment. Some of these illusions appear highly comparable to the tilt and width after-effects, and have therefore invited neurophysiological speculations about their causes comparable to the Békésy-type explanation of after-effects.

Consider, for example, **100**. The two central gratings of circular shape are in fact composed of vertical lines, but the lines look tilted away from vertical in the opposite direction to the angle of the lines that surround them. Could the surrounding lines have somehow induced a change in the activity profile dealing with the central lines, a change of shape comparable to that described for the tilt after-effect? Blakemore, Carpenter and Georgeson have proposed that this might be so, and have further suggested that *inhibition* between line detectors might play the same role in this context that adaptation played in the equivalent successive illusion, i.e. the tilt after-effect.

The idea here is that orientationally-tuned cells might inhibit one another, with each cell attempting to reduce the activity in its 'orientation-neighbours', so to speak. This inhibition would disturb the activity profiles produced in response to the vertical lines of **100**, and produce profiles characteristic of non-vertical lines: hence the illusion. The Zöllner illusion [**101**] provides another example of an orientation illusion which can be explained in the same general way.

What might be the purpose served by such inhibition? Why introduce such a curious process, one which is capable of generating distortions? In chapter 3, inhibition (combined with excitation) did the job of giving certain brain cells pattern-sensitivity, of making them possess optimal stimuli for activation, and this proved very useful in creating detectors as a step on the way to building up a feature description of a scene. But here inhibition is given a rather different role. What good might it serve?

Blakemore, Carpenter and Georgeson, like previous authors who have considered the role of inhibition in sensory systems, suggest that it might serve in certain circumstances to improve the 'sharpness' of the peak in the activity profiles of orientation detectors. In this way, it might facilitate the discovery of the activity peak which, in their theory of orientation perception, determines the perceived orientation of the line features in question. But if, as in this chapter, one regards the important aspect of an activity profile as not its peak but rather its overall shape, upon which the processes of interpretation execute the job of interpolation (see p. 53), then it is not so clear why a process which makes peaks more prominent (as inhibition certainly can do) is very advantageous. Within our general viewpoint, if inhibition is here doing a valuable job then it must be to improve the *whole* activity profile in some way, and not just help locate the peak. Thus it might be preferable to think of the proposed inhibition as re-creating a 'proper' activity profile, one which is better fitted overall for the interpolation/interpretation process. So perhaps what the inhibition is doing is 'taking out' of the activity profile, as it were, unwanted components of the total



102 (a) The Ponzo illusion The upper horizontal line appears longer than the lower one, perhaps because of a misapplication on the part of the brain of size-perception mechanisms appropriate to normal scenes (see railway lines figure) but not to simple line drawings.
(b) Richard Gregory writes of this illusion: 'The two rectangles superposed on [the] photograph of railroad tracks are precisely the same size, yet the top rectangle looks distinctly larger. [I regard] this illusion as the prototype of visual distortions in which the perceptual mechanism, involving the brain, attempts to maintain a rough size constancy for similar objects placed at different distances. Since we know that the distant railroad ties are as large as the nearest ones, any object lying between the rails in the middle distance (the upper rectangle) is unconsciously enlarged. Indeed, if the rectangles were real objects lying between the rails, we would know immediately that the more distant was larger.'

activity, components which might have been put in by a certain sloppiness in the orientational tuning of the line detectors. Engineers sometimes call this kind of process 'deblurring'. Perhaps the visual system has evolved certain deblurring mechanisms, using inhibition, which work very well for most scenes it has to deal with but which carry with them the price of creating illusions in other, less common scenes. All engineering systems are designed to work within a certain restricted range of situations. Perhaps illusions such as **97** and **98** are instances of the system working outside its specification, as it were, the specification drawn up by evolution via the process of natural selection in order to cope with natural and commonplace scenes, not artificial line drawings.

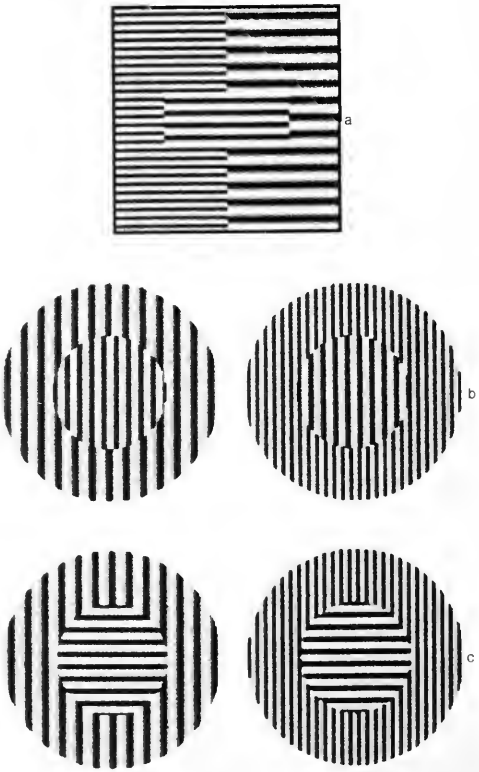
Other Geometric Illusions

There are a vast number of illusions which stem from collections of lines (the so-called *geometric illusions*), and I can present only a selection here. Moreover, it is important to note that whereas it has been convenient to use certain line illusions to illustrate the theme of this chapter (illusions as the psychologist's microelectrode), a large number of theories of these illusions exist, most of them couched in non-physiological terms. Limitations of space prevent me going into these theories, though **102** and its caption illustrate one much-debated theory of this kind. The interested reader is invited to consult the bibliography at the end of the book which will direct him to suitable sources describing these theories, should he want to follow them up.

An illusion which rounds off very nicely the major thrust of this chapter is **103**a. This shows a simultaneous width-of-line illusion invented by Donald MacKay. The horizontal lines in the inset central region are in fact of the same width throughout but they look strikingly different in the left and right halves of this region. That is, the region placed against the thin-line background appears to contain thicker lines than does the region placed against the thick-line background. This is a simultaneous version of the line-width after-effect **[94]**. An explanation in terms of inhibition between width-specific detectors can be offered for it, just as inhibition between orientation-specific cells was posited for the simultaneous tilt illusion of **100**.

It is an interesting fact that MacKay's simultaneous width-

103 Size-contrast effects (a) The central inset lines are the same width right across the figure but where surrounded by thick lines they appear narrower than where surrounded by thin lines. The effect can be enhanced if a length of thin black cotton is laid over the vertical midline of the figure.
(b) Another version of the effect described in (a). The lines in the small inset discs are of the same width in both halves of the figure but appear of different thicknesses.
(c) The size-contrast effect is at least much reduced if the inducing lines in the surround are set at right angles to the central lines. This shows that the effect is orientation-specific. See text



of-line contrast illusion is abolished, or at least substantially reduced, if the lines in the central region are set at 90° to those in the surround: compare **103**b with **103**c. This suggests that line-tuned processes are involved and, moreover, that if the effect is caused by inhibition, then this inhibition takes place only between cells with the same preferred orientation. In terms of hypercolumn machinery, this suggests that the inhibition is restricted to slabs with the same or similar orientation tuning.

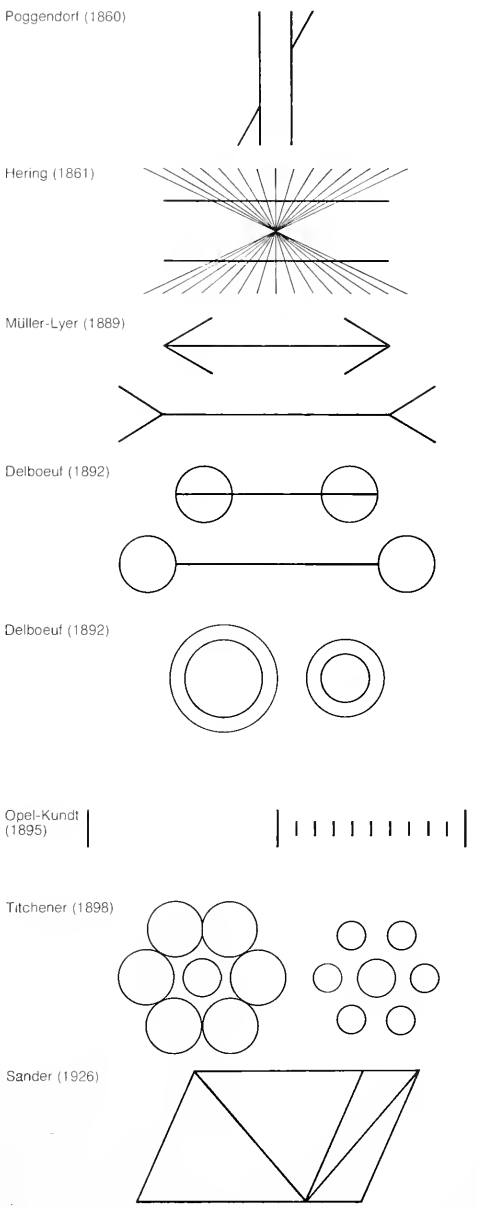
A range of other geometric illusions can be seen in **104**. The diversity of effects is impressive and gives ample scope for clue-hunting about visual mechanisms.

Three Approaches to the Problem of Seeing

This chapter has shown how the psychological approach to the study of seeing blends superbly with the neurophysiological and computational approaches described in earlier chapters. The logic of the psychological approach is: let us treat visual phenomena as the *output* of a device (the visual system) which takes as its *input* the retinal image, and ask what mechanisms must be present in the device in order to explain the output, given the input. Most frequently, the phenomena chosen for study are illusions, in the hope that when a system breaks down (i.e. when an illusion occurs), this is a good opportunity to gather clues about how the system works both normally and abnormally.

This psychological approach to visual mechanisms is the natural complement to the computational and neurophysiological approaches. The computational approach attempts to tackle the problem of seeing by trying to build an image-processing system, taking advantage of the powerful resources of the computer. The neurophysiological approach studies perceptual mechanisms directly by physiological and anatomical studies of the visual system. All three approaches have become blended in recent years, so that it is sometimes difficult to know who is the computer scientist, who the psychologist, and who the neurophysiologist. Thus it is increasingly common for any one visual scientist to wear any one of these hats, depending on the particular problem he is studying. Certainly it has become essential for the psychologist to familiarise himself with the literatures of computational and neurophysiological research so that he can take advantage of the tremendous strides forward which these disciplines have made over the past two decades or so.

But all in all, the current excitement in vision research stems from the amalgamation of all three approaches into one. This chapter has emphasised one aspect of this combination, the interrelationship between psychological studies of certain illusions, particularly but not exclusively after-effects, and the neurophysiology of hypercolumns. But of course this represents only a tiny fraction of a vast research literature.



104 Geometric illusions

5 SEEING OBJECTS

We are amazingly good at recognising objects by sight. Have you ever pondered, for example, our extraordinary ability for recognising cartoons of public figures? A few well-chosen strokes of the artist's pen capture a 'likeness' of the person, or his photograph [**105**]. We can see all three 'objects' – person, photograph and cartoon – as alike in some fundamental respect even though they differ enormously in their details.

At first sight, our success at recognising cartoons suggests that the information crucially required for recognition consists of the sharply-defined contours of an input image, all else not being used for the job. But this is not so, as **106** demonstrates. We can be just as good at recognising a blurred photograph as we can at recognising one which preserves only the sharp contours of the original.

Another feat of recognition is the ease with which we read different typefaces. We can easily see all the patterns in **107** as upper case Ts, despite wide variations in the nature of their constituent features. Our ability to cope with different specimens of handwriting is another achievement in the same high class. We have little difficulty in seeing which handwritten numbers are which, despite considerable differences in the details of numerals of the same value [**108**].

As noted in chapter 1, the visual system's fluent ability to recognise objects obscures its great achievements in this regard, and can mislead us into thinking that the task is a simple one. But its true complexity is so great that understanding how it is done at the brain's level of sophistication has so far defeated all those who have ever studied it, be they psychologists, engineers, neurophysiologists, mathematicians, or whatever. This failure is all the more noteworthy when it is remembered that large sums have been spent on investigating the 'object-recognition problem', as it is sometimes called, because of the immense industrial and military potential which successful understanding would bring. Competent visual robots, if competence is measured on anything like the human scale of achievement, are a long way off. None the less, advances in recent years have been sufficient to enable us at least to see more clearly the nature of the problems to be tackled, and to sketch in a conceptual framework within which research can be better directed. The aim of this chapter is to elucidate this framework.

Structural Descriptions

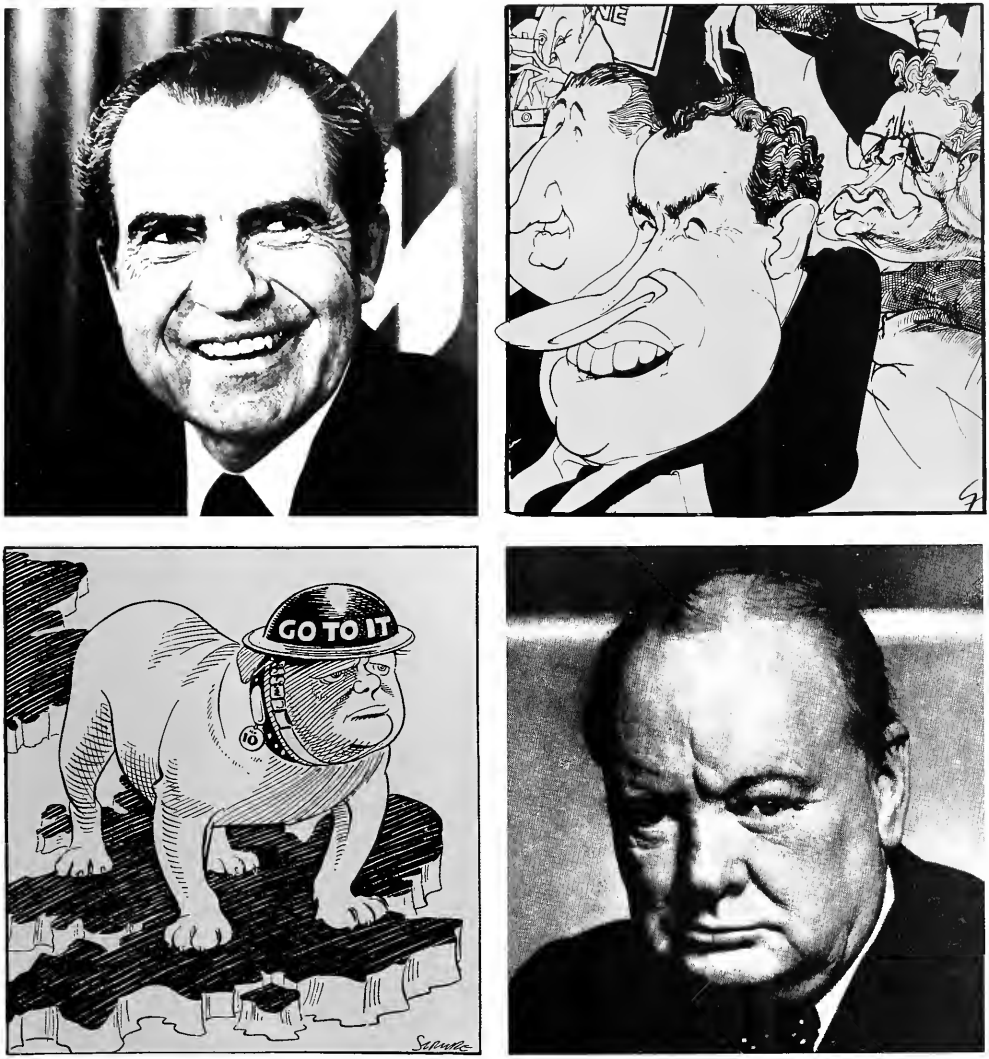
What, exactly, is it to 'recognise' something? When we recognise all the Ts in **107** as the 'same thing', in what does their 'sameness' lie?

Stuart Sutherland and others have argued that the best way of talking about the recognition of 'sameness' is in terms of *structural descriptions*. All the patterns in **107** are Ts because they all share a common *structure* which can be defined abstractly as follows: a T is a 'vertical bar' which is joined at its upper end to the middle of a 'horizontal bar'. As long as the terms 'vertical bar' and 'horizontal bar' in this structural description are themselves flexibly defined (see below), then all the Ts in **107** fit the description as stated, and therein lies their T-ness. Thus the structural description for a T is, if you like, a general formula for a T.

For Sutherland, what happens when we recognise any one of the patterns in **107** as a T is that our visual system examines the pattern, makes explicit its structure as a symbolic description, and then finds that this structural description matches a *stored description* for a T. The stored description for a T is lodged somewhere inside our heads as a permanent memory which we built up when we learnt to read. Of course, this whole recognition process is not something that we are consciously aware of, in that when we look at **107** and see all its patterns as Ts, we do not need deliberately and consciously to scrutinise each pattern and check that it does indeed fit what we were told long ago constitutes a T. Rather, the whole process is automatic and unconscious, at least in fluent adult readers dealing with reasonably legible text. But although we are unaware of what is going on, a stored description of a T must exist somewhere inside our brains and the act of recognising a T must involve matching the structural description obtained from the visual input against this stored description.

To take another example, when we recognise the cartoon of Nixon in **105** what happens is that we arrive at a structural description of this cartoon and find that it matches a stored structural description for 'Nixon' which we laid down as a memory when we first took note of him. This description might include such items as: 'Small eyes with bushy eyebrows, large protruding nose, receding hairline, frizzy hair, swollen cheeks, shaven but dark chin' etc. But these particular characteristics would need to be tied somehow to the structure of a face, which might itself be described as: 'Oval outline in which is set a horizontal bar [mouth] in the lower region, a vertical bar [nose] in the central region, and two blobs [eyes] in the middle-to-upper region' etc. When we recognise the cartoon of Churchill, the structural description for 'face' will be the same, but with a different set of details attached to it, leading finally to the recognition of 'Churchill'.

Expressing structural descriptions in terms of words



105 Photograph – cartoon pairs

should not be taken to mean that the symbols in our heads which provide these decriptions are themselves in wordlike form. They are not verbal symbols as such, although it is possible to use the language analogy and call these symbols 'visual words', as each stands for a particular seen 'thing', just as verbal words stand for 'things' (as well as much else, of course). It is not easy to think about 'visual word' symbols, as we discovered in chapter 1 when the whole business of seeing as a symbolic scene description was introduced. But we have no choice in the matter: quite simply, seeing is a matter of building up symbols for entities in the scene before us, and that is that. We just have to get used to the idea that when we look at a T, or a picture of Nixon, or whatever, myriad symbols ('visual words') become activated inside our brains, and that amongst this multitude is one standing for T-ness, if a T is present, or one standing for Nixon, if he or a

106 Distorted photographs It is easy to recognise Groucho Marx even when his photograph (a) is blurred (b) or reproduced in outline (c) (technically low and high pass spatial frequency filtered respectively).

107 The letter T in different typefaces

picture of him is present. These symbols might be regarded as 'high-level' ones as they derive from the activation of 'low-level' ones that represent features. Look around at the scene before you and reflect that for everything you 'see', be it a simple feature, a cluster of features, or a battered armchair, the 'seeing' must have been mediated in each case by some kind of symbol becoming selected or activated inside your brain. This symbol stands for the visual feature or object and constitutes its explicit description. When the symbols which match the structural descriptions emanating from the scene before you have been chosen from all the vast number of possibilities, most of which lie dormant inside our heads most of the time, then the collection of activated chosen symbols quite simply *is* our visual experience at the relevant moment. It is a strange, even fantastic, realisation, but an inescapable one.

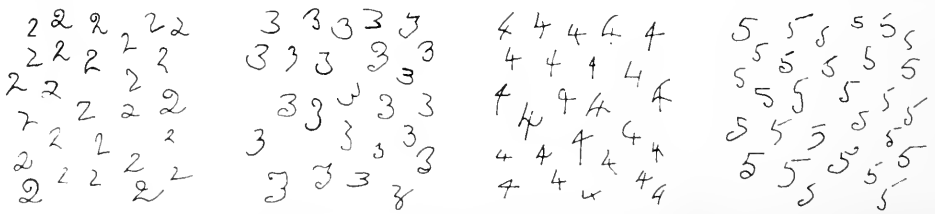
Talking about recognition as the business of matching structural descriptions is not so much a full-blown theory of recognition as a clear way of stating what a successful theory must accomplish. First, it must explain how suitable abstract symbolic descriptions of the structure of an input image are built up. Second, it must tell us how such descriptions can be stored. And third, it must tell us how any given structural description obtained by looking at an input image can be matched to the 'right one' of all the countless stored descriptions which we possess – and how this can be done, usually, in a fraction of a second. These are tall orders for any theory and perhaps you can begin to appreciate why the 'object-recognition problem' continues to be such a baffling one.

Template Matching

The idea of recognition via structural descriptions needs a great deal of development, then, before it could be said to constitute an adequate theory. But it does have the merit of indicating certain approaches to the recognition problem which are *in*adequate, if recognition performance is to match human capabilities. One such inadequate scheme is *template matching*. Chapter 2 outlined the difficulties of building a corner detector template, and all these difficulties would apply equally to building object templates. But in addition, it would prove quite impossible to build a template which could cope, for example, with all the variations and distortions of the Ts shown in **107**. One could never arrange a cluster of photocells, however cleverly or cunningly, which would give a strong combined signal when a T was present, and not otherwise, for all the different types of T which we can indeed recognise.

Template pattern-recognition, however, does play a valuable role in many man-made recognition systems in industry. Recognising account numbers on bank cheques is a good example [**109**]. The number-recognising machine has in it a set of templates, one for each numeral, and works by seeing which one best fits the number on the cheque which it is trying to read. The process is in principle entirely similar to that tried for the corner template in chapter 2, except that the details of the machinery are different. The system is made to work effectively in this context by using a specially designed set of numerals whose shapes are very different from one

108 Handwritten numerals



another to the man-made eye, and by not allowing matters to be complicated by any variations in brightness, contrast, shape and size of numerals, etc. So here is a perfectly good recognition system for the task it tries to tackle, but this task is a very simple and mundane one by human recognition standards. The advantage of this template system is, of course, that it can do its simple task extremely quickly and very economically.

From Feature Descriptions to Object Descriptions

At this point let us recapitulate what has been said so far about 'scene descriptions' in this book.

In chapter 1, the idea was first introduced that seeing is a symbolic scene description which makes various aspects of a scene explicit. In chapter 2, the notion of a feature description was explained and some of the problems in arriving at a set of symbols standing for features were outlined. In chapter 3, the business of obtaining a satisfactory scene description was pursued further, via an account of the visual machinery of striate cortex which seems to be devoted to a computation leading to the activation of neural feature symbols. In chapter 4, psychological approaches to studying feature description mechanisms were explained in connection with an account of various illusions. Now we have reached the point of wanting to proceed from symbols standing for features to symbols standing for objects. This general idea was briefly referred to in **31**, which showed how in principle one might try to obtain a symbolic object description from a symbolic feature description, using as an illustration the possibility of obtaining a scene description 'rectangle present' by noting that four corners were present, suitably connected together by four edges.

Somehow, then, the visual system proceeds from its 'low-level' feature description to its 'high-level' object description. This process does not take place all in one step, however. When we said a T was a vertical bar joined at its upper end to the middle of a horizontal bar, note that the concept of a 'bar' is itself quite a general one and not to be equated with a simple feature, as a glance at **107** will confirm. A simple feature description of the kind dealt with in chapter 3 was something like: 'Line present in location dealt with by the hypercolumn concerned, orientation 25°, contrast low, fuzziness medium.' Clearly, this type of feature statement is very different from one involving the concept of 'bar' required to cover all the diverse Ts in **107**. A 'bar' is itself a kind of object and requires a symbolic description all of its own at a much higher level than that dealt with by simple features. Thus a 'bar' might be defined as 'an aggregate of features that has one axis about five times as long as the other axis'. And a symbol representing such a perceptual entity must be activated as the first step in recognising a T, because the definition of T-ness is couched in terms of such bars.

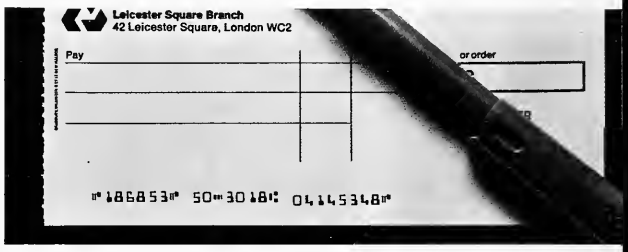
A similar consideration applies to the structural description of Nixon. An eye is really a very sophisticated object-concept in its own right, as well as serving in the definition of even more sophisticated object-concepts such as 'Nixon', and it needs to be recognised as a prior step in recognising a face of (almost) any kind. Moreover, the structural description used in recognising an eye could contain terms such as 'bars' (for eyelashes), 'circles' (for pupils), and so on – terms which again need to be 'unpacked' as just described for 'bars'.

Thus object recognition must involve the activation of symbols at many levels of complexity, so that some can help in the construction of others. In explaining something of how this is done, it is best to begin at the beginning and consider what happens immediately after a low-level feature description has been obtained.

Structures within Feature Descriptions

In chapter 3, we left the neurophysiological story of vision at the point where each part of the input image had been analysed by a hypercolumn, a piece of neural machinery which 'looked' at its particular part of the input (which we called its hyperfield) and assessed what feature type was present (line, edge, slit), and what its characteristics were (orientation, contrast, fuzziness). The next problem is to find what structures are present within the mass of feature information provided by all the hypercolumns together, so that these structures can then be recognised by comparing them with stored object descriptions.

The problem of finding structures within a feature description is illustrated by **110**. The scene being observed contains a teddy bear, and a grey level image of this scene is shown. The grey level image was obtained from an input image, and you should refer back to **2** if you need a reminder on this sequence. The next item in the series is a feature description. This is shown by a mass of small line segments found all over the image by hypercolumn-type machinery. (In point of fact, this feature description comes from a computer vision system devised by David Marr. However, the theory described in



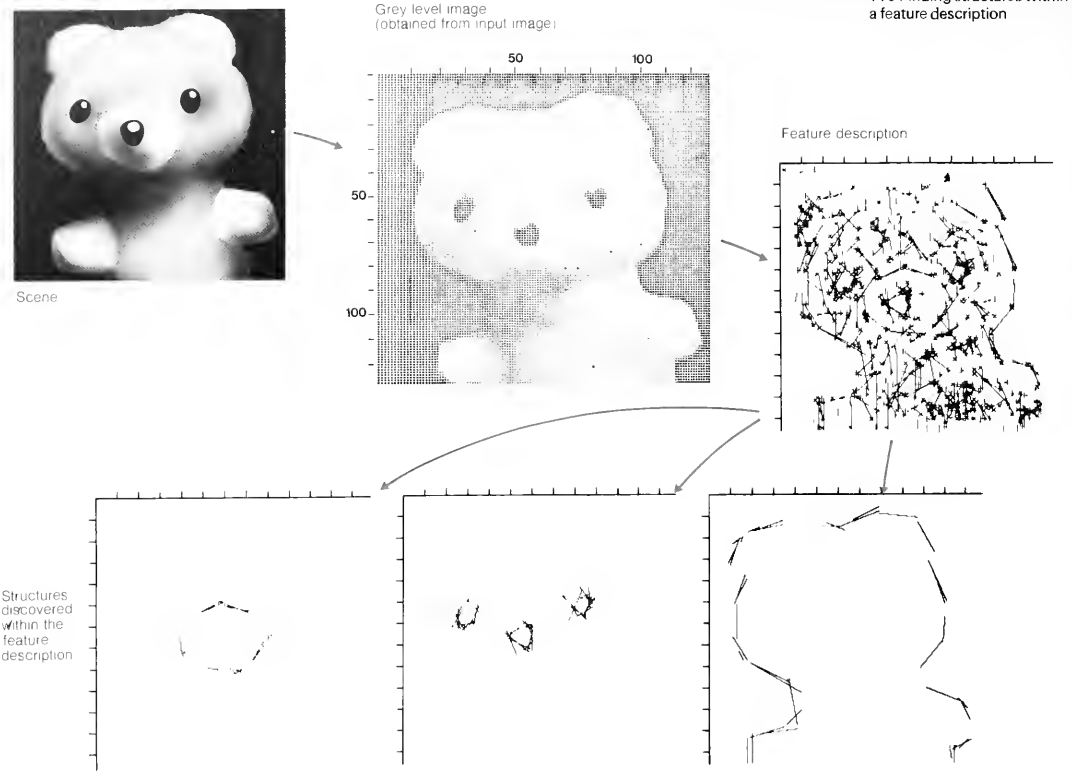
109 Special numerals on a bank cheque

chapter 3 about the way hypercolumns might work was heavily dependent on Marr's ideas and so we can reasonably assume that hypercolumns would produce a feature description rather like the one shown.)

The feature description illustrated in **110** is only an abbreviated version of the full description. That is, the picture simply shows each line segment as having a location and an orientation, but in fact each segment would also have associated with it a statement of its type (line, edge, slit), its fuzziness value, and its contrast. The picture would become impossibly complicated if all this information was somehow included, but we must think of it as being present. Thus each line segment must be regarded as representing for illustrative purposes the existence of a feature symbol which is much more richly defined than a simple printed line can possibly suggest.

The first thing to notice about the feature description, even in this simplified form, is its messy and confused state. Feature statements appear almost everywhere. This is perhaps not so surprising when you look carefully at the teddy bear itself and note that your own visual system also can detect many features with varying properties in lots of different places. But the pro-

110 Finding structures within a feature description



fusion of these different features is not usually immediately apparent to us because our visual system can readily *group together* clusters of features, and thereby see structures within the otherwise unwieldy mass of low-level feature statements: and the awareness of these larger structures, as it were, pushes the awareness of their component features into the background. Three such feature clusters have been separated out in **110**, and each is shown below the feature description as a structure in its own separate picture. These structures – 'perceptual groupings' if you like – are the overall outline of the teddy bear, the teddy bear's eyes and nose, and the teddy bear's muzzle. Each structure was found, by Marr's computer vision system, 'hidden' within the confused low-level feature description. The manner of their discovery represents a considerable achievement because it goes a long way towards solving a fundamental difficulty in object recognition, called the *segmentation problem*.

The segmentation problem is the usual title given to the very difficult job of taking a mass of feature statements and finding out what regions go together and hence form structures. Our visual system is very good at it but it has proved a big stumbling block for machine vision systems. By way of explaining Marr's approach to segmentation, it is best to turn next to the *principles of grouping* which seem to operate within our own visual system.

Principles of Grouping

Look at the op art painting by Victor Vasarely in **111** and observe that your eye is immediately 'caught' by certain structures within it, even though from one point of view all the picture contains is numerous individual elements of varying shape in a regular pattern. For example, you instantly see the round elements as a distinct pattern standing out from the straight-edged ones. Also, a group of diamond shapes about a third of the way down the painting stand out as a cluster, contrasting with the square shapes surrounding them. These are instances of *perceptual groupings*, each one a 'bringing together' into a perceptual structure of certain elements which share a common property of some kind. In the two examples just given, the common property is *similar shape*, here either roundness or angularity.

Notice also that as you continue to look at the picture other perceptual groupings become manifest. For example, you probably see the round elements further divided into strings.

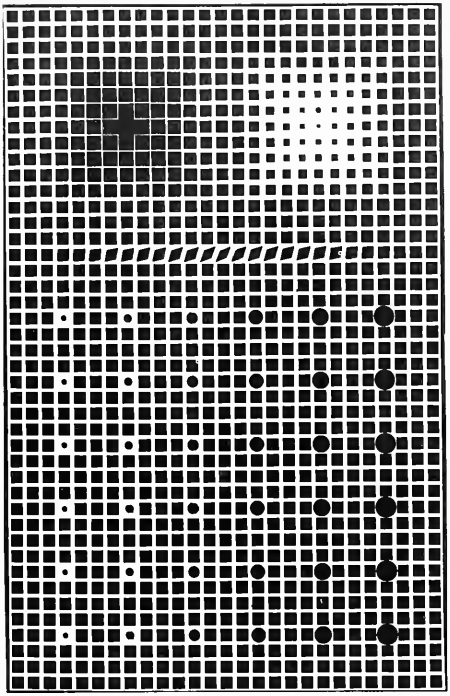
Indeed, many readers might well have seen these strings first and the whole group of round elements second. The grouping principle which brings the constituent elements of the strings together is that of *similar size*. So we have six vertical strings of round elements, each set apart from the others by virtue of the size of its members. The strings on the right of the painting, composed of the largest discs, are the ones which most clearly group themselves together on this basis. The grouping principle of size can also be seen at work in the upper part of the painting, where large square elements form one perceptual cluster and small square elements form another.

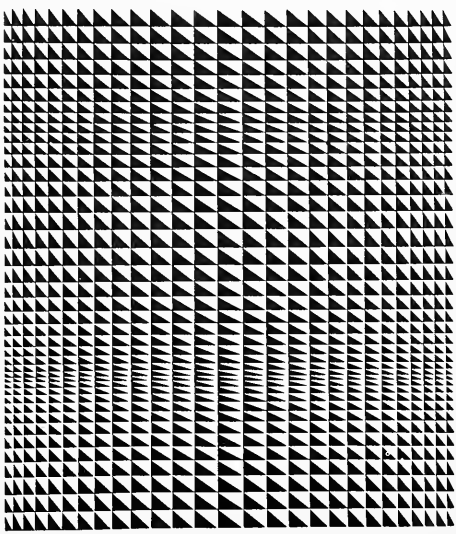
Another principle of grouping, that of *continuity*, is clearly illustrated in Bridget Riley's painting *Straight Curve* [**112**]. The painting is composed only of small black triangles, all with straight sides. The curves are the result of your visual system grouping together edges of roughly similar orientation into higher-level perceptual structures, so producing the apparent presence of curved lines sweeping across the painting. The curves are 'put in' by your visual system, which links individual elements together on the grounds of continuity in their orientations. The power of this grouping principle is well attested by Fraser's spiral [**4**], the first illusion shown in this book and one of the most dramatic of all. The illusory spiral is seen because the visual system finds that the best continuity between individual elements is that given by a spiral, rather than by the concentric circles which are really present.

A very simple principle of grouping, and yet one of the most important, is that of *proximity*. In **113**a, for example, we see the row of discs broken up into three pairs, each one a result of the proximity of pair members to one another. In **113**b, all the discs are equally spaced one from another and we can therefore see rows of discs or columns of discs with equal ease. If the matrix of dots is flattened, however, as in **113**c, then columns become perceptually dominant because the flattening has brought certain discs into close proximity and the visual system picks up this fact and 'reports columns' accordingly. A dominance of rows [**113**d] is equally easy to produce by an opposite alteration of the shape of the matrix of discs, again showing the proximity principle at work.

The grouping principle of *closure* is illustrated in **114**. In the upper row of this figure are a set of brackets which can be grouped in various ways. In the lower row, lines have been drawn between the tips of the brackets to give shapes with a 'closed' contour. In the left half of this lower row, the closed-contour shapes are outline television screens, and in the right half they are columns. In each case, the principle of closure, of seeing elements which produce a completed form as a single structure, dictates that the ambiguous vertical brackets are grouped to give enclosed entities.

These and other grouping principles were first emphasised and studied as important clues about perceptual mechanisms by the *Gestalt* psychologists, working in the 1920s and 1930s. The German word *Gestalt* is not readily translated but roughly it means 'form' or 'configuration'. The Gestalt psychologists noticed that many perceptions exhibit the active grouping of elements together, as just described, and from this realisation they developed their favourite saying: '*The perceptual whole is more than the sum of its parts.*' That is,





111 [top] Victor Vasarely. *Supernovae*

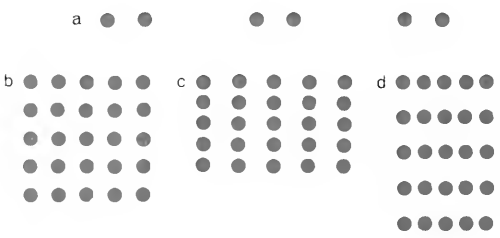
112 [bottom] Bridget Riley. *Straight Curve*

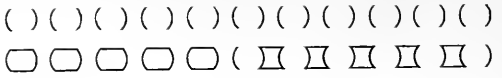
parts are *not* treated as separate and isolated entities in perception. Rather, parts interact to produce a *Gestalt* which can differ markedly from what would be expected if parts did not affect one another.

Interestingly, Marr has found that grouping principles of the type discovered by the Gestalt psychologists are very valuable in 'segmenting', or dividing up, a low-level feature description. Most importantly, he has found that general principles of this sort can go a long way towards achieving a successful segmentation without the need of any special advance information about the nature of the scene from which the feature description comes. Thus his computer system 'found' the structures shown in the lower part of **110** using a host of these grouping principles and without 'knowing' that it was looking for a teddy bear or parts of a teddy bear. The outline-of-the-bear structure is a grouping of features achieved primarily on the basis of continuity of orientation [cf. **112**]; the eyes and nose demonstrate a grouping of features primarily on the basis of proximity and similarity [cf. **113**], and the muzzle shows the principle of closure at work, features being brought together which make a completed oval shape [cf. **114**]. In fact, obtaining each structure did not uniquely depend upon any one principle of grouping because all principles were used at once and therefore combined to give the end results shown. Moreover, it must be remembered that the feature description given in **110** is very abbreviated in what it shows, and that in fact there was much more scope for bringing together similar features, either on the basis of type (edge, line, slit), or of shared characteristics (such as contrast and fuzziness), than is readily apparent from the illustration (which emphasises only location and orientation of features).

Actually, the feature description shown in **110** is not quite the 'raw' feature description which we have so far regarded it as being. Even to arrive at this stage, Marr's computer program used certain grouping processes of a primitive kind to bring together the very low-level feature statements produced by hypercolumn-type machinery. Thus if you look carefully, you can see that some of the feature elements shown in this description are quite large in extent and transcend what would be expected from each individual hypercolumn looking at just its own limited hyperfield. We will not describe in any detail

113 The grouping principle of proximity (a) The row of discs appear as three separate pairs. (b) Rows and columns seen with equal ease. (c) Columns dominate because of proximity of discs along vertical axis. (d) Rows dominate because of proximity of discs along horizontal axis.





114 The grouping principle of closure (see text)

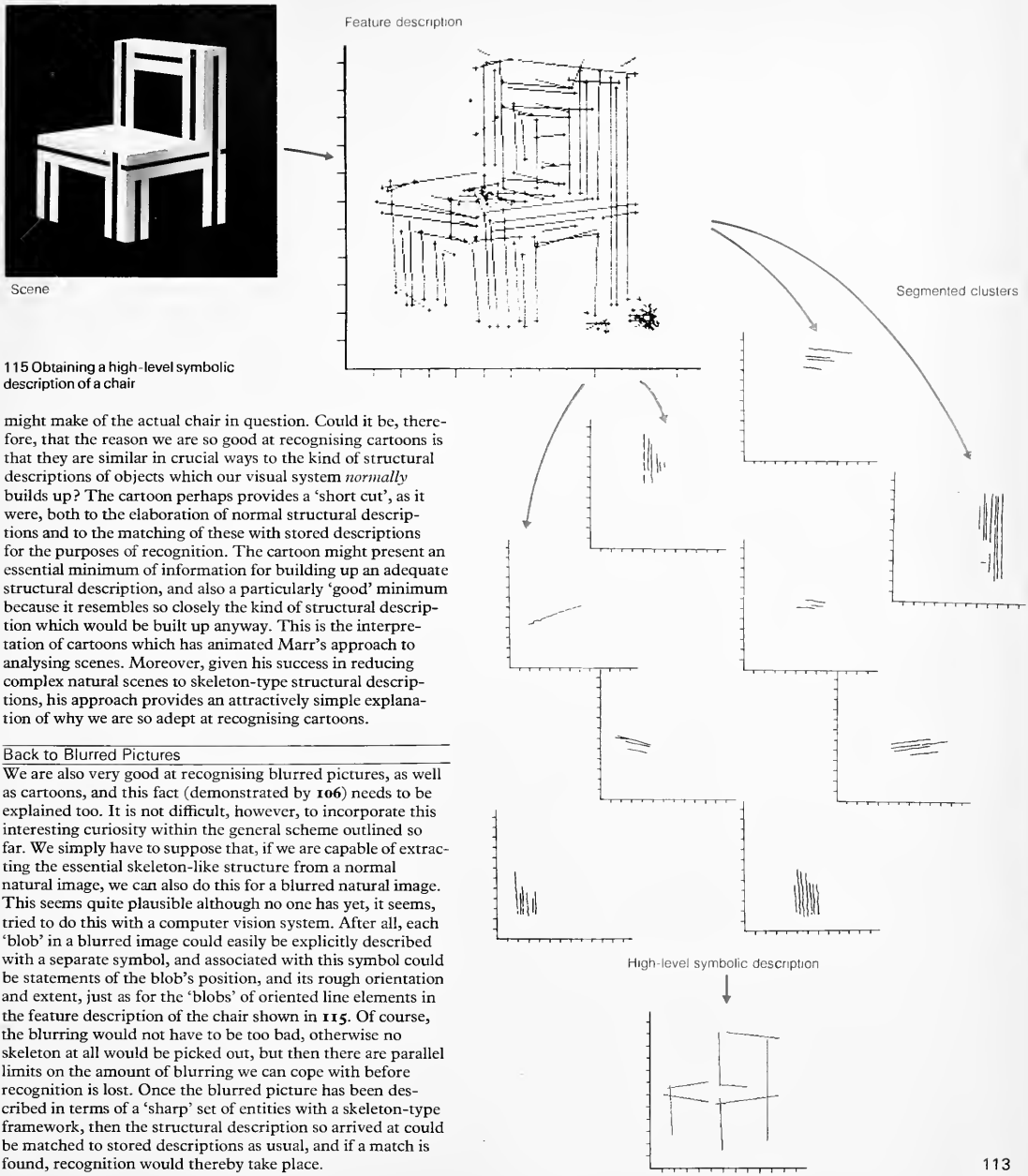
these early grouping operations which Marr has found it convenient to embed in his system, but simply note that they are similar in type to those already described, but more restricted in scope (e.g. they link adjacent features of almost exactly similar type and orientation). Marr uses a special term for a feature description produced by just a bare minimum of grouping processes: he calls it the *primal sketch*. It is a symbolic description of the various intensity changes in the input image which is, he suggests, at a level of complexity roughly similar to that of the low-level features in the 'visual image' of which we are conscious. All subsequent analysis and interpretation use the primal sketch as their supply of information and not the preceding measurements on which this sketch was itself based. Thus the primal sketch, or the point where 'visual experience' begins, is the sole input to later processes of object recognition, thinking, problem solving, etc.

Remember that although for the purposes of illustration in **110** the structures the system discovers are printed out as a collection of their constituent features, in Marr's program there would be a symbol representing each one. Or as Marr would put it, each would be *explicitly named* and thus capable of being referred to by subsequent object-recognition processes. In other words, each structure would be represented in Marr's computer program with a label. The label would have associated with it a list of its constituent features, its location, etc., but the label – the name – would be the unifying symbol which gives the cluster of features its structural identity. Thus we can equate the 'naming' by Marr's program of a discovered structure with our visual system 'noticing' one of the structures in, say, the Vasarely painting [**111**]. Our 'noticing the structure' is an act of explicit description, and there must be some symbol inside our heads which *is* this description. This is not an easy concept to grasp, but hopefully enough has been said by now about explicit descriptions and seeing for you to understand the central point being made. In any event, I will give another example in the next section to make things clearer.

Back to Cartoons

In **115** a chair is present in the scene being observed, and Marr's program produces its customary elaborate feature description (in this instance the by now familiar grey level description is left out). At the next stage clusters of features are separated out as belonging together in different regions of the scene: these are shown in **115** printed out in separate pictures. The clusters were found mainly by grouping together elements with similar orientation. Each cluster is then described as a unit with a given position, orientation and extent. This unit would be labelled with a symbol, and the characteristics just mentioned would be attached to this symbol. If the 'things' represented by these symbols are printed out, then a skeleton outline of the chair appears. Each 'thing' is a perceptual unit, if you like, and Marr has shown us how they can be extracted (i.e. explicitly described) from the input image of a scene.

You will doubtless notice at once the similarity between the skeleton chair of **115** and the kind of sketch which a cartoonist



115 Obtaining a high-level symbolic description of a chair

might make of the actual chair in question. Could it be, therefore, that the reason we are so good at recognising cartoons is that they are similar in crucial ways to the kind of structural descriptions of objects which our visual system *normally* builds up? The cartoon perhaps provides a 'short cut', as it were, both to the elaboration of normal structural descriptions and to the matching of these with stored descriptions for the purposes of recognition. The cartoon might present an essential minimum of information for building up an adequate structural description, and also a particularly 'good' minimum because it resembles so closely the kind of structural description which would be built up anyway. This is the interpretation of cartoons which has animated Marr's approach to analysing scenes. Moreover, given his success in reducing complex natural scenes to skeleton-type structural descriptions, his approach provides an attractively simple explanation of why we are so adept at recognising cartoons.

Back to Blurred Pictures

We are also very good at recognising blurred pictures, as well as cartoons, and this fact (demonstrated by **106**) needs to be explained too. It is not difficult, however, to incorporate this interesting curiosity within the general scheme outlined so far. We simply have to suppose that, if we are capable of extracting the essential skeleton-like structure from a normal natural image, we can also do this for a blurred natural image. This seems quite plausible although no one has yet, it seems, tried to do this with a computer vision system. After all, each 'blob' in a blurred image could easily be explicitly described with a separate symbol, and associated with this symbol could be statements of the blob's position, and its rough orientation and extent, just as for the 'blobs' of oriented line elements in the feature description of the chair shown in **115**. Of course, the blurring would not have to be too bad, otherwise no skeleton at all would be picked out, but then there are parallel limits on the amount of blurring we can cope with before recognition is lost. Once the blurred picture has been described in terms of a 'sharp' set of entities with a skeleton-type framework, then the structural description so arrived at could be matched to stored descriptions as usual, and if a match is found, recognition would thereby take place.

Figure and Ground

The segmentation problem is closely related to what psychologists call the *figure-ground phenomenon*. When we look at **116** we can see either a white cross standing out against a black surround, or a black cross on a white surround. To put this another way, we see **116** as a figure-on-a-ground, and the reason this is interesting is that it shows us that sorting out a figure from its ground is an *active* perceptual process, as it must be if for **116** the visual system can come up with two different answers at different times. We could just as well describe this as the visual system 'segmenting' **116** differently at different times.

The cue used for segmentation in **116** is brightness: either the white elements are brought together, or the black ones. This is a very straightforward basis for grouping, but a valuable one. Colour can also form a good basis for grouping. Indeed, colour vision may have initially evolved for this very reason.

116 The figure-ground phenomenon: we can see either a white cross on a dark ground, or a black cross on a light ground.

W	E	D	N	E	S	D	A	Y	E	A	R	U	O	H	Y
I	E	V	E	N	I	N	G	E	S	U	N	S	E	T	A
N	A	E	Y	A	D	N	U	S	G	S	E	C	O	N	D
T	U	P	K	N	I	G	H	T	N	A	R	E	T	O	D
E	G	S	U	M	M	E	R	E	E	T	U	N	I	M	I
R	U	Y	A	D	E	K	H	R	H	U	M	T	M	A	M
D	S	L	E	N	T	T	V	D	G	R	O	U	E	Y	W
E	T	U	U	X	Y	L	I	A	D	D	M	R	E	V	E
C	R	J	A	N	U	A	R	Y	R	A	E	Y	W	E	N
E	T	S	A	U	T	U	M	N	F	Y	N	C	V	O	O
M	O	N	D	A	Y	N	O	O	S	C	T	W	A	N	O
B	D	L	L	H	A	N	P	Q	E	L	C	Y	C	D	N
E	A	I	C	X	D	A	T	E	A	R	E	T	S	A	E
R	Y	R	E	S	I	R	N	U	S	P	R	I	N	G	I
B	A	P	Z	N	R	T	O	M	O	R	R	O	W	O	N
M	L	A	T	E	F	O	R	T	N	I	G	H	T	M	L

118 How many words can you discover?

Many of the ambiguous figures in chapter 1 illustrate figure-ground phenomena very nicely (e.g. **17**). Whenever figure and ground oscillate, as in these ambiguous pictures, what is probably happening is that application of the various grouping principles does not produce a single 'solution', but competing perceptual structures are 'discovered' at different times. In an ordinary scene, however, it is almost always the case that a unique answer is found to the problem of segmenting the various elements of the feature description. This is the 'true' answer in the sense that the structures found are usually those which actually correspond to reality.

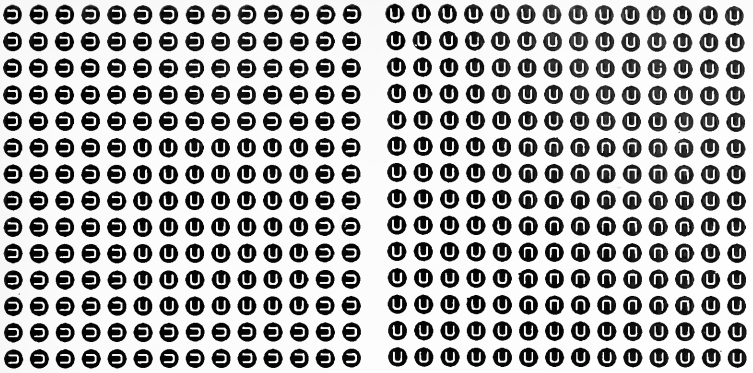
Figure-ground phenomena are nowhere better illustrated than in a recently introduced type of doodling pad [**117**, plate 7]. The patterns provided by each page of these pads are capable of endlessly different interpretations, as the two examples provided in **117** by my young daughters make clear. And as one looks at the patterns, a large variety of different perceptual organisations emerge as the grouping principles incorporated in our visual system constantly produce new inventive figure-ground solutions to the multiply ambiguous input. What pattern is 'really' there? There is no answer to this question. And indeed, even in normal scenes, where it is quite clear which pattern it is 'sensible' to see, there are always other theoretical possibilities which go unnoticed.

The automatic character of figure-ground separation tends to make us take it for granted. But another kind of puzzle-pad currently on the market provides its fascination by taking away grouping clues, leaving a task of finding hidden words which is surprisingly difficult [**118**]. Normally we group together letters belonging to a word on the principle of proximity. But in **118** this clue is removed, because all the constituent letters are equally spaced. This forces us to use high-level knowledge about words to group individual letters together, and the inefficiency of this approach is clearly brought about by the difficulty of finding the hidden words.

Segmentation by Texture

Principles of grouping, and related matters to do with figure-ground separation, are often studied by psychologists and others using patterns with carefully controlled *visual textures*. For example, look at **119** and you should be able to see fairly easily within it a 'hidden' square area. The square is segmented as a region separate from the rest of the figure by its slightly different texture. The difference in this instance is that the small U shapes within the square region are upright (U), whereas in the surround they are lying on their sides (⊃). Our visual system detects the difference and we see the square separated out as a discrete perceptual entity.

Bela Julesz and his colleagues, who devised **119**, have tried to work out how the visual system goes about identifying regions on the basis of their textures by finding instances of texture differences which we *cannot* discriminate, and using these as clues about how we deal with those which we *can* manage. Consider **120** for example. Can you see the hidden square this time? Probably not, unless you look very carefully and deliberately, using high-level cognitive processes (e.g. memorising some elements, then checking others against them, and so on). In other words, the square in **120** does not immediately stand out as figure against ground, as does the square in **119**, and this gives a clue about mechanisms of immediate texture vision. In particular, it suggests that differences in orientation can be used (because such differences exist in **119**), whereas simple shifts in the position of



119 [left] A square of texture elements differing from those around it can be seen fairly easily.

120 [right] A square of texture elements differing from those around it also exists here, but now it can be detected only with careful scrutiny.

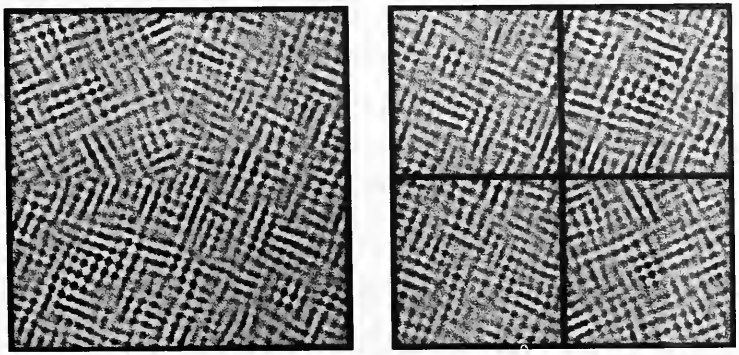
certain textural elements are not so helpful (the square in **120** is made of ∩s, the surround of Us, so the only difference between them is in the position of the 'cross piece' of the component shapes).

John Mayhew and I have recently discovered, however, that orientation differences between two regions can sometimes be quite marked and yet still go unnoticed. Consider for example **121**. It is easy to see a square standing out as figure, its 'woven texture' having a different orientation from that of the surround. But can you readily see the selfsame square standing out in **122**? Almost certainly not, and yet the *only* difference between these two figures is that in the latter the boundaries between sub-regions of the picture have been obscured with a black line. Look carefully and check this surprising fact for yourself. The conclusion seems to be that segmentation by analysis of visual texture may sometimes depend crucially on detecting differences at the *boundaries* between regions, and that when these are masked, even by quite a thin line, figure-ground perception via texture perception may fail. This is surprising because by far the largest part of the texture regions in **122** are quite unaffected by the black borders, and one would have thought in advance that 'region finding' should have been able to proceed quite happily on this basis.

Camouflage and Constraints

Principles of grouping take advantage of certain typical properties of objects in the actual world. That is, it usually *is* the case that objects or sub-regions of objects actually *are* made up of similarly textured elements, or that their boundaries *are* defined by features with continuity of orientation, and so on. This is how everyday objects are made up, and the reason why grouping principles can be used so successfully by Marr is that these principles take advantage of this fact. Putting this in technical language, one can say that grouping principles take advantage of certain *constraints* on feature relations shown by typical objects.

Not all objects show these constraints, camouflaged ones being a good example to the contrary. In **123** the reason the aircraft is so difficult to spot is that deliberate attempts have been made to make its real boundaries conflict with normal rules. Thus the principle of continuity, for example, groups together the edges of the camouflage blobs on the sides of the aircraft and makes these perceptually dominant at the expense of the real edges of the object. Many examples of animal camouflage can be interpreted in the same general way, i.e. as creating boundaries within and across the animal which conflict with the normal constraints on how features from



121 [left] The differentiated quadrant is easy to discern, the orientations of its components being twisted with respect to those of its surround.

122 [right] Which is the odd quadrant out? The pattern is identical to 121 except for the black lines, and except for being rotated as a whole to a new position. See text for details.

123 Military camouflage

objects cluster together. The examples given in chapter 1 [20, 21] further illustrate the point.

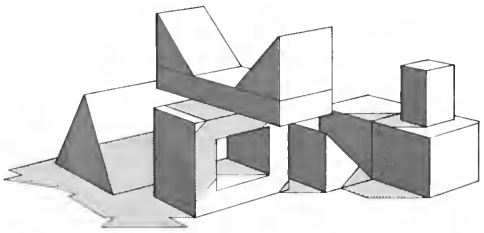
Computer Vision and the Blocks World

The best worked out example to date of constraint analysis in the field of computer vision is the work of David Waltz in connection with the so-called *blocks world*. A typical blocks world scene is shown in **124**. It consists of a jumble of building blocks of the kind a child might play with. The problem that Waltz tackled (following some pioneering but less detailed analysis by Roberts and Guzman) was how to segment the scene into its component parts ready for recognition. Our visual system does this task very readily. We can easily report the existence of the arch, the triangular block, the wedge resting on cubes, etc. But a closer look indicates that this segmentation is a clever achievement of visual processing, and not something immediately 'obvious'. How do we actually come to see adjacent regions as belonging to one block or another?

Waltz has devised a computer program which solves this problem in a very thorough fashion. It can segment the scene shown in **124** into its constituent blocks, despite the presence of shadows. Indeed, the shadows are actually helpful to Waltz's program, which works because it takes advantage of constraints which exist between regions belonging to the same object in block scenes. For example, if the program finds an arrowhead junction, such as that circled in **125**, then it knows that only three sensible possibilities exist for labelling its constituent lines (**125**b,c,d). All other labellings would depict edge junctions which simply could not exist in a real scene. The program then uses other constraints to choose which of the three possible labellings is the right one. For example, advantage is taken of the constraint that a convex edge at one junction cannot change along its length so that it becomes a concave edge at a neighbouring junction: again, this would be physically impossible. From this and other constraints (e.g. those imposed by shadows) Waltz's program excludes mutually incompatible line labellings, thereby discovering the correct ones, and from this point the segmentation of the scene into its constituent blocks is easy, using boundary labellings.

But therein lies the snag. Waltz's program uses constraints which are valid for the blocks world, but these do not solve the segmentation problem posed by many natural images (for example, the teddy bear dealt with by Marr's program: **110**). That is, the constraints employed by Waltz are more specialised than the grouping principles of Marr, and hence less generally applicable. Even so, Waltz's achievement in writing his extremely clever program represents an impressive milestone on the path towards high-grade computer vision. As an example of what can be achieved by constraint analysis, it has no peer at present.

124 A 'blocks world' scene of the kind easily segmented by Waltz's program

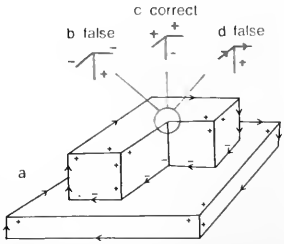


Segmentation via Object Recognition

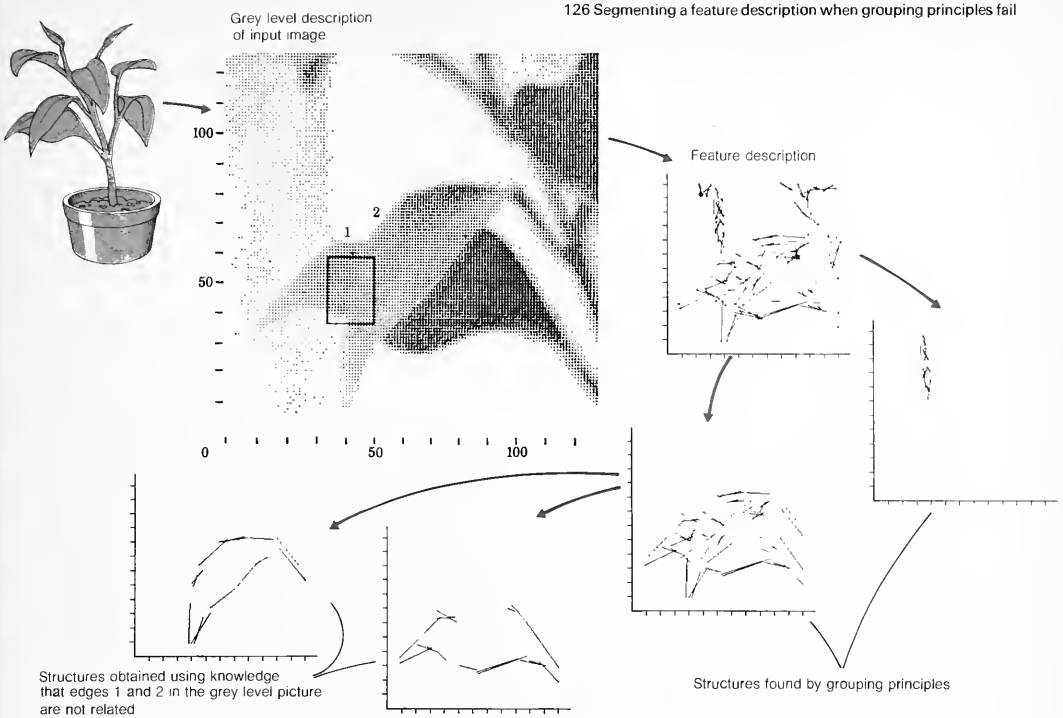
Marr's program attempts to segment its feature description without help from object-recognition processes. So the teddy bear outline was found as a 'perceptual structure' without the program being told to 'look for' a teddy-bear shape. Nor was the success of the program dependent on its discovering halfway through that a teddy-bear was present, and then using this information to search out his eyes, nose, etc. All these structures were found without the program 'knowing' what it was doing. There are occasions, however, when ambiguities in segmentation crop up, and then the use of information from higher-level processes concerned with object recognition can be helpful.

An example of this given by Marr is shown in **126**. The scene here is a bowl of flowers, and the part of the grey level description concerned with the leaves is shown as the part being segmented into constituent structures. The feature description shows the by now familiar mass of feature statements which requires interpretation. The structure in the upper left area is easily discovered using rules of proximity and similarity. But the leaves themselves remain stubbornly grouped together, and not separated out as the two distinct objects which they so obviously are – to our visual system. The problem is that in the area of overlap of the two leaves (shown as an inset rectangle in the grey level description), the initial feature description cannot find enough line features to demarcate the two leaves clearly, so closely matched are the grey levels in these areas. Marr found that he could get his

125 A line drawing of a blocks scene without shadows Object boundary lines are labelled with a >. Lines separating adjoining regions are labelled either concave (−) or convex (+) according to surface geometry. Three possible types of arrowhead labelling (b, c, d) are shown as candidates considered by Waltz's program for the circled arrowhead junction. The two false ones are eliminated because they do not fit in with constraints imposed by other junctions (forks, Ls, Ts, and other arrowheads) in the scene.



126 Segmenting a feature description when grouping principles fail



program to solve this problem if he 'told' it that the line segments deriving from regions labelled 1 and 2 in the grey level description come from different objects. Given this clue, the program then had no further difficulty in separating out the two leaves, as **126** illustrates.

In this instance, the information about regions 1 and 2 came from Marr himself, who told his program what to do about them. In principle, however, one can imagine this information coming from higher-order object-recognition processes added on to Marr's program. These processes might have managed to recognise, from such segmentation as had already been achieved, that a bowl of flowers, say, was present in the scene, and therefore that leaves could be expected in various locations. Knowledge about the likely shape of leaves might then be fed down to the lower-level segmentation processes and provide the essential clue about regions 1 and 2. Alternatively, and perhaps more probably in the present instance, higher-level processes might discover that they were having no success in recognising the large feature cluster, and as a result direct the visual system to have a 'closer look'. Thus eye movements might be directed to this region to maximise the chances of picking up helpful new features, or perhaps the scene might be looked at from a new vantage point altogether, with the observer moving his head or body position to achieve this. Alternatively, attention might be directed to other types of clues, for example distance information, colour, etc.

These are the ways in which help can come from what is called 'downward flowing information', a good term because it clearly brings out they key attribute of the information, namely that higher-level processes are influencing the operation of lower-level processes. If the downward flowing information provides help deriving from knowledge about objects (e.g. the probable shapes of leaves in the above example), then it is usually termed *conceptually driven processing*. That is, a concept (e.g. 'leafness') drives (i.e. guides) the interpretation of information provided by the input image. On the other hand, if the downward flowing information merely guides the new application of low-level processes (e.g. generates a head movement to bring new information into play), without any reliance on object concepts or knowledge about objects, then it is customarily said simply to be exerting a *control function*. In the latter case, all processing is said to be *data driven*, rather than conceptually driven, because high level knowledge about objects plays no part, even though there are high level control processes 'noticing' how well the low-level processes are getting on, and guiding their operations accordingly if any 'difficulties' (e.g. absence of recognition) arise.

Exactly when and how in vision knowledge about the world in terms of object concepts takes over from or joins in with knowledge-free processes is an active controversy. Some feel that conceptually driven processing must be involved in the

very earliest stages of vision, even in those concerned with building up a feature description. Others, like Marr, argue that a great deal can be achieved without concepts of objects entering into the process until quite late on, and he brings forth his primal sketch and its segmentation to buttress his claim.

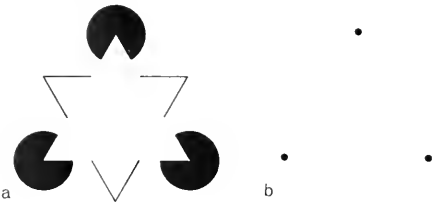
An example of this controversy in psychology concerns an illusion called Kanizsa's triangle [**127**a]. This figure consists of three sectored discs and some lines, lying on a ground of even light-intensity. We see, however, an illusory bright triangle, seemingly lying over the other pattern elements. Richard Gregory has suggested that the illusory bright contour is 'created' by our visual system to 'make sense' of the input image. The various black pattern elements suggest the object concept 'triangle', and this is enough for bright edges to be 'invented' to give this triangle a complete boundary contour. If this explanation is correct, then high-level object concepts ('triangle present which masks discs and an outline triangle') have not only led to a segmentation of the figure but also to the creation of edges in the feature description. Thus Gregory's theory goes beyond simply saying that the various bits and pieces of Kanizsa's figure have been grouped together, and then a structural description for 'triangle' arrived at. This latter type of process happens in **127**b but it is not associated with illusory brightness edges. So Gregory is here proposing a very deep involvement of high-level concepts in the very first stages of seeing.

But an alternative view of the illusory triangle is possible. This is, as I have argued in conjunction with Jeremy Clatworthy, that the effect is essentially a low-level one to do with brightness-contrast effects, and depends very little upon object-recognition processes, even though the illusory brightness engendered in the figure will of course create feature descriptions of contours which will subsequently be used for the symbolic description 'triangle present'. I will explain this low-level account fully in chapter 6, where I tackle the problem of the perception of lightness. I mention it at this stage just as an illustration of one side of the controversy about where, exactly, knowledge and conceptually driven processing in general come into the sequence of operations which constitute seeing.

That knowledge must play a part at some stage is, of course, not in doubt. Remember, for example, **19**. Until you were told that a Dalmatian dog was present, you probably had great difficulty in seeing this figure as anything other than a set of blobs. A similar effect was present in **18**, the rider-on-horseback. Moreover, the interpretation of the upside-down figures incorporating Little Lady Lovekins and Old Man Muffaroo [**14**] is undoubtedly conceptually driven in part. That is, these ambiguous figures provide a collection of feature clusters, which can be grouped together into more than one set of structures, and which structures become dominant at any one time depends on what sense is made of the whole figure. If the structural description for a large bird is matched with the structure found in the input image [**13**], then a large bird is duly seen and the associated structure becomes perceptually dominant. If, on the other hand, the figure is turned upside-down and the structural descriptions for a boat, an island, a man etc. are matched, then the separate structures associated with these objects become the dominant ones. Here, what is recognised in the scene seems to determine crucially what is finally grouped together, so much so that it is difficult at first to force your visual system to the alternative groupings, which are obtained with ease just by turning the figure upside-down.

The role of conceptually driven processing in seeing is very evident in reading. Whether we see 'clay' or 'day' in **128**, as the sample sentences show, depends very much on the *context* within which the word appears. Just how 'locked in' to one given interpretation of text we can become is illustrated in the pamphlet cover of **129**. I leave you to discover for yourself the printing error which this cover contains. It is a quite genuine cover, proof-read at various stages in its production by the Printing Unit of my University, and yet the error was missed by all who saw it until it was too late and it was widely distributed. Here we have a fine example of conceptually driven processing at work, dictating what is 'seen', in this case at the expense of what is really there.

127 (a) Kanizsa's triangle showing enhanced brightness and illusory contours, as opposed to (b) triangle with inferred 'boundaries' but no illusory brightness and no illusory contours. (Note that most demonstrations of illusory contrast-contours can be enhanced by placing tracing paper over the displays and/or viewing them from a distance.)



a b

It's hard work digging clay.
Save it for a raining day.

128 What is the last word written in each sentence? (See text for details.)

Block Portraits, Cubist Art and Structural Descriptions

Look at the block pattern shown in **130** [plate 8]. What do you see? Most people report nothing more than a mass of blocks of varying shades of blue. Having read chapter 1, you might choose to describe it as a grey level description with a very coarse pixel size, so that each block has a grey value which is the average grey level for quite a sizeable area of the input image. Compare the third grey level image in **3**. But even now that you have the 'sophistication' to be able to say this, you probably still cannot guess what the original input image was.

Now try looking at **130** from a distance of about 2 metres (6 feet) with your eyes screwed up so that your vision is blurred. If you wear spectacles you might find that you can blur your input image satisfactorily by looking at the picture without spectacles. In any event, if you are successful in blurring your vision you will almost certainly now be able to recognise the object which **130** contains. In case you have difficulty in obtaining an adequate blur, a suitably blurred image is provided on p. 120 [**131**a], with a caption telling you what the object is.

Figure **130** presents a very curious state of affairs. Normally we are used to sharp vision giving us the best hope of recognising objects, not the worst. Millions of people wear spectacles for this very reason. So what is going on here?

First of all, an account has already been offered (page 113) of how we recognise blurred pictures, and so that particular aspect of this demonstration need not be discussed again. The problem to be faced is: how is it that the portrait is 'hidden' in **130** when the image is sharp, and 'released' when the image is blurred?

To understand the answer to this problem, it is necessary to understand that the block portrait is really made up of two pictures added together [**131**]. The first of these is a coarse picture of Abraham Lincoln, and this is what appears when the image is blurred. The second picture portrays the sharp edges of the blocks, and this is the information which is removed by blurring. Thus the explanation of why blurring **130** improves the visibility of the object it contains is quite simple. When both the coarse portrait and the sharp edges are present together, the picture produces in us a structural description of blocks of various shades of grey, and this structural description simply does not match the structural description for Abraham Lincoln. Consequently, we do not recognise him in **130** when we view it well-focused. On the other hand, when the information about the sharp edges is is stripped away by blurring, a proper structural description for Abraham Lincoln is now built up, and recognition duly occurs. The mystery is resolved!

You may like to know how it was that the block portrait came to have its two component pictures 'sewn together', as it were. Here one simply has to state that it is a fact of physics that when a block picture is made, by averaging light-intensity over sub-regions of an image to obtain the grey levels for the various blocks, this averaging does not disturb very greatly the coarse structure of the original image (as long as the blocks are not too big), although it does destroy the sharp detail of the original and replace it with new sharp detail (the edges of the blocks). Given that the blocking process does not disturb the coarse information, and given also that blurring does not affect this either, it follows that both ways of transforming the image leave behind an unmolested coarse picture of Abraham Lincoln (even if in the case of the block portrait this picture is concealed) – and this is the picture we finally 'see' and duly recognise.

Some more block portraits are given in **132**. See if you can recognise them by blurring your vision and viewing them from 2 metres or so. Check your 'asnwers' to these problems by inspecting the photographs from which the portraits were derived (p. 122).

It is interesting that a block picture of a very ordinary original scene can have considerable aesthetic appeal. Or at least, so it appears to John Mayhew and myself. We took a very mundane visual scene in **133** (the chimney of my house) and reproduced it as a block picture. The aesthetic improvement effected by this transformation is, we think, remarkable, and reminiscent of certain styles in twentieth-century art. Consider, for example, the series of paintings of trees [**134**] created by Mondrian in 1912. The first in the series is fairly realistic but the final one is rather block-like. Indeed, in the latter case one might be forgiven for not realising that a tree was indeed Mondrian's subject, even though one might still enjoy the aesthetic effect engendered by the block-like pattern.

129 Can you spot the printing error?
135 Charcoal drawing by Pablo Picasso

A further illustration of cubist art appears in **135** (see above), this time a drawing by Picasso. Can you guess what the subject of Picasso's picture is? Try blurring your vision and see if this helps matters. Failing that, look at **136** (p. 122) and you will see a blurred version provided for you. The improvement in visibility is quite remarkable, just as it was for Harmon's block portrait of Abraham Lincoln.

It is tempting to speculate on the basis of these various demonstrations that one strand in modern art has been, perhaps unwittingly, to stretch to the limit the visual system's capability for building up a satisfactorily structural description for recognition of what a painting contains. Perhaps it is when the visual system is hovering on the brink of successful recognition, but not quite able to achieve it, that certain aesthetic experiences become manifest.

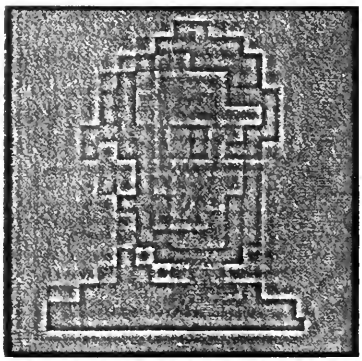
Representations of Object Descriptions

How are object descriptions represented in our brains? What is the neural language for objects? In chapter 3 (page 53), our present uncertainty about the neural symbols for features was described. The feature code could be at the level of the cell, the level of groups of cells, the level of biochemical events, and/or the level of impulse patterns. It is probably true at present that most workers in this area favour the idea of the single cell as the appropriate level for the feature code, and Horace Barlow, an eminent neurophysiologist, has considered at length both this question and the follow-up question: might the single cell also be the appropriate level of coding for objects? That is, when we recognise all the patterns in **107** as

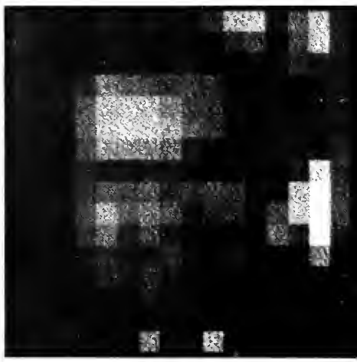
131 [top row, left and centre] The block portrait of 130 [plate 8] really has two constituents: (a) a blurred picture of Abraham Lincoln and (b) a set of sharp edges outlining the blocks. Blurring removes (b), leaving (a) visible.

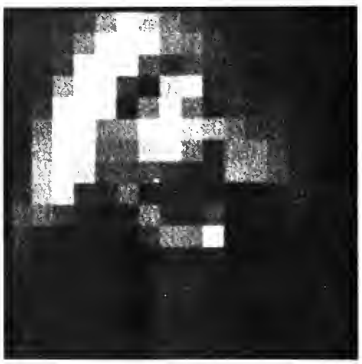
132 [top row, right, centre row, bottom row, left and centre] What do these block portraits conceal? See text for viewing details. The solutions are given on p. 122.

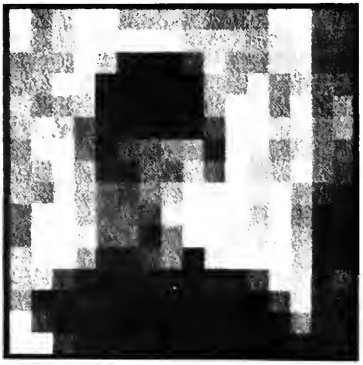
a

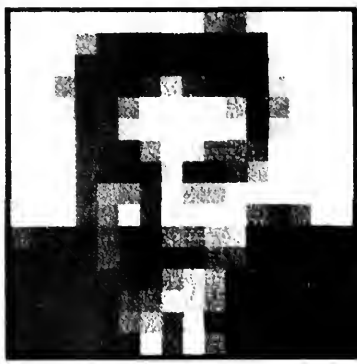


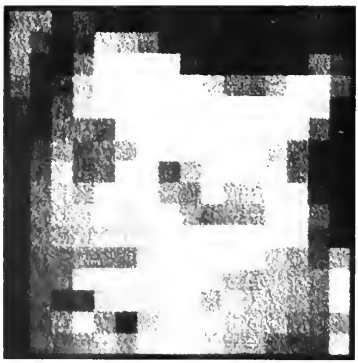
b

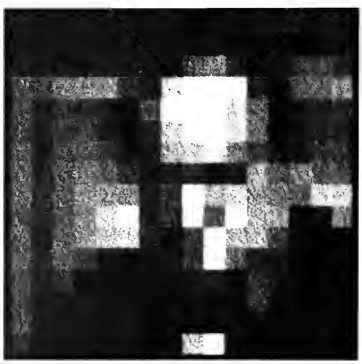






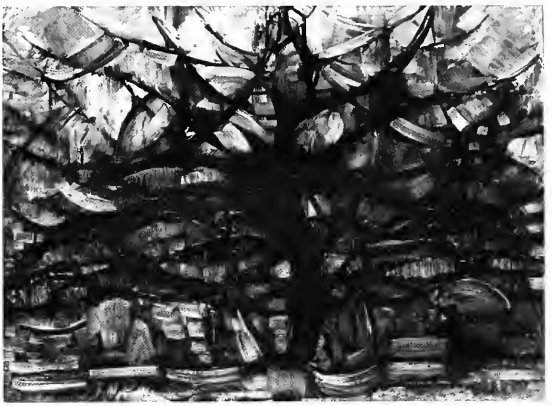


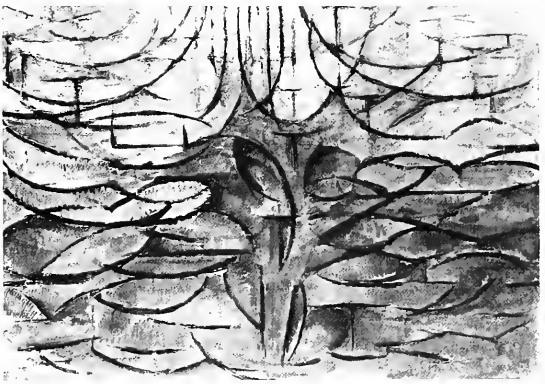




a

134 Tree paintings by Piet Mondrian of increasing levels of abstraction [135 is on p. 119.]





Ts, is the neural code for T-ness a certain cell firing away in our heads? And equally, when we recognise our grandmother as just having entered our field of view, is this act of recognition represented in us as a 'grandmother cell' becoming active? Of course, even if this were true, lots of other cells would also be active at one and the same time, for example all the cells representing the various features of which she was made up. The population of active feature cells would change from moment to moment as she changed her position, took her spectacles off or put them on, picked up a book or put it down, and so forth. But despite the shifts and changes in the feature descriptions being made at any one moment, we see her always as 'grandmother'. Is this grandmother-ness represented by a single particular cell being active throughout the period we see her, one which represents the unchanging aspects of grandmother despite changes in the details of the features described at any one time? Have we, in short, a 'grandmother cell' for representing the structural description of grannie?

The example I chose for explaining the single-cell theory of object representations was not accidental. This is the customary example given, so much so that the whole idea of single-cell object-representations is often called the 'grandmother cell theory', as a kind of quick shorthand title.

This idea has some curious aspects to it. First, it suggests that we have a separate cell for every recognisable object, so that we can ask: are there enough cells in the brain to go round? It is difficult to answer this question definitively, but the probable answer is yes. There are over 10,000,000,000 cells in the human brain and the chances are that a very large proportion of them are given over to object recognition, broadly considered. The question then becomes: how many different objects, scenes, patterns etc. can we recognise? It is

b

133 [left and above] The chimney of the author's house (a) Original photograph; (b) block version

137 Examples of shapes used to stimulate an inferotemporal cell which apparently has very complex trigger features The stimuli are arranged from left to right in order of increasing ability to 'drive' the neuron, from none (1) or little (2 and 3) to maximum (6).

terribly hard to estimate this, but note that there are only about 464,000 words in *The Oxford English Dictionary* (including its *Supplement*), of which a normal speaker of English can recognise only about 20,000. Note also that the grandmother-cell idea does not require a different single cell of this kind to be active for each distinct scene as a whole. Rather, there will be a collection of many such cells active at any given moment, each one of which will represent a particular object or, at the lower descriptive level, a particular feature. Obviously, if every unique scene that we could 'see' was to require a unique cell, then we would soon run out of cells as we proceeded through life. But this would not be the consequence if every unique scene was represented by a unique collection of active feature cells and object cells. Individual cells would be used over and over again in representing different scenes, but each time they would be in different company, so that the overall pattern of activity would remain unique.

A second curiosity of the grandmother-cell theory is that because we cannot grow new brain cells, but have to make do with the number we were born with, there must be brain cells lying dormant, waiting to cope with the new recognitions which we learn as we go through life. If you are introduced to someone, then on the grandmother-cell theory it seems that you give over a cell to his description and this is the cell which becomes activated when you later recognise him. Do we then have vast numbers of dormant cells, all awaiting their moment of life, doing nothing until they are somehow incorporated into an object description? Or is it that cells coding 'old' recognitions can somehow be disconnected and given a new lease of life coding new ones? We have no answer to these questions at the moment. Indeed, our ignorance about the neural mechanisms that mediate object representation is so great that we do not even know whether such questions are sensible ones.

Even so, some support for the idea that single cells are devoted to representing object descriptions comes from remarkable neurophysiological results reported by Charles Gross and his colleagues. Recording from single cells in the inferotemporal cortex (an area in front of the prestriate cortex: see page 40), they discovered some cells with highly complex and specific trigger requirements. One particularly famous cell of this kind has become known as the 'monkey-paw detector' because it seemed that it would only fire if a monkey's paw was present in the input image. Various other stimuli gave only weak responses [137]. Cells in the inferotemporal cortex often have very large receptive fields, sometimes encompassing most of the field of view, a fact which further supports the idea that they become active whenever and *wherever* their required object is present.

Interpreting Gross's findings in terms of a monkey-paw detector may not in fact be correct, and certainly other scientists have doubted that this is their real significance. We must always remember, as I argued at length in chapter 3, that it is dangerous to assume that a *property* of a neuron can be directly equated with its *function*: 'line detectors' have the property of responding optimally to a line of a given type but this does *not* mean that they are 'line detectors' in the full and proper sense of this term, that is, that their function is to detect lines. Even so, Gross's units could be an example of Barlow's ideas: single neurons representing complex objects. In our terms, the monkey-paw neuron would then be the neural representation of the stored structural description for 'monkey paw'. Whether this speculation will be confirmed by future neurophysiological research remains to be seen. Some people doubt that neurophysiology can tell us much about complex brain events, arguing that these must depend on complex interactions between cells, and that the neurophysiologist's present tools are too limited to look at these complex interactions. For those of this persuasion, the proper way forward now is to concentrate on the computer-vision approach. When we can actually build a machine which can see objects in the way that we can see them, then perhaps we will have a better idea of what questions to ask neurophysiologically. Either way, however, the prospects for the future are exciting indeed.

136 [left] The subject of 135 revealed by blurring: a seated nude woman

Solutions for block portraits in 132 top row Richard Nixon, Queen Victoria, Charlie Chaplin, bottom row Groucho Marx, J F Kennedy, Sir Winston Churchill

6 SEEING LIGHTNESS AND BRIGHTNESS

Imagine you are looking at a writing-desk illuminated by a table lamp and on the desk lies a black-edged blotter holding a white sheet of paper, as in **138**. If you came across such a scene in real life, you would have no trouble seeing that both edges of the blotter were black and that the paper was white. But this apparently simple matter is in fact deeply mysterious. The reason is that the black edge of the blotter lying directly under the lamp sends *more* light to the eye than the far edge of the paper – which yet appears white! This is odd, to say the least. How can a surface appear black when it is sending more light to the eye than one which appears white? If the blackness/whiteness of a surface was simply seen according to the amount of light entering the eye from the surface in question, then the physically *black* surface under the lamp should appear *whiter* than the white surface distant from the lamp!

The scene shown in **138** is a perfectly usual one and no tricks are being played. Indeed, all the clever tricks here happen inside the viewer's visual system. Somehow this system has taken notice of the fact that the black edge under the lamp is more highly illuminated than the white paper distant from the lamp, and that once the factor of illumination is allowed for, it is 'proper' to see the surfaces as respectively black and white because this is how they really are.

The problem which the visual system faces in this kind of situation is depicted in **139**, which shows in schematic form the likely *luminance profile* across the blotter. *Luminance* refers to the amount of light coming from a surface, and it is dependent on two factors: the amount of light falling on the surface (its *illumination*) and the proportion of this incident light which is bounced off the surface (its *reflectance*).

The perceptual correlate of reflectance is termed *lightness*, and this varies from black (low reflectance) through grey (medium reflectance) to white (high reflectance). The visual system is remarkably good at computing lightness irrespective of illumination, and it is this fact to which **138** draws attention. Somehow we see a brightly illuminated black surface as black even though the luminance of this surface is greater than that of a dimly illuminated white surface. The visual system has to work from an intrinsically ambiguous luminance profile [**139**, upper], and yet from this it manages to extract the two distinct stimulus attributes [**139**, lower] of surface reflectance (perceived as lightness) and illumination (perceived as brightness: we can discriminate between a brightly-lit white and a dimly-lit white, or a brightly-lit black and a dimly-lit black).

The Computation of Lightness

How can the visual system extract from the luminance profile of a scene the reflectance of the various surfaces, given that the luminance profile is contaminated with the unwanted contribution made by the varying illumination over the scene? The trick used by the visual system in solving this problem seems to take advantage of the fact that variations in illumination are relatively gradual whereas variations in reflectance are rather sudden. This can be seen in **139** (lower), where the transitions from black to white and vice versa are steep whereas the fall-off in illumination from left to right is relatively gentle. Given this consideration, one way to 'filter off' the component due to varying illumination is as follows:

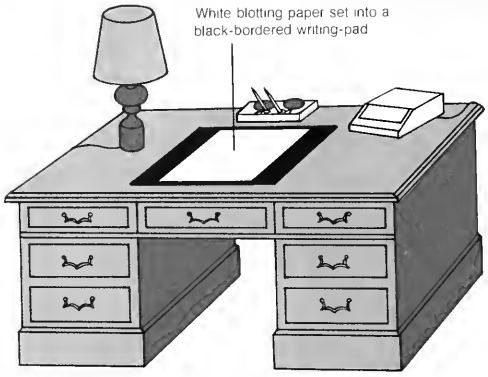
First, *detect edges*, marked by reasonably large and steep jumps in luminance between the areas to either side of them. This means discarding all information about luminance differences between adjacent points in the input image if these differences fail to exceed a certain threshold value. In this way, the gradual luminance transitions are eliminated while the sudden ones are preserved.

Second, build up the required lightness profile by *reconstituting between edges*. This amounts to 'joining up' areas between above-threshold edges, giving these areas lightness values determined by the size of the luminance differences forming the edges.

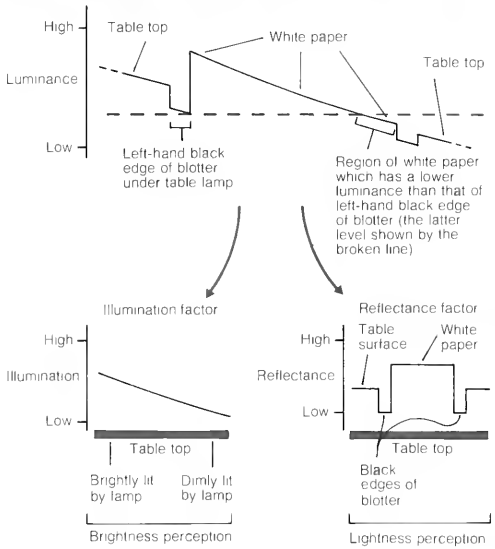
This general approach to the computation of lightness was suggested by Edwin Land following his psychological studies on the perception of lightness. Berthold Horn has implemented it in a successful and elegant computer program which, as David Marr has pointed out, has many features which make it ideally suited to be carried out by known mechanisms in the retina. I will describe briefly how this computation works, because it illustrates particularly well the current state of visual science – a fascinating interplay, as we have already seen, of psychology, neurophysiology, neuroanatomy and computer programs.

The Craik-Cornsweet-O'Brien Illusion

To begin with the psychology, consider the two stimuli shown in **140a**, b. They look very similar: a light grey area adjoining a dark grey area in each case. But they are in fact very different, as their associated luminance profiles reveal [**140**c, d]. Thus the left-hand stimulus really is composed of adjoining light and dark grey areas, but the equivalent areas in the right-hand stimulus are in fact *the same* except for the presence of an



138 A commonplace form of lighting, but one which presents the visual system with a difficult problem (see text).



139 Luminance is determined by illumination and reflectance.

'edge' at the boundary between them [**140d**]. Check this surprising fact for yourself by covering the boundary in **140b** with a pencil, which will enable you to see the two areas in their true lightnesses.

This illusion, called the Craik-Cornsweet-O'Brien illusion after its joint discoverers, is a dramatic effect, one of the most impressive of all visual illusions. It represents a particularly startling instance of how the visual system can sometimes be fooled by using a computational stratagem which is usually successful but which is occasionally quite inappropriate to the prevailing input. Like so many illusions, it offers a valuable clue about how the visual system works in normal cases. And the illusion suggests that the lightness computation carried out by the visual system is indeed as Land has proposed, because its obvious explanation runs as follows:

In *both* **140a** and **140b**, the visual system first detects the above-threshold edge separating the two grey areas, discarding all other luminance information. This produces the edge-detection profiles shown in **140e** and **140f**, which can be seen to be very similar. Next, the visual system in each case 'reconstitutes' the areas on either side of the edge to produce the perceived lightness profiles illustrated in **140g** and **140h**. Thus the reasons **140a** and **140b** look alike is that the visual system applies to both a lightness computation designed to eliminate problems due to varying illumination, had any been present, and this lightness computation is misled into thinking that the edge profile in **140d** comes from the boundary between two regions of different reflectance. It therefore produces a lightness profile accordingly, but an illusory one in this instance.

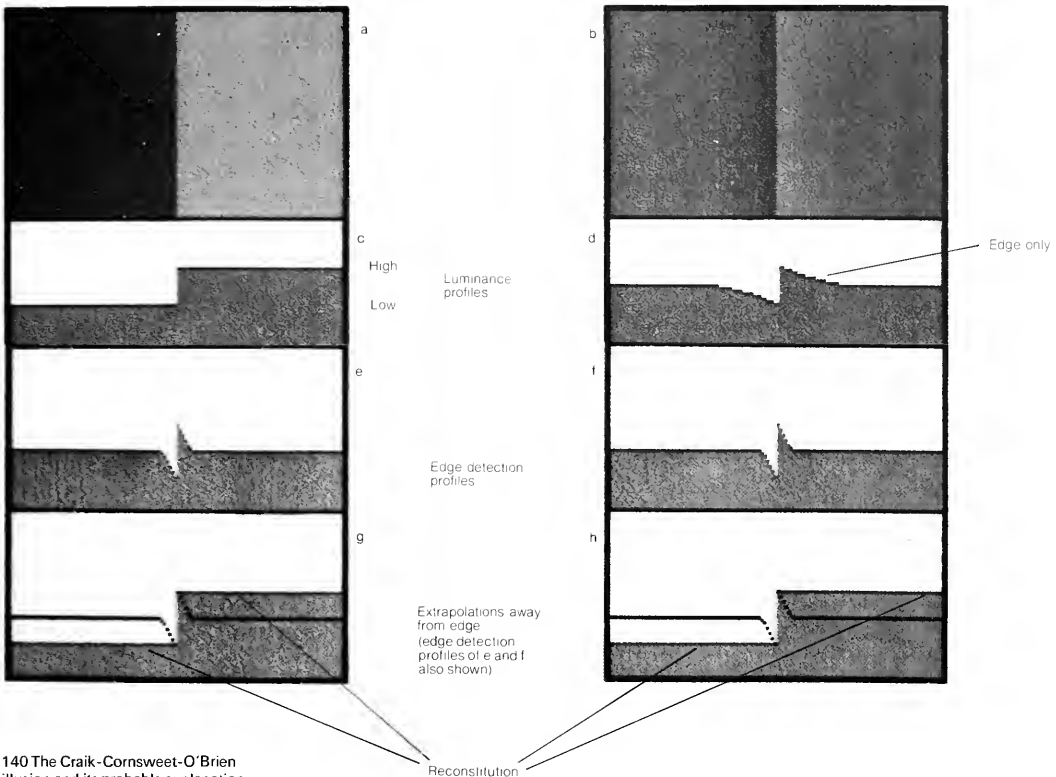
The Craik-Cornsweet-O'Brien illusion shows a clever way of tricking the visual system by turning one of its own tricks against it. Similar inputs sometimes come about in quite ordinary real-life situations, however, as Floyd Ratliff has pointed out. Consider, for example, the piece of Ting white porcelain shown in **141**. Its glaze is a uniform white, but the incised lotus design appears lighter than its background because the incisions have edges shaped rather like those of the Craik-Cornsweet-O'Brien illusion, so producing shadows of just the required kind to generate an illusory lightness effect.

Of course, it must be remembered that we are here talking solely about the computation of lightness, and so when mention is made of discarding all luminance information except that at edges, this should be understood as applying to the lightness computation alone. Illumination too must be noted, for the purposes of brightness perception, and here luminance gradients other than those marking edges presumably play a crucial part. Also, gradual transitions in luminance might be independently detected and used for the perception of three-dimensional shapes. Refer back to **16** for an example of how shadows and shading, instances of illumination differences, can be powerful cues to depth.

Detecting Edges for the Computation of Lightness

A great deal of attention was given in chapters 2 and 3 to the problem of detecting features such as edges. It was seen to be a much more complicated business than one might have guessed, and required a good deal of intricate computational machinery in the striate cortex. It turns out, however, that the edge-detection requirements for the computation of lightness are much less stringent, and that rather simple mechanisms can suffice, mechanisms almost certainly located in the retina. But before going on to describe the neural basis for this kind of edge detection, we will first introduce some of the general ideas involved, so that the neural mechanisms can be better understood.

Edge detecting for the purposes of lightness perception depends, both in Horn's computer program solution to the problem and in the visual system itself, on the process of *convolution*. Much was said about convolution in chapters 2 and 3, but to remind you of the general idea, **142** shows a receptor mosaic coupled by various sets of connecting fibres to a *convolution array of cells*. Each convolution cell extracts from



140 The Craik-Cornsweet-O'Brien illusion and its probable explanation

the receptor mosaic a certain limited type of information according to the design of the excitatory and inhibitory connections which feed into it. In any given convolution array, all the cells are tuned to the same type of information, but they all 'look' at different parts of the input image. In the instance shown in **142**, each convolution cell receives inputs from an approximately circular cluster of receptors. The clusters for just four convolution cells are picked out with heavy outlines to help illustrate what each such cell is 'looking at'. In the previous convolutions described in this book, the receptor cluster was corner-shaped, or line-shaped, etc., but the same basic idea of convolution applies here just as it did earlier.

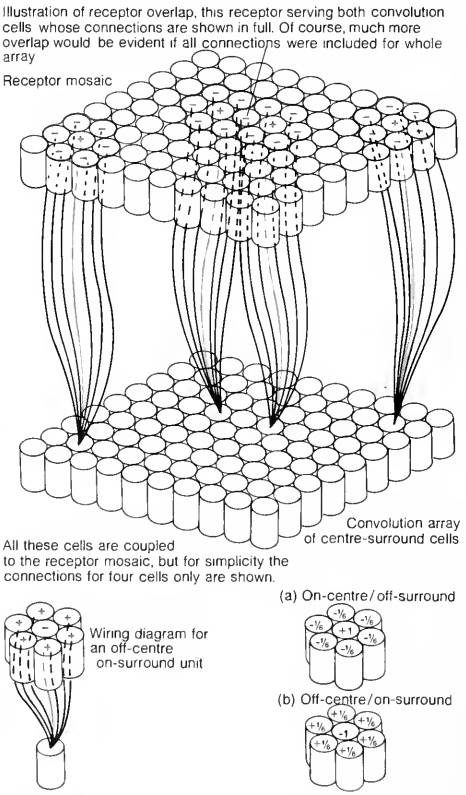
Note that the central receptor in each cluster feeds a convolution cell whose position in the convolution array exactly matches that of this central receptor. The two sheets of cells, receptor mosaic and convolution array, are thus neatly lined up, one over the other. In **142**, only the fibres connecting up a few convolution cells are shown, for simplicity, but in fact every convolution cell would be connected to the receptor mosaic in a similar fashion, each with a set of inputs coming from receptor clusters in slightly

141 Chinese Ting Yao saucer is an example of the famous Ting white porcelain produced in the Sung dynasty of about A.D. 1000. Although the entire surface is covered with only a single creamy white glaze, the incised lotus design appears brighter than the background because of the incisions, which have a sharp inner edge and a graded outer edge, producing exactly the kind of contour that creates an apparent difference in brightness of the Craik-Cornsweet-O'Brien kind.

different positions in the receptor mosaic. Because of these multiple connections, each receptor has to feed many different convolution cells. But, again for reasons of simplicity, **142** shows only one small overlap – that between the receptive fields of the two central convolution cells whose wiring diagram is shown in full, which share one receptor in common.

As in our previous convolutions, the receptor connections carry either excitatory or inhibitory influences to the cells in the convolution array, which then proceed to count up the sum of these two types of inputs. In **142**, the central receptor in each cluster (shown with a brown connecting fibre) feeds excitation to its convolution cell, whereas those in the surround feed inhibition. These two types of influence are marked on the receptor mosaic with +s or −s respectively. Because the centre and surround connections are antagonistic in this way, the convolution cells are customarily called *on-centre/off-surround* units. It is possible to have cells wired up with the reverse arrangement, and then these are called *off-centre/on-surround* units, as you would expect [**143**].

142 A convolution network for on-centre/off-surround cells



143 [left] Wiring diagram for an off-centre/on-surround unit

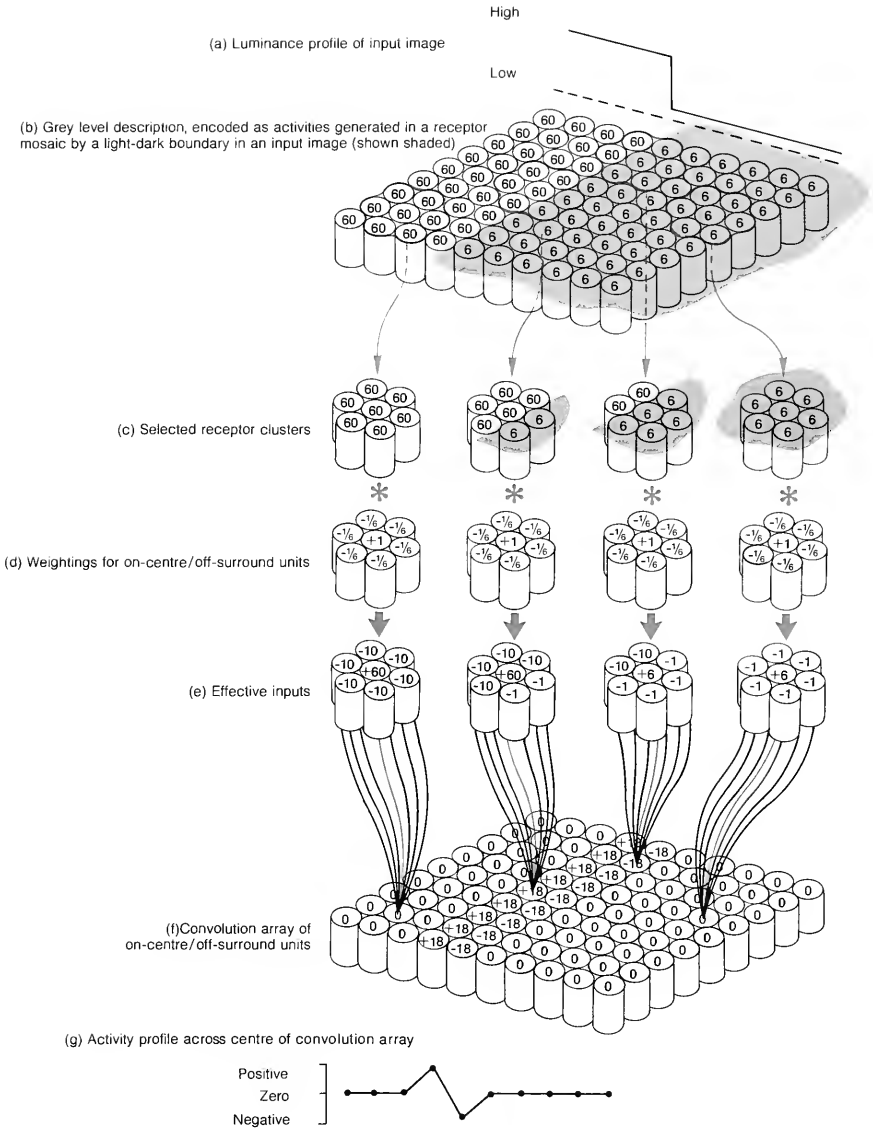
144 [right] Weighting diagrams for centre-surround units

The basic objective is to use convolution machinery of the kind just described to detect edges. Therefore it is desirable to make the excitatory and inhibitory influences on a cell add up to zero if no edge is resting on a receptor cluster serving the cell. This can be done by giving each receptor in the cluster a certain *weighting* in its influence, so that if all the receptors are active to the same extent, because they are simulated by an area of even illumination, then the net influence of all the receptors is zero. Suitable weighting for achieving this with our simple centre/surround clusters are $+1$ or -1 for the centres and $+\frac{1}{6}$ or $-\frac{1}{6}$ for the surrounds, depending on the type of unit in question. Weighting diagrams designed on this basis are shown in **144** for the two types of unit. You can readily appreciate that the $+1$ of the on-centre is cancelled by the six receptors of the off-surround, each feeding $\frac{1}{6}$ of their activity as inhibition. Equally, the -1 of the off-centre is cancelled by the six 'doses' of surround excitation, each with a weighting of $+\frac{1}{6}$. Thus for each type of unit, the set of weightings given to centre and surround receptor activities ensures that a region of uniform illumination produces a zero output, as required. Let us now see how things work out if this convolutional machinery is faced with the type of input it is supposed to detect – an edge.

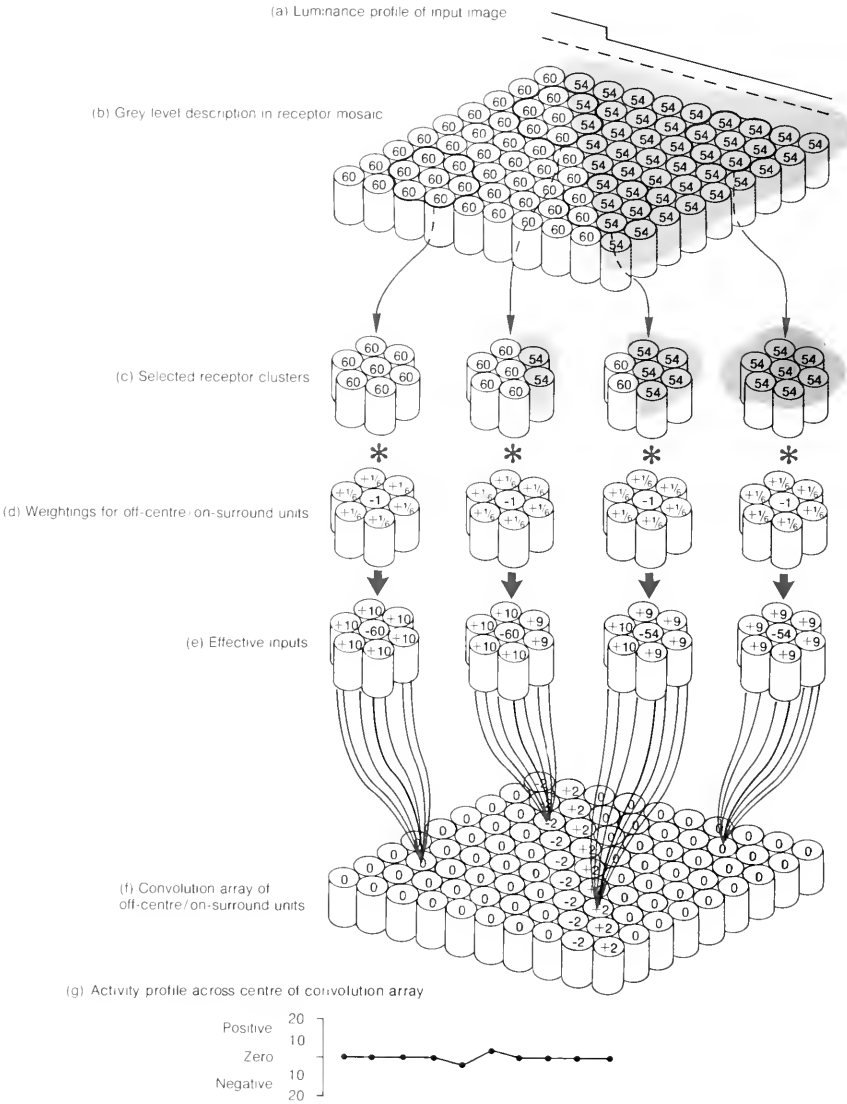
An input image containing a steep luminance profile [**145a**] is shown resting on a receptor mosaic in **145b**, with the dark portion of the edge shaded as grey and the light portion left as white. The dark region sets up only weak receptor activity, here for illustrative purposes shown as a figure 6 in each receptor, which indicates '6 units of activity'. The light side of the edge in **145b** sets up receptor activity of 60 units, and the whole pattern of numbers throughout the entire receptor mosaic constitutes a grey level description of the input image.

Having obtained a grey level description via the receptor mosaic, the convolution array of **145f** proceeds to 'look' for edges. Each convolution cell 'inspects' one particular region of the receptor mosaic, as described for **142**, and counts up the excitatory and inhibitory influences coming from this region. Receptor clusters feeding four convolution cells are shown picked out with a heavy border in **145b**, and then again as isolated clusters in **145c**. Each receptor's activity has to be weighted in its influence, and weighting diagrams for on-centre/off-surround units are shown in **145d**. Each receptor's activity is multiplied by the appropriate weighting, an operation symbolised by the asterisks between **145c** and **145d**. The results of the multiplications are shown in **145e**, which is the set of effective inputs (i.e. post-weighting inputs) fed through to the convolution cells for counting. The counts obtained are shown in **145f**, where you can see that most of the convolution array is inactive (the zeros), but that high positive and high negative counts are present along the two sides of the edge boundary. These edge counts are $+18$ and -18 respectively. An activity profile through the convolution array is shown in **145g**, and this emphasises the fact that the array has succeeded in locating the edge.

A question which might occur to you at this stage is: although **145** is not intended to show biological convolution machinery, how would the negative numbers be coded in a biological equivalent built up from nerve cells? The general problem of keeping track of negative quantities in biological



145 Edge detection by convolution with centre-surround units



146 Convolution detection of a faint edge

convolutions was discussed earlier (p. 32) in connection with the design of white-on-black and black-on-white corner detectors. Suffice it to say at this point that later on in this chapter we will present a similar solution to that proposed previously, namely different populations of cells dealing with the quantities which in **145** appear as positive and negative numbers.

Another question which might also be bothering you is: where are the weighting multiplications actually performed? It is easy enough to imagine several cables leaving each receptor and ending at different convolution cells: but these different cables have to carry different messages, sometimes positive, sometimes negative, sometimes strong, sometimes weak, according to the weighting diagram. Does the receptor 'decide' what to send to each convolution cell, or does each convolution cell receive the receptor messages unweighted in the first place, and then weight them appropriately before proceeding with its count? Either of these possibilities is as good as the other, in principle, but it turns out that the centre-surround computational machinery of biological systems seems to use a method quite different from either of these, and I will describe it in due course.

For the present, keep in mind the basic objective of our lightness computation: to find out where the edges are in the input scene, to accept only those above-threshold in size of intensity step, and then to 'join up' the areas between the edges so obtained to produce the required lightness profile. Let us then see how the convolutional edge machinery copes with a weak edge, one which can be regarded as probably having come from an illumination difference rather than a reflectance difference.

An edge with a relatively small luminance difference between the areas to either side of it is shown as a profile in **146**a, and lying on a receptor mosaic in **146**b. The two areas bounded by the edge stimulate receptor activity at levels of 60 and 54 units. These numbers have been chosen to keep the arithmetic simple, but of course an even weaker edge would be dealt with in an essentially similar way. To illustrate other relevant aspects of the lightness computation, I have set the weak edge at an angle to the vertical, and shown it being convolved with *off*-centre/*on*-surround units: this differentiates this example from the convolution illustrated in **145** in two further ways, apart from the faintness of the edge. But despite these differences, the various steps in the edge analysis follow closely those explained in detail for **145**, and you should have no difficulty in following through the arithmetic for yourself, and understanding how the edge comes to be represented in the convolution array as a strip of cells responding at levels of activity of $+2$ and -2. But there are certain new features to note:

First, the fact that the edge was not vertical has made no difference. The circularly-symmetric centre-surround units have the valuable design feature of being sensitive to edges regardless of their orientation. This is useful because it means that the lightness computation can proceed without the need of many different convolution arrays, each serving a different orientation (as in the case of feature detection for edges, where orientation was important). Orientation is unimportant for centre-surround units, and that is fine for present purposes. Of course, the particular examples of edges chosen for **145** and **146** lie neatly on the boundaries between receptors, and other orientations (as well as the same orientation, but simply shifted slightly sideways) could cause a minor problem of deciding when a partially-covered receptor was or was not deemed to be active. But this is a detail we can ignore, and simply note that in a real receptor mosaic of very fine resolution (such as the human retina contains) the problem is essentially immaterial.

Second, having *off*-centre units perform the convolution, rather than *on*-centre ones as in **145**, has also made little difference. It has simply led to the boundary being represented by a negative-to-positive change in convolution-cell activity as one proceeds across the convolution array from light to dark, rather than by a positive-to-negative change as in **145**. The question might therefore be asked: why bother with both types of unit? One answer is that in a computer system one need not have both types, because negative quantities can easily be dealt with by the computer's arithmetical machinery; but in a biological system composed of nerve cells, negative quantities pose problems for representation. A cell can either be active or inactive: it cannot be 'negatively active'. Of course, it would be possible in principle to have 'zero' represented not by inactivity but by a medium level of activity, and then reckon that 'positive' quantities are shown by activity above this medium level and 'negative' ones by activity below it. Biological systems sometimes seem to use this option, but in the case of the computation of lightness it is most probable that positive quantities are sent down one channel of cells (here the on-centre cells) and negative ones down another (here the off-centre cells). This issue was briefly referred to earlier in this chapter and there was a fuller discussion in chapter 2 (p. 32). One advantage of using separate negative and positive channels, rather than using variations about a medium level of activity, is, Marr suggests, that it makes so much easier the setting and re-setting of the threshold which determines when an edge-signal is to be regarded as coming from a genuine edge rather than from an illumination gradient. Alterations in threshold are needed to cope with problems posed by changes in the overall illumination of a scene (see below).

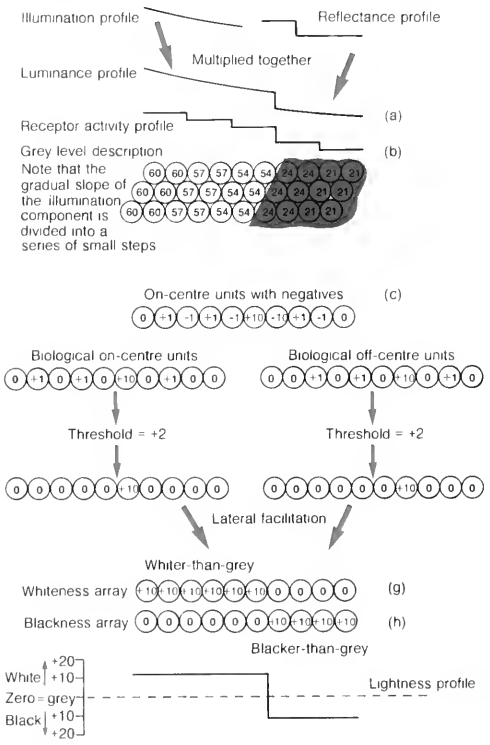
The third thing to notice about **146** in comparison with **145** is that the faint edge has produced a weak edge-signal in the convolution array. Thus in **145** an edge represented in the grey level description by adjacent receptor-activities of 60 and and 6 produced convolution signals of $+18$ and -18, whereas in **146** the 60/54 edge came out at $-2/+2$. This is exactly as required. The first step of the lightness computation is to find edges, but allied to this is the need to ignore gently sloping luminance gradients, which in turn means ignoring 'very faint edges', on the assumption that these come from illumination differences rather than from reflectance differences. Of course, this critical assumption carries with it the penalty of 'missing' genuine but slight reflectance changes (i.e. genuine but very faint and gradually sloping edges), but this is a price the visual system has to pay for 'buying' the proper computation of lightness in most circumstances. In any event, to be able to decide whether an edge is or is not above threshold, one first needs to know its strength, and this is given by the size of the convolution signal. If the threshold was set, say, to plus or minus 3, then the first edge (plus or minus 18) would be accepted as caused by a reflectance difference, whereas the second one (plus or minus 2) would be disregarded (for the purposes of the lightness computation, that is: as noted earlier, it might be desirable to keep track of it for something else).

Reconstitution by Deconvolution

The final step of lightness computation is to build up a lightness profile by 'filling in the gaps' between the edges which have been discovered by convolution, and checked as being above-threshold and therefore 'noteworthy' for lightness purposes. This is a bit more tricky to achieve, but is essentially just the opposite of the original centre-surround convolution, and is therefore sometimes called *deconvolution*. It can be performed by arrays of units which *facilitate* each other adjacently. The centre-surround convolutions described above had as their key feature the antagonistic influences of excitation and inhibition (sometimes called lateral inhibition). But the final step in lightness computation uses simply excitation, so that activity can spread out from the edge, as described in brief for the Craik-Cornsweet-O'Brien illusion.

Consider, for example, a luminance profile made up of a gradual illumination change superimposed upon a sudden reflectance change [**147**a]. Clearly, the objective of lightness computation is to extract the reflectance profile from this ambiguous input. A grey level description is shown in **147**b, in the usual form of levels of activity in the receptor mosaic (albeit a smaller mosaic, for reasons of space). Its numbers show that the illumination component appears in the form of a set of small steps, of 60/57, 57/54 and 24/21. On the other hand, the reflectance step is much greater: 54/24.

147 Lightness computation by biological centre-surround units



The next thing to note is that the grey level description is convolved with on-centre units. A central strip of convolution cells from a convolution array is shown in **147**c. For this convolution, negatives were allowed, and so some negatives appear in the convolution cells. Try working out for yourself why the numbers are as they are, referring back to **145** for guidance.

In **147**d, convolutions with 'biological' centre-surround units are shown, their distinguishing feature being that no negatives are allowed. Thus any cell which would have been negative is set to zero. Convolutions with on-centre and off-centre units of this type are illustrated, again with the results shown from just one strip across a full convolution array. As can be seen, only positive numbers appear in each strip, but where they appear is different in each case. Thus the negative numbers of **147**c appear as positive numbers in **147**d for the *off*-centre units. And the positive numbers of **147**c appear as positive numbers in the on-centre strip of **147**d. This might sound complicated, but the complexity is more apparent than real, as **147**d shows.

Having obtained our on-centre and off-centre convolutions, the next step is to apply the threshold. In **147**e, the threshold is set to +2. (In real situations the threshold would need to be dependent on the exact viewing circumstances. For example, if the general level of illumination is high then variations in illumination are also likely to be high, and if the threshold was left low these variations would then masquerade as genuine reflectance changes. But how the threshold might be set and re-set to suit the prevailing illumination conditions is a detail we must leave aside here.) The result of applying the +2 threshold is shown in **147**f: it leaves just the large edge-measurements of +10 as the only ones appearing in each convolution array.

Once the threshold has been applied, the final step is to build up the required lightness profile by extending the activity outwards from the above-threshold edges. This operation is done in two quite new sets of arrays, termed the *whiteness* and *blackness arrays* respectively in **147**g and **147**h. The whiteness array shows the results of extending out from the edge recorded by the on-centre units, and the blackness array does likewise for the off-centre units. Somehow activity in the whiteness array is not allowed to spread in the wrong direction, across the white/black border, vice versa for the blackness array. This essential restriction is perhaps achieved by coupling together in an opponent-process fashion (p. 101) each white/black pair of cells dealing with the same part of the grey level description, so that whichever cell is more active 'wins out' and inhibits the other one to a zero level. Thus any facilitation passed across the edge within either array would never exceed the value of the inhibiting opponent cell, and so it would be 'vetoed'. Without going into too much detail, it is easy to see that whiteness and blackness arrays of the kind displayed in **147**g and **147**h can between them 'carry' the lightness profile shown in **147**i. Note the clever way in which the use of whiteness and blackness cells means that activity is always positive: it is just that positive activity means something different in the two cases.

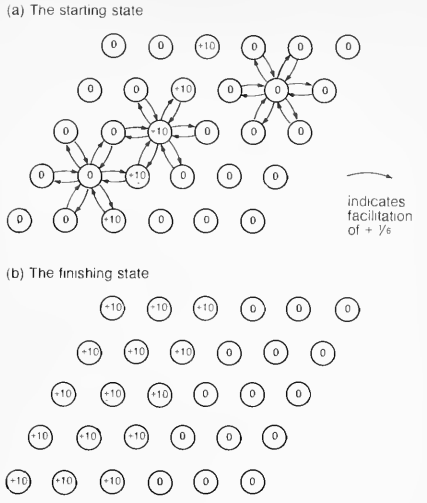
The remaining task is to explain how the spread of activity necessary for deconvolution is achieved within the whiteness and blackness arrays by the process of lateral facilitation. The

network of connections required for doing this is shown in **148**a. This time the connections are within the array itself, as contrasted with what happened in the centre-surround convolution case, where connections went from the receptors to the convolution cells (but note later remarks about retinal machinery for deconvolution). Here, to mediate the process of deconvolution, the fibres linking up cells run between adjacent cells within each 'deconvolution array'. Note that in **148**a, each deconvolution cell is shown with a number in it – the post-threshold edge information fed into the array by the preceding stage. The array shown in **148**a could be either a blackness or a whiteness array: it does not matter for the purposes of exposition, as they both work in a similar fashion and both deal solely with positive numbers.

Each cell in **148**a both influences its neighbours and is influenced by them. Thus 'double' connections are shown between cells. Not all connections are given in **148**a, for reasons of simplicity, but you should imagine all adjacent cells connected up as indicated for the three chosen examples. The $+\frac{1}{6}$ fractions shown for each connection indicate the weighting of the lateral facilitation mediated by the connections. That is, each cell excites its neighbours by $\frac{1}{6}$ of its own activity level. Moreover, each receives excitation from its neighbours of $\frac{1}{6}$ of their level. So each cell is 'helping out' its neighbours and in turn being 'helped out' by them: hence the suitability of the term 'lateral facilitation'.

An important point to grasp is that the process of mutual facilitation goes on and on repeatedly until a steady state is achieved by the network. For example, consider the cell in the lower left-hand corner of **148**a which is shown with its connections to neighbouring cells. It starts off from a level of zero because this is what was fed into it by the preceding thresholding stage. It thus begins by offering no excitation to its neighbours because, of course, $0 \times \frac{1}{6} = 0$. On the other hand, it receives excitation from two neighbouring cells which started from a level of $+10$ because they are 'on' an edge. As $+10 \times \frac{1}{6} = 1.67$, the total facilitation received by this unit is 3.34. Having now taken on an activity level of 3.34, it can proceed to facilitate its neighbours more distant from the edge, cells which initially received no excitation at all because when the deconvolution started they were entirely surrounded by cells at zero levels of activity. Perhaps you can grasp intuitively that with this kind of progressive updating of facilitation, a spread of activity would rush through the network like a brush-fire. Soon all the cells would assume the same value on the same side of the border – and the lightness profile would be built up as required, as shown in **148**b.

But you might reasonably ask at this point: what stops the facilitation rushing across the edge boundary? Why do not *all* cells end up on $+10$? The answer must be that the opposing deconvolution array stops this happening by inhibiting such a spread. Remarks were made above to this effect and one must suppose that if, for example, **148**a is a whiteness array, then the blackness array provides the vital inhibition which stops the spread going too far. The blackness array would also have to provide some inhibition to stop the whiteness array going 'mad' and rising via the process of mutual facilitation up to 'super-white' levels of activity from starting points of a much more modest kind. After all, the cells on the edge would initially receive no facilitation from their neighbours, but once these neighbours had been brought alive, as it were, they would facilitate the very cells by which they had themselves been activated, and, if these latter were not 'held down' by



148 Deconvolution by lateral facilitation (see text)
(a) The starting state Connections for just a few cells are shown but all cells are in fact connected up identically.
(b) The finishing state All connections have been removed here for simplicity.

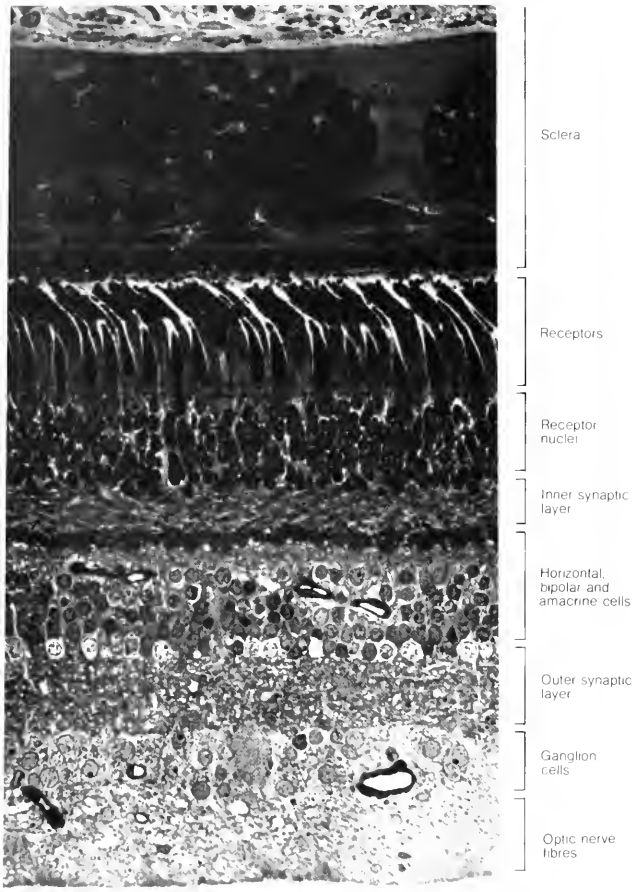
inhibition from the blackness array, the whole thing would get out of hand. Thus although the visual system solves the problem of not being allowed negative levels of activity by inventing a blackness channel to complement a whiteness channel, the price that has to be paid is that the two arrays need to 'speak to one another' so that silly results do not come about. The 'speaking' probably takes the form of inhibitory (subtractive) cross-couplings, as already noted, of the opponent-process kind.

This section has explained the probable basis of how the visual system computes lightness. We consider next which cells might perform each of the steps of edge detection, thresholding, and then deconvolution.

The Retina

The retina is a layer of cells located on the inner surface of the eyeball [**149**, plate 9] and it contains the light-sensitive *receptors* which initiate the whole business of symbolic scene description. The lens system of the eye focuses an image of the observed scene on to the receptor layer, and the receptors 'notice' when light shines on them and trigger activity in various nerve cells in the retina, which in turn send messages to the brain.

A cross-section through a human retina is shown in **150** [plate 10], enlarged microscopically so that certain details of its structure can be seen. The first and most curious point is that the receptors are tucked away *inside* the retina and are not on the surface which receives the light. This is an odd arrangement because the light which stimulates them has first to pass through many (albeit near-transparent) layers of cells, as well as blood vessels, before it can activate the receptors and thus set going the processes of seeing. But



150 Cross-section through the back of the human retina: magnification approximately ×350

although it is odd, this arrangement obviously does not provide a crucial handicap, as our remarkably good vision testifies. Indeed, it might even provide some desirable structural or protective advantage, perhaps helping to shield the receptor layer from possible deformations due to movements of the eyes, etc. On the other hand, squids and octopuses have eyes which are very similar to our own except that their receptor layer *is* on the outer surface of the retina, where light can strike it directly, and so either way the advantages and disadvantages do not seem crucial.

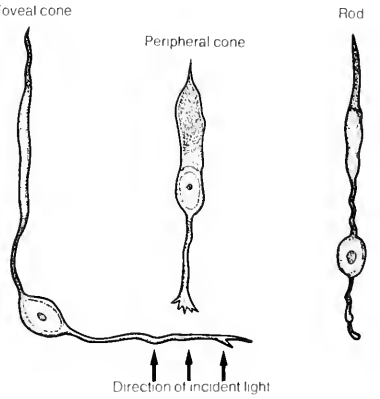
The receptors form a sheet of cells, a mosaic which has already been shown in plan view [**42**]. The mosaic is made up of two types of receptors, called *rods* and *cones* on the basis of their different anatomical shapes [**151**]. By no means all creatures possess both types of receptors. For example, the pigeon has an all-cone retina, whereas the cat has a retina made up almost entirely of rods. It may be that certain animals never evolved the capability for both rod and cone vision, or it could be that different animals have found it better for their own particular problems of survival to concentrate on just one type. For example, the cat is essentially a nocturnal animal and its rod retina might be specially well adapted to the low light levels characteristic of night life (see below).

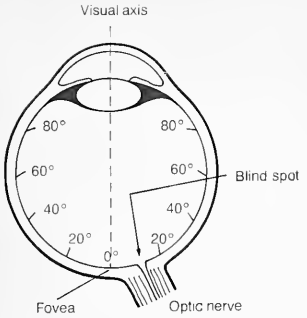
Rods and cones contain different types of *photosensitive pigment* which enable them to specialise in different tasks. The rods have a substance called 'visual purple', or *rhodopsin*, and as a result are about 500 times more sensitive to light than cones. On the other hand, cones contain pigments which make colour vision possible. So really the rods and cones are two distinct light-sensitive systems packaged together into a single 'camera' – the eye.

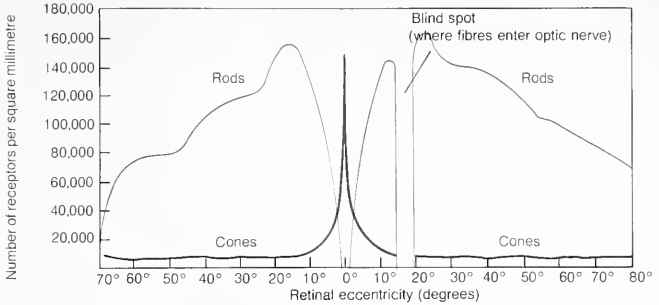
There are about 120 million rods to be found in each human eye, and they lie all over the retina except in the fovea and in the blind spot [**152, 153**]. Cones number about 6 million in each eye, and are concentrated into the central region of the retina. Indeed, in the *fovea*, a tiny recessed region at the very centre of the retina, only cones are to be found, packed in at the very high density of about 150,000 cones per square millimetre. The very central region of the fovea, where the cones are smallest and packed tightest, is very small – only about 1/10,000 of a millimetre across, and serving therefore a visual angle of no more than 20 minutes of arc, or one third of a degree. Visual angle is a convenient way of describing the size of a retinal image irrespective of the particular size or distance of the object which gave rise to it. The angle in question is that between the imaginary straight lines drawn from the eye to the outer edges of the object which is said to subtend that angle. For example, a 1 cm line viewed from about 57 cm subtends a visual angle of 1°, and so does a 2 cm line viewed from about 114 cm. The very central region of the fovea just referred to, therefore, would be filled by an image of a disc measuring about 0.3 cm viewed from about 57 cm, as this disc subtends a visual angle of about one third of a degree.

This central foveal region probably contains only about 2000 cones in all, but even within this densely packed region of retina, further variations in the size of the cones can be

151 Rods and cones







152 [left] Cross-section through eye showing how retinal regions can be labelled according to their angle of eccentricity from the fovea

153 [right] Distribution of rods and cones according to retinal eccentricity

detected with careful scrutiny, such that the very centre of our gaze, served by the very smallest cones packed most tightly together, may depend on a total of only a couple of dozen receptors. It is a remarkable thought that when we stare at something very carefully in an effort to maximise our ability to detect fine details, we might be relying on just 20–30 cones, each with a diameter of about 1 micron or one millionth of a metre. This biological machinery far outclasses any detector surface yet used in machine-vision systems By way of contrast, it may be noted that a television camera usually produces a picture of only about 400,000 spots (or pixels). This is an extremely small number compared with the number of elements in the picture provided by the human eye.

There are, then, about 126 million receptors (equivalent to pixels) in each eye; but only about 800,000 nerve fibres leave each eye, carrying messages to the brain via the optic nerve. Clearly, each receptor is not served by a single cable to the brain, since the average ratio of receptors to optic-nerve fibres is about 15 to 1. In the fovea, the ratio is much closer to 1 to 1, but it is misleading to think of each foveal receptor having its own 'private telephone line' to the brain, because in between each receptor and the output cables of the retina are many cells which link together receptor signals in various ways. It is to these cells that we now turn.

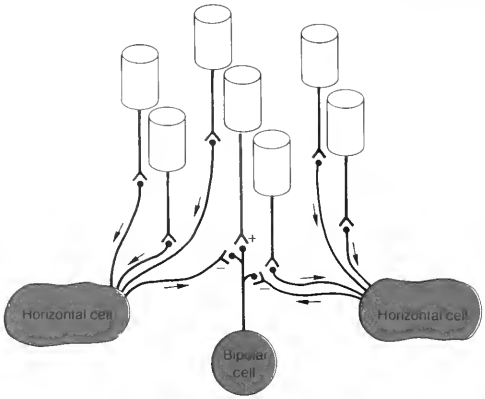
Once stimulated, the receptors pass on a signal to *bipolar cells*, which in turn pass on signals to *ganglion cells*, which then send messages to the brain along fibres in the optic nerve. This sequence of connections is illustrated schematically in **154** [plate 10]. Note that each receptor can serve more than one bipolar cell (which is why the idea that receptors have private lines to the brain is misleading), and that ganglion cells can collect messages from more than one bipolar cell. The receptor-bipolar-ganglion cell pathway is often referred to as the 'vertical organisation' of the retina, to distinguish it from the 'horizontal organisation', which consists of cells and their fibres running laterally. The latter cells are of two types: *horizontal cells*, which carry lateral messages at the level of the receptor-bipolar junctions, and *amacrine cells*, which do likewise at the bipolar-ganglion cell junctions. It must be emphasised that **154**, though adequate and not misleading for our purposes, is a highly schematic outline of retinal anatomy. Subdivisions can be distinguished within the various cell classes, just as the receptors can be divided into rods and cones. Indeed, retinal anatomy is a highly developed field of knowledge. What is less well understood is how the retina functions and what computational job it performs. Why do so many cells intervene between the receptors and the cells of the brain concerned with building up a scene description – the simple cells of the striate cortex, and so on?

The answer to this question in general terms is that the retina does a good deal of visual analysis itself before it sends messages on to the brain. It is, if you like, a visual brain itself. Interestingly, the retina develops in the foetus as an outgrowth of embryonic brain tissue; so calling it a brain is not far wide of the mark. But if the retina is to be thought of as a mini-brain, what is it 'thinking' about? What are its tasks of visual analysis? It turns out that the answer is different for different animals. Those that have well-developed visual mechanisms within their brains proper, such as men, monkeys and cats, have rather simple-minded retinas. Other animals, without the same extensive central brain machinery for vision, rely more upon their retinas for visual analysis, which therefore become more complicated. In the neurophysiological account which follows, we will content ourselves with describing the typical workings of the more straightforward retinas.

The Neurophysiology of the Retina

Some important clues about retinal function have come from neurophysiological studies using the technique of single cell recording described earlier for striate cells (chapter 3, p. 42). The general idea is to insert a microelectrode into a retinal cell and record from it while stimulating the retina with a pattern of light (compare **56**). The pattern of light is manipulated in various ways, and the processing job performed by the cell is then inferred from the nature of its responses, its optimal stimulus, and so on.

The creature which has contributed most to our understanding of retinal function in recent years is the *mudpuppy* (*Necturus maculosus*), a curious fish that lives in the muddy depths of silt-laden rivers. Ironically, the environment of the mudpuppy is such that it can hardly depend much on vision for its survival. And yet this animal has proved a boon for neurophysiologists interested in the retina because it is blessed with relatively huge retinal cells. John Dowling and Frank Werblin have exploited this fact to insert the fine tips of microelectrodes into every type of cell, even receptors,



155 Highly schematic and abbreviated wiring diagram for a bipolar centre-surround cell The + and − symbols show excitatory and inhibitory synapses impinging on the bipolar cell

penetrations which could not be made in most animals without causing significant damage. Once the electrode has been so inserted, it is possible to record from the cell as it responds to stimuli flashed on a screen viewed by the animal.

With such techniques, Dowling and Werblin have found that the receptors respond to the luminance of the input image in just the way required of a grey level description. That is, their response (a change in voltage across their cell membrane, in fact) is proportional to the intensity of incident light, and lasts for as long as the stimulus lasts. The voltage level is thus wholly equivalent to the numbers in a computer's grey level description of an input image. But remember that the concept of 'grey level' is used very broadly in this book, and includes a cone responding selectively to a particular colour (or more properly, light of a particular wavelength – see later). This might prove confusing to some readers, who might therefore prefer to think in terms of a cone coding 'blue level', 'red level' or 'green level' as the case may be. Space forbids my taking much account of colour perception, though I will mention it briefly later on.

The receptors, then, are the site of the grey level description. What comes next? The answer is that the bipolars are the site of the centre-surround edge measurements. The wiring diagram is shown in highly schematic and abbreviated form in **155**, which shows the arrangement for an on-centre bipolar. This unit receives excitation from a central receptor via a junction called a synapse (p. 32). Other receptors surrounding this central receptor feed inhibition to the bipolar, but not directly: instead, the receptors of the surround feed into horizontal cells which then proceed to inhibit the bipolar. The horizontal cells thus look like the ideal site for providing the required weighting of surround receptor signals, a very neat solution to the question posed earlier about how this might be done.

It is very important to realise that **155** is *highly* schematic, and that for the sake of simplicity lots of interesting details are left out. For example, several receptors might feed excitation into the bipolar, which would thereby become more sensitive, albeit at the price of loss of detailed accuracy, because the bipolar would not then be able to provide a varying signal if a very small spot of light was moved about over this enlarged central region. It would always 'think' that the spot was in the same place. In the more realistic, although still schematic, retinal wiring diagram of **153**, most bipolars can be seen collecting inputs from several receptors – the second bipolar from the right is the only one shown receiving an input from a single receptor.

Another interesting complication which is not made evident in **155** is the fact that in addition to the receptor-horizontal-bipolar connections there are also receptor-horizontal-receptor linkages. Dowling and Werblin suggest that these latter pathways enable receptors to be adjusted in their sensitivity according to the general prevailing illumination of a scene.

The last but very important complication I will draw attention to is that the receptors can feed two bipolars simultaneously. One bipolar of each pair might be an on-centre unit, the other an off-centre one. Certainly about equal numbers of bipolars of each type are found in the mudpuppy, and this makes good sense, of course, the two types of bipolars being the beginning of the whiteness and blackness channels respectively. The fourteenth receptor from the right in **154** is shown feeding two different bipolars, although most are shown feeding just one. An off-centre bipolar would have the same basic wiring diagram as shown in **155** for an on-centre bipolar, but with the location of excitatory and inhibitory influences reversed. For these off-centre bipolars, that is, the central receptor would feed inhibition and the horizontal cells would feed excitation.

If the bipolars are the site of the first step in lightness computation, that of edge detection, where are the sites of the next two processes – thresholding and deconvolution? Marr has suggested that the bipolars themselves are well suited to operate in a threshold manner, which would mean that they respond only if the edge they are dealing with is sufficiently prominent. How their threshold for responding is adjusted, as it must be to meet the requirements posed by variations in the overall level of illumination, is not clear, but it could be set somehow by horizontal cells or by amacrine cells.

But the most interesting suggestion of Marr's concerning retinal function is that the deconvolution operation is performed at the bipolar-ganglion cell junction, and is mediated by the lateral connections provided by the amacrine cells. I will not go into the details, but the general idea is that there are two sets of ganglion cells, one set carrying the whiter-than-grey lightness information and another set dealing with the blacker-than-grey. Each pathway would be fed by bipolars of matching type, and the need for the two pathways to be kept closely coupled in their operations, for reasons mentioned earlier, might be met, according to Marr, by connections provided by yet other types of amacrine cells. Look back to **154** and note that the inner synaptic layer, where the bipolar-ganglion-amacrine cell junctions are located, is richly provided with lateral fibres supplied by the amacrine cells, and such a network seems well suited in principle to effect the deconvolution operation which requires just such lateral connections (refer back to **148**).

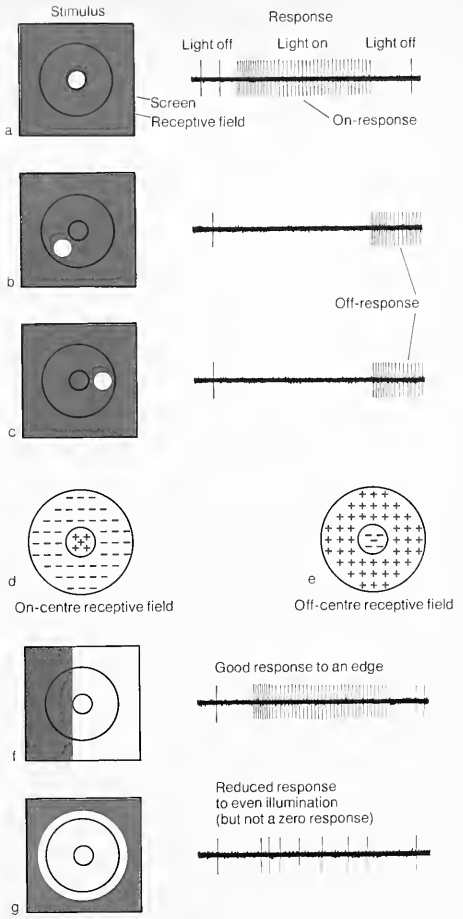
Marr's suggestion about deconvolution being effected at the ganglion cell layer is, however, a controversial one and not yet confirmed. The conventional view is that ganglion cells are edge-detecting units, much as described above for the bipolars. That is, they are customarily viewed as units which

help to detect contrast changes despite variations in general luminance. Certainly ganglion cells show an antagonistic centre-surround receptive field organisation when studied using the single cell recording technique in conjunction with simple stimuli such as spots or edges. This has been known since the early pioneering days of single cell recording, from work on the retinas of such diverse creatures as cats, frogs and rabbits. And centre-surround organisation at the ganglion cell level has been the usual finding for most cells. Typical recordings are shown in **156**, obtained from a cat using the general single cell recording set-up described in connection with the monkey shown in **42**. For this ganglion cell, a spot of light which was shone on the screen in such a position that it landed on the centre of the receptive field caused the cell to emit a vigorous burst of impulses [**156**a]. On the other hand, a spot shone anywhere in the periphery of the receptive field (e.g. **156**b, c) caused inhibition, with elimination of the resting (sometimes called 'maintained') discharge rate. Clearly, such a cell is an on-centre/off-surround unit. Note that when the spot is switched *off* in the surround, the cell then emits a burst of impulses, as though the inhibition produced by the spot had 'held it down' in some way, only for it to 'bounce back' briefly as soon as this inhibition ended with the spot being switched off. It was this kind of response which gave rise in the first place to the terms *off*-surround and (where appropriate) *off*-centre, the post-stimulus burst of impulses being called an *off*-response. Note that the centre and surround regions are concentric and can be plotted out by exploring with a spot of light to give the receptive field diagram shown in **156**d. An off-centre equivalent diagram is shown in **156**e.

Retinal ganglion cells respond very well to suitably positioned edges. For example, in **156**f a light-dark edge is shown falling on the receptive field in such a position that the excitatory centre receives a high light intensity all over it, whereas the inhibitory surround is only partly covered. This stimulus arrangement produces a lively response because insufficient inhibition is generated from the surround to cancel the excitation produced from the centre.

What happens if a patch of light is used which is large enough to cover the whole of the receptive field which is discovered using small spots of light? Most retinal ganglion cells seem much less responsive in these circumstances [**156**g], but it is not true, as is so often stated in introductory textbooks on vision, that such a stimulus produces no response at all, at least, not true for all types of retinal ganglion cells. Many of them are quite responsive to even illumination, albeit not so responsive as to a suitably positioned edge.

How do these response characteristics of retinal ganglion cells fit in with Marr's idea about their being the site of deconvolution? The arguments are highly technical: suffice it to say here that some trace of antagonistic centre-surround organisation would be expected on Marr's view if these cells are studied using spots of light, because they receive their input from the bipolars which, it is commonly agreed, initiate the centre-surround computation. The responsiveness to large areas of illumination is also to be expected, of course, if the ganglion cells are indeed carrying the output of the lightness computation: we see the lightness of surfaces over their whole extent, of course, and not just at their edges. Presumably this responsiveness to even illumination comes from facilitation fed to them from their neighbouring ganglion cells, perhaps via the amacrine cells (see above), with



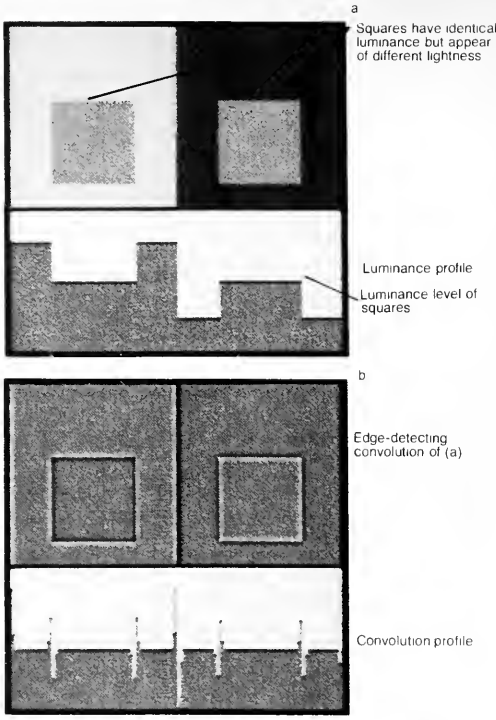
156 Single cell recordings from a cat's ganglion cell

these neighbours receiving the input either from other neighbouring ganglion cells (again fed through by amacrine cells), or from bipolars responding at a high level due to their discovery of an edge. But all this would be so only given Marr's theory of how the Land-Horn lightness computation might actually be performed in the neural machinery of the retina. This theory, while attractive in many ways, has yet to be confirmed, and it could be that the lightness deconvolution is in fact performed somewhere in the brain, rather than in the retina. John Mayhew and I are trying to investigate this problem, as well as the whole lightness computation scheme, by building a computer simulation of Marr's theory to see just what output would or would not be expected for various types of stimuli. Without actually implementing such a theory in a computer program, it is difficult to know exactly

what its detailed performance is like on all types of stimuli. Moreover, as Horn has suggested, if the deconvolution process was not perfect (i.e. not the exact inverse of the initial centre-surround operation which removes the illumination component of luminance variation), then some preferred sensitivity for edges would be expected at the ganglion cell level. Indeed, a lack of perfect precision at the deconvolution stage might be essential if we are to explain various *lightness contrast illusions*.

But before turning to these and leaving the topic of retinal neurophysiology, let me mention one curious property of some retinal ganglion cells which does seem to fit in with Marr's theory. This is the *periphery effect*, discovered by James McIlwain. He found that if an edge is presented well to one side of the receptive field of a retinal ganglion cell on which is falling a just sub-threshold stimulus, then the cell can become active and 'notice' the sub-threshold input which previously went undetected. At first sight this seems surprising: here we have a stimulus falling *outside* the conventionally-plotted receptive field (using small spots of light) which is none the less capable of influencing the cell. However, if retinal ganglion cells really are the site of deconvolution, then some such effect as this would be expected, with the facilitation being mediated by amacrine cells passing on lateral facilitation between retinal ganglion cells. Whether this is indeed the correct interpretation of McIlwain's periphery effect remains to be seen, but it certainly seems in keeping with expectations based on retinal ganglion cells being the site of deconvolution. Further studies of this effect can now be expected, guided by ideas of what 'should' be found if the Land-Horn-Marr theorising is valid.

At this point, it is necessary to emphasise that the account of retinal ganglion cells given here is necessarily very abbreviated and simplified. There are in fact many different types of retinal ganglion cells (just as there are different types of receptor, bipolar, horizontal and amacrine cells), and it is almost certainly the case that not all of them are involved in deconvolution, even if some are. The ganglion cells most likely to be those subserving deconvolution are ones which respond in a *sustained* fashion to a stimulus, producing a response for as long as the stimulus lasts. These are termed X cells. Other ganglion cells respond in a *transient* manner to a stimulus, producing a burst of impulses only when it first appears and/or when it goes off. These are called Y cells and might well be involved with movement detection (see earlier discussion on p. 62). Yet another class are the W ganglion cells, some of which at least are characterised by the total absence of a centre-surround organisation: so they might conceivably have something to do not with edge-detection and the lightness computation, but rather with the detection of overall illumination level. Moreover, cells of all types differ in their receptive field sizes and in their distribution across the retina. X cells are found mainly in the central retina, Y cells in the periphery of the retina, and ganglion cells in general show increasing field sizes as one moves from centre to periphery. Certainly the retina is a complex structure which fully justifies the title of 'mini-brain'.



157 An explanation for a simple contrast illusion

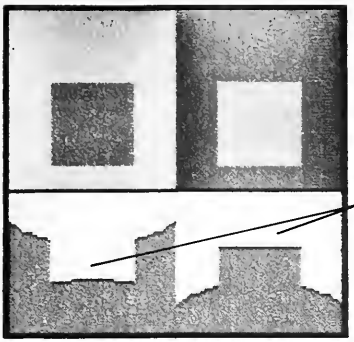
Contrast Illusions

I began explaining the probable basis of lightness computation in the visual system by introducing the Craik-Cornsweet-O'Brien illusion. This is a fine example of a contrast illusion, i.e. a stimulus possessing certain light and dark regions which combine to give an illusory lightness/darkness effect of some kind. There are many such illusions and, as usual with illusions, they offer valuable clues about how visual mechanisms work.

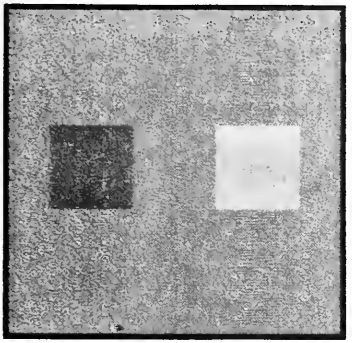
Let us begin by considering perhaps the simplest contrast illusion of all, one first shown in chapter 1 in **11** and now repeated here in **157**a. The small inset grey squares are of equal luminance, difficult though this is to believe because the one set on a white ground looks so much darker than the one set on a black ground. Why should this illusion, this failure of 'correct' lightness perception, come about? In fact, an illusion of this type is a necessary consequence of the Land-Horn lightness computation, and its existence adds credibility to the idea that our own visual system works in a similar fashion. The illusion is, if you like, a necessary penalty that has to be paid if the strategy adopted for obtaining reasonably truthful lightness perception in most circumstances is one which begins with an edge-detecting phase, followed by a deconvolution phase. I will now explain in some detail why this must be so.

First, consider **157**b: it shows the result of subjecting **157**a to an edge-detecting convolution of the kind thought to be performed by bipolars. **157**b is therefore a convolution image (p. 38) and is to be interpreted in the following way: think of it as composed of a mass of dots, so small and so closely packed together that each individual dot cannot be distinguished as a separate entity. Each dot represents the

c



d



output from a pair of bipolar cells, one an on-centre unit and one an off-centre unit. If both bipolars are firing at zero level, then the dot representing them is set to a mid-grey level. On the other hand, if one or both is firing at a non-zero level, then the dot is set either to whiter-than-mid-grey, if the on-centre unit is dominant (more active), or to blacker-than-mid-grey, if the off-centre unit is dominant. The convolution image as a whole is therefore signalling where the edges are, and this can easily be seen both in the convolution image itself and in the profile of activity across a central slice of this image which is shown just beneath it. This edge profile is, of course, very reminiscent of the edge profiles shown in connection with the Craik-Cornsweet-O'Brien illusion [**140**].

Second, consider **157**c: this shows the result of deconvolving the edge convolution image of **157**b, using procedures similar to those described earlier in this chapter. Think of **157**c as illustrating the activity in retinal ganglion cells, viewed as deconvolution devices. Each dot of the image (again the dots are too small and closely packed to be picked out individually in the illustration) is set to mid-grey if an on-centre and off-centre pair of cells dealing with the image point in question are both either inactive or active at the same level. But if the on-centre unit 'gets on top', as it were, then the dot in the image is shown as whiter-than-mid-grey. If the off-centre unit wins out then the dot is set to blacker-than-mid-grey. The computer program which did this deconvolution was written by John Mayhew and its upshot is shown at a point in the computer's deliberations where full and proper deconvolution has not yet been obtained.

The important point to note about **157**c, of course, is that it embodies a contrast illusion of the type we see ourselves when we look at the input stimulus – **157**a. It is necessary to be careful here to look at the lightness levels of the inset squares of **157**c in similar surrounds, otherwise the visual system will quite happily treat **157**c as just another input figure for its own lightness computation and give us another contrast illusion which we might be misled into thinking was a 'physical' event rather than a psychological one. So a luminance profile of **157**c is given just below it and, in addition, the squares are printed out separately on a mid-grey surround. In both cases, the 'genuinely' different lightnesses of the inset squares can be appreciated. Thus the Land-Horn lightness computation produces a contrast illusion, and it does so as an inevitable consequence of its stratagem of using an edge-detecting phase followed by a deconvolution phase. This is an impressive parallel with what we experience, and certainly supports the idea that the visual system itself operates in at least a broadly similar fashion.

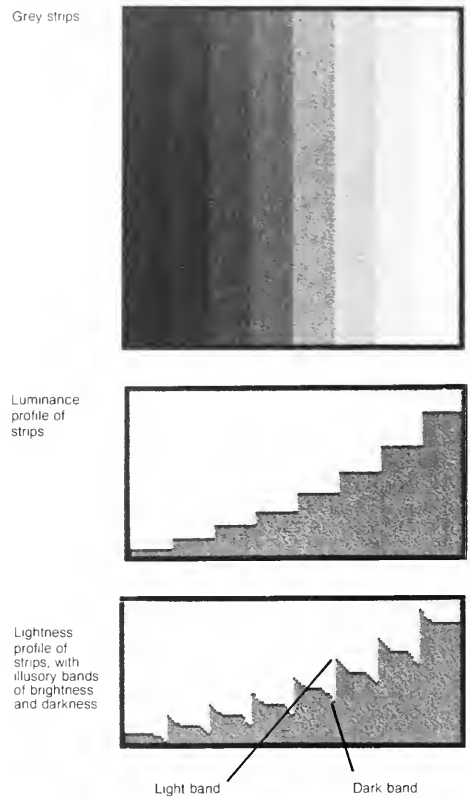
Mach Bands

Perhaps the most famous contrast illusion of all goes under the title of *Mach bands*. Ernst Mach was an Austrian sensory physiologist whose investigations of various visual phenomena led him to conclude that there must be processes of lateral inhibition operating within the visual system – and he arrived at this conclusion in the 1860s, well before the modern weapons of microelectrodes were available. (Remember, centre-surround units are built on the basis of lateral inhibition: e.g. an off-surround can be said to laterally inhibit an on-centre unit.) The kind of phenomenon that impressed Mach, and to which his name is now firmly attached, is illustrated in **158**. Each strip of grey is in fact of uniform luminance, as the luminance profile of the figure shows, but appearances are very different. Thus the boundaries between strips are apparently bounded by light and dark bands, illusory Mach bands, so that the perceived lightness profile is as shown in the lower part of **158**.

How can Mach bands be explained? It might be that they are an inevitable consequence of a Land-Horn type of lightness computation, just as was the contrast illusion dealt with above. On the other hand, it might be necessary to regard them as a failure of lightness computation which is not intrinsic to the Land-Horn approach, but instead the product of a rather poor embodiment of this stratagem in the visual system. For example, it may be that the deconvolution network is not perfectly designed to cope with the edge information it receives from the bipolars, so that this network, be it in the retina or the brain, is not quite up to the job of extending out from edges in the circumstances which give rise to Mach bands. But why exactly the failure should crop up in certain figures and not others is a tricky research question still being tackled.

The Hermann Grid

Another famous contrast illusion is shown in **159**, an illustration taken from a book written by John Tyndall in 1869 on sound. While reading this book, Hermann noticed that dark shadowy dots appeared at the intersections of the component figures. Thus it was that Tyndall's book on sound came to make an indirect contribution to vision.



158 Mach bands

Hermann reported his observation in 1870, and thenceforth the so-called Hermann grid illusion has been much studied and debated by visual scientists. The explanation most favoured at present is one which sees the illusion as an unwanted side-effect of centre-surround analysis. The general idea is shown in **160**, where it can be seen that an on-centre unit dealing with an intersection [**160**a] receives four 'doses' of inhibition from the surrounding arms of the grating, whereas a similar unit dealing with a zone in between intersections [**160**b] would receive only two 'doses' of inhibition. The net result would be that a weaker centre/surround signal would be generated at the intersections than in between them. It can thus be appreciated why the intersection appears grey in comparison with the regions in between which appear white.

One interesting property of the Hermann grid is that the grey spots do not appear at the intersection which is being viewed directly, but only at those intersections which fall on peripheral retina. At least, this is so in **159**, but one can obtain a Hermann grid effect in central vision if the size of the lines is reduced sufficiently. For example, you should be able to see small grey spots in **161** where the lines are thin even with central (or at least near-central) fixation of the intersections. The reason for this is almost certainly related to the fact that the size of centre-surround receptive fields varies across the retina, with small ones found in central regions and progressively larger ones out towards the periphery. This obviously fits in well with the explanation of the illusion in terms of centre/surround analysis, because this explanation relies on a fairly 'neat' fit of size of intersection to size of on-centre [**160**]. Given this explanation, one can turn the tables on the Hermann grid, as it were, and instead of regarding it as an illusion to be explained, one can regard it as a device for measuring centre-surround receptive-field sizes in human observers, for whom microelectrode penetrations are obviously inappropriate. When this is done, by finding the size of intersection which is necessary to give an illusory grey spot, for a whole range of different eccentricities of fixation (i.e. distances from central gaze of the intersection in question), then one can plot a graph of 'size of intersection required' against 'eccentricity of viewing' and use this to estimate the sizes of field centres in different retinal locations. This is the very first visual psychophysics experiment which is performed by students attending our course on seeing in the University of Sheffield. Use of a visual illusion to track down a neurophysiological mechanism whose computational significance is now beginning to be understood is a particularly good way to start the course.

Back to Kanizsa's Triangle

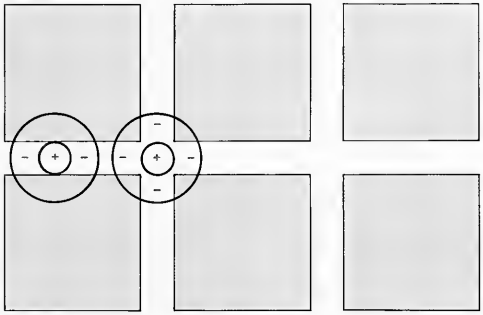
In chapter 5, an illusion called Kanizsa's Triangle was introduced [**162**], together with an explanation of its illusory light-dark contours as the result of the visual system's attempt to interpret the input stimulus in terms of the 'object hypothesis of triangle masking discs, etc.'. This kind of 'cognitive' or 'high-level' explanation, preferred by Richard Gregory and others, stands in marked contrast to the 'low-level' or 'data driven' one which treats the contours as a side-effect of the visual system's lightness computation. This latter view, espoused by myself, Jeremy Clatworthy and others, attempts to explain the contours as unusual instances of lightness contrast illusions. The basic idea here is that the input is analysed by centre-surround units, and then the deconvolution process extends out the edges cut into the discs, forming the apices to the triangle, beyond their 'proper' bounds, so causing the illusion. John Mayhew and I are currently trying to test this idea with a rigorous computer simulation of the machinery for lightness computation described earlier in this chapter, and this work is not yet finished. But there is some indirect evidence in favour of the idea, as follows.

First, it is not necessary to have an 'incomplete' masking object in order to get the lightness illusion. Even if a clearly visible masking triangle is substituted [**163**], an illusory lightness increment is still apparent. It is not so easy here to see why the illusion should come about if the correct explanation is a high-level one to do with object hypotheses. That is, there just seems no reason for the error to come about in **163** if the effect is due to processes of high-level interpretation. Of course, it could be argued that **163** is simply a *different* illusion from **162**, despite some superficial similarities. But in any event, **163** shows how Kanizsa's triangle can in principle be considered as a lightness contrast illusion with an incomplete border (compare **164**) and a white inside area that none

the less suffers the lightness contrast effect.

Second, the prominence of the illusory contours in Kanizsa's figure are greatly reduced, in my opinion to zero, if the luminance of the inducing discs and lines are made about the same as that of the ground on which they lie. This can be done in various ways, but in **165** [plate 11] washed-out pastel shades are used. Here the input 'evidence' still exists for a masking object, so why is it not 'hypothesised' and an illusory light/dark contour made manifest accordingly? This seems to me a difficult version of Kanizsa's triangle for the cognitivist theory to cope with. Another difficult figure for this theory is **165**b, which presents sectored discs made out of dots, rather than solid black. An illusory triangle is not present here, although the figure gives masking cues, and so the expectations of the cognitive viewpoint are not fulfilled. On the other hand, dot textures of the kind used in **165**b generate poor lightness contrast effects, and so on that theory the absence of an illusory triangle is sensible.

But Kanizsa's triangle is right in the centre of a controversy at present, about what kind of theory is appropriate, and under

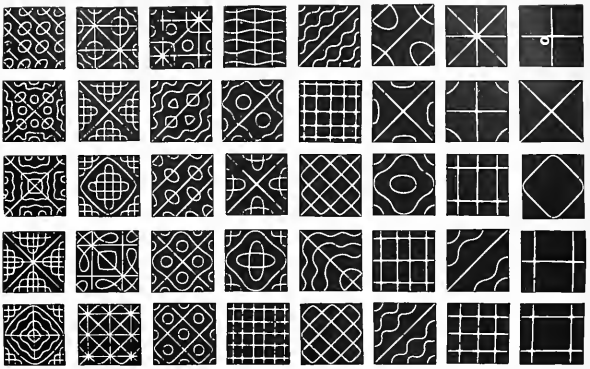


160 The possible influence of black and white areas upon the activity of on-centre receptive fields Greater inhibition would be operating on those stimulated at the intersections, as compared with others.

what circumstances, for illusory contours, and it would be unwise to be dogmatic about its 'correct' explanation. My own clear preference is to push the low-level theory as hard as possible to see what its limits are before bringing in high-level factors. The future, however, might decide that *both* types of explanation have a contribution to make, a familiar end result of so many scientific controversies.

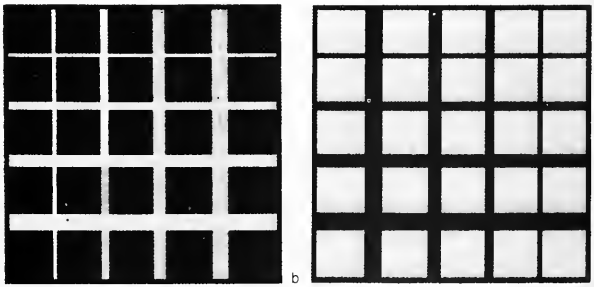
Colour Perception as a Lightness Computation

I have said already that space forbids a proper treatment of colour perception in this book. But I will mention briefly Land's view that colour perception is a matter of combining the results of lightness computations conducted on *three* different grey level descriptions – one provided by cones sensitive to blue light, one by red-sensitive cones and one by green-sensitive cones. The idea that these cones can be thought of as providing different grey level descriptions, each built from light of a different wavelength (wavelength is the physical correlate of colour: long wavelengths appear red, medium wavelengths green, and short wavelengths blue) is illustrated in **166** [plate 11]. The four pictures shown are of



159 The matrix of Chladni figures in which Hermann first observed the dark dots at the unfixated intersections

the same scene, but one is produced with all three wavelength ranges and hence appears coloured, whereas the other three show the scene in greys which correspond to the luminance of the objects in the scene when viewed by light of one particular wavelength. Thus each of these three pictures is, if you like, a grey level description of a coloured input. What Land has suggested is that the three different types of cones provide just such descriptions of the observed scene, that each is then subjected to its own more or less independent lightness computation (to get rid of variation in illumination), following which the information on the three lightness scales is combined into the colours which are one of the most prominent aspects of our visually perceived world. Whether the lightness computations really are fully independent is a controversial question, and we know very little about how the lightness information on each scale is brought together as a neural representation of some kind. But Land has pointed the



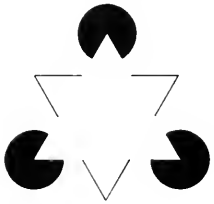
a b

161 Hermann grids with intersections of varying sizes Try looking just to one side of the intersections and note that you need to look much farther away from thick-line intersections to obtain an illusory spot than you do from thin-line intersections. This effect forms the basis of the laboratory class experiment described in the text. Note also that whereas (a) produces darker-than-white spots at the intersections (b) produces whiter-than-black. You might find it instructive to try to explain this result in terms of centre-surround analysis.

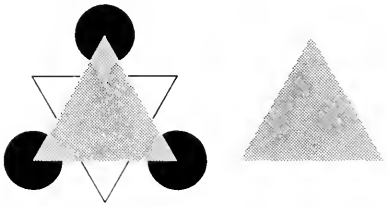
way forward, basing his so-called retinex theory on some remarkable phenomena of colour perception which the interested reader can pursue further by consulting the reference list at the end of the book.

Conclusion

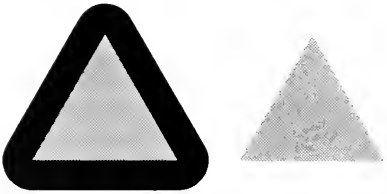
Perhaps you have been rather more surprised by this chapter than by the other ones so far. After all, how many non-specialists would suspect that there was much problem about anything so mundane as seeing blacks and whites? And yet some remarkably intricate machinery is necessary to extract even this elementary information from the inherently ambiguous retinal image. Indeed, the processes are so intricate, and even now so poorly understood, that it has seemed wise to postpone until this late chapter a description and discussion of the neural machinery which we know to be crucially involved – that of the retina. This is in itself odd. Who would have thought that a book about seeing should postpone almost until the last a discussion about the eye?! But this has been done because the business of seeing blacks and greys and whites is a difficult matter and best coped with when the reader has been through the easier material on seeing objects and features.



162 Kanizsa's triangle



163 Kanizsa-type figure The grey triangle enclosed by sectored discs and interrupted lines appears lighter than the physically similar triangle to its right.



164 Classical simultaneous brightness contrast demonstration The left-hand triangle appears brighter than the right-hand triangle.

One of the curious things described in this chapter has been that the visual system detects blackness as a property of surfaces in its own right, just as it might detect blueness, redness or whatever. This is the opposite of most people's intuitions. The 'inner screen' theory of 1 reflected the commonsense view that blackness is registered simply as the *absence* of activity in nerve cells, but this natural expectation is in fact far from the truth. Blackness is coded by *activity* in a certain population of nerve cells, with greater activity coding a deeper black. Of course, we do not know as yet where this coding finally takes place – we do not know the brain site whose activity is the neural correlate of conscious awareness of blackness. The cells in the retina that seem related to blackness perception can only be a first step: no one supposes that the retina completes the whole business of lightness perception. Somehow its outputs have to be tied in with a whole host of other visual processes carried out by the brain, and what this chapter has been concerned with is but an early stage of the mechanisms for seeing lightness and brightness. But this is a fair reflection of current knowledge: we have little idea about brain mechanisms concerned with these perceptual attributes.

If you find it difficult to believe that blackness is coded by activity in certain cells, rather than simply by *in*activity in cells which cover the whole black–grey–white range, you might find it helpful to consider the following curious fact. A television screen appears a pale grey when switched off, and yet it has no trouble presenting us with blacks if the image being transmitted requires them. This is odd because the electronics of the television set have no way of dimming the screen, only of brightening it up! So the blacks we perceive in television images are 'created' by our visual system as a product of its lightness computation. Blackness is 'discovered' where only grey physically exists. And if our brain cells responded simply to the physical luminance of points on the screen, as proposed by the 'inner screen' theory of 1, then we should never be able to see blacks on the screen at all – only grey through to white. Most odd!

It is difficult enough, in my experience, to convince people about the problem of seeing objects or seeing features. But it is even more difficult to persuade them that seeing lightness is problematic. The example given in the previous paragraph, plus the various contrast illusions shown in this chapter, might help undermine the commonsense view and so lead you to an appreciation of the difficulties. But if you need a final shove in this direction, it is perhaps as well to end with a famous truth which almost all psychology textbooks present to students. This is that when you descend into that dark cellar to collect some coal which is stacked up against a white-painted wall, and then return to a sunlit room with a bucketful of the substance, the intensity of the light entering the eye from the sunlit coal is greater than that which came from the dimly-lit white cellar wall. And yet the coal still looks black, and the wall still looked white. This illustrates the computational problem of lightness perception *par excellence*. Black remains black even if it has greater luminance than a poorly illuminated white – a superb piece of visual computation.

7 SEEING WITH TWO EYES

Have you ever wondered why we have two eyes? If you think the answer is obvious, try the simple experiment of closing or covering one eye and then walking around the room, or wherever you are reading this book. Did the world look very different? Did you bump into anything? I doubt it. The visual system is so richly endowed with a variety of mechanisms for explicit scene description that you were able to see the room's various objects quite well and in their proper spatial relationships despite being limited to one eye. Indeed, this simple experiment shows that we can get by so well without two-eyed or *binocular* vision that some of the celebrated cases of one-eyed or *monocular* achievements become thereby rather more understandable. For example, the Nawab of Pataudi played first class cricket with great skill after losing the sight of one eye, and Wiley Post, who made the first solo flight around the world, was similarly one-eyed. Such high-class performances demonstrate that whatever the benefits bestowed by binocular vision, they are not so crucial as to make monocular vision a desperate handicap. This conclusion is supported by the estimate that approximately 2 people in every 100 are essentially monocular, in that although each of their eyes works well enough alone, the eyes do not cooperate as they should. But these people get by remarkably well, so much so that many of them have no idea that they suffer a binocular visual deficit until it is detected by routine clinical screening. Such considerations as these suggest that the proper answer to the question with which I began this chapter is that the prime benefit given by two eyes is quite simply that having two is a good insurance against losing one!

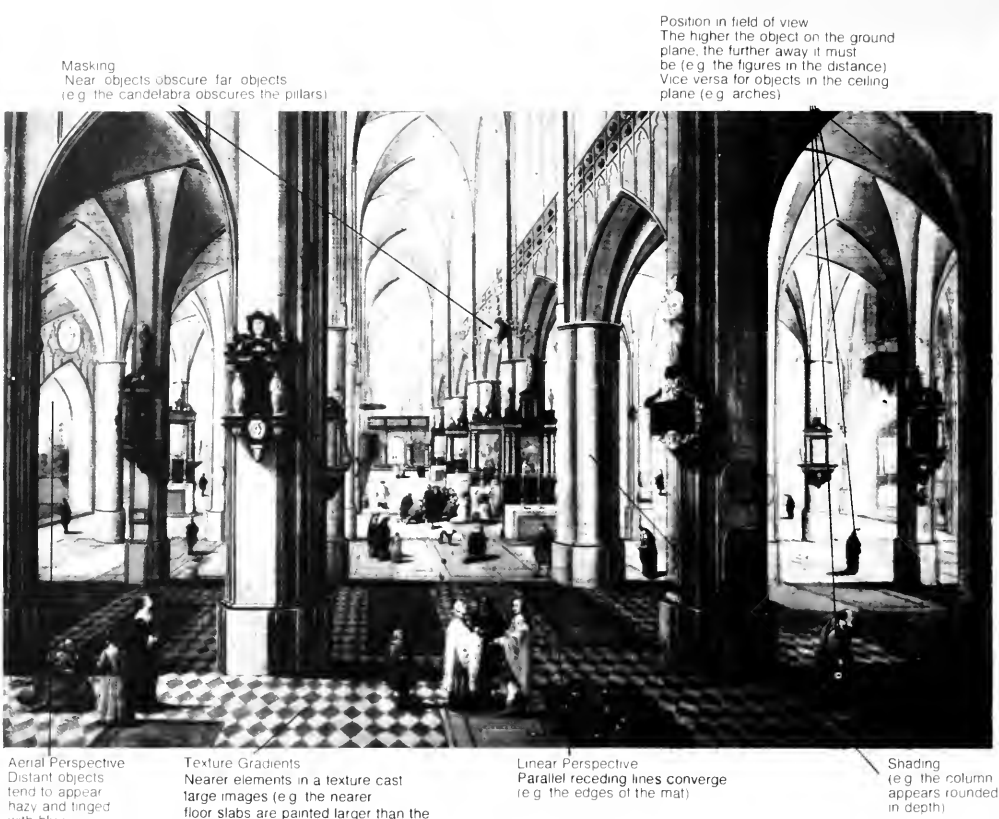
But if we choose the viewing conditions more carefully, a one-eyed viewing experiment *can* give us a clue about a possible reason for having evolved binocular vision. Try the experiment again, but this time look first with two eyes, and then with just one, at a vase of flowers, a tree, a bush, or some similar object which has lots of parts arranged at different distances (or 'depths') from you. With one eye only it is difficult to discriminate the relative depths of the leaves, petals, etc., but with two eyes it is easy. The appearance of depth when the second eye is opened is called *stereopsis* (named from the Greek for 'solid sight'). Repeat the experiment several times to be sure that you see the no-depth/depth transition as you open the closed eye. Be sure also to look for several seconds with one eye only before looking again with two. This precaution is necessary because the three-dimensional (3D) effect existing prior to closing one eye, and itself built up with two-eyed vision, takes that long to disappear. Be careful not to move your head at any stage. Objects at different depths produce retinal images which move at different rates across the retina when the head is moved, a clue (or 'cue') to depth called *head-movement parallax*. The brain is quite up to the task of using this cue to generate vidid depth perceptions which can appear equally as good as stereopsis. This is why in the opening paragraph you were invited to *walk* around the room with one eye covered, rather than looking at it from a single stationary position. Walking ensures that the depth cue of head-movement parallax is available to the visual system, and hence walking helped my purpose of showing just how well we can cope when forced to rely on just one eye.

But although walking helped ensure a 'depthful' visual world in the initial one-eyed viewing experiment, it is by no means the case that the world appears totally flat with stationary monocular inspection. To be sure, certain sorts of objects, such as vases of flowers, lose a certain kind of subjective 'depthfulness' (i.e. that called stereopsis) under these conditions, but the world as a whole is still seen in 3D. Some of the *monocular depth cues* which the brain uses for this kind of distance perception are illustrated in **167**, which also demonstrates how adept painters have become in deploying these depth cues for their own ends.

Of course, a picture presents us with a paradoxical perception in that we see depth within it and yet at the same time we also see that it is printed on a flat piece of paper. If cues supporting the latter flat aspect of the perception are eliminated by viewing the picture monocularly through a tube which obscures everything other than a one-eyed view of just the picture itself, then the paradox is removed and, moreover, the depth impression is heightened - a trick well worth trying on your next visit to a picture gallery.

Anaglyphs

Towards the top of **168** [plates 12–13] are shown two rather similar pictures. Just underneath them appears an *anaglyph* of this pair created by printing the picture on the left in red and the picture on the right in green, one on top of the other. The name 'anaglyph' comes from the Greek for a carving in relief, and if you try looking at the anaglyph through the red/green spectacles provided on the inside back cover of this book, you will see why the name is so appropriate. If you normally wear glasses for reading, do not take them off, but hold the red/green spectacles in front of them. Make sure that the **R**ed filter is in front of your **R**ight eye, the green filter in front of



167 Monocular cues to depth

your left. If your two eyes work together as they normally should, the anaglyph will appear three-dimensional. If you find the 3D effect does not come at first, be persistent. Many people take a few minutes to discover it, but still find it perfectly vivid when they succeed. Once you think you have experienced the effect, try closing or covering one eye. The illusion of depth will then vanish, which demonstrates that it depends on binocular vision and not on any of the monocular depth cues present in either the red or green component pictures.

The way an anaglyph works is also illustrated in **168**. The basic idea is that the two component pictures are the two halves of a *stereogram*, also called a *stereo pair*, and that the red green printing coupled with the red/green spectacles is a way of presenting one component picture to one eye and the other component picture to the other eye.

Some people, when they are shown anaglyphs for the first time, are worried that they may not be able to experience the 3D effect if they are colour-blind. But there is nothing inherently necessary about using red and green colours: the depth effect seen in the anaglyph has nothing whatever to do with colour vision as such and colour-blindness does not affect it at all.

Binocular Disparity

Look carefully at the two separate halves of the stereogram shown in **168** (upper). Although the two pictures are similar they differ in their details. These differences, technically termed *binocular disparities*, mimic differences which ordinarily exist between the views of genuine 3D scenes which are received by the two eyes.

The most prominent binocular disparity between the two component pictures is perhaps most easily seen when the anaglyph itself is inspected without the spectacles: a disparity in the position of the figures in the red and green images is then clearly visible. But of course, this particular disparity is a deliberately artificial one, carefully created photographically to give the curious illusion of these figures floating in mid-air behind the page when the anaglyph is viewed with the spectacles.

The fact that the two eyes do not receive an exactly similar view of a 3D scene is illustrated in **169** [plate 13] by a picture of an observer looking at a solid object – a pyramid with the top sliced off. The left and right views are shown above the observer, and you can see that the top of the pyramid falls in a slightly different position in each one: it is off-centre in both views, to the right in the left view, and vice versa. This difference is simply a straightforward geometrical conse-

quence of that fact that the two eyes are located in different positions in the head. You can easily observe the fact of binocular disparity for yourself by looking at objects not too distant from you, and noting the slightly different relative positions of objects in the views of the two eyes: these differences show up clearly if you close or cover each eye in turn.

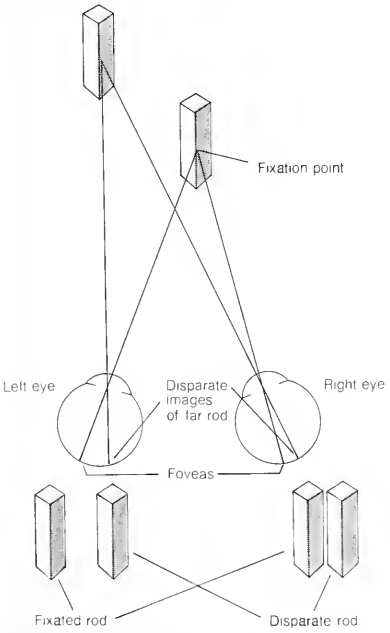
Another way of illustrating how it comes about that the two eyes receive different views of what we are looking at is shown in **170**.

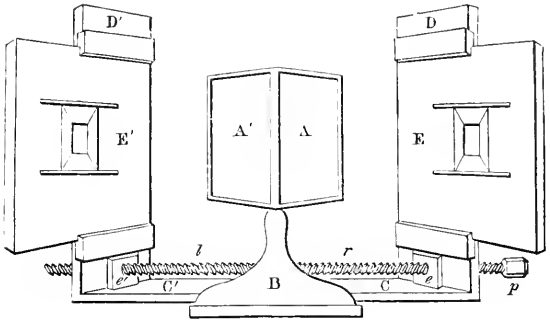
Stereoscopes

Most readers probably did not need this book to tell them that binocular vision has something to do with the perception of depth, because they learnt this in childhood by playing with one or other form of *stereoscope* (from the Greek for 'solid viewing'). Stereoscopes are optical instruments which enable a different picture to be presented to each eye, and so serve the same basic purpose as anaglyphs, but use such components as mirrors, prisms and lenses.

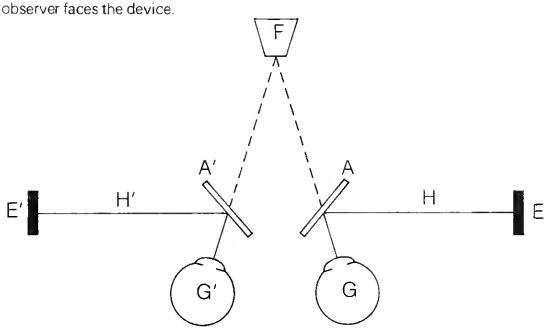
It was Wheatstone who invented the first and simplest type of stereoscope in about 1833. His device was based on two mirrors and is illustrated in **171** by his own drawing and in **172** by a schematic view. It was while using this instrument that Wheatstone discovered that binocular disparities of the type described in the previous section, which had previously been just speculated about by authors, are indeed used by the brain for depth perception. His stereoscope works as described in the captions to **171**, **172** and **173** [plate 13].

170 Binocular disparity The eyes are fixated on the nearer rod, which therefore casts an image in the middle of each retina (the fovea). The image of the more distant rod falls on different or *disparate* locations in the two eyes – further to one side of the fovea in the left eye. The extent of the resulting binocular disparity is illustrated by the pairs of rods under each eye.





171 Wheatstone's stereoscope The panels EE' hold disparate drawings which are reflected by the mirrors AA', one into each eye, as the observer faces the device.



172 A plan diagram of Wheatstone's stereoscope This makes it clear that the fused object F (here the sliced-off pyramid) is seen lying in depth directly in front of the observer's eyes GG' and behind the mirrors AA'. The lines HH' represent light rays from the disparate drawings EE'.

There are yet other ways of solving the problem of presenting different pictures to the two eyes. One type of stereoscope, popular in Victorian times, is shown in **174**. The two photographs placed in the device are of the same scene, but taken from slightly different camera positions chosen to imitate the different viewing positions of the two eyes. Thus the pictures might have been taken with a single camera which was moved sideways between shots by about 6 cm. Alternatively, a special stereoscopic camera equipped with two lenses about 6 cm apart, and capable of taking two shots simultaneously, might have been used. Either way, binocular fusion of the pair of photographs is made possible in the stereoscope by a prism-cum-lens serving each eye [**175**]. This ingenious arrangement, invented by Brewster in 1847, uses the prismatic function of the prism-cum-lens to 'bend' the paths of light rays coming from the two photographs so that they seem to both eyes to originate from a common source.

The lens aspect of the prism-cum-lens ensures that the viewer can focus his eyes by using a setting of his own eye-ball optics (technically called his state of *accommodation*) which is comfortable for the angle of *vergence* between his eyes that is required for fusion. This latter angle is the one between the two dotted lines in **175**, and it varies according to whether we are fixating objects near to us (large vergence angle, large accommodation effort required for the normal person) or far from us (small vergence angle, small accommodation effort). In any event, the net result is that the observer sees a single fused scene, with the various objects in the scene appearing at vividly different depths similar to those which would have been seen in the original scene from which the photographs came, had this scene been viewed normally.

The Brewster-type arrangement is the one used in modern stereoscopes available in any good toy-shop, and it is well worth having a look at one. If you have not seen the stereopsis effect produced by a stereoscope before, it can be quite startling, with the general vividness of the stereopsis exceeding that obtained with the anaglyphs in this chapter.

The Poor Man's Stereoscope

The simplest and cheapest way of getting the two halves of a stereo pair to the two eyes separately is to cross the eyes. This technique is illustrated in **176**. Not many people, however, can voluntarily cross their eyes in the controlled way required for this technique to be successful – at least, not without considerable practice. But the practice is well rewarded because it enables you to expand enormously the range of doodling which you can engage in during boring moments in talks, seminars, lectures etc.! Once you can cross your eyes suitably, you can explore an endless variety of home-made stereo effects by drawing your own stereo pairs. Precision in drawing is not required. The squiggles shown in **177** [plate 14] will happily fuse binocularly, despite the many mismatches of their various parts, so that the upper fused 'blob' appears nearer to you than the lower one.

174 A Victorian stereoscope based on Brewster's design

One way to help you learn how to cross your eyes in the way required is illustrated in **178**. Hold a pencil or some other marker about half way between you and the two halves of the stereo pair shown in **169** (upper). Fixate on the top of the pencil and you are now crossing your eyes to about the right extent. Keep the tip of the pencil below the stereo halves, so that it does not obscure them, and try gradually to pay attention not to the pencil but to the stereogram. With effort, and a bit of luck, the stereo halves will eventually 'snap' into fusion, and once this has happened you can take the pencil away and stare at the glorious stereo effect quite unaided. At first, you will find that your ability to hold fusion and focusing will wax and wane, and your eyes will get tired: so do not overdo it. But eventually, if you are persistent, you should be able to obtain stereo fusion without need even of the pencil marker. (Note that the pyramid in **169** will recede, not protrude, with crossed-eyes fusion; see below.)

Some readers might prefer to relax the angle of vergence of their eyes, rather than over-converge as happens when the eyes are crossed, to obtain the stereo effect [**179**]. This is a more unusual ability, however, and may indicate that your eyes have a natural tendency to swing outwards anyway. Note that if the stereo halves are separated by more than about 6 cm (the approximate separation of the eyes), then it will be necessary to have the eyes pointing outwards from parallel if fusion is to be obtained. Some ophthalmologists believe that it is impossible for the person with wholly normal eyes to enlarge the angle of their vergence voluntarily beyond parallel, i.e. that the most that can be achieved is a relaxation of *con*vergence to the parallel position. In view of this possibility, it seems best to practice over-convergence rather than relaxation of convergence, so that one is not limited to fusing stereo halves separated by no more than about 6 cm.

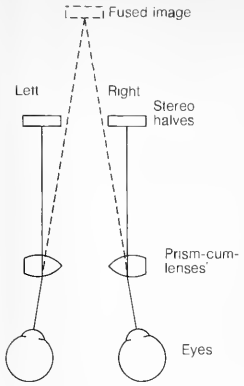
It might help you as you practise to say that one of the major problems to be overcome whether you try to over-converge or under-converge is that you have to 'decouple' your usual link between vergence angle and accommodation. That is, you have to get used to focusing your eyes for a distance which is different from the one normally matched with the required vergence angle. So when practising, do expect to see a rather fuzzy image during the early stages. Later on you will find it will become clear and sharp.

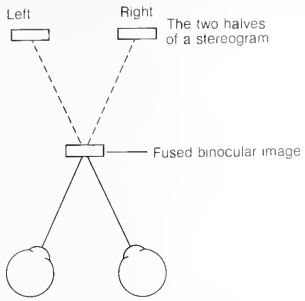
Finally, if you decide to practice under-convergence, you should expect to see the depth effect described in the text for each associated anaglyph. This is because under-convergence will result in the left stereo half going to the left eye and the right stereo half going to the right eye, as happens in the case of the anaglyph (when the spectacles are worn in their usual position – **R**ed filter in front of **R**ight eye). On the other hand, if you practise over-convergence, expect to see a reversed depth effect, because the right eye receives the left stereo half, and vice versa (see **183** and also remarks on p. 146).

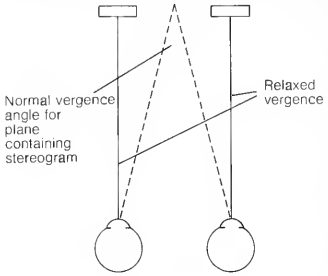
Random-dot Stereograms

The last two decades have witnessed great strides forward in our understanding of stereopsis on all three fronts – psychological, physiological and computational. Central to these developments has been the exploitation by Bela Julesz, a radar engineer turned visual psychologist, of the computer-generated *random-dot stereogram*.

Like any other stereogram, one of Julesz's stereograms consists of two halves which need to be presented separately to the two eyes, using a stereoscopic viewing technique of the







Note that the 'left half' enters the right eye with the crossed-eye method, vice versa for the 'right half'. This means that the depth effect is reversed - see 183

175 [top left] Schematic diagram illustrating the principle of the Brewster stereoscope

176 [top centre] The crossed-eye method of fusing a stereogram. Note that the 'left half' enters the right eye with the crossed-eye method, vice versa for the 'right half'. This means that the depth effect is reversed – see 183.

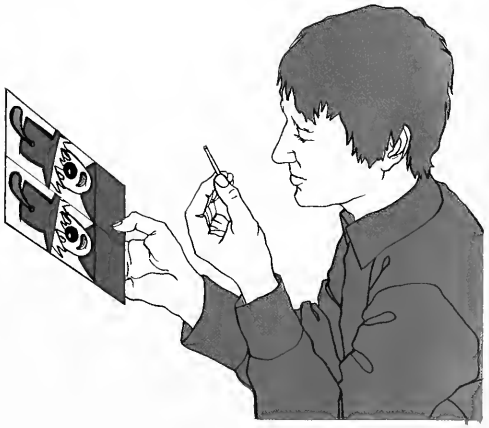
179 [top right] Relaxation of convergence to obtain stereo fusion

kind described above. An example is shown in **180** [plate 15], which should be inspected with the red/green spectacles, making sure that the **R**ed filter is in front of your **R**ight eye (the two halves of this stereogram are printed in black and white below the anaglyph version). If you have normal binocular vision, you will experience depth as illustrated in the small diagram below the anaglyph. In the centre of the anaglyph you will see a square (A in small diagram) floating a few centimetres above its surround (S). The surround too will appear to be lifted off the page, again by a few centimetres. This latter effect is due simply to the two stereo halves being printed on the page in a slightly offset fashion, so giving a disparity to the whole figure.

The experience of depth is for many observers rather more difficult to achieve with a stereogram of this type than it is with stereograms of natural scenes (e.g. **168**). So if you have difficulty in getting the effect, do not give up trying too soon.

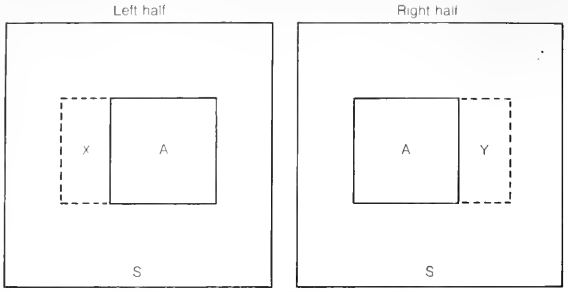
How does this illusion of depth come about? Despite the fact that the two halves of the stereogram look identical, they in fact contain a depth cue of binocular disparity which mimics that which would be produced by a genuinely protruding square surface. It is as if a speckled square was held in front of a speckled background in such a way that the boundaries of the square merged so well into the texture of the background that this boundary could not be seen with either eye alone. In fact, each half stereogram has a square within it which is shifted in relation to the corresponding area in the other half, much like the squares within the anaglyph of the sliced-off pyramid [**173**], except that the squares of the random-dot stereogram are hidden to monocular view. This will become clearer if I describe how the stereogram was made.

The first step in the manufacture of a random-dot stereogram is to create a piece of 'random visual texture'. In the



178 A helpful trick in learning to cross your eyes to obtain binocular fusion of a stereo pair

stereo pair of **180** a computer was used to 'draw' a random chequer-board of small black and grey squares (or 'dots' – hence the name 'random-dot stereogram'). But other components could have been used to equally good effect, such as round dots or small lines, as illustrated in the anaglyphs of **181** [plate 14]. The important thing in a random-dot stereogram is not the shape of the component elements but the fact that they are distributed in a random way, which camouflages, in each eye's individual view, the area which finally appears in depth. That is, the randomness of the texture of which the picture is composed ensures that the area where there is disparity cannot be picked out when each half-stereogram is inspected on its own. (Note that when an anaglyph is inspected *without* the red/green spectacles, either with one eye or both, the disparate area often *can* be detected. But this is a by-product of the printing of one picture on top of the other, and as soon as the anaglyph is viewed through the spectacles,

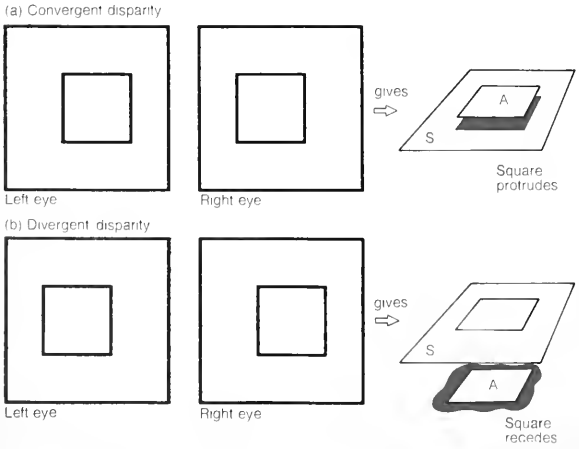


182 Making a random-dot stereogram The disparate area is labelled A, its surround S. Shifting area A in each half results in a certain portion of the texture being 'covered up' and thus lost to view. The shifts also result in 'holes' (X and Y) being created in each half-stereogram, but these are filled in with new random texture so that they do not exist in the finished products. The shifted area thus remains completely hidden in each member of the stereo pair – and the key property of a random-dot stereogram is thus ensured.

this by-product disappears, and the only remaining clue to the shape of the area which is intended to be seen in depth is the disparity between the two half-stereograms. Often, though, there is no trace of the shape to be seen in depth even when the anaglyph is viewed *without* the red/green spectacles.)

Once a piece of random visual texture of some kind has been created, the next step is to make two copies of this piece of texture which look identical but are in fact subtly different. These two copies become the two halves of the stereo pair. In one or both, a certain patch of random texture is shifted slightly horizontally, usually in a different direction in each half. This process is illustrated in **182** for a random-dot stereogram which has a square-shaped disparate area, as in **180** (but remember that the disparate area can be almost any shape required: see later examples). The disparate area is labelled A, its surround S. Shifting area A results in a certain portion of the texture being 'covered up' and thus lost to view. The shifts also result in 'holes' (X and Y) being created in each half

183 Types of disparity



stereogram, but these are filled in with new random texture so that they do not exist in the finished product. So it is impossible to see, by inspecting each member of the stereo pair on its own, which area has been shifted, and the key property of a random-dot stereogram is thus ensured.

Some people do not find it easy to visualise the steps in the creation of a random-dot stereogram, and as an attempt to help them, the following alternative description is offered: First, obtain two identical copies of a piece of random visual texture: these will be used as the surround. Second, obtain two identical copies of another similar but smaller piece of random texture: these will be used as the disparate areas. Third, lay the small pieces of random texture on top of the first, larger pieces – one small piece on top of each large piece – to obtain the two members of the stereo pair. Make sure that the small pieces are precisely aligned on the larger pieces so that they cannot be distinguished as a different entity when each stereo half is viewed alone. Also, place the small pieces in shifted positions in the two halves, to give the disparity cue. You now have a random-dot stereogram.

Now try the effect of reversing the spectacles so that the red filter is in front of your left eye, the green one in front of your right. If you hold the spectacles in this way and look at **180** afresh, you should now see in the centre of the anaglyph a 'window' through which you can see a square about a centimetre *behind* the frame of the window. The surround too seems to be further away from you than the page on which it is printed. This reversal in depth occurs because when you reverse the spectacles you exchange the views which the two eyes receive, so that the disparity now mimics that which would be produced in the two eyes by a genuine window with a square seen through it.

The same effect could have been achieved if, in the manufacture of the stereogram, the disparate squares were shifted not towards the mid-line, as shown in **183**a, but instead towards the outer edges, as in **183**b. As can be seen by comparing these two figures, this is wholly equivalent to exchanging the two eyes' views.

When we see something lying *further away* than the point we happen to be fixating (e.g. the receding square seen as a surface behind a window in **180**, if we fixate the surround with the glasses reversed), then the disparity cue is said to be *divergent* because we would have to diverge our eyes to transfer our fixation to this further-away object. Equally, a *convergent* disparity cue is one which generates the perception of something *nearer* to us than the point we happen to be fixating (e.g. the protruding square seen in **180** if we fixate the surround with the spectacles used normally).

Whether a convergent or divergent disparity cue is present with respect to the surround of a random-dot stereogram depends simply on the shift imposed on the disparate region. If **180** is viewed with the **R**ed filter over the **R**ight eye so that the disparate square protrudes, the shifts in the position of this square in each stereo half are as illustrated in **183**a (convergent disparity). Reversing the spectacles reverses the shift, giving a receding effect (**183**b: divergent disparity). The reason shifts in opposing directions give opposite effects is simply geometrical. Receding surfaces just do cause one sort of shift, protruding the other, in retinal images of genuine 3D scenes. The brain takes advantage of this fact and generates depth perceptions accordingly.

Of course, in a stereogram with a protruding square [**183**a], the square has a convergent disparity with respect to the

surround as just explained, but if after fusion one chooses to fixate the square, then it becomes the case that the surround has a divergent disparity with respect to the square. In real life, our vergence angle changes all the time as we look around us, so the disparity signals landing on our retinas are also constantly changing, even for a stationary scene.

An important point to grasp about random-dot stereograms, and the one which makes them so attractive for studying stereopsis, is that they present the disparity cue to depth in a very 'pure' form, unaccompanied by any other cue to depth. None of the monocular cues exemplified in **167** can play any role, because the two stereo halves completely lack any perspective, masking, shading etc.

Stereopsis is not Dependent on Object Recognition

The first important conclusion which Julesz drew from his research into random-dot stereograms was that stereopsis can be computed by the visual system without need of a prior stage of object recognition. We cannot see any 'object' within either stereo half of **180** viewed alone, but this does not prevent us fusing the two halves to achieve vivid stereopsis. Indeed, it is only *after* stereopsis has been computed that we can say 'Ah! There is a square' and thereby succeed in the business of object recognition. These considerations lead to the overall scheme shown in **184**a for the sequence of processing operations that take place when a random-dot stereogram is fused.

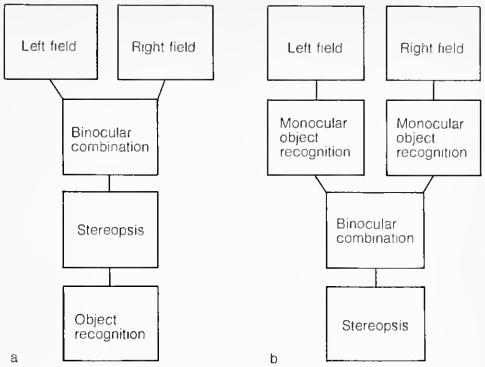
This conclusion was surprising because before Julesz's work the generally acknowledged scheme of processing was as shown in **184**b. Thus it was thought that a full scene description was computed for each eye's view separately, including the stage of object recognition, and that it was only then that binocular combination occurred, leading to stereopsis. Random-dot stereograms proved this theory wrong, at least as a full account of stereopsis, because they showed that binocular combination *need* not happen after object recognition. In the terms of this book, we can express this by saying that building up high-level structural descriptions of the left and right views and then matching these with stored structural descriptions for objects is not a necessary prerequisite for obtaining stereopsis.

Because one sees nothing but a random texture before stereopsis occurs, random-dot stereograms have found a practical application in eye clinics. They are useful screening stimuli because it is impossible for a patient who is being tested for the presence/absence of binocular vision to 'cheat' with them, either intentionally or otherwise. If he can see the 'hidden object' and say what it is, he *must* have stereopsis – quite unambiguously.

But if binocular combination can be achieved before object recognition, sometimes at least, at what level of processing is it done in these circumstances? The obvious answer is to say that it occurs at the level of feature description. That is, individual small points, dots, lines, edges etc. are described as being present in each view, and then the features in the two eyes' views are fused together in some way. I will return to this possibility later on, but for the present I will simply allow this theory to take us to the next important conclusion deriving from Julesz's work.

The Problem of Global Stereopsis

The second important conclusion soon drawn by Julesz from random-dot stereograms concerns the immense problem of *ambiguity* which confronts the brain when it tackles the task of



184 Flow chart representing two different theories of stereopsis
Random-dot stereograms show that stereopsis can occur according to theory (a).

combining left and right images. If binocular fusion takes place at the level of feature description, how does the brain 'know' which are the correct left/right pairs of features to fuse together? The most obvious answer is to propose that it fuses features of similar shape, size, contrast etc. But consider the stereogram shown in **185** [plate 16]. As you can see from its separate stereo halves, its random texture is made up of tiny dots, each one very similar to all the others in shape, size, contrast etc. Because of this similarity, any given dot in one image could in theory be fused with any one of a large number of dots in the other image. So how can the brain possibly decide which dot in the left stereo half is to be fused with any given dot in the right stereo half? Indeed, if it turned out that stereo fusion of this figure was impossible it would make a great deal more sense than the result which actually occurs – which is that the brain fuses the figure with ease and produces the correct result of a central rectangle protruding in depth! At the very best, we might reasonably have expected no more than a random 'fog' of dots at many different distances from us, each one resulting from the more or less arbitrary matching of dots in the left and right images. And yet the brain solves the 'ambiguity problem' and comes up with the 'right' answer. It gives us exactly the perception of depth which was in fact intended when the stereogram was made. This is a truly amazing achievement by the brain, and one which went unrecognised until random-dot stereograms were invented.

The problem of ambiguity is often called the problem of *global stereopsis*, because the brain must find the correct overall or 'global' matching of features from among the wide variety of possible sets of matches between individual local elements. Each possible individual match between a feature in the left image and a feature in the right image is called a *local match*, and the problem of arriving at the correct overall match for the two fields, such that only those local matches which 'should' be made are in fact made, is called the *global matching problem*.

It is important to realise that the global stereopsis problem is not confined to unusual and artificial stimuli such as computer-generated random-dot stereograms. Consider, for example, natural scenes such as a leafy tree, or a vase of flowers, or even just a carpet, receding from you into the distance. How

does the brain decide which leaf, or petal, or tuft of carpet, in one image to match with the appropriate one amongst all the myriad possibilities in the other image? Answering this question is perhaps *the* central theoretical challenge which current research on stereopsis is confronting.

Complex Random-dot Stereograms

The global stereopsis problem and its inherent difficulties are perhaps best brought out not by simple stereograms containing just a square or a rectangle in depth, problematic as these are, but by stereograms containing a much more elaborate 'hidden object'. The random-dot stereogram shown in **186** [plate 16] is a case in point. Have a look at it with the red/green spectacles as usual and see if you can make out the intricate 3D shape which it contains. But be patient! Many people take quite a long time to see stereopsis in a complex stereogram of this kind, and may even need several attempts each lasting a few minutes. But as before, effort and patience are well rewarded: the final depth effect is truly remarkable and very beautiful. Once the global matching problem has been solved by your brain, you will see a saddle-like shape with a ring around it.

Another complex stereogram is shown in **187** [plate 17]. In this case the hidden object is a spiral staircase. A similar stereogram, but with the outline of the staircase drawn in to help you see it in depth, is shown in **188** [plate 18]. This latter stereogram is not a true random-dot stereogram because it breaks the rule that the shape designed to be seen in depth must not be visible in either the left or the right image by itself, but the monocularly-discriminable contour helps unpractised observers to get going! An interesting thing about both **187** and **188** is that you only realise gradually, as you look at them, that the spiral is made up of stairs. At first, the surface of the spiral looks smooth – like a helter-skelter. But bit by bit the stairs reveal themselves, as the brain continues to refine the matching which it achieves between the dots in the left half-stereogram and those in the right. Some people never manage to see the steps, because their binocular vision is not sufficiently acute to pick up the very tiny differences in disparity between the dots in one step and the dots in the next.

Try reversing the red/green spectacles while looking at **187**, so that the red filter covers the left eye. You should now see a receding spiral staircase, corkscrewing down through the page. The effect is a marvellous one and, incidentally, one which some people find easier to see than the protruding staircase, so if you had to give up on **187** with the spectacles in their normal position, do have another try with the spectacles reversed.

In complicated stereograms like the one of the saddle [**186**] or the one of the spiral staircase [**187**], different parts of the picture appear to be at a whole range of different distances from you, while in the earlier examples (e.g. **185**) only three distances were used – the page, the surround, and the square or rectangle. The wider selection of distances is easily achieved by shifting different parts of the random texture by different amounts – the bigger the shift, the greater the illusion of distance which results. This relationship between the size of the shift and the amount of the illusory distance is clearly illustrated by the anaglyph of **189** [plate 18]. The upper small square area has been shifted in each eye's view by about twice the amount of the shift in the lower square. The result is that the upper square seems to protrude by about twice as much.

Learning to See Random-dot Stereograms

The tactic, shown in **188**, of drawing a monocularly-prominent line around the area which is designed to be seen in depth is especially helpful if the disparity between the left and right images is rather large, as in the anaglyph of **190** [plate 18]. Most people find it very difficult to see a shape in depth in this stereogram, but if an outline of more closely packed dots is added, as in **191** [plate 18], then the illusion is much easier to see. This outline probably helps by giving the brain appropriate cues for the control of eye movements. One stratagem the brain uses to help in overcoming the problem of global stereopsis is to avoid fusing left and right elements with disparities larger than a certain limiting size. But this restriction means that features which fall in grossly different depth planes, and so create large disparities, cannot be fused unless suitable eye movements are made to fixate the two features successively. The idea here is that first one left/right feature pair is fused and thereafter held 'locked' together while a new eye movement enables the next feature pair, with a very different disparity due to its very different depth, to be fused. Thus there is a limit on the size of disparity only in the case of initial fusions: once made, a fusion can survive despite being supported by a much larger disparity (although there are limits on this also). Ann Saye and myself have found evidence supporting the idea that prominent monocular contours help produce stereopsis by facilitating the required eye movements. This evidence takes the form of showing that the contours help when the disparities are large – when eye movements are necessary for fusion – and not when disparities are small [**192**, plate 19] – when eye movements are not required for fusion.

Indeed, it is my belief that the fact that we need to learn how to see random-dot stereograms *in general*, regardless of the size of disparity, is due to those stimuli providing poor cues for eye movements. When the naïve reader was struggling initially to see even the simple square-in-depth stereogram of **180**, whose disparity was modest, this was probably because he kept his eyes firmly fixated on the page, as well he might, because before binocular fusion there is nothing to 'see' in any other depth plane. As you get better at fusing random-dot stereograms, what is probably happening is that you are unconsciously learning to relax your fixation somewhat, so that you converge and diverge without the need for first seeing something in each monocular image to converge or diverge upon. The monocular contours of **191**, however, do provide 'something' to converge/diverge on, and this is why they help in achieving stereopsis.

The monocular contours might of course play a helpful role in other ways as well. For instance, they might provide high-level shape information which could in principle guide the solution of the global stereopsis problem. But I personally doubt that much benefit is in fact bestowed along these lines, for two reasons. First, Ann Saye and I have shown that the monocular contours can still provide a helpful asset even if their shape bears no particular relationship to that of the disparate zone (e.g. **193** [plate 19]). And second, Jeremy Clatworthy and I, in an experiment which gave different groups of subjects different cues about the staircase stereogram of **187**, could find no evidence of benefit from high-level cues such as telling subjects what they 'ought' to see, or even showing them a 3D model of the staircase before presenting the stereogram itself.

A Computational Solution to the Problem of Global Stereopsis

A schematic illustration of the problem of global stereopsis is given in **194** [plate 19]. The left and right eyes are shown viewing the left and right halves of a random-dot stereogram, presented to them in a Brewster-type stereoscope. The whole of **194** is a partial, selective view, so one has to imagine its stereo halves as just slices taken from a full random-dot stereogram. For simplicity, just three representative dots are picked out for special attention in each stereo half, and light rays (solid lines) are drawn from these to the left and right retinas. Somehow the brain fuses these dots appropriately: the *correct fusions* are shown as filled-in black dots in the visual world built up by binocular fusion. But notice that the correct fusions, whereby the three dots all lie at the same distance from the observer, are not the only possible ones. For example, there is nothing in principle to stop, say, the central dot in the left image fusing, not with the central dot in the right image as it 'should' do, but instead with either the left or right dot in the right image. These particular *false fusions*, or ghost fusions as they are sometimes called, plus all other possible ones, are depicted in **194** with outline circles. Thus the problem of global stereopsis can be described as the job of ensuring that only the correct (black disc) fusions are selected from amongst all nine possible fusions in **194**, six of which are false (outline discs). Of course, **194** underestimates the global matching problem enormously because, for simplicity, it shows only three dots in each stereo half, and the possible confusions they create. Consider the number of confusions which would be possible in a richly-textured stereogram such as **187**.

One computational approach to solving the problem of global stereopsis tackles it in two stages. First, all possible local fusions are established, including both correct and false ones. These local fusions are represented in a network of computer elements, rather like idealised neurons, with one active element for each local fusion discovered. Second, false fusions are 'killed off' (i.e. the elements representing them are made inactive) by the combined action of inhibitory and excitatory connections between elements throughout the network. This process leaves the correct fusions 'alive' (i.e. active), and so the final state of the network gives a solution to the problem of global stereopsis. It is a solution because it has made *explicit* which dots lie in which depth planes. This general approach was first tried out by Dev, but I will present here a version due in its essentials to Marr and Poggio which is in several ways more developed.

The lower part of **194** shows the first stage, identifying all possible local fusions, for our simplified case of a three-dot random-dot stereogram. A network of elements is shown, each one an idealised neuron, and each one receiving an input from each eye. One neuron is provided for each and every possible local match. Note that each input cable from each retina serves three cells, those which would deal with fusions along the same *line of sight* from the eye in question. Follow the various cables through to check this fact yourself, and note that the network of neurons is really just a simple replica of the visual world depicted out in front of the eyes, but one cast in terms of neurons rather than perceived fusions. Thus the general idea here is that each active neuron stands in a one-to-one relationship to a particular local fusion, so that if a feature is described as being at a certain location and in a certain depth plane, then the representation of this description is activity in a particular cell.

The depth-processing network shown in **194** readily lends itself to a neurophysiological interpretation, so much so that it is shown as a network of neurons drawn within an outline of a brain. Note that signals from the right eye are shown crossing over to the left hemisphere at the optic chiasma. If the row of three dots had fallen on the other side of the fovea from the one shown, then it would have been necessary to display the depth network in the right hemisphere rather than in the left one – separate networks exist in each hemisphere for dealing with different regions of the field of view (see **51** [plate 4]).

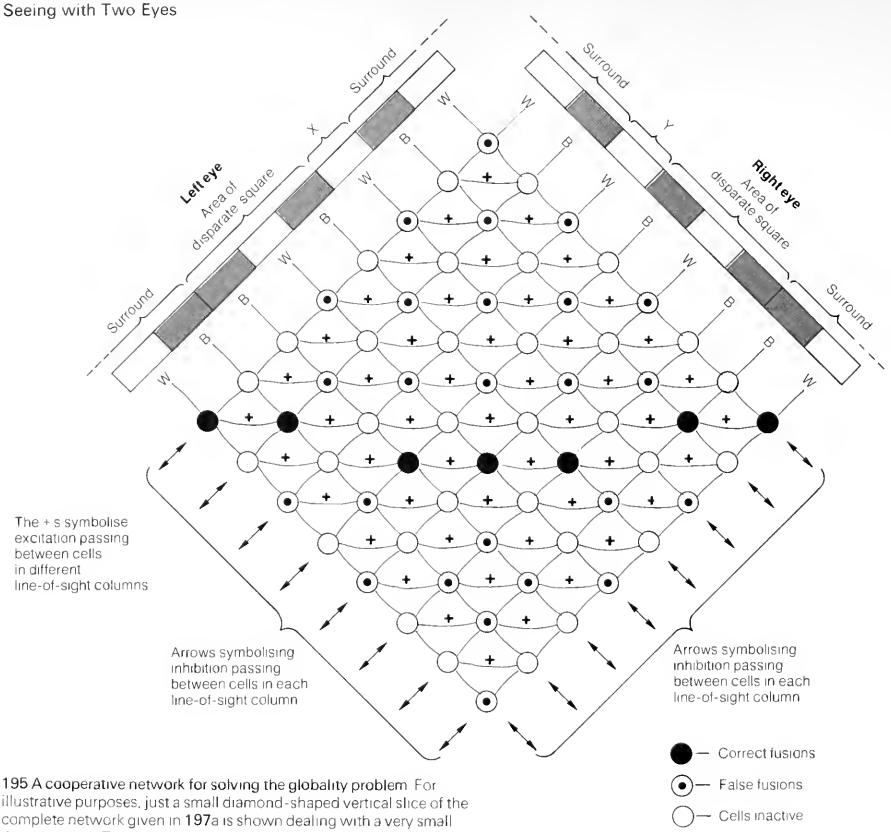
To understand what happens next it is necessary to show a rather larger network of depth-processing cells, as in **195**. This network is wholly similar in structure to the smaller one shown in **194**, but now there are $8 \times 8 = 64$ cells, rather than just 9. Cables are shown coming from the left and right eyes to feed the network. These cables carry either information about a black point in a certain retinal location or information about a white point. This is because the eyes are viewing a black/white random-dot stereogram. Thus each eye's input can be thought of as a slice from a random chequer board, as shown in **195**.

The first thing to appreciate is that those cells which become active initially are those which receive a suitable input from both eyes, i.e. white left/white right inputs, or black left/black right. Thus each cell cares not about whether its inputs are black or white, only about whether its two inputs match each other. This requirement ensures, of course, that not all cells become active in the network. Those which do become active are shown either as black discs, indicating that they are correct local fusions, or as outline discs with a dot in them, to show that they are false local fusions. Try following down an input cable from one eye through the network, and note that either a black disc or a dotted disc occurs only when this cable meets a cell whose input cable from the other eye is carrying similar information.

Look next at the chequer-board patterns coming from the two eyes. Note that each one has an 'area of disparate square': this is a slice of the square in the random-dot stereogram which appears to stand out in depth because of the disparity of its position in the two eyes. Check for yourself that it is disparate in location in the chequer-board slices shown in **195** for each eye. Thus it appears *two* elements from the left in the left eye but *three* elements from the left in the right eye. Small patches of surround area are shown with zero disparity (i.e. their positions match exactly in the two eyes). And finally, note the X and Y areas – the bits of the pattern in each eye which have no matches in the other eye because of the shift imposed on the disparate squares (refer back to **182** for an explanation of these areas).

Having explained how the network is set up, and having drawn attention to the disparate area, I must next explain how it is that the black disc fusions come to be selected from all those initially registered. The key feature of the Marr/Poggio scheme for achieving this is to have *inhibitory connections* between cells lying within the same line-of-sight columns and *excitatory connections* between cells representing the same depth plane. The arrows just outside the network in **195** represent inhibitory influences passing up and down the line-of-sight columns, and the $+$s on the lateral connections between cells indicate the passage of excitation between adjacent cells coding the same depth.

Why try to select the correct fusions in this way? Marr and



195 A cooperative network for solving the globality problem For illustrative purposes, just a small diamond-shaped vertical slice of the complete network given in 197a is shown dealing with a very small disparate area This latter feature makes the correct fusions appear less extensive than some areas of ghost fusions, a result unrepresentative of larger areas (see 197b)

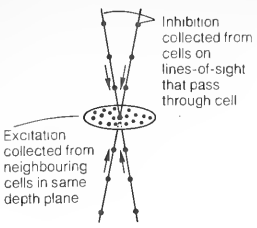
Poggio reasoned that along any given line-of-sight from either eye, one and only one point can be seen, and it is seen in just one depth plane. Thus it seems sensible to set all cells coding depths along a line-of-sight to 'fight' to see which one is the 'strongest' (most active) and therefore justifies selection against the competition provided by all the others. But how can some turn out to be stronger than others? The answer is that some active cells get a boost in activity from the excitation passed to them from 'friends' in the same depth plane from neighbouring line-of-sight columns.

But why, you might ask, should help from lateral depth neighbours be the way to promote the interests of the correct fusions? The answer is that in most scenes the correct fusions will always have helpful neighbours because surfaces tend to lie in the same or similar depth planes. Of course, sudden switches or discontinuities in depth are quite common also, but not nearly so common as smooth gradations. This leads on to the point that the facilitation need not be limited to exactly similar depth neighbours. It could be that each cell gets excitation passed to it from, say, exact depth neighbours plus those neighbours one depth layer away. Another related point to realise is that the excitation arrives from depth neighbours *encircling* each cell. In **195**, the facilitation is shown coming from only two neighbours because the network is shown as a two-dimensional slice, not as a complete 3D entity. The effective network connections for an individual cell are shown in 3D in **196** to make this point clearer.

Having explained something of the design of the network, it is now time to see it at work on a genuine random-dot stereogram. In **197**a, a 3D model of a 7-plane depth network is illustrated (in **195**, remember, just part of a vertical slice of the network was shown). Each layer in this structure represents one layer of depth, the central layer being the one for zero disparity (i.e. the depth at which we are fixating), the upper three layers those for convergent (near) disparities, and the lower three layers those for divergent (far) disparities. For simplicity, excitatory and inhibitory connections are shown for a few cells only, but in fact all cells in all depth planes have similar sorts of linkages. Note also that **197**a shows simply the network structure prior to input of a stereo pair for computation.

In **197**b, all connections between cells are omitted for

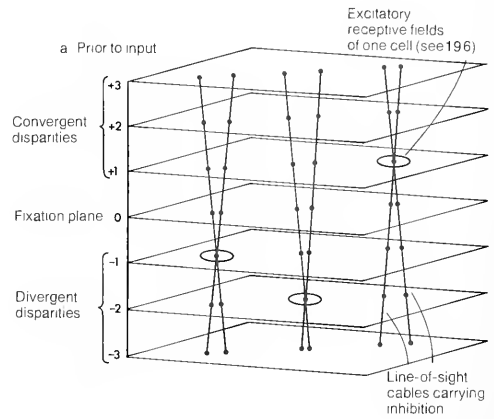
196 Network connections for an individual cell in the Marr/Poggio computation

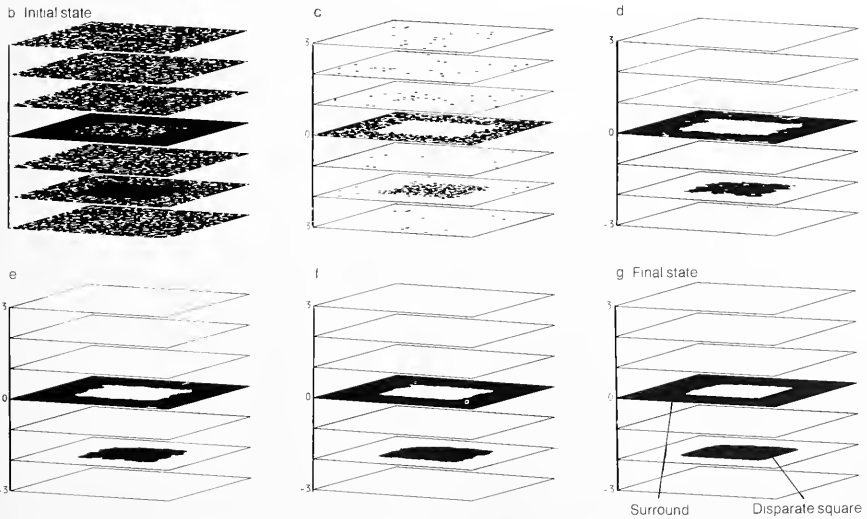


clarity, and now the tiny black dots in each depth plane represent active cells, each dot thus depicting a cell which has become active because its left and right eye inputs match, either as black/black or as white/white. Be careful to note that each black dot shows an active cell *regardless* of the nature of its match: do not get confused by thinking that the black dots represent only black/black matches (although the original Marr/Poggio computation, unlike the present version, did in fact start from black/black matches only). The inhibitory cables are now left out for the sake of clarity.

The initial state of the network before any inhibition or excitation has passed between cells is thus the one shown by **197**b. It is the first step in the global stereopsis computation – the identification of all local fusions, be they correct or false. And note just how many false ones there are! Black dots appear almost everywhere, showing just how extensive the globality problem is. But note also that the regions of completely dense black representing the surround of the random-dot stereogram and its disparate square area are also discernible, in the zero disparity layer and the second divergent-disparity layers respectively. The task of the computation is to leave these regions intact, at least in the final state of the network, while killing off the myriad false fusions.

The results of the first round of excitation and inhibition are shown in **197**c. To get to this state, each cell had to do the following sum: it had a certain number of units of activity





197 A 3D picture of the Marr/Poggio network dealing with a random-dot stereogram (based on a program designed by Peter Gale)

according to whether or not it was active in the first place; it then subtracted from this number the total of inhibitory units sent by its competitors down the line-of-sight columns which passed 'through' it; and it added on the units of excitation given to it by its depth-plane neighbours. It worked out this total and then 'asked' whether or not it was above a pre-set threshold value. If it was, it became 'active' for the next round; if it failed to exceed the threshold, then the cell was 'inactive' for the next round.

The results of the next round of excitation and inhibition are shown in **197**d. You may be wondering why more than one round is necessary. In a *biological* system, a set of network connections of the kind we are talking about can settle fairly quickly to its final stable state, with all the connections constantly and *simultaneously* passing inhibition or excitation according to the influences they are receiving at any one time. But in a serial computer simulation of such a network, it is necessary to approach the end state in a series of steps, called *iterations*, each one doing just a bit of the overall computation for each cell in turn, so that the result of each iteration represents a 'snapshot', if you like, of the network as it passes through a particular stage on the way to its final state. Each iteration does exactly the same arithmetic as any other one, but its input is different: it works out the sum of inhibition and excitation using the state of the cells produced by the preceding iteration.

The state of the network at various stages of its computation is shown in **197** by the products of various iterations. Note that a lot of 'killing off' takes place immediately, and then the successful solution begins to grow back from the rather tattered remnants of the first battle. But it grows well enough, and in the end [**197**g] the desired state, in which surround and disparate square are picked out in different depth planes, is achieved. The battle is won!

The network I have just described is said to exhibit the property of *cooperativity*. That is, it reaches a state of global organisation via local but highly interactive processes which together cooperate to produce the required solution. Of course, the local processes have no knowledge about what is going on elsewhere in the network, so theirs is a 'blind' form of cooperativity: they simply do their own thing according to the rules, and the global state of organisation 'pops out' at the end of it all as an inevitable but highly desirable consequence. Julesz was the first to propose that cooperativity in this sense was a property of the mechanism of stereopsis. If he is right, then the type of network just described gains added credibility as a model of the actual neural computation of stereopsis performed by our visual system.

Neurophysiological Mechanisms for Stereopsis

If the Marr/Poggio type of computation is the one used by the visual system for obtaining global stereopsis, then the first requirement is neurons which can detect the local fusions. Possible candidates for such cells have been reported by many neurophysiologists. The single cell recording technique (p. 42) has revealed cortical neurons in the cat which become active only if their optimal stimulus is positioned very carefully in the two eyes so that it possesses the degree of disparity required for the particular cell in question.

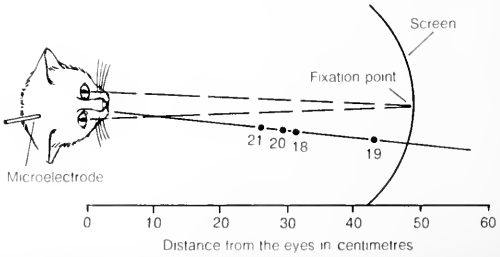
Some results of this kind, obtained by Blakemore from the cat, are particularly interesting in the present context because he found what he called *direction columns* [**198**]. That is, if his microelectrode stayed perpendicular to the surface of the cortex, so that all the cells he recorded from were in just one column (refer back to p. 47 for a reminder on cortical columns), then he sometimes found that cells varied in their required depth (disparity) along a line-of-sight from one eye. This raises the tempting possibility that Blakemore has discovered the neurophysiological machinery for providing the line-of-sight inhibition required by the Marr/Poggio computation, because it would clearly be a straightforward matter for inhibitory influences to pass up and down these columns along known fibre tracts within the columns [**54**]. Lateral connections between columns might then be the mediators of the lateral excitation provided by other cells in the same depth plane. If so, this is an encouraging and exciting link between a computational analysis on the one hand, and a set of neurophysiological and neuroanatomical findings on the other.

What Exactly is a 'Point'?

But there are difficulties with the above schemes, both computationally and neurophysiologically. These centre around the deceptively simple question about what it is exactly that is fused to create a local match. We began by suggesting that it was a feature of some sort, by implication one of the kind arrived at by the hypercolumn machinery of chapter 3. Such a feature might be a small line segment, an edge, etc., and the idea was that binocular combination would *follow* this stage of processing. But the known neurophysiological disparity units seem much like the various cortical cells described in chapter 3 in their general level of processing sophistication. And if we were right in chapter 3 to say that such cells are but *measurement precursors* to building up a proper feature description, and that they are not part of the representation of the feature description itself, then the known disparity cells seem poor candidates for mediating local feature matches directly.

Note that for convenience, while explaining the Marr/Poggio computational solution to the problem of global stereopsis, we slid gently and surreptiously into treating a point-for-fusion as a small black or white zone in either the left or right retinal image. Such zones might conceivably have as their neurophysiological representation active retinal ganglion cells that signal blackness or whiteness (i.e. neural elements coding one small zone of the image after the lightness computation). But this view of what constitutes a point-for-fusion is quite at odds with our present neurophysiological

198 Disparity-sensitive neurons. The numbered dots represent cells whose optimal monocular stimuli had to have a disparity in position in the two eyes such that they were most sensitive to binocular stimuli appearing in the depth locations shown.



knowledge about disparity cells. These are all orientation-selective, which suggests that when fibres stemming from retinal ganglion cells are used to build up the activity of disparity cells, the property of orientational selectivity is built in at the selfsame moment (see **70** for the possible wiring diagram for obtaining 'oriented' cells from fibres emanating from a receptor mosaic).

The question therefore about what stimulus attributes in the left and right images form the basis for local fusions is a tricky one, and a thorough discussion of it is beyond the scope of this book. Instead, in the remainder of this chapter I will present a wide range of different stereograms so that you can at least appreciate visually the kinds of fusions which are possible and thus some of the phenomena which any fully satisfactory theory of stereopsis must be able to cope with. In discussing these stereograms, the term 'point' will be used loosely to refer to whatever type of stimulus element it might be which is combined for the purposes of obtaining local fusions. Some of the stereograms [**204–210**] are not referred to in the text, but have self-contained captions: see plates 21–2.

Illusory Contours in Depth

In **199**a [plate 20] an ordinary random-dot stereogram is divided in two by a white strip. Look at the anaglyph using the red/green spectacles and you will see that the white strip cuts a central protruding square in half. But if you look more carefully, you will notice that the square has not really been cut in two at all, at least not as far as its apparent depth is concerned. Rather, the part of the white strip which cuts across the square seems to be in the same depth plane as the square itself. It is as though this central region of the strip gets 'sucked up' with the square, despite the fact that it contains no texture, and so offers no disparity cues to justify its allocation to the same depth as the square.

A possible explanation of this curious depth effect, whose boundaries are marked out by illusory contours defining the depth boundary, is in terms of lateral excitation of the kind already described in connection with the Marr/Poggio computation of global stereopsis. Perhaps cells in the depth-processing network feed out lateral facilitation to disparity units dealing with the region of the central strip and 'bring them alive'. This is possible because the region of the central strip is essentially ambiguous as far as disparity processing is concerned – it would give rise to white/white matches in all possible depth planes. Thus it makes sense that those cells receiving help from their depth neighbours win out, to give protruding depth to the strip when it passes through the central square, and no depth when it passes through the surround (the latter region 'pulling down' the strip, just as the square-in-depth 'pulls it up').

In **199**b, a low density of dots produces a similar effect, with the black dots in any one depth plane 'taking with them' the part of the white ground on which they lie.

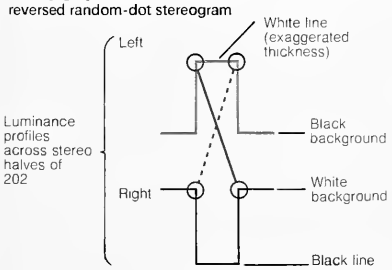
Stereopsis Survives Wide Contrast Differences

It is possible to fuse quite readily left and right stereo halves which differ greatly in contrast, as **200** [plate 20] shows. Thus the rule for matching left and right points is not one which insists that, say, a 'white' point of a given brightness in the left half can be matched only with a white point of similar brightness in the right half. Rather, it seems that as long as a white point in one field can find a whiter-than-mid-grey point in the other field, then it will fuse with it. What it refuses to do is fuse with a blacker-than-mid-grey, as is shown by **201** [plate 20], in which contrast has been reversed. Each element of this stereogram is of opposite black/white 'colour' in the two halves and stereopsis is impossible. So it seems that fusions can take place between stimulus elements coded by the whiteness array of the last chapter (p. 130), or between elements coded by the blackness array. What cannot happen, it seems, is the fusion of an element coded by the blackness array with an element coded by the whiteness array.

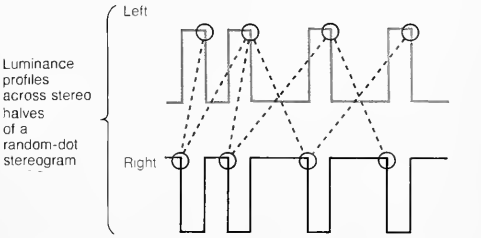
Of course, expressing this conclusion in connection with the whiteness and blackness arrays of the retina does not necessarily mean that the conclusion favours the view that what constitutes a point-for-fusion is necessarily a zone of whiteness or blackness. It could be that the white-for-white and black-for-black rule applies to line features, for example, and is mediated by orientationally selective units.

Interestingly, it has been known for a very long time that reversal of contrast in a simple line-stereogram can be tolerated quite well [**202**, plate 21]. But here, as Paul Whittle has argued, it could be that the brain is fusing white edges (or black edges) with one another [**203**a]. If so, perhaps the reason the system fails for a random-dot stereogram is that reversal of contrast in this case disrupts rather more seriously the locations of such edges in each field, producing too great a variety of different and inconsistent fusions of the edges of randomly-determined blocks of white or black cells [**203**b].

203a [below] Whittle's explanation for stereopsis from 202
203b [right] Luminance profiles across stereo halves of a contrast-reversed random-dot stereogram



Fusion might take place between white-to-black edges in each half, as shown by dotted line and circles. Alternatively, fusion could be between black-to-white edges, as shown by continuous line and circles.



Perhaps fusion is impossible here because there are too many conflicting possibilities - possible fusions are shown only for white-to-black edges.

Stereopsis from Stereograms with Rivalrous Texture

If two textures do not fuse happily together when presented stereoscopically, then the resulting perception is said to be *rivalrous*. That is, first one eye's view and then the other succeeds in becoming dominant, the two seeming to be in a state of rivalry for 'possession' of visual awareness. But it is possible to obtain stereopsis in the face of such rivalry, as **211** [plate 23] demonstrates. The details of the textures forming each square are rivalrous because they do not match up systematically in the view of the left and right eyes, and yet the squares as a whole have a disparity shift which successfully yields stereopsis. We see a staircase in depth of the three squares, which are binocularly combined overall despite being rivalrous in their details.

Interestingly, if the squares are marked out with textures of different spatial frequencies, then rivalry is much more pronounced and stereopsis is much more difficult if indeed not impossible [**212**, plate 23].

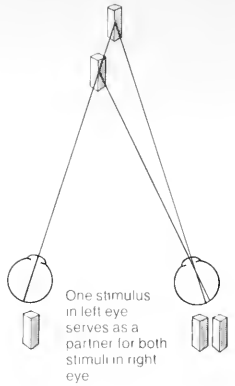
Perhaps the most extraordinary effect of this kind, recently discovered by John Mayhew and myself, is shown in **213** [plate 23]. Here the background is matched in the details of its texture in the left and right eyes (and so is non-rivalrous when the stereogram is binocularly combined), but the monocularly-discriminable shapes in the left field have no corresponding shapes in the right field whatsoever – and yet stereopsis is obtainable! Moreover, the degree of perceived depth is a function of the width of the shapes, the narrowest seeming to protrude least (or recede – the depth effect can spontaneously reverse), and the widest seeming to protrude farthest. This is weird indeed! How can depth appear when there is nothing in one eye for the shape-in-depth to fuse with?

This latter effect probably has something in common with Panum's limiting case [**214**, plate 23]. Here the supposedly minimal requirements for stereopsis are provided: a single line in one eye and a pair of lines in the other. Stereopsis is possible, the usual interpretation being that the single line serves as a left-field partner for both right-field lines [**215**]. But it is difficult to apply this interpretation to **213**, and it may be that quite different mechanisms operate *both* there *and* in Panum's limiting case itself.

Does the Computation of Lightness Precede Stereopsis?

The last chapter made out a case for the computation of lightness being completed in the retina. If this is true, then one would expect various lightness illusions to occur even if the components that induce these illusions are eventually seen to lie in different depth planes. Gilchrist has recently argued the converse case and proposed that the shade of grey we perceive depends primarily on the luminance relationships between surfaces seen to lie in the same depth plane, and not between surfaces that are merely adjacent in the retinal image. He thus regards depth perception as a precursor of lightness perception; that is, we have to work out what surfaces are in which depth planes before proceeding to the lightness computation. Without going into the details of this debate, I can just mention that the curious change in perceived lightness (almost a fluorescence) which comes about when the ambiguous folded paper object in **30** flips to the incorrect perceptual organisation suggests that the computation of lightness is rather more intricate than I suggested in chapter 6, and can be influenced, in some way at least, by the final perceptual organisation of the scene being observed.

215 Usual explanation for Panum's limiting case



In any event, John Mayhew and I have recently tested Gilchrist's hypothesis with the stereogram in **216** [plate 24]. Each stereo half viewed alone produces a lightness-contrast effect, as one would expect. But upon binocular combination the lightness illusion is still present, despite the fact that the small grey inserts are now seen in different depth planes from their immediate surrounds, and in the same depth plane as the field on the opposite side of the figure. As this opposite field is of opposite lightness to the one which surrounds them on the retina, the binocular lightness illusion should be the reverse of the monocular one if Gilchrist's hypothesis is correct. To our eyes it does not reverse, but stubbornly maintains the same basic nature binocularly as it has monocularly. You can decide for yourself what you see, but we doubt that Gilchrist's ideas can be the whole story, to say the least, in the face of our stereogram.

Viewing Distance Alters the Amount of Depth Perceived from a Stereogram

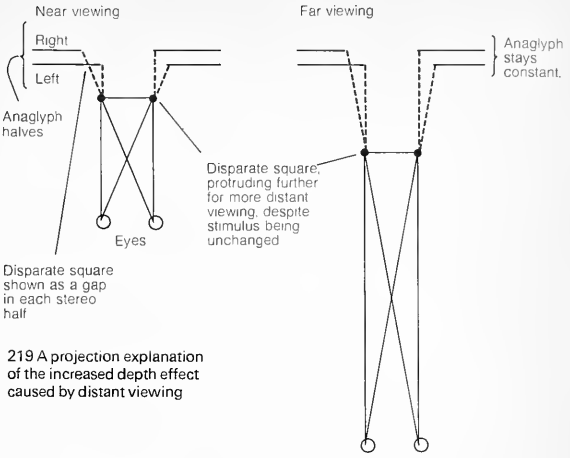
Prop the book open on some suitable surface and view the anaglyph of **218** [plate 24] from about 30 centimetres (about 1 foot). Its elements have been printed in a large size to make them clearly visible from this relatively large distance. Note the size of the gap which seems to exist between the central protruding square and its surround – about 5 cm or so. Now walk backwards (carefully!), all the while continuing to look at the anaglyph. You will find that the apparent distance between square and surround increases, the further away you go. For example, if you look at the anaglyph from about 3 metres, then the distance between square and surround will appear to be about 20 cm.

This is an interesting effect to observe, but not perhaps quite as mysterious as one might at first think, in so far as an increase in depth is exactly what would be expected simply from the optics of the situation, as **219** makes clear. Thus, since the stimulus itself does not change, of course, as the viewing position alters, the location in depth which the disparate square 'must' occupy is different for different viewing distances. But how the brain interprets any given disparity signal according to the viewing circumstances is not so well understood. There is probably some subtle link between vergence angle and the amount of depth which any given disparity signal is taken to convey, so that the disparity

signal is scaled, as it were, to mean different things for different vergence angles.

It is also interesting to try the effects of moving your head to and fro sideways while viewing **218**. You will find that the disparate square moves with your head movement. This seems very odd at first: why does it not stay still, as a normal 3D object would seem to do in this situation? Understanding why it moves is again helped by considering the changed optics of the situation, this time those created by a sideways change in head position. Thus **220** shows that the disparate object simply 'has' to move as it does because this is the only perception consistent with the retinal image.

Finally, it is worth experiencing the effect of running a finger across the surface of a page on which an anaglyph is printed, and also of bending up the corner of the page so that its surface is curved rather than flat. These manipulations can cause interesting distortions of the various depth effects, again for optical reasons due to altered disparity signals sent to the eyes.



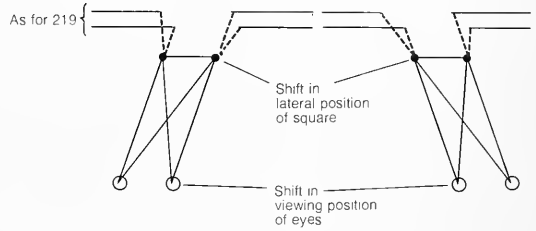
219 A projection explanation of the increased depth effect caused by distant viewing

But Why Two Eyes?

Once again, I have tried in this chapter to display the fruitfulness of linking computational, neurophysiological and psychological approaches to a perceptual problem. But also, I have placed before you a wide range of stereo effects, many surprising and most quite beautiful. Limitations of space have prevented my discussing thoroughly what is known about them. The area of binocular vision is perhaps the best developed one in the whole of visual psychology, and yet I have been able only to scratch its surface. But if you find the stereo effects particularly interesting, and are tempted to extend your knowledge of them further, then there are suggestions at the end of the book for further reading.

A fitting way to end this chapter is to return to the question with which it opened: why two eyes? Now that we have seen so many depth effects, the immediate and obvious answer is: we use the depth cue of binocular disparity, provided by virtue of the fact that the two eyes look at the world from different positions, to tell us about the relative depths of objects in the scene before us. But this eminently plausible and sensible answer to our question may not in fact capture all the benefit which stereopsis conveys. To begin with, we might note again as we did at the outset that the kind of situation in which disparity provides the only adequate depth cue is the stationary inspection of such things as a bunch of flowers, or a tree. In the vast majority of situations, we can get by with other cues. Indeed, we might note that these other cues seem to be given extra weight by the visual system, in that they tend to override stereopsis if they are placed in conflict with it. If the prime benefit of having two eyes is seeing depth, it is clearly not the case that this depth information is so stressed by the visual system that it wins out over other depth cues, come what may.

This consideration prompts a somewhat different answer to our question, 'Why two eyes?' Could it be, perhaps, that the depth effect is a secondary advantage, and that the prime one is giving the visual system a superb way of grouping together features for the purposes of building up an explicit scene description at the level of object recognition? It was explained in chapter 5 that grouping of features is a vital step in scene description, and various grouping rules were described. But these rules failed in certain circumstances: refer back, for example, to the problem posed by the overlapping leaves in



220 Shift in apparent location of square caused by change in head position.

126. Now, with stereopsis available, this ambiguity could have been settled by grouping together those features with a similar depth. That is, features belonging to one leaf could have been separated from those belonging to the other leaf, without any need of conceptually driven processing to disambiguate them. From this viewpoint, perhaps the initial evolutionary advantage of having two eyes was as a solution to the problem of decoding camouflage. Perhaps two-eyed vision really came into its own when it provided a means of grouping together stripe features belonging to the tiger (or other predator, or desirable but hidden prey), and separating them from stripe features produced by the branches, twigs and leaves of the tree in which he was hiding, ready to pounce. This speculation is certainly in keeping with the discovery of random-dot stereograms, because they show just how superb a camouflage-breaking system stereopsis is: only after their binocular fusion can any object whatsoever be seen. Interestingly, stereo photographs are taken from aircraft in order to break down the camouflage of military installations on the ground, a trick which takes advantage of stereopsis as a system for decoding camouflage. So perhaps, with the special kind of depth perception which is stereopsis in its armoury, the visual system is very much better able to break up a scene into its constituent regions, and thereby to get on with the job of seeing what is present.

8 DESCRIPTIONS IN OUR HEADS

What goes on inside our heads when we see? That was the question which opened this book, and an imaginary 'ordinary person' produced his photographic or 'inner screen' theory in reply. Hopefully, the 'ordinary reader' for whom this book is intended could now have a better shot at it, perhaps answering along the following lines:

> Seeing is a matter of building up an *explicit symbolic description of the scene* observed. The photographic analogy is no good because a photograph simply reproduces one image by another, leaving all the information about the scene implicit within the new image. That is hopeless. When we see things, we are engaged in a process of identification – of features, objects, and other attributes of the scene. Sight enables us to point to things, to pick them up, to talk about them, in a word to *act* in relation to them. It does so because it makes *explicit* what the visual scene contains in a *description* cast in a language of *symbols*. Seeing must be a symbolic process because the world itself obviously does not exist inside our heads, and so our 'internal' visual world must be a collection of symbols standing for the scene and its various attributes. Uncertainty surrounds the nature of the brain's symbols for seeing, but one prominent theory suggests that each separate attribute which we see is coded by a separate brain cell. On this theory, activation of any one 'seeing cell' is the physical event determining a conscious visual perception, and most scenes would require very many such cells to be active if they were to be explicitly described in full. But other theories exist about the brain's visual symbols too, and little that is definite can be said at this stage. What we can be sure about is that arriving at an explicit scene description is not a straightforward business. Each eye's *retinal image*, which initiates the whole process of seeing, is inherently *ambiguous*. Various *measurements* are taken from it, and these are *interpreted* to give the required identification of attributes of the scene. The interpretative mechanisms are often embodied in *low-level* networks of cells which do not rely on *high-level* information about what objects the scene might or might not contain. These networks, in which the cells are interconnected with excitatory and inhibitory couplings, incorporate *computational strategies* which are appropriate to many scenes, but not all. Where they are inappropriate, a visual illusion results, so that illusions can afford valuable clues about the ordinary mechanisms of seeing.

We might round off this summary from our attentive and imaginary reader by adding (see **221**) that the initial step in building an explicit scene description is that the retinal image is first encoded as a *grey level description* in the receptors of the retina (a description which is perhaps the closest our visual system ever gets to a photographic type of representation). Then, certain networks of retinal cells make allowance for variations in the illumination of the scene, so that the brain receives information about the *lightness* of points in the scene which is relatively free from complications about whether they are in the shade or in bright light. Next, a *feature description* is built up by using the hypercolumns of the striate cortex. This description is expressed in a vocabulary of low-level symbols which encode all the useful information about changes of lightness within the scene in terms of symbols for such things as edges, lines, blobs etc. The feature description is *segmented* into collections of features which 'go together' in that they come from the same visual structure. Segmentation relies on processes of grouping, texture discrimination, colour perception, movement perception, depth perception etc. Finally, the visual system proceeds to the high-level problem of *object recognition*, probably by checking a structural description of each feature cluster against stored structural descriptions for known objects, and recognising what each feature cluster is by noticing when a match occurs. This process can be helped by high-level contextual information.

Little more need be said by way of summary in this final chapter, but a last reminder should be added that the literature of visual science is vast and that necessarily only a smattering has been referred to here. All topics have been simplified to communicate essentials, but I hope not to the point of over-simplification, and thus not to the point of serious distortion. The reader who finds his appetite whetted for further information about seeing is once again directed to the reference section which follows this chapter.

The recurring theme throughout the book has been the claim that the problem of seeing is best tackled by a combined assault using psychological, physiological and computational methods in unison. This does not mean that it is necessary nowadays to be a jack of all trades to study vision, but it does imply that a familiarity with all three literatures is important, and that a team approach is probably the best way forward.

The *psychologist* provides methodologies for studying the input-output (retinal-image-to-perception) performance of the best visual systems presently known, those of man and other animals. Often, the phenomena chosen for study are

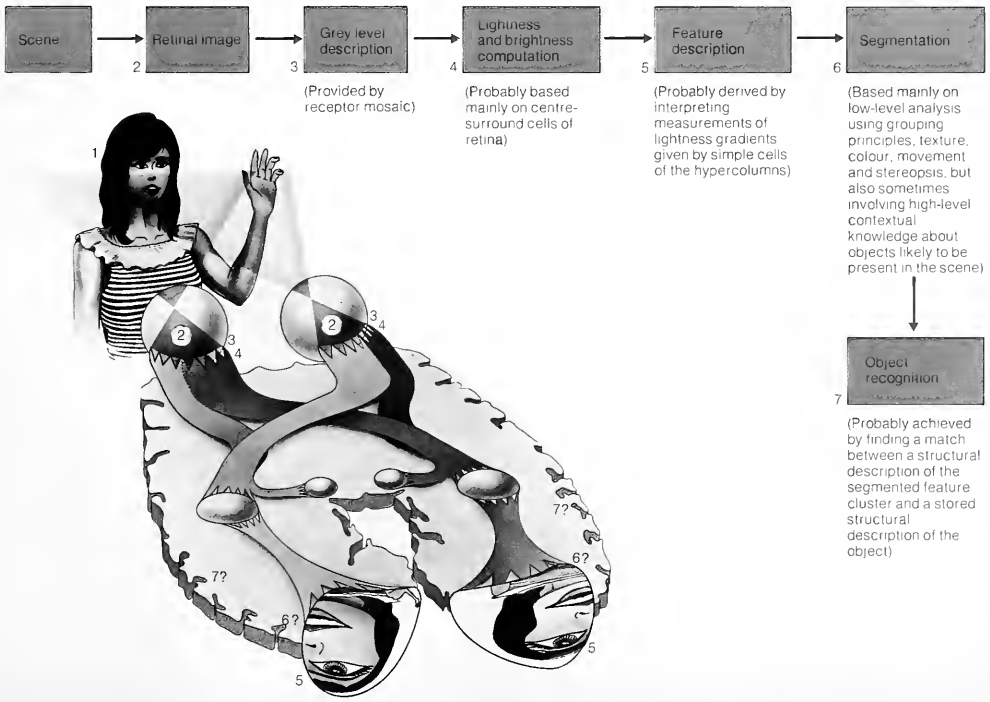
illusions, each illusion being regarded as the result of a misapplication of an interpretative stratagem (low-level or high-level) to a scene for which it is not wholly suitable. Illusions thus provide clues about strategies which the visual system has found it profitable to employ in interpreting the retinal image, strategies picked out by natural selection during the course of evolution.

The *physiologist* studies the hardware of biological visual systems directly, using such techniques as microelectrode recordings, often allied to neuroanatomical probings. He provides information on the biological 'nuts and bolts' of seeing, as it were, and great strides forward have been taken in this direction over the past two decades or so.

The *computational scientist* takes on the job of actually trying to build a visual system. He tends to study the fundamental processing requirements of a given visual processing task, if only because his 'first shots' tend to misfire badly, and he is therefore thrown back to thinking hard about what *exactly* the problem is that has to be solved, and then what computational strategies might be open to him to solve it. It is at this point that a knowledge of the psychology and physiology of seeing can be valuable. But it is also true that many applied workers in the area, trying to invent machine vision for industrial applications where they can control rather precisely the inputs with which they have to deal, can adopt *ad hoc* solutions (e.g. specially designed numbers on cheques). These are adequate for their particular problem, but they have scant relation to what goes on inside biological visual systems. The latter have to make do with natural images of natural scenes, given to them, moreover, by (often) low-quality eyes, and so have developed a sophisticated range of strategies to cope with this problematic input.

If we use this threefold approach – psychological, physiological and computational – the prospect before us, perhaps within the lifetimes of younger readers, is of a seeing machine built to match human visual performance. Or at least, that is my guess. Others regard the task as so difficult as to be centuries away, or perhaps impossible altogether. For them, neural tissue, with its tiny components and richly interconnected networks, is the only material up to the job. But my own belief is based on the fact that we have not yet had much more than three decades of developing the electronic computer, and even less of what could properly be called an attempt to build a machine-vision system, and yet already progress has been good. To be sure, progress has not been as rapid and as spectacular as originally hoped. The problems of mimicking all aspects of mind, of creating artificial intelligence, have turned out to be much more difficult than at first suspected by many enthusiasts, and this bitter experience has substantially moderated the optimism of many. Yet we are constantly witnessing the advent of new computer technology, and there is no sign yet of levelling-off in machine perform-

221 The human visual system Some steps in building up an explicit symbolic scene description



ance, in terms either of hardware (the computer's electronic circuitry) or of software (the computer programs which use the circuitry).

As far as the hardware is concerned, the most exciting recent development is that of the 'distributed processor'. This is a computing system composed of many more or less independent processing sub-units, each able to get on with its own job, but also able to 'talk' to one another and to other parts of the machine. This looks like the true dawn of genuinely parallel processing, hitherto the province *par excellence* only of biological brains. In most computers to date, there has been just one 'central processor', a single device through which all parts of a computation had to proceed, one by one, in a so-called serial mode: any particular piece of computation has to wait upon the completion of the one before it, or else interrupt it. Such a serial device can, awkwardly, perform certain computations that require parallel processing, for example by using an iterative technique (p. 152). But that way is slow and clumsy and we are now on the point of having parallel-processing computers composed of myriad micro-processors, each one 'doing its own thing' at any given moment.

The distributed processor is remarkably reminiscent of the hypercolumns of the visual cortex. Each hypercolumn 'looks' at its own particular part of the retinal image and processes whatever information it finds there, thus contributing its bit to the feature description of the entire scene. All hypercolumns work in parallel, although they probably 'speak' to one another, as neighbours, during segmentation, just as distributed micro-processors can be made to do. Hypercolumns feed other brain sites (other micro-processors), and so on, in one extensive parallel-processing network which culminates in the symbolic scene description which constitutes sight. Thus, with the development of the distributed processing computer, perhaps we are in a position to implement, say, a version of Marr's theory and run it on natural images in 'real time' (computer jargon for dealing with inputs as they actually come about in a real-world setting, rather than on the basis of a memory of the inputs stored at the time of their original production and then used subsequently). In any event, this new kind of computer machinery must have enormous implications for those interested in simulating parallel-processing networks of the kind found in biological vision.

Suppose we were able to build a seeing machine capable of matching human vision, in the sense that if confronted with a natural scene of the kind we are accustomed to, it could print out on its typewriter, or 'speak' on some other output device, an explicit description of what the scene contained: would such a machine have conscious visual experience? Presumably we will never know. After all, how do you know that I am conscious? Obviously because I tell you so, and because I am built rather like you, and you are therefore happy to extrapolate from knowledge of your own conscious experience and accept that I too share this attribute. But such an argument would not satisfy a determined sceptic, as debate between philosophers throughout the ages on this time-honoured question testifies. It seems therefore that we will have to remain agnostic about machine consciousness. All we can be sure of, given a perceptual ability which matches that of humans, is that machines will *appear* to be conscious. Already computer programmers are typically anthropomorphic in their everyday dealings with their machines. They use phrases like 'It's thinking hard right now', 'It got confused then', 'It's suffering an illusion', 'It thought I wanted x but in fact I wanted y', etc. This is the natural way to talk about clever machines, just as it is the natural way we talk about clever animals.

But it is my belief that the anthropomorphism will become much more marked, even rather spooky ('Is there somebody in there?'), when we start to deal with high-class perceiving machines. Indeed, I suspect that clever perceiving machines will prove far more impressive than, say, clever chess-playing ones. (Although it is probable that the development of the latter will hinge upon the former, in that to play good chess requires 'seeing' in the positions of the pieces certain possibilities, a form of thinking closer to vision than perhaps we normally acknowledge, despite the frequent use of the word 'seeing' to refer to understanding, for example 'Ah! I see what you mean.')

In fact, as the pursuit of artificial intelligence proceeds, I am sure we will have to adjust our notions about the nature of man, just as the Victorians had to adjust theirs in the face of Darwin's theory of evolution. 'Man as an animal? Rubbish!' was the irrational, all too common, but also very understandable, reaction to Darwin's ideas. Today the parallel response is: 'Man as a machine? Ridiculous!', quickly followed by remarks revealing some sadly ignorant myths – 'Machines can't think', 'Computers are no more than large, electronic arithmetic calculators', 'Machines do only what they are told to do', and so on. Machines are simply not necessarily like that, certainly not present-day sophisticated computers, but this fact is not widely recognised. Even science fiction and the television space sagas have not yet convinced many people that it is sensible and proper to consider man as a type of machine, albeit a very special kind of machine (just as he is a very special kind of animal). Perhaps this is because a frequent science fiction theme, perhaps *the* most frequent, is that 'Man is more than a robot'. So often, at the last moment, the plot is resolved not by the appearance of a *deus ex machina* but, so to speak, by a *homo ex machina*! Man's body is machine-like, yes, but his intellect – no! His mind has some special non-machine-like quality which usually saves the day. This was particularly true of *Star Wars*, the enormously successful space classic, where the hero (human, of course) won the day by switching off his computer and beating the enemy with his own bare hands – or, perhaps one should say, his own bare mind. Indeed, the dramatic success of this film might in part be ascribed to the fact that it panders to man's desire to see himself as a cut above computers, just as he craved to see himself as a cut above animals. But surely this warped view will fade as people come to be aware that their ideas of what constitutes a machine are absurdly limited. Machines need need not be rigid slaves bound for ever to follow their instructions to the letter, if this means showing no creativity, no learning, no adaptability, no perceiving. Already we know how to instruct computers to exhibit these qualities in certain situations. But I suspect that the full consequences of artificial intelligence will not become manifest to the ordinary man until the problem of seeing is solved. Only when the computer is given its own eyes, and its own capacity for explicit symbolic scene description, will it reveal its true potential for life-like action.

FURTHER READING

It would have been out of character for an introductory book such as this to give detailed references to sources in the text, and it would be equally out of place to give an extensive bibliography now. Instead, I confine myself to making just a few suggestions for further reading in connection with each chapter. Anyone wanting to track down references to a particular problem area should have little difficulty in doing so from these starting points, with the general texts given under chapter 1 applying to the whole book.

Chapter 1 Pictures in Our Heads

GREGORY, R.L. (1971) *The Intelligent Eye*, and GREGORY, R.L. (1977) *Eye and Brain : the Psychology of Seeing*. These are two marvellously stimulating books written for the non-specialist but also presenting original ideas.

HELD, R., and RICHARDS, W. (1972) *Perception : Mechanisms and Models*, HELD, R. (1974) *Image, Object and Illusion*, and HELD, R., and RICHARDS, W. (1976) *Recent Progress in Perception*. These three books are *Readings from Scientific American* and contain papers which are often excellent and always superbly illustrated.

LINDSAY, P.H., and NORMAN, D.A. (1977) *Human Information Processing*. A good introductory undergraduate textbook covering perception.

KAUFMAN, L. (1974) *Sight and Mind : An Introduction to Visual Perception*. A good advanced undergraduate textbook.

Chapter 2 Seeing Features

ULLMAN, J.R., and ROSENFELD, A. (1977) 'Picture recognition and analysis.' *The Radio and Electronic Engineer*, Vol. 47, No. 1/2, pp. 33-48. This is a good introductory review article on pattern recognition by machine and, amongst much else, explains various ways of tackling problems with feature templates. It has an extensive bibliography.

Chapter 3 The Visual Machinery of the Brain

ROBSON, J.G. (1975) 'Receptive fields: neural representation of the spatial and intensive attributes of the visual image.' In CARTERETTE, E.D., and FRIEDMAN, M.P. (1975) (Eds) *Handbook of Perception. Vol. 5 : Seeing*, pp. 81-117. This is an excellent review article for those who wish to pursue in depth neurophysiological knowledge about the visual pathway, but the general reader should be warned that it is technical. It provides a good bibliography of classic papers in the field.

HUBEL, D.H., WIESEL, T.N., and STRYKER, M.P. (1978) 'Anatomical demonstration of orientation columns in macaque monkey.' *Journal of Comparative Neurology*, Vol. 177, No. 3, pp. 361-80. This is one of the latest papers in Hubel and Wiesel's brilliant series on the neuroanatomy of the striate cortex. It is difficult but necessary reading if you want to understand more about this area than explained in this book, and it reports the best evidence yet for the existence of hypercolumns.

MARR, D. (1976) 'Early processing of visual information.' *Philosophical Transactions of the Royal Society of London*, Series B, Vol. 275, pp. 483-524. This is Marr's seminal paper which has guided much of the material of this book.

MARR, D., and NISHIHARA, H.K. (1978) 'Visual information processing: artificial intelligence and the sensorium of sight.' *Technology Review*, Vol. 81, October, pp. 1-23. A useful non-technical review of Marr's approach to perception.

MAFFEI, L., and FIORENTINI, A. (1977) 'Spatial frequency rows in the striate visual cortex.' *Vision Research*, Vol. 17, pp. 257-64. The authors describe results suggesting that in the cat's visual cortex different rows of cells are concerned with analysing different spatial frequencies.

INGLE, D., and SPRAGUE, J.H. (1975) (Eds) *Sensorimotor Function of Mid-brain Tectum. Neuroscience Research Program Bulletin*, Vol. 13. Technical, but a good source of work on the superior colliculus.

Chapter 4 After-effects - The Psychologist's Microelectrode

MOLLON, J.D. (1974) 'After-effects and the brain.' *New Scientist*, 21 February, pp. 479-82. This is an admirably clear and succinct summary at an introductory level explaining why after-effects are of such interest to psychologists studying perception.

BLAKEMORE, C., and CAMPBELL, F.W. (1969) 'On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images.' *Journal of Physiology*, Vol. 203, pp. 237-60. This is one of the classic papers of the past decade of visual research and is the source of the data presented on p. 96.

CAMPBELL, F.W., and MAFFEI, L. (1974) 'Contrast and spatial frequency.' *Scientific American*, Vol. 231, November, pp. 106-14. A useful introduction to the concept of spatial frequency analysis.

BLAKEMORE, C., and COOPER, G.F. (1970) 'Development of the brain depends on the visual environment.' *Nature*, Vol. 228, pp. 477-8. Reports the kitten-in-a-drum experiment described on p. 95.

GREGORY, R.L. (1968) 'Visual illusions.' *Scientific American*, Vol. 219, November, pp. 66-76. This is the source for my quotation from Gregory on p. 104.

GREGORY, R.L. (1973) 'The Confounded Eye.' In GREGORY, R.L., and GOMBRICH, E.H. (1973) (Eds) *Illusion in Nature and Art*, pp. 49-96. This is perhaps Gregory's most definitive statement to date on how he chooses to categorise and consider visual illusions.

ROBINSON, J.O. (1972) *The Psychology of Visual Illusion*. This is a marvellous source book for anyone trying to track down an illusory figure (Who invented it first?) as well as a very useful summary of theories of various illusions.

Chapter 5 Seeing Objects

SUTHERLAND, N.S. (1973) 'Object Recognition.' In CARTERETTE, E.D., and FRIEDMAN, M.P. (1973) (Eds) *Handbook of Perception. Vol. 3 : Biology of Perceptual Systems*, pp. 157-206. A good short review article on the psychology of object recognition.

ULLMAN, J.R., and ROSENFELD, A. (1977) See under chapter 2.

MARR, D. (1976) See under chapter 3.

HARMON, L.D. (1973) 'The recognition of faces.' *Scientific American*, Vol. 229, November, pp. 71-82. Discusses computer-generated block portraits and other ways of studying face recognition.

JULESZ, B. (1975) 'Experiments in the visual perception of texture.' *Scientific American*, Vol. 232, April, pp. 34-43. Includes many fascinating examples of texture differences which can and cannot be easily discriminated.

WINSTON, P.H. (1977) *Artificial Intelligence*. The early chapters are a useful starting point for tackling the 'blocks world' literature.

BARLOW, H.B. (1972) 'Single units and sensation: a neuron doctrine for perceptual psychology?' *Perception*, Vol. 1, pp. 371-495. Barlow here argues the case for active single nerve cells as the code for elements of the perceptual world.

WADE, N.J. (1978) 'Op art and visual perception.' *Perception*, Vol. 7, pp. 21-46. An interesting and broad review of perceptual effects used by op artists.

Chapter 6 Seeing Lightness and Brightness

HORN, B.K.P. (1974) 'Determining lightness from an image.' *Computer Graphics and Image Processing*, Vol. 3, pp. 277-99. Explains clearly the logic and the mathematics behind Horn's approach to the lightness computation.

LAND, E.H. (1977) 'The retinex theory of colour vision.' *Scientific American*, Vol. 237, December, pp. 108-28. A clear account of Land's theory of colour vision.

MARR, D. (1974) 'The computation of lightness by the primate retina.' *Vision Research*, Vol. 14, pp. 1377-88. Describes how the retina seems well equipped with structures suitable for carrying out the Land/Horn lightness computation.

ROBSON, J.G. (1975) See under chapter 3. A good review of receptive field properties of retinal ganglion cells is included in this paper.

FRISBY, J.P., and CLATWORTHY, J.L. (1975) 'Illusory contours: curious cases of simultaneous brightness contrast?' *Perception*, Vol. 4, pp. 349-57. This paper discusses the idea that Kanizsa's triangle might be explicable partly in terms of low-level contrast mechanisms.

GILCHRIST, A.L. (1979) See under chapter 7.

Chapter 7 Seeing with Two Eyes

JULESZ, B. (1971) *Foundations of Cyclopean Perception*. This is the major recent work on binocular vision and has many beautiful anaglyphs.

PETTIGREW, J.D. (1972) 'The neurophysiology of binocular vision.' *Scientific American*, Vol. 227, August, pp. 84-95. A clear account of work on nerve cells selectively sensitive to binocular disparity.

MARR, D., and POGGIO, T. (1976) 'Co-operative computation of stereo disparity.' *Science*, Vol. 194, pp. 283-7. A full account of the network described here on pp. 149-52.

GILCHRIST, A.L. (1979) 'The perception of surface blacks and whites.' *Scientific American*, Vol. 240, March, pp. 88-97. Relevant to the debate on stereopsis and lightness perception referred to on p. 154.

Chapter 8 Descriptions in Our Heads

BODEN, M. (1977) *Artificial Intelligence and Natural Man*. A splendid introduction to the field of artificial intelligence.

WINSTON, P.H. (1977) See under chapter 5.

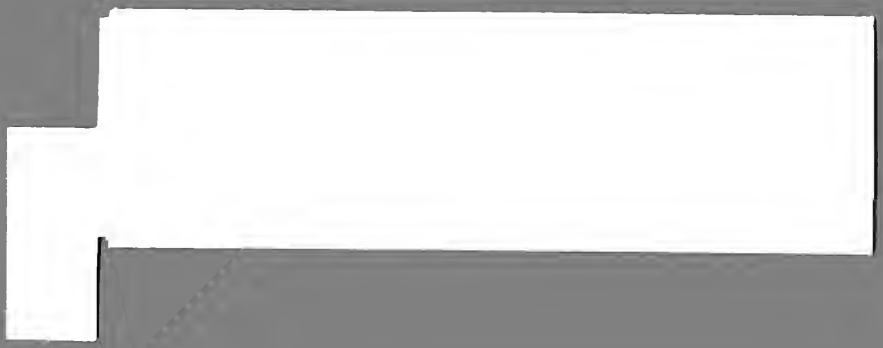
MARR, D., and POGGIO, T. (1977) 'From understanding computation to understanding neural circuitry.' *Neurosciences Research Program Bulletin*, Vol. 15, No. 3, pp. 470-88. A useful review for anyone wishing to consider further the problems and dangers in extrapolating from properties of nerve cells to their functional role in a computation. This issue of the Bulletin also has many other articles of interest to students of perception.

INDEX

References are to page numbers, except that references in bold type are to figure numbers, with the relevant plate number following in parenthesis where applicable. All references are given in a single sequence, in the order of their occurrence in the text.

Subjects

Authors



The Author: John Frisby is Professor of Psychology at the University of Sheffield, and a leading authority on vision, especially the two-eyed 'stereoscopic' vision which is a major topic of this book. **Seeing** represents the fulfilment of a long-standing wish to convey to a wide public the fascination and significance of the study of seeing – especially of illusions.

Professor Frisby has worked with Bela Julesz, the inventor of the computer-generated 'random-dot stereogram', and has written many papers about seeing. One of his special interests is the art of M.C.Escher, some of whose work is illustrated in this book.

Front cover: Stand the book upright and try looking at the front cover from about 2 metres (6 feet) away, through half-closed eyes, so that your vision is blurred. You will find that this makes it easier to see what is portrayed there, not more difficult as one might expect. This curiosity is discussed in detail on p.118, and is just one of the many visual effects, illusions and paradoxes in the book.

Back cover: Fraser's spiral. Try following the course of the spiral with a finger. You will find that only concentric circles exist in the printed image. The spiral is a creation of your visual system, a dramatic instance of how what we see can differ dramatically from what is actually present. Illusions like this give us valuable clues about how seeing works.

OXFORD UNIVERSITY PRESS