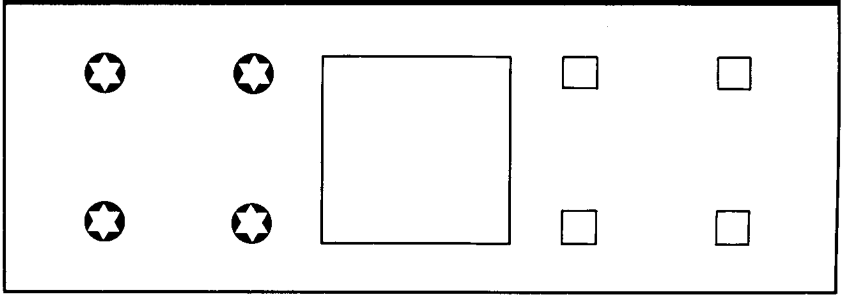
# 6

# Perceptual Organisation

In Chapter 5 we considered how the primal sketch—a description of edge and line segments, terminations and other key features—may be derived from an array of intensities in the retinal image. In this chapter we turn to consider how such low-level descriptions may be organised into larger perceptual "chunks". When we view the world we do not see a collection of edges and blobs—unless we adopt a very analytical perceptual attitude—but see instead an organised world of surfaces and objects. How is such perceptual organisation achieved? How do we know which parts of the visual information reaching our sensory apparatus belong together? These are the questions addressed in this chapter. The first part of the chapter concentrates on human perception, because it was through the study of this that many of the principles of perceptual organisation became established. We return to the broader perspective of animal vision when we consider how such perceptual principles may be exploited in natural camouflage and advertisement. In the final part of the chapter we turn to artificial intelligence approaches to perceptual organisation.

As we discussed in Chapter 4, the psychology of human visual perception during the late 19th and early 20th century was dominated by associationism. It was assumed that perception could be analysed in terms of its component sensations, and that complex ideas were the result of associating together simpler ones. However, as the Gestalt psychologists pointed out, an analysis of perception into discrete sensations overlooks some important aspects of form and structure. Each of the arrangements shown in Fig. 6.1 possesses the quality of "squareness" despite being composed of

FIGURE 6.1



Each of these three forms is seen as being square, despite being composed of quite different elements.

quite different elements. A tune is recognisable despite being played in a different key or at a different speed. The spatial and temporal relationships between elements are as important as the absolute size, location, or nature of the elements themselves, and a sensation-based account of perception fails to capture this.

Even Wundt (1896) recognised that a simple structuralist analysis failed to capture certain perceptual phenomena (Wundt, 1896, trans. 1907, p.368):

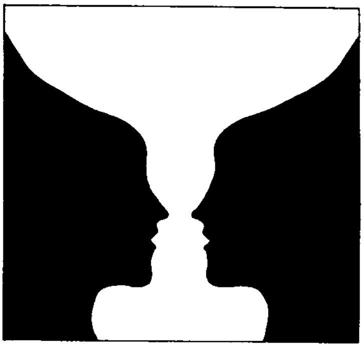
> A compound clang is more in its ideational and affective attributes than merely a sum of single tones.

But it was the Gestalt psychologists, notably Koffka (1935), Köhler (1947), and Wertheimer (1923), with whom the catch-phrase "the whole is greater than the sum of its parts" became identified. We will first describe the Gestalt ideas about perceptual organisation, and then consider more recent accounts.

## AMBIGUOUS PICTURES

The world that we view appears to be composed of discrete objects of various sizes seen against a background of textured surfaces. We usually have no difficulty in seeing the boundaries of objects, unless these are successfully camouflaged (see later), and there is generally no doubt about which areas are "figures" and which comprise the "ground". However, it is possible to construct pictures in which there is ambiguity about which region is "figure" and which "ground". Edgar Rubin, one of the Gestalt psychologists, used the face–vase picture (Fig. 6.2) to illustrate this. The picture can be seen either as a pair of black faces in profile, or as a white vase, but it is impossible to maintain simultaneously the perception of both the faces and the vase. The contour dividing the black and white regions of the picture appears to have a one-sided function. It "belongs" to whichever region is perceived as figure. People viewing this picture usually find that their perception of it shifts

**FIGURE 6.2**



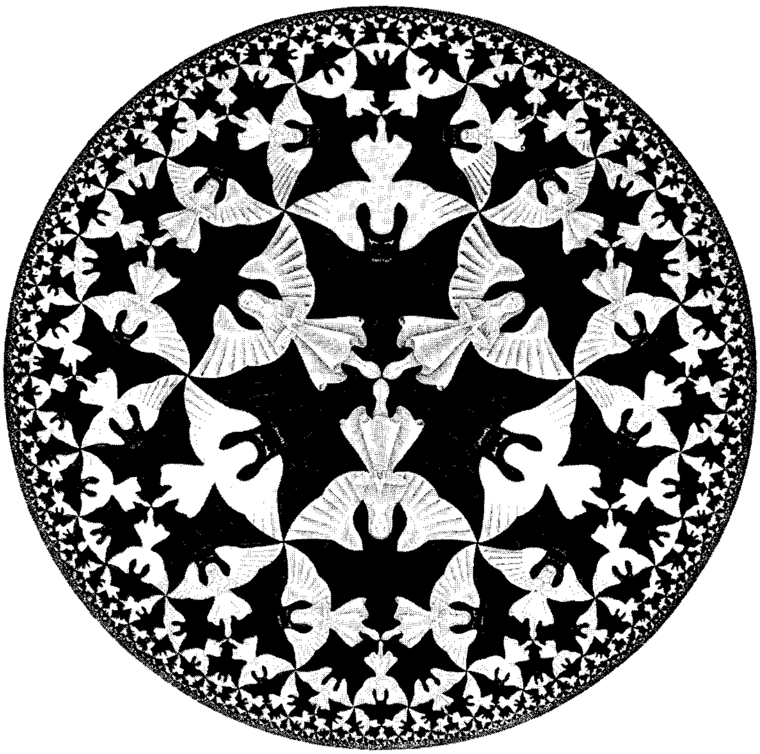
This picture, devised by E. Rubin in 1915, can be seen *either* as a pair of faces in silhouette, *or* as a white vase.

from one interpretation to the other, sometimes quite spontaneously. The artist M.C. Escher exploited this principle of perceptual reversibility when he produced etchings in which there is figure/ground ambiguity (see Fig. 6.3).

It is also possible to construct pictures so that the internal organisation of a particular figure is ambiguous. Jastrow's duck–rabbit picture (Fig. 6.4a) may be seen as a duck (beak at the left), or a rabbit (ears at the left), but not both simultaneously. Even a figure as simple as a triangle turns out to be perceptually ambiguous, as Fig. 6.4b shows (Attneave, 1971). The triangles appear to "point" in any one of the three possible directions, and when they appear to change direction they all change together, implying some spatially extended organising process that is applied to all the individual triangles simultaneously (Palmer, 1992). Some abstract and "op"-art may be perplexing to view because no stable organisation is apparent (see Fig. 6.5).

The perception of such ambiguous displays is interesting in its own right, and psychologists have investigated the factors influencing which organisation of an ambiguous display will be preferred, and the factors determining perceptual reversals (for example see Attneave, 1971; Hochberg, 1950; Pheiffer, Eure, & Hamilton, 1956). In all these examples, the perceptual "data" remain the same, but the interpretation varies. It

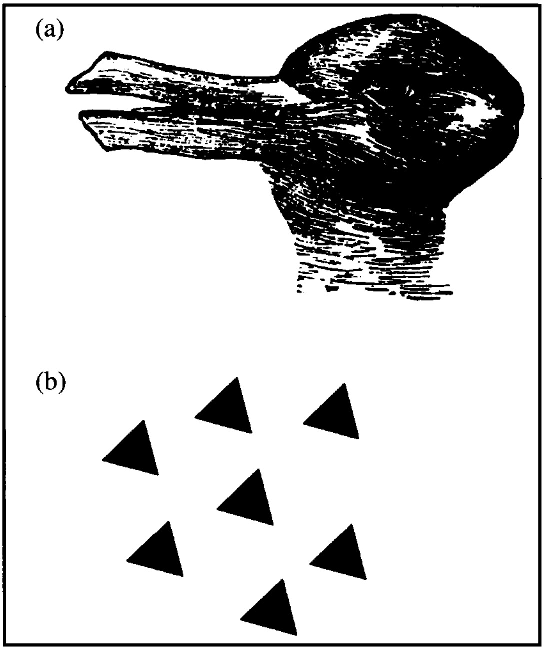
FIGURE 6.3 (top); FIG. 6.4 (bottom left); FIG. 6.5 (bottom right)

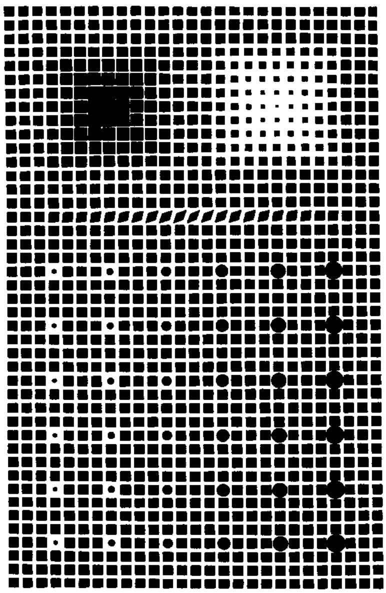


M.C. Escher's "Circle Limit IV".


Fig. 6.4 (below left)
(a) Duck or rabbit? This ambiguous picture was introduced to psychologists by J. Jastrow in 1900.
(b) Triangles as ambiguous figures. Note how the whole group of triangles appears to point in one, then another of the three possible directions. After Attneave (1971).

Fig. 6.5 (below)
"Supernovae" 1959–1961 by Victor Vasarely.



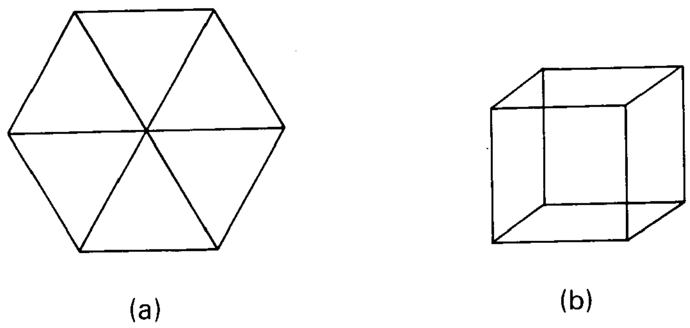
seems as though there must be a strong "top-down" component in such perceptions. Higher levels of perceptual interpretation appear to be continually constraining and guiding the lower levels of image analysis.

However, these ambiguous pictures have been cleverly constructed, and our perception of them is not necessarily typical of normal processing. Ambiguity generally does not arise in the real world, nor in most pictures. Rather than having constantly shifting interpretations, we usually see a stable and organised world. For example, viewing Fig. 6.6a in isolation, most people would report seeing a hexagon, whereas those viewing Fig. 6.6b report seeing a picture of a three-dimensional cube, even though Fig. 6.6a is an equally legitimate view of a cube, viewed corner on. Figure 6.7 is seen as a set of overlapping circles, rather than as one circle touching two adjoining shapes that have "bites" taken out of them. Why, given these possible alternative perceptions, do we see these pictures in these ways?

## GESTALT LAWS OF ORGANISATION

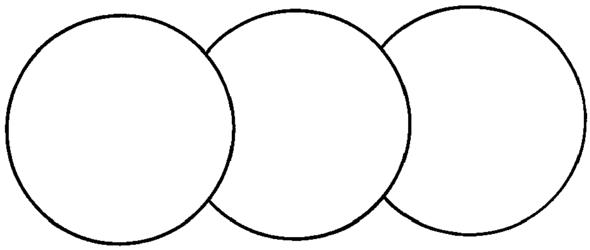
The Gestalt psychologists formulated a number of principles of perceptual organisation to describe how certain perceptions are more likely to occur than others. Some of their principles were primarily to do with the grouping of sub-regions of figures, and others were more concerned with the segregation of figure from ground. However, as sub-regions of a figure need to be grouped in order for a larger region to be seen as "belonging

FIGURE 6.6




The form at (a) looks like a hexagon, whereas that at (b) looks like a cube. Of course (a) is also a legitimate view of a cube.

FIGURE 6.7



Most people would see this as a set of overlapping circles, although two of the shapes might have "bites" taken out of them.

together" as a figure, we will discuss all these principles together.

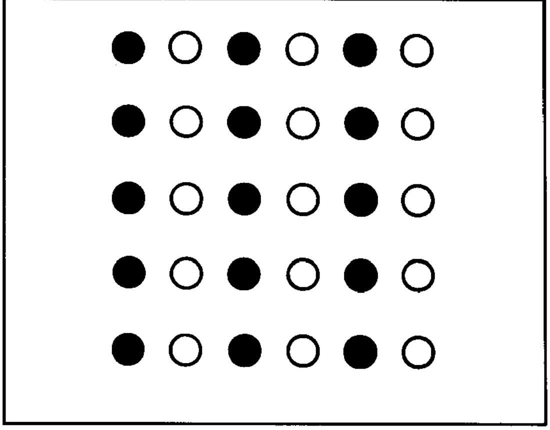
### Proximity

One of the most important factors determining the perceptual organisation of a scene is proximity of the elements within it. Things that are close together are grouped together. In Fig. 6.8a the perception is of columns, because the horizontal spacing of the dots is greater than their vertical spacing. In Fig. 6.8b we see rows, because the horizontal spacing of the dots is the smaller, and Fig. 6.8c is ambiguous: the dots are equally spaced in both directions. Proximity in depth is a powerful organising factor. The central square in a Julesz random-dot stereogram (see Ch.7, p.144) is not visible until the two halves of the stereo pair are viewed in a stereoscope. Dots with the same disparity values are then grouped together and the square is seen as a distinct figure floating above its background.

### Similarity

Things that look "similar" are grouped together. The examples shown at the top of Fig. 6.16 (p.112) appear to consist of two distinct regions, with a boundary between them. The elements on one side of this boundary have a different orientation from those on the other. In Fig. 6.9 the perception is of columns, even though the proximity information suggests rows, illustrating that similarity may override proximity information. The question of *how* similar items must be in order to be grouped together is an empirical one to which we will return.

FIGURE 6.9

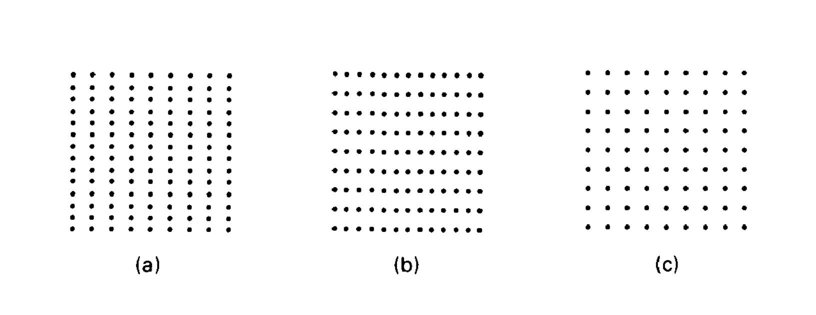


This picture is seen as columns. Similarity in brightness of the dots overrides proximity.

### Common fate

Things that appear to move together are grouped together—think of a flock of birds or a school of fish. A camouflaged animal will remain well-hidden only if it remains stationary. As soon as it moves it is easier to see. Gibson, Gibson, Smith, and Flock (1959) illustrated grouping by common fate with a simple demonstration. They sprinkled powder on two sheets of glass, and projected an image of the powder onto a screen. While the sheets were held still a single collection of powder was seen. As soon as one sheet was moved across the other, viewers saw the powder segregated into two independent collections, by virtue of the movement in the display. Johansson (1973) has produced an even more dramatic demonstration of the power of movement to confer organisation. He attached

FIGURE 6.8




The dots in (a) form columns because they are closer vertically than horizontally. At (b) we see rows, the dots here are closer horizontally; (c) is ambiguous, the dots are equally spaced in both directions.

lights to the joints of a darkly clothed actor and filmed him as he moved in a dark room, so that only the lights were visible. When the actor was at rest, observers reported perceiving a disorganised collection of points. As soon as the actor walked, their perception was that of a moving human figure, whose actions, gait, and even gender could be discerned from the pattern of moving points. Johansson's demonstrations suggest that "common fate" involves much more than simply grouping together elements that have a common speed and direction, and we shall return to discuss the perceptual organisation of such complex displays in Chapter 15.

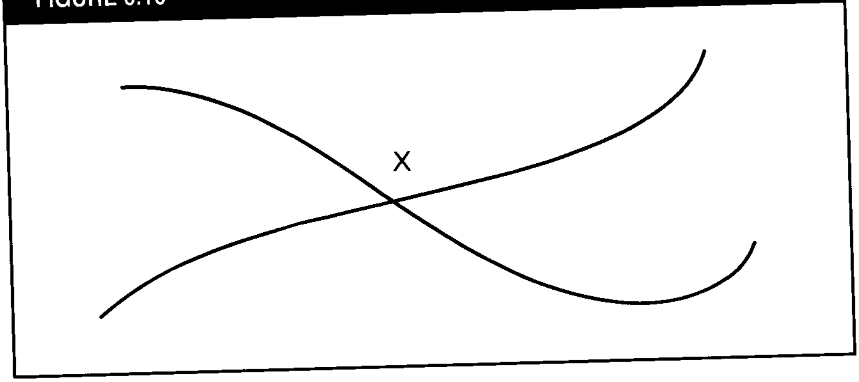
## Good continuation

In a figure such as Fig. 6.10, one tends to perceive two smooth curves that cross at point X, rather than perceiving two irregular V-shaped forms touching at X. The Gestaltists argued that perceptual organisation will tend to preserve smooth continuity rather than yielding abrupt changes. Quite dissimilar objects may be perceived as "belonging together" by virtue of a combination of proximity and good continuity (see Fig. 6.11). Good continuation may be considered the spatial analogue of common fate.

## Closure

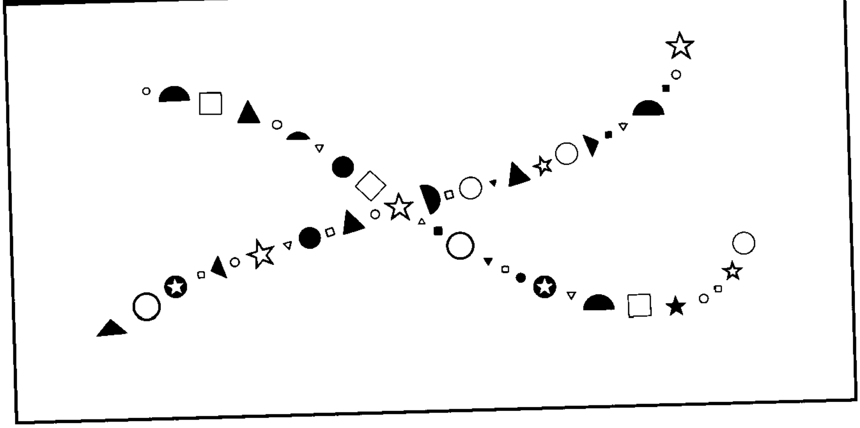
Of several geometrically possible perceptual organisations, that one will be seen which produces a "closed" rather than an "open" figure. Thus the patterns on the left and right of Fig. 6.1 are seen as squares rather than as crosses, because the former are closed. The Gestaltists suggested that the stellar constellation "the plough" might be seen as a plough because of closure and good continuation.

FIGURE 6.10




This is seen as two smooth lines crossing at X, rather than as two V-shapes touching at X.

FIGURE 6.11



Quite dissimilar shapes may be grouped together through a combination of proximity and good continuation.

### Relative size, surroundedness, orientation, and symmetry

Other things being equal, the smaller of two areas will be seen as a figure against a larger background. Thus Fig. 6.12a will tend to be perceived as a black propellor shape against a white background because the black area is the smaller. This effect is enhanced if the white area actually surrounds the black as in Fig. 6.12b, because surrounded areas tend to be seen as figures. However, if we orient the figure so that the white area is arranged around the horizontal and vertical axes then it is easier to see this larger area as a figure (Fig. 6.12c). There seems to be a preference for horizontally or vertically oriented regions to be seen as figures. Also note that both these sets of patterns are symmetrical. Symmetry is a powerful perceptual property, and may be more salient perceptually than nonreflected repetition (Bruce & Morgan, 1975). Examples of symmetry and repetition are shown in Fig. 6.13. Symmetrical areas will tend to be perceived as figures, against asymmetrical backgrounds. Figure 6.14 shows how relative size, orientation, symmetry, and surroundedness may all operate together so that it is difficult if not impossible to see anything other than the black areas as the figures in this picture. The reader will note the perceptual stability of this picture compared with the ambiguity of Fig. 6.2, where the relative sizes, surroundedness, and symmetries in the display favour neither the "faces" nor the "vase" particularly strongly.

### The Law of Prägnanz

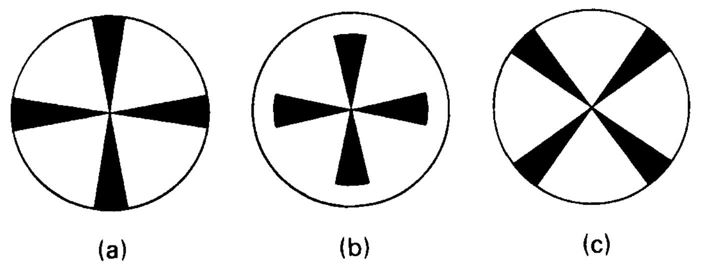
For the Gestalt psychologists, many of these laws were held to be manifestations of the Law of Prägnanz, introduced by Wertheimer. Koffka (1935, p.138) describes the law:

> Of several geometrically possible organisations that one will actually occur which possesses the best, simplest and most stable shape.

Thus an organisation of four dots arranged as though they were at the corners of a square (Fig. 6.1, right) will be seen as a "square" because this is a "better" arrangement than, say, a cross or a triangle plus an extra dot. The square is a closed, symmetrical form, which the Gestaltists maintained was the most stable.

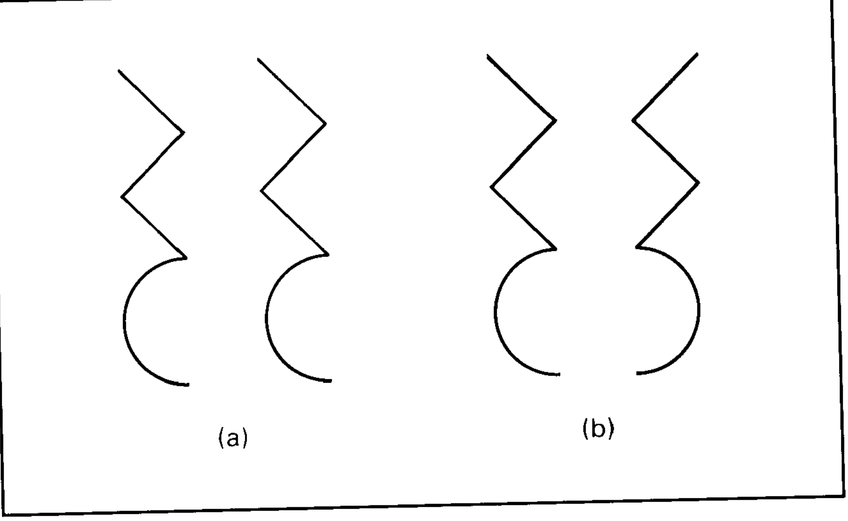
Although the Gestaltists accepted that familiarity with objects in the world, and "objective set", might influence perceptual organisation, they rejected an explanation solely in these terms. A major determinant of perceptual organisation for them was couched in terms of certain "field forces" that they thought operated within the brain. The Gestaltists maintained a *Doctrine of Isomorphism*, according to which there is, underlying every sensory experience, a brain event that is structurally similar to that experience. Thus when one perceives a circle, a "circular trace" is established, and so on. Field forces were held to operate to make the outcome as stable as possible, just as the forces operating on a soap bubble are such that its most stable state is a sphere.

FIGURE 6.12



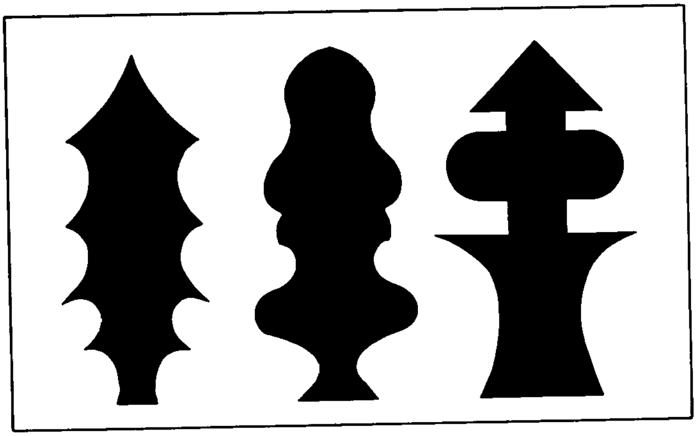

The preferred perception of (a) is a black propeller on a white background. This preference is enhanced if the white area surrounds the black, as at (b). If the orientation of the forms is altered, so that the white area is oriented around the horizontal and vertical axes, as at (c), then it is easier to see the larger white area as a figure.

FIGURE 6.13




At (a) one form is repeated without reflection around a vertical axis. This arrangement is not as perceptually salient as the arrangement shown at (b), where repetition with reflection around the vertical axis produces bilateral symmetry.

FIGURE 6.14



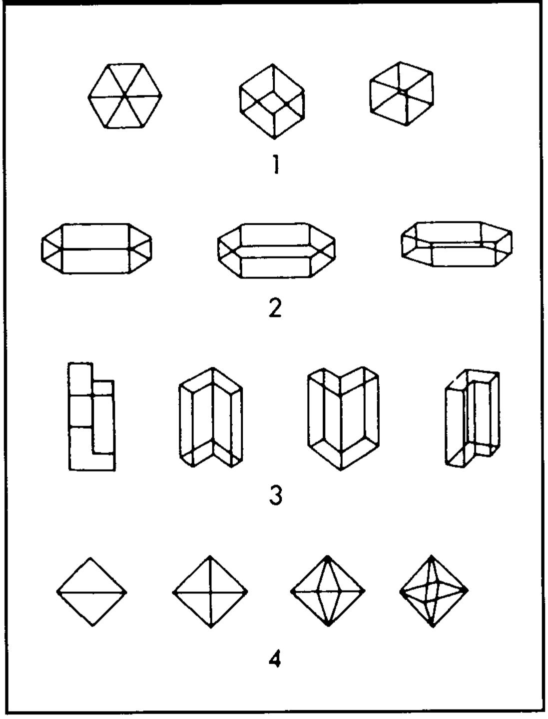
This picture clearly shows black shapes on a white background. The black shapes are vertically oriented, symmetrical, small (relative to the background), and surrounded by the background.

Unfortunately, no evidence has been provided for such field forces, and the physiological theory of the Gestaltists has fallen by the wayside, leaving us with a set of descriptive principles, but without a model of perceptual processing. Indeed, some of their "laws" of perceptual organisation today sound vague and inadequate. What is meant by a "good" or a "simple" shape, for example? Recently workers have attempted to formalise at least some of the Gestalt perceptual principles.

## RECENT APPROACHES TO PERCEPTUAL ORGANISATION

Hochberg and Brooks (1960) tried to provide a more objective criterion for the notion of "goodness" of shape by presenting subjects with line drawings (Fig. 6.15) and asking them to rate the apparent tridimensionality in these figures. They argued that as the complexity of the figures

FIGURE 6.15



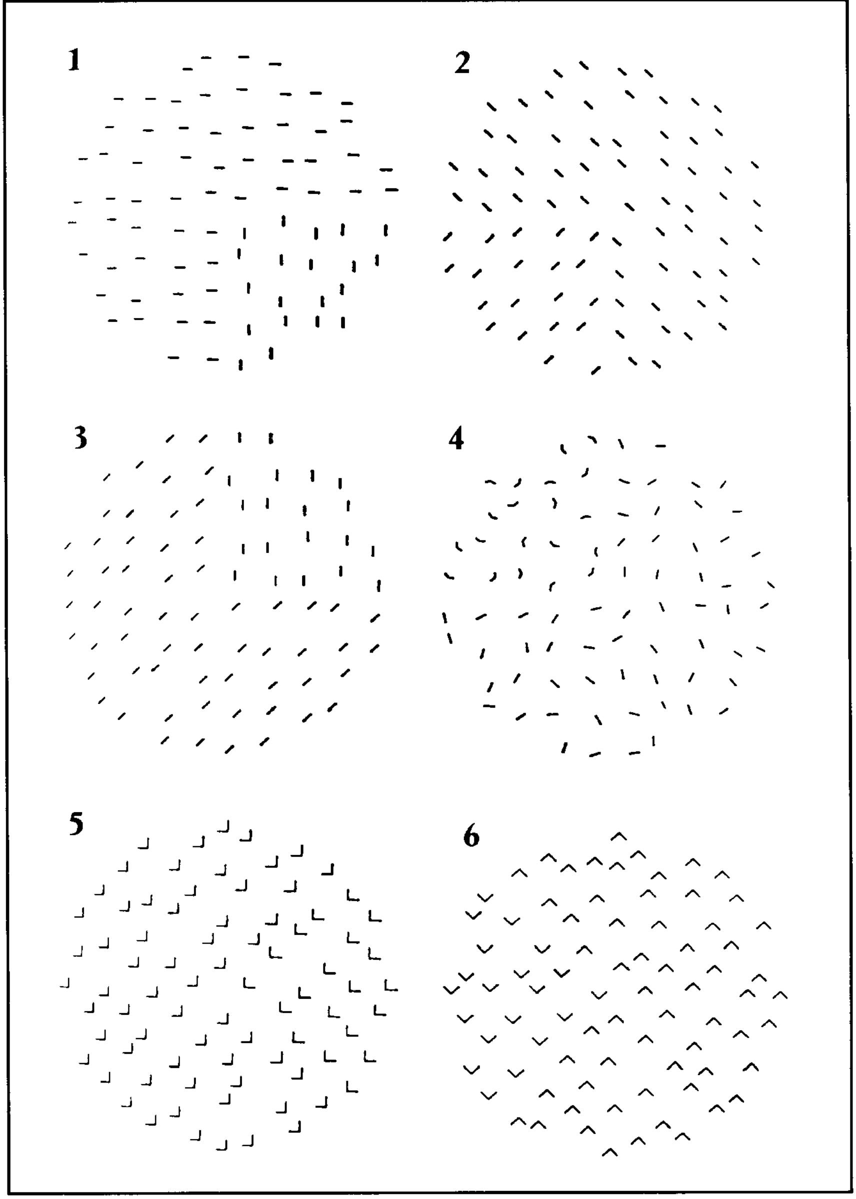

Examples of the forms used by Hochberg and Brooks (1960). In each of rows 1–4, the figure at the right is most likely to be seen as three-dimensional. From Julian E. Hochberg, *Perception*, 2nd edn., p.142. Copyright © 1978. Prentice-Hall Press, New York.

as two-dimensional line drawings increased, so there should be a tendency for the figures to be perceived as though they were three-dimensional objects. They made a number of measurements on the figures and looked for those that correlated well with perceived three-dimensionality. The best measure was the number of angles in the figure. This measure seems to represent "complexity". The more angles the figure contains, the more complex it is in two dimensions, and the more likely it is to be perceived as a representation of a "simpler", three-dimensional object. A second measure that correlated well was the number of differently sized angles. This reflects the asymmetry in the 2-D figure, as a figure in which many of the angles are of the same size is more likely to be symmetrical than one in which many differently sized angles are present. A final measure was the number of continuous lines. This reflects the discontinuity present, as the more continuous lines there are, the more discontinuities must be present between each. Thus the more complex, asymmetrical, and discontinuous the 2-D pattern, the more likely it was to be perceived as the projection of a 3-D figure. Hochberg and Brooks then applied their measures to a set of new figures and found they correlated well with the perceived three-dimensionality in these.

Thus it is possible to express Gestalt ideas such as "good shape" more precisely. In similar vein we now consider recent attempts to tackle the problem of grouping by similarity. How similar must items be before they are grouped together? It is unlikely that they must be identical, because no camouflage can ever perfectly match its surroundings, yet we know that camouflage can be remarkably successful. But if identity is not required, what are the important variables that determine grouping by similarity? This has been investigated by seeing how easily two different regions of a pattern, or more naturally textured image, segregate perceptually from each other. The logic of this is that the more the elements in two different regions cohere with one another, by virtue of the perceptual similarity that exists between them, the less visible will be the boundary between these two regions.

Olson and Attneave (1970) required observers to indicate where the "odd" quadrant lay within a circular display of simple pattern elements (see Fig. 6.16). They found that the quadrant was most easily spotted if the elements within it differed in slope from those of the rest of the display (e.g. < V) and was most difficult to find if the elements differed in configuration, but not in the slopes of their component parts (e.g. > <). Similar conclusions were reached by Beck (1972) who asked his subjects to count elements of one type (e.g. <) that were distributed randomly within a display containing elements of a different type (e.g. >). Again he reasoned that the more the odd elements stood out from the background elements, and grouped with each other rather than with the background, the easier they would be to isolate and count. Like Olson and Attneave, Beck found that slope differences led to faster counting than configurational differences.

FIGURE 6.16




Some of the displays used by Olson and Attneave (1970) to investigate grouping by similarity. In the displays marked 1, 2, and 3, the lines in one region are of a different orientation to the rest, and the odd region is easy to spot. In display 4, odd elements are curved, and the odd region is reasonably evident. In displays 5 and 6, the configurations, but not the slopes, of the elements differ from one region to the next. Here it is much harder to spot the odd quadrant. Reprinted from Olson and Attneave (1970). Copyright © 1970 by the Board of Trustees of the University of Illinois.

Such findings are interesting because they demonstrate that the variables that influence grouping by similarity are not necessarily the same as those that would influence the judged *conceptual* similarity of the same elements viewed individually by humans. Thus L and < might be considered more similar (both letter L, with one example tilted) when viewed as a single pair, than would L and ⌉ . However, when large numbers of these elements are combined, it is the difference in the common orientations of two populations that is perceptually more salient than the difference between the conceptual identity of the individual members. Julesz (1965) stresses that in similarity grouping we are looking at spontaneous, pre-attentive visual processing, which precedes the identification of patterns and objects, and which is quite different from deliberate scrutiny. In some textures it is possible to discern that an odd region is present by carefully examining and comparing individual pattern elements, just as one can find a camouflaged animal by careful inspection. Such processes of scrutiny, however, are at a much higher level than the grouping mechanisms we are discussing here. These "pre-attentive" grouping mechanisms appear to be implemented at a relatively early stage of processing; indeed, DeYoe, Knierem, Sagi, Julesz, and van Essen (1986) have observed cells in areas V1 and V2 of the monkey cortex that respond if texture elements differ in orientation between the centre and surround of their receptive fields.

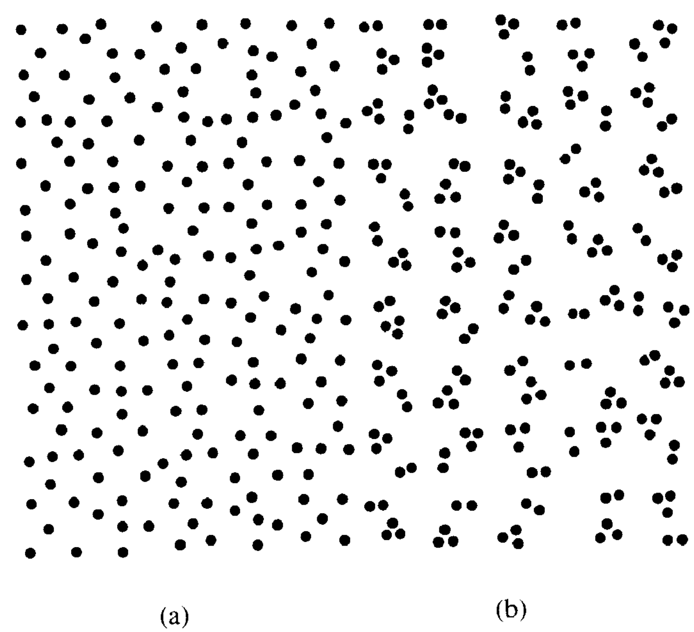
Julesz (1965, 1975) extended the study of grouping by similarity to include more naturally textured images in which brightnesses and colour were varied as well as slope and configuration of elements. First, and most simply, he noted that two regions would be segregated if there was a clear brightness or colour difference between them, and that brightness and colour grouping appeared to operate by "averaging" rather than taking detailed account of statistical differences in the brightness distributions. If two halves of a pattern are constructed so that one half contains mostly black and dark grey squares (with a few light grey and white ones), and the other contains mostly light grey and white squares (with a few dark grey and black ones), then a clear boundary is seen between the two regions. If, however, one region contains mostly black and light grey squares, and the other contains mostly dark grey and white ones, then no clear boundary is perceived even though the composition of the two pictures in respect to the relative frequencies of the different types of square is still quite different. Here it is the average brightnesses in the two halves of the pattern that matters, not the details of the composition of these average brightnesses. Similarly, if a region is composed mostly of red and yellow squares (with a few blue and green ones) and the adjacent area is mostly green and blue (with a few red and yellow squares) then good segregation is achieved. If one region is mostly red and green, and the other mostly yellow and blue, then the segregation is not as clear. Julesz (1965) proposed that the perceptual system imposes a "slicer" mechanism. A region of similar brightness or similar wavelength could be grouped together as distinct from another region where the "average" brightness or wavelength differed from the first.

However, the spatial distribution or "granularity" of different regions is also important. If two regions have the same overall average brightness, but with the pattern elements distributed differently, so that they are spaced apart in one region and clumped together in the other (see Fig. 6.17), a perceptual boundary will be evident. Finally, in other work on region discrimination, Julesz confirmed the findings of Olson and Attneave, and Beck, in demonstrating the importance of differences in slope in perceptual segregation.

Julesz tried to tie together a number of observations on perceptual grouping in terms of the formal statistical properties of the patterns being viewed. He initially made the strong claim that two regions can not be discriminated if their first- and second-order statistics are identical (Frisch & Julesz, 1966; Julesz, Frisch, Gilbert, & Shepp, 1973). Differences in the first-order statistics of patterns capture differences in their overall brightness. Differences in second-order statistics capture differences in granularity and slope.

Julesz's initial attempts to capture the variables determining similarity grouping in formal mathematical terms were thwarted by counter-examples, and this led him to devise his theory of "textons" (e.g. Julesz, 1981). Julesz suggested that texture discrimination depended on whether or not differences could be detected in local *features*, where these features or "textons" were the basic elements of pre-attentive (i.e. early) vision. Textons are elongated blobs or line segments (with associated parameters such as aspect-ratio and orientation) and their terminators, and thus correspond to the representation in the primal sketch as conceived by Marr and others. Julesz (1981) suggested that textures would only be discriminable where there were differences in the first-order statistics of the textons. For example, two regions of texture containing different elements with the same numbers of terminators

FIGURE 6.17




The average brightness in region (a) is the same as that in (b), but the dots in (b) are clumped together more than those in (a). A clear boundary is seen between the two regions.

(see Fig. 6.18) do not segregate; only the number of textons (elongated features and terminations) seems to be important, whereas their spatial arrangement, and plausible candidate features such as closure and connectivity, appear to be unimportant.

> The pre-attentive visual system, evidently, cannot determine the location of terminators, but can count their numbers (or density) or their first-order statistics (Julesz, 1981, p.95).

The work of Julesz and associates emphasises that pattern elements need not be identical in order to be treated together by grouping processes. Such grouping processes seem to operate between elements of similar brightness, wavelength, slope, and granularity. These properties correspond to those that we know are extracted early on in visual processing (see Chapters 2 and 3), and correspond to the properties captured by the descriptions in the primal sketch (Chapter 5). We return to consider Marr's work towards the end of this chapter.

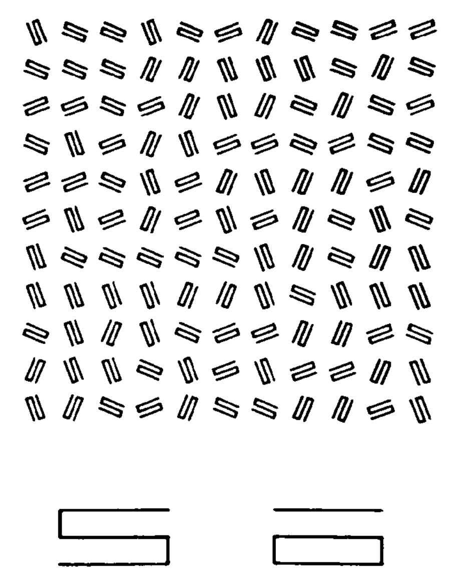
Thus we have seen how recent work in the area of perceptual grouping has quantified the Gestalt principles of "good shape" and "similarity". All the work described earlier has made use of artificial patterns and textures, however. Can these laws of perceptual organisation be demonstrated in more natural settings?

## CONCEALMENT AND ADVERTISEMENT

In this section of the chapter we return to the broader perspective of animal vision, to demonstrate how the Gestalt principles can give some insights into the ways in which the colouration and shapes of animals can help to conceal or to reveal them. The study of camouflage in nature also provides us with a way of exploring the ways in which perceptual grouping processes in other species may be similar to, or differ from, our own.

Animals that remain concealed from predators have a greater chance of surviving and

FIGURE 6.18



At the top is shown a texture made up of two different regions. The larger region is composed of "S-shaped" elements, with a smaller region composed of "10-shaped" elements. The odd region cannot be discriminated without close scrutiny, showing that if the number of lines and number of terminators agree, their exact positions are ignored. From Julesz (1981). Reprinted with permission from *Nature* (Vol.290, pp.91–97). © Macmillan Magazines Limited.

reproducing, and a predator also stands a better chance of obtaining food if it is not easily visible to its own prey. In order to remain concealed an animal should not stand out as a figure against its background, but needs instead to blend with it. On the other hand, for the purposes of breeding or defending territory, animals may need to be conspicuous to potential mates or competitors. Under such circumstances the animal may need to stand out as a distinctive figure against its surroundings. Whether an animal will be camouflaged or conspicuous depends on its behavioural needs and the habitat in which it lives. Some animals may need to remain hidden for much of the time but have the potential of occasionally giving a highly distinctive display or warning sign. This can be achieved by temporarily revealing distinctive surface features that are normally hidden beneath wings or tails, or by changing skin colour, coat, or plumage to meet changing circumstances. Thus, to understand why a creature is coloured in a particular way we need to consider ecological factors as well as perceptual ones.

## Merging and contrasting

An animal that needs to be hidden should avoid standing out as figure against its background. An animal that needs to be seen should stand out as a figure distinctly from its background. The way in which this is achieved will depend on the nature of the background habitat.

An animal can blend with a uniform background by being of similar average colour and brightness. Many species are coloured fairly simply to match the habitat in which they live—for example, tropical tree-snakes are green and polar bears are white. Some animals show seasonal variation in their colouration to match their changing habitats. The Arctic fox is white in winter and brown in summer. In other species, different animals may be coloured differently in differing environments. The peppered moth is darker in urban than in rural areas, for example.

A more radical way to prevent standing out as a figure, and one that is more effective against nonuniform habitats, is to break up the perceptual cohesiveness of the surface of the body by *disruptive colouration*. Dissimilar surface areas are less likely to be grouped together as a single figure. If some of the patches of surface colour are in turn similar to elements in the background this leads the animal's surface to be grouped together with its habitat. Disruptive colouration along with *background picturing* is an efficient method of camouflage where the habitat contains different coloured and shaped elements and variations in light and shade. The tree frog shown in Fig. 6.19 is a good example of a creature whose surface markings incorporate these features. An artificial example is the green, brown, and black mottled pattern painted on tanks and combat jackets to achieve camouflage for military purposes.

In these examples of camouflage we see exploitation of the principle that grouping by similarity (with the background habitat) may

FIGURE 6.19

This African frog is well camouflaged when seen against the bark of the tree on which it habitually rests. Its asymmetrical markings show disruptive colouration along with background picturing. However, the shadow cast by the frog's head on the bark could reveal it. Photograph by P. Ward. Copyright © P. Ward/Bruce Coleman Ltd. Used by permission.

override grouping by proximity (of adjacent parts of the animal's surface). Effective disruptive colouration must match the background in terms of average brightness, colour, and density of markings. Where the background texture elements have some intrinsic orientation, as in Fig. 6.19, surface markings must be similarly oriented.

Conversely, in order to be conspicuous, an animal needs to have high brightness or colour contrast with its background. The all-black crow is a distinctive form against grass-land or stubble field. In addition, the outline of an animal's body or of significant signs or structures can be enhanced by *outlining*. Many butterflies have contrasting borders around the edges of their wings, and some fish may have black edges to emphasise their distinctive fin shapes. Because flat structures are less conspicuous as figures than solid ones (see later), and solid bodies have no single contour line (the particular contour that will correspond to the animal's outline in an image of it will depend on vantage point), it makes sense that only "flat" structures such as butterfly wings or fish fins should be outlined in this way (Hailman, 1977). More frequently we find local outlining to emphasise the shape of a patch that serves as a courtship or warning signal. Such patches are often outlined in white or black in fish and birds.

### Symmetry and regularity

Symmetrical forms stand out more readily as figures against their backgrounds than do asymmetrical ones. As animals are bilaterally symmetric, those that need to be hidden must reduce their apparent symmetry. Camouflaged snakes may rest in irregular coils so that the symmetry of surface markings is not evident. The frog in Fig. 6.19 has asymmetrical markings that make it less visible as separate figure. Conversely, symmetry and repetition may be employed to enhance an animal's outline or internal features. Butterflies with dramatic markings on their wings become highly salient symmetrical figures when their wings are spread. Signal patches can also be made conspicuous by virtue of their shape. Regular geometric forms such as circles, squares, and triangles are perceptually salient figures, and they are also rare in the habitats of animals, so that surface markings that are geometrically regular will be additionally distinctive due to their dissimilarity to background elements. Hailman (1977) suggests that the common use of circular signal patches may be because they are regular, and hence unusual, rather than functioning as eye mimics. Triangular patterns can be seen on the breeding plumage of some male birds such as peacocks, and rectangular patches are found on the wings of some ducks. In many birds and fish we find another kind of regularity in the form of repeated markings—the series of tail spots on cuckoos and head stripes on sparrows, for example.

### Counter-shading and reverse counter-shading

A creature might be well camouflaged in terms of matching the colours and contrasts in its background, but could still be apparent as a distinct figure by virtue of its solidity. There will be a distinct depth difference between the upper part of a cylindrical body and the surface on which it rests, which may be revealed to an observer by stereopsis or motion parallax (see Chapter 7). This will lead to grouping by proximity in depth and by common fate, just as central regions of texture can be revealed in random-dot stereograms (see Chapter 7). Many birds use sideways movements of their heads to reveal prey by motion parallax. A solution to this problem is to be as flat as possible, either behaviourally, by crouching, or structurally, by becoming flat during the course of evolution. Moths and flatfish, by virtue of their flat shapes, are at minimal depth differences from the surfaces on which they rest. Crouching also reduces the likelihood that an animal will be revealed by the shadow it casts. The head of the frog in Fig. 6.19 casts such a shadow on the bark.

Creatures that inhabit environments where there is a strong light source have the additional problem of unequal illumination of their surfaces, leading to self-shading, which again will tend to reveal. This can be compensated for by *counter-shading*. A counter-shaded animal has its darkest surface areas where the most light strikes its body, and is lighter where less light is incident (Thayer, 1918). The zebra's stripes may in part serve a counter-shading function. The black stripes are at their broadest (and hence the coat on average is at its darkest) where the body receives most light.

The clearest examples of counter-shading are found among fish, who often have dark dorsal and light ventral regions. This means that they are relatively concealed from air-borne predators where their darker backs will be seen against the murk of the water, and also concealed when viewed by a predator swimming beneath them, as their light undersides are now seen against the brighter sky above. The counter-shading evident in an animal can usually be explained in terms of the direction of the habitual light source and the shape of its body. Caterpillars that live on the undersides of leaves have counter-shading, with their undersides, which receive the most light, darker than their backs.

If appropriate counter-shading can serve to conceal an animal then *reverse* counter-shading could act to reveal it. There are some animals, who lead their lives in upright posture, who are lighter dorsally and darker ventrally. In the male bobolink (a kind of blackbird) reverse counter-shaded plumage, in which the head and back are white, is adopted for the breeding season, where the bird needs to be distinctive, but discarded during winter when the bird's plumage is darker dorsally (Hailman, 1977).

### Immobility and camouflage

However well concealed a stationary animal may be, grouping by "common fate" would tend to reveal it if it moved. It therefore benefits a camouflaged animal if it can remain still for a large proportion of the time. In many species, the adults may be brightly coloured for courtship or aggressive purposes, whereas the young may have quite different plumage or coats that merge with their backgrounds. Whether the young of a species are camouflaged or not will depend on the habitat in which the nest or den is sited, and on the behaviour of both young and parents. For example, if the nest is on open ground, and both parents leave it to forage, then camouflage of the young is more important.

However, as Cott (1940) pointed out, although immobility is advantageous to concealment, it is not essential. A green tennis ball is harder to follow on a grass court than a white one. Thus, even if an animal is active, it will be harder to spot or track if it merges with its background. Some creatures have markings that appear to make it harder to track their movement. Many snakes that flee in defence (Jackson, Ingram, & Campbell, 1976), and some fish, have longitudinal stripes that may deceive observers because they appear to remain still as the animal moves forward. Of course, movement is one of the easiest ways for an animal to reveal itself when it needs to be conspicuous. Some make use of temporal redundancy by making repetitive or stereotyped movements of their bodies in their displays, and others repeatedly flash signal patches beneath their tails or wings.

## PERCEPTUAL ORGANISATION IN OTHER SPECIES

These examples of animal colouration thus illustrate how the Gestalt laws may be useful to help understand camouflage and concealment principles, at least when assessed by human vision. However, a successful camouflage for a particular species is not necessarily that which prevents its detection by a human visual system. It is the properties of the predator's, the prey's, and the conspecific's visual systems that are important. Some species may not appear well-hidden to us because their colours are different from those of their habitats. However, provided they need to be hidden from colour-blind species, only the brightness levels are important. Conversely, crab spiders, which match the flowers on which they live, may be well concealed to our eyes and to the eyes of many of their predators, but they may be detectable to any insect prey that have good sensitivity to ultraviolet radiation (Eisner, Silberglied, Aneshansley, Carrel, & Howland, 1969). How can we find out whether an animal's colouration is having its apparent (to our eyes) effect of hiding the animal or making it conspicuous?

It is possible to examine how accurate our own perceptual intuitions are about the relative degrees of concealment attained by camouflaged animals by observing the "success" that different surface markings confer to an animal in terms of its survival. This can be done through natural observation or through experiment. For example, a radical change in the predominant colouration of the peppered moth has been observed in areas where industrial pollution is present. At one time, darker members of the peppered moth species were rare. Over the last 200 years or so, in areas where buildings and trees are polluted with soot and grime, the predominant colouration in the moths has changed. Darker members are much more frequent than lighter ones, whereas in rural areas the lighter moths are still common. This suggests that the avian predators that feed on such moths find light moths distinctive on dark backgrounds in the same way that we do. In industrial areas the gene for darker colouration has conferred an advantage on those possessing it, whose chances of surviving and reproducing have therefore been enhanced (Kettlewell, 1973).

As well as such "natural" experiments, it is possible to conduct controlled experiments in which members of a prey species are placed against different backgrounds and then exposed to predators. The success of a particular camouflage can be assessed in terms of the number of prey that survive! Sumner (1934), for example, reared mosquito fish in differently coloured tanks. These are fish that, in common with many others, adjust their colours to tone in with their surroundings. After seven to eight weeks those fish reared in a black tank were very dark, whereas those raised in a white tank were a much paler buff or grey. Equal numbers of the "black" and "white" fish were transferred into experimental tanks that were painted black or pale grey, and exposed to the Galapagos penguin as predator. Of the fish that were appropriately colour-adapted 32% were eaten, as compared with 68% of those that were inappropriately colour-adapted. Thus the Galapagos penguin seems to find fish with high contrast to their background more easily than those with low contrast, again in agreement with our own perceptions. More dramatic experiments can involve artificially colouring the prey species before exposing them to predation (e.g. Croze, 1970).

Experiments such as these are not always ethically acceptable (most people would be unhappy if the prey used in such studies were mammals or birds rather than fish or insects). A further experimental way of assessing degree of concealment without the sacrifice of too many animals is illustrated by Pietrewicz and Kamil (1977). They conducted operant conditioning experiments in which blue jays were trained to detect moths in colour slides. Interestingly they showed that the birds were sensitive to the orientation of the moths (whether their heads were pointing up, down, or horizontally), as well as to the degree of visual similarity existing between the moth's markings and those of the bark against which it was photographed. Only the latter aspect is noticeable to us. Similarly, Dittrich, Gilbert, Green, McGregor, and Grewcock (1993) trained pigeons to respond to photographs of wasps and then tested their transfer to photographs of various hoverfly species with black and yellow striped markings. The pigeons' ranking of hoverflies in terms of their similarity to wasps was different from that made by humans.

Observing the "success" of various surface markings is, of course, not the only way to study perceptual organisation in other species. Hertz (1928, 1929), for example, describes some delightful experiments with jays and bees in which she investigated aspects of their figural perception directly as they searched for food from arrangements of objects. She concluded that for the jay birds (though not for the bees), perceptual organisation was very similar to our own. To appreciate fully the adaptive significance of animal colouration requires continued research along these lines.

## WHY DO THE GESTALT LAWS WORK?

We have shown that many of the Gestalt laws are useful descriptive tools for a discussion of perceptual organisation in the real world, but we are still some way from having an adequate theory of *why* the principles work and *how* perceptual organisation is achieved. We mentioned earlier how the Gestalt psychologists themselves attempted to answer both these points with their model of brain field forces. What alternative answers would contemporary theorists provide?

Marr's (1976, 1982) approach to vision emphasises that we should always consider what general assumptions about the world can be brought to bear on visual processing to constrain the range of possible interpretations for any particular image. The Gestalt principles of organisation may work because they reflect a set of sensible assumptions that can be made about the world of physical and biological objects. Because the same kind of surface reflects and absorbs light in the same kind of way, the different sub-regions of a single object are likely to look similar. Because matter is cohesive, adjacent regions are likely to belong together, and will retain their adjacency relations despite movement of the object. The shapes of natural objects tend to vary smoothly rather than having abrupt discontinuities, and many natural objects (at least those that grow) are symmetrical. A solid object stands on (and hence is at a different depth from) the surface on which it rests, and objects tend to be small compared with the ground. A perceptual system that made use of such assumptions to interpret natural images would generally achieve correct solutions to perceptual organisation, unless deceived by a camouflage exploiting these very same assumptions. It is perhaps not surprising that in our perception of unnatural displays (such as the patterns used by experimental psychologists or the authors of textbooks), we employ the same set of assumptions that serve us well in interpreting natural images.

However, having a set of descriptive principles, even if we know why they work, is still only a starting point for a full information-processing theory of grouping processes. We need to know how such principles can be applied to primitive elements recovered from images—edges, blobs, and so on—in order to recover the potentially significant structures present. It is research in artificial intelligence (AI), which has attempted to provide such a *process* theory of perceptual organisation, that is much more powerful than a purely *descriptive* theory, such as that of the Gestaltists or more recent workers like Julesz. Marr's (1976) early visual processing program implemented such a process theory and made extensive use of Gestalt principles to achieve perceptual organisation. Before describing Marr's work, we digress briefly to introduce other research in AI that has attempted to formalise organisational processes by making use of a rather different set of constraints.

## ARTIFICIAL INTELLIGENCE APPROACHES TO GROUPING

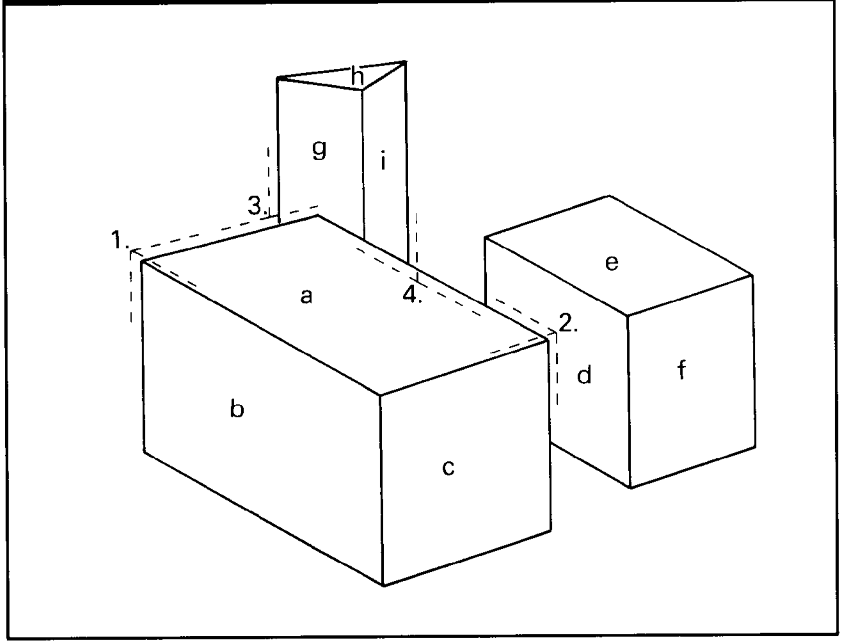
### Scene analysis programs

Many researchers in AI during the 1960s and 1970s attempted to solve what became known as the "segmentation problem". This is the problem of dividing a visual scene into a number of distinct *objects*. Most researchers avoided the complexities of natural images, and restricted their programs to a world of matt, white prismatic solids that were evenly illuminated. Figure 6.20 shows a line drawing of the outlines of such a collection of objects. Viewed analytically, Fig. 6.20 is just a collection of straight lines in a variety of orientations. However, our spontaneous perception of such a scene is more likely to be of a collection of distinct objects. Thus this scene is readily described by our visual apparatus as being made up of two blocks and a wedge, with one block partially occluding the other two structures. Here again we have an example of perceptual organisation. Somehow we know that the regions labelled a, b, and c belong together as one structure, distinct from d, e, and f, which belong to another. The Gestalt psychologists might argue that the perception of regions a, b, and c as belonging to a "cube" provides a closed, simple, and symmetrical interpretation, but this does not really address the question of how such a solution is achieved by visual processing.

This was the kind of problem tackled by Guzman (1968), Clowes (1971), and Waltz (1975), among others, who set out to produce computer programs that could "see" objects from collections of lines such as these. The common principle in all their work was a consideration of the *junctions* present in these figures. A junction is a point where two or more lines meet. Different junction types have different implications for the possible arrangement of surfaces within the picture. Thus Guzman suggested that the presence of an *arrow* junction would generally imply that the edges that formed the fins of the arrow belonged to a single body, whereas a T junction generally implied that the shaft and the cross-bar of the T belonged to different bodies. Figure 6.20 shows how these principles apply in our example.

Guzman's program SEE considered only junctions, and incorporated his own informal

**FIGURE 6.20**



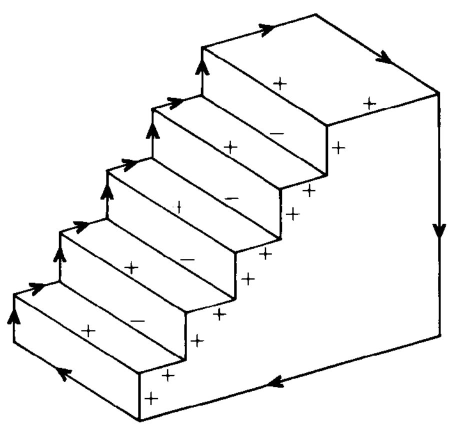

We have no difficulty in seeing that regions a, b, and c belong together (likewise d, e, and f; g, h, and i). Guzman's program SEE interpreted pictures like this by examining the junctions present. Examples of arrow junctions are shown at 1 and 2, and T junctions are shown at 3 and 4. What other kinds of junction are there in this picture?

intuitions about the interpretations of different junction types. Clowes tackled the problem of junction specification more systematically, and employed a more sophisticated notion of how different junction types in the image relate to the organisation of *objects* in the "scene" depicted. This was achieved by considering the nature of the *edges* depicted by the junction lines (Clowes, 1971; Huffman, 1971), as well as the nature of the intersection of these lines. Edges may be convex, concave, or occluding (see Fig 6.21). Only certain combinations of edge types are compatible with a particular configuration of lines at a junction. By ensuring that edges were consistently labelled along their entire length, Clowes' program OBSCENE was able to interpret pictures successfully provided that no more than three lines met at a single junction. The program was also able to "reject" certain pictures as "impossible" (see Fig. 6.22, for example), whereas SEE would simply accept such examples as objects.

The most elegant example of work of this type was that of Waltz (1975), who introduced a fourth edge type, the crack, and whose program accepted pictures of scenes containing shadows. Once shadows are introduced, the possible labellings for a particular type of junction increase dramatically, as a number of different types of edge can now be present. Nevertheless, Waltz's program was able successfully to parse scenes containing shadows. His work illustrates how adding more information in the form of light and shading may actually aid the interpretation of a scene by providing additional local constraints.

Although such A.I. programs are intrinsically interesting, and point out the complicated processing that may underlie our everyday ability to perceive patterns such as these, they are of limited importance. The programs work by incorporating the constraints of their visual worlds, but the particular constraints employed are specific to the world of white prismatic solids—an artificially manufactured world for whose perception our visual systems did not evolve. The principles embodied within these segmentation programs would fail to recover the significant structures in natural images. Natural objects may have internal markings, texture, and shading (see Ch.5, p.76). Straight lines and angular junctions are rare. Indeed, A.I. segmentation programs of the type described either start with a line drawing as input, or make use of initial programs to find the edges in images of prismatic solids by using the assumption that edges are straight, along with higher-level knowledge about "likely" places to find lines (e.g. Shirai, 1973).

FIGURE 6.21

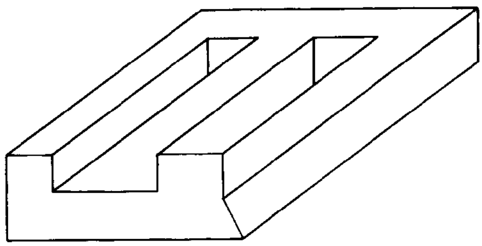



Three different kinds of edge are shown: concave (–), convex (+), and occluding (>).

## Marr's program

Of more interest to our discussion is a processing model that aims to recover structures from natural images of everyday objects and surfaces, despite

FIGURE 6.22

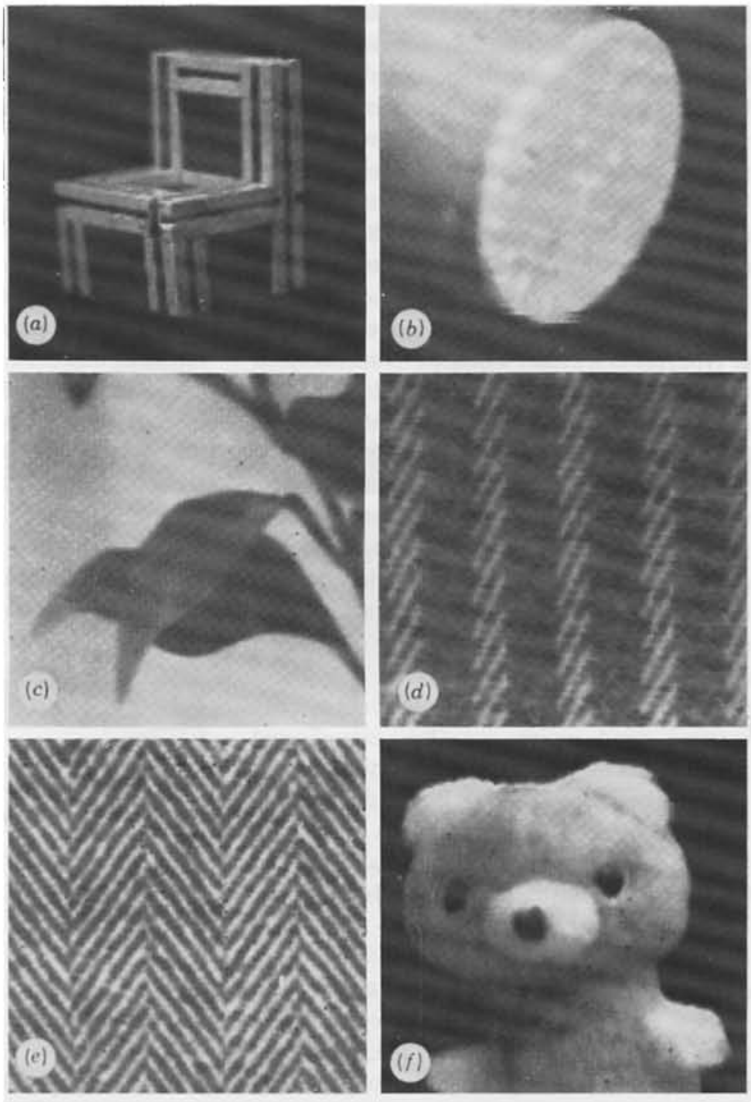


An impossible object. Reprinted from Clowes (1971) with kind permission of Elsevier Science, The Netherlands.

their noise, texture, and shadow. Marr's (1976, 1982) early visual processing program finds occluding and internal contours from images such as those shown in Fig. 6.23. We have already considered some of Marr's ideas in Chapters 4 and 5. He proposed that cells in the retina and visual cortex of mammals function to locate *zero-crossings* (see pp.78–82) in the spatially filtered retinal image, which serve as the first step towards recovering information about edges in the world. A comparison of the zero-crossings found by sets of cells with different receptive field sizes leads to a set of assertions about the "features" present at each location in the image. This set of assertions is the raw primal sketch.

The primitives in the raw primal sketch are edges, bars, blobs, and terminations, which have the associated attributes of orientation, contrast,

FIGURE 6.23




Examples of the images analysed by Marr's (1976) early vision program: (a) a chair; (b) a rod; (c) a plant; (d) and (e) textures; and (f) a teddy bear. Reproduced from Marr (1976) with permission of The Royal Society.

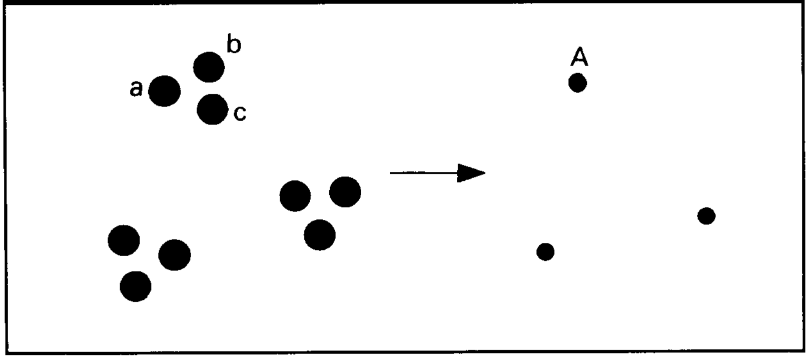
length, width, and position. The representation of a straight line would consist of a termination, then several segments having the same orientation, then a final termination. The raw primal sketch is a very complex, messy affair (see Fig. 5.6), from which we need to recover global structures as well as internal structures and surface texture.

Marr proposed that this is achieved in the next stage of early visual processing by the recursive assignment of *place tokens* to small structures, or aggregations of structures, in the raw primal sketch. These place tokens are in turn aggregated together to form larger units, in a cyclical manner. Place tokens can be defined by the *position* of a blob, or of a short line or edge; by the *termination* of a longer edge, line, or elongated blob; or by a small *aggregation* of tokens. Aggregation of these place tokens can proceed by clustering nearby place tokens on the basis of changes in spatial density (see Fig. 6.24), by curvilinear aggregation, which produces contours by joining aligned items that are near to one another (see Fig. 6.25), and finally by theta aggregation. Theta aggregation involves the grouping of similarly oriented items in a direction that relies on, but differs from, their intrinsic orientation. Theta aggregation can, for example, be used to recover the vertical stripes in a herring-bone pattern where all the individual texture elements are oriented obliquely (see Fig. 6.26).

The grouping together of place tokens thus relies on local proximity (adjacent elements are combined) and similarity (similarly oriented elements are combined), but more global considerations can also influence the structures detected. For example, in curvilinear aggregation, a "closure" principle could allow two edge segments to be joined even though the contrast across the edge segments differed due to illumination effects (see the image in Fig. 6.27). Marr's theory therefore embodies many of the Gestalt principles that we discussed at length earlier.

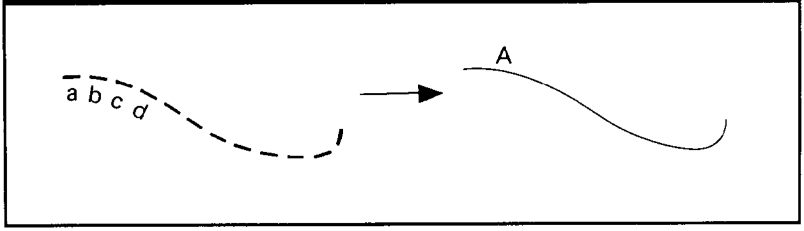
The grouping procedures use the construction of tokens at different scales to locate physically meaningful boundaries in the image. It is essential that different scales are used in order to recover different kinds of surface properties.

FIGURE 6.24



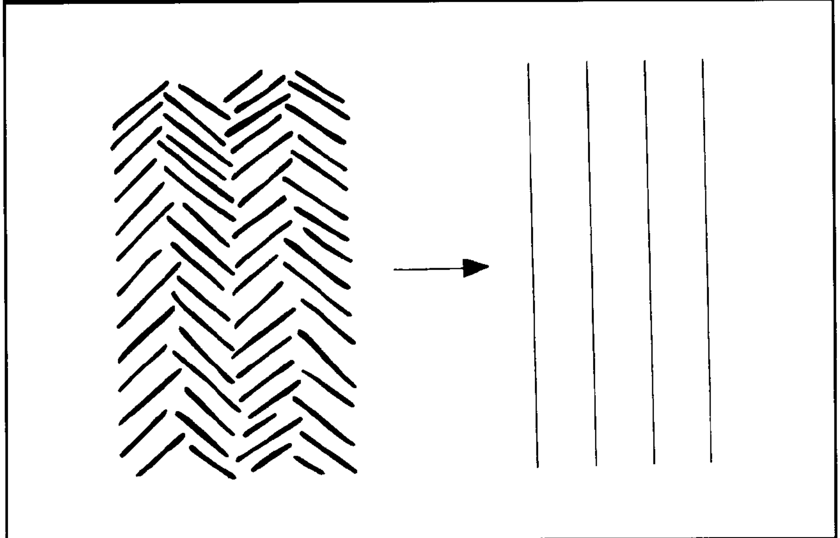

Place tokens corresponding to small dots can be grouped together by proximity to yield higher-order place tokens. Here, place tokens at a, b, and c are grouped to yield a place token at A, and likewise for the other dots in this figure.

FIGURE 6.25




Curvilinear aggregation will group place tokens at a, b, c, d, and so on to yield a single structure A.

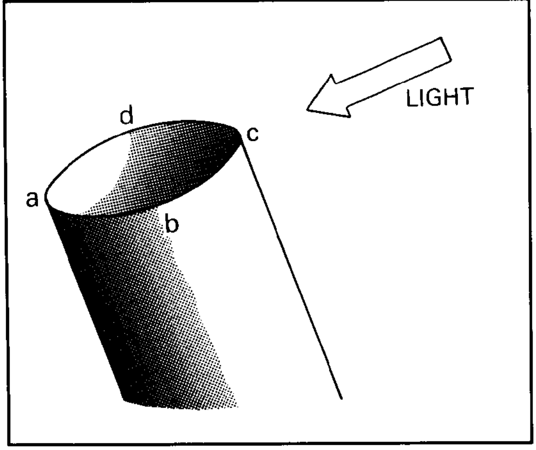
FIGURE 6.26



Theta aggregation can recover the vertical orientation of the stripes of a herring-bone pattern.

Thus if the image was a close-up view of a cat, the raw primal sketch might yield descriptions mostly at the scale of the cat's hairs. At the next level the markings on its coat may appear ... and at a yet higher level there is the parallel stripe structure of these markings (Marr 1982, p. 91).

FIGURE 6.27



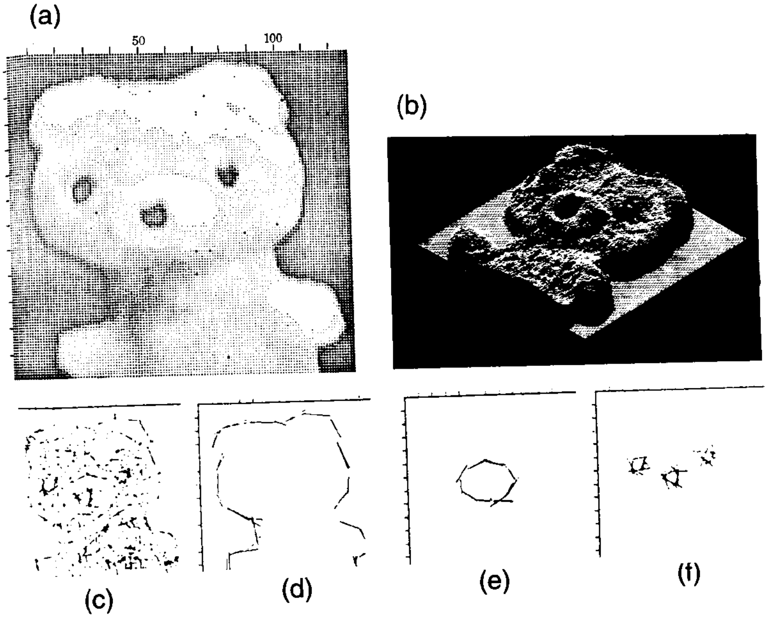

Curvilinear aggregation along with the application of a closure principle could reveal the contour a-b-c-d, despite the different contrasts of the edge segments along this contour. This pattern of shading might arise if a tube was illuminated in the direction shown.

In a herring-bone pattern we know that the "bones" are short parallel segments oriented at 45°, and that, at a larger scale, these form vertical stripes (Fig. 6.26).

Boundaries arising from changes in surface material (where two different objects overlap, for example), or from abrupt changes in surface orientation or depth, can be revealed in two ways. First, boundaries may simply be marked by place tokens. The elliptical boundary perceived in Fig. 5.19C, for example, could be produced by the curvilinear aggregation of the place tokens assigned to the termination of each line. We saw earlier (Fig. 5.20) how complex cells in V2 may be organised to link such line-ends to implement the kind of aggregation process that Marr had in mind. Second, texture boundaries may be revealed by discontinuities in parameters that describe the spatial organisation of an image. Changes in the local density of place tokens, their spacing, or their overall orientation structure could all be used to reveal such boundaries. Although the boundary in Fig. 6.29 is not defined by the spacing of place tokens, it is revealed by discontinuity in the spatial distribution of orientations of the small elements in the image.

This last example also illustrates how Marr's theory can be applied to the problem of texture and region discrimination tackled by Julesz (see

FIGURE 6.28

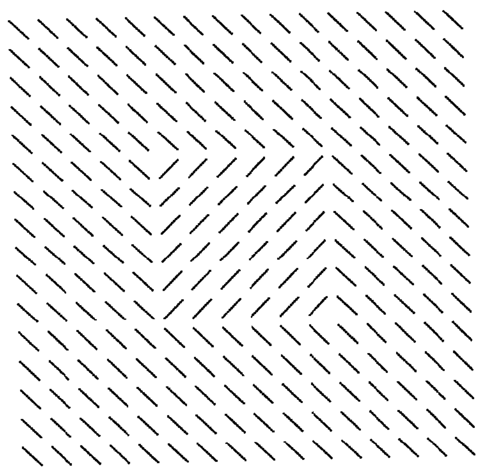



The image of a teddy bear (Fig 6.23f) is printed at (a), and shown as an intensity map at (b). The location of all the small edge segments in the raw primal sketch is shown at (c). The structures that emerge after grouping operations are shown at (d), (e), and (f). Reproduced from Marr (1976) with permission of The Royal Society.

pp.113–114). Julesz tried to arrive at a universal mathematical formula to explain why some texture boundaries were perceptually evident, although others were invisible without scrutiny, but Marr provided a process theory that offered a more powerful explanation. Julesz's explanation was purely descriptive; Marr showed how a set of descriptive principles could be used to recover texture and larger structures from images.

The success of these organising principles in Marr's (1976, 1982) early visual processing program can be judged by its ability to recover the occluding contours from realistic images such as the teddy bear (see Fig. 6.23f), and to reveal the internal contours of the bear, which correspond to eyes, nose, and muzzle outlines (Fig. 6.28). Such structures are recovered without recourse to high-level knowledge. The program knows nothing of the usual shape of a teddy bear's head, and does not find the contours that correspond to its eyes because it "expects" to find them. Marr's

FIGURE 6.29



Example of texture segmentation, defined by "orientation contrast". The centre and background regions have the same mean luminance but differ in element orientation.

theory of early visual processing thus contrasts strongly with some computer models, or more general theories of visual perception where expectations and "object-hypotheses" guide every stage of perceptual analysis (e.g. Gregory, 1973; Roberts, 1965). The processing of natural images by Marr's program works because the program embodies grouping principles that reflect general properties of the world. Things that are oriented similarly, or lie next to each other, are more likely to "belong together" than things that are oriented dissimilarly and spaced far apart.

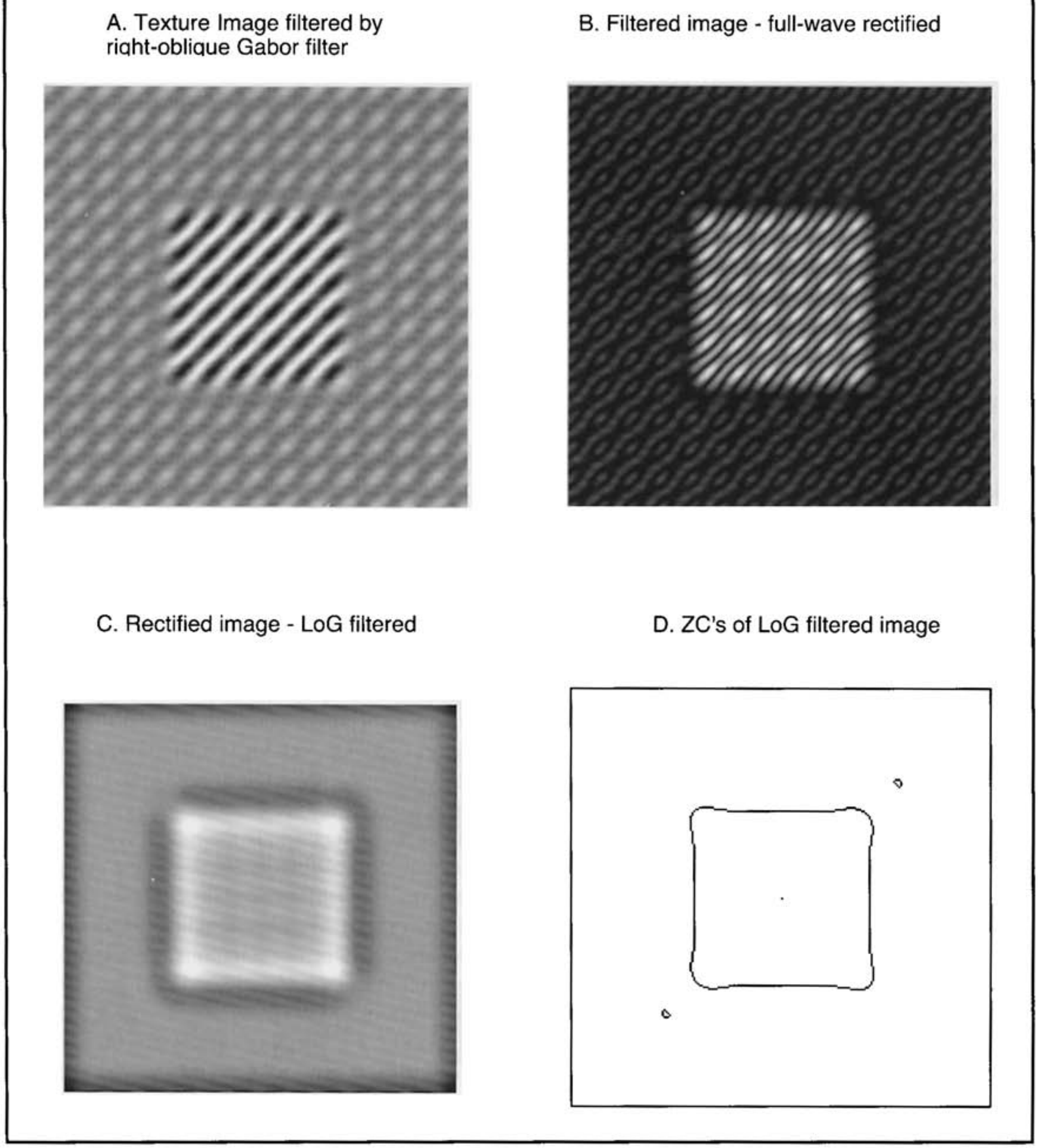
## ENERGY MODELS FOR TEXTURE SEGMENTATION

In Marr's scheme, local oriented features are found first, then subjected to a variety of grouping operations, described earlier. Recent thinking about texture segmentation in the context of spatial filtering has suggested, on the other hand, that boundaries between regions differing in texture may be detected rather more easily and directly than previously suspected. Consider the image of Fig. 6.29 in which the centre square and background regions are defined only by a difference in element orientation. A spatial filter tuned to right-oblique orientation would respond strongly to the centre region, but rather little to the background, as illustrated in Fig. 6.30A. At this point, the texture border has been transformed into a boundary between low- and high-contrast areas in the filter's response pattern. How might this boundary be found? A simple, but important idea is to do further processing that converts this contrast difference into an intensity difference, and then to find that intensity edge by standard methods (cf. Chapter 5).

To convert the contrast difference into an intensity difference requires a *nonlinear* operation, and there are several, roughly equivalent, possibilities that the modeller, or evolution, could select. To get the flavour of this, let us suppose that the filter response values of Fig. 6.30A were either +10 or −10 (arbitrary units) in the centre region and +1 or −1 in the background. The mean response is 0 in each region, and so local averaging would lose the border altogether. *Full-wave rectification*, however, will do what we want. By definition, a full-wave rectifier sets negative values to positive. All values in the centre region would then become +10, and the surround values would be +1. Contrast difference has been converted to intensity difference as required. *Half-wave rectification* sets negative values to zero, leaving positive values unchanged. The centre responses would then become 0 or 10 (mean 5), while the background values were 0 or 1 (mean 0.5). Again a mean intensity difference emerges, but accompanied by greater variability of local values. A third option is *squaring* of the filter responses. As in the full-wave case, all values go positive, but in addition high values are emphasised relative to low ones. Centre responses would go to +100, and surround values to +1. These nonlinear, rectifying operations are both simple and physiologically plausible. Figure 6.30B shows an example of the effect of full-wave rectification on the pattern of filter responses. A mean intensity difference is clearly achieved, but with some residual variation at a fine scale. An appropriate degree of spatial smoothing in the edge-detection operator allows this variation to be ironed out (Fig. 6.30C) before the boundary is located (Fig. 6.30D).

In summary then, the foundation for several recent *energy models* of texture segmentation (Bergen & Adelson, 1988; Bergen & Landy, 1991; Bovik, Clark, & Geisler, 1990; Malik & Perona, 1990) is a chain of simple ideas: (i) regions that are differently textured necessarily differ in their spatial structure; (ii) if we apply a bank of spatial filters to the image, at different orientations and different scales, then at least some of those filters will be more strongly activated by one texture than the other; (iii) the difference in activation can be converted into a simple intensity difference in response energy across space; (iv) this energy difference is greatest at the texture boundary and can be found by procedures already familiar from the detection of luminance edges. In short, spatial filtering (step ii) and a suitable nonlinearity (step iii) can convert the problem of texture segmentation into a much simpler problem of edge detection.

FIGURE 6.30



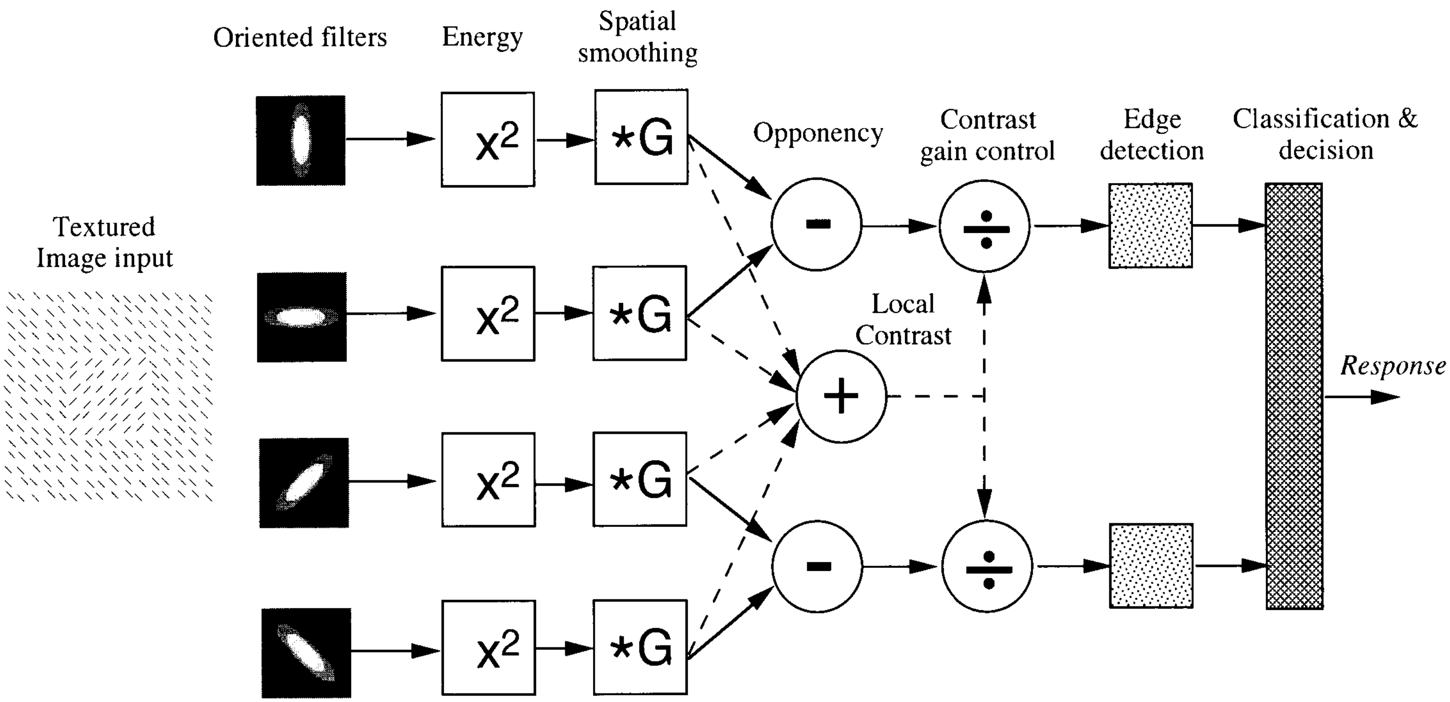

Worked example of the energy approach to finding texture boundaries. The input image is a texture with the centre region defined by orientation contrast (Fig. 6.29). (A) Output image from a small, right-oblique filter. Note large response to right-oblique lines, as expected. In A and C, large positive and negative responses are represented by light and dark, respectively; zero response is mid-grey. (B) Filter output (as A) after full-wave rectification, i.e. negative values in A set to positive. This converts the difference in contrast to a difference in mean intensity of response. Zero response is black in this plot. (C) Rectified image (as B) after a second stage of filtering by a smoothing, second-derivative filter (LoG; see Figs. 5.3, 5.4). This image should have zero-crossings at the texture edges, and this is confirmed by the zero-crossing contour shown in D.

How well does this approach fit with human perception of textured regions? Nothdurft (1985) found that people could recognise the shape of texture regions such as that in Fig. 6.29 especially well when the line elements were densely packed and relatively long. Short lines, sparsely spaced, yielded no visible shapes at all. Shape discrimination also improved with the difference in orientation between lines in the "figure" and lines in the background, being fairly poor for a 10° difference and much better for differences of 30° or 90°. All of these factors would be expected to enhance the differential activation of filters between the two regions, and so the energy models are broadly consistent with Nothdurft's results. Furthermore, we can ask whether the *edge* is vital in segmentation or whether we need consider only the general difference in activation between the regions. Nothdurft (1985) offers a demonstration suggesting that the edge is crucial, and specifically that *orientation contrast* at the boundary is the important factor. When element orientation changed abruptly at the region boundary, segmentation was very clear, but when the same elements were redistributed within the centre and background regions, segmentation disappeared. When the variability of orientations in the background increased, the orientation change at the boundary also had to be increased to maintain segregation (Nothdurft, 1991). The importance of feature contrast at boundaries also seems to hold for segmentation based on changes in motion direction and colour (Nothdurft, 1993).

These principles for finding texture boundaries were applied in a full, multi-channel model by Malik and Perona (1990). Their model used both circular and oriented filters at a range of spatial scales, in order to capture the variety of textural differences that might occur in images. Interestingly, they argued in favour of half-wave rectification rather than full-wave or squaring, and for the use of only even-symmetric filters, not odd-symmetric; see their paper for details. The model also incorporated some inhibitory interactions between filter responses to accentuate the larger values, followed by smoothing and edge detection, as described earlier. The model performed well when tested against a variety of texture pairs for which good psychophysical data exist. The rank order of the model's predictions matched the rank order of human performance. An important test case is where adjacent regions are formed from randomly oriented "+" and "L" elements. The two regions have the same mean luminance, and the line segments have the same length and the same (random) distribution of orientations. As segmentation is very clear for this texture pair, Julesz (e.g. 1984) argued that line-intersections (in the "+" elements) acted as one of the primitive atoms ("textons") of texture vision. The success of filter-based energy models on this and related tasks clearly challenges the need for texture vision to make features or textons explicit (Bergen & Adelson,1988; Malik & Perona, 1990).

Bergen and Landy (1991) developed a model with a similar flavour to the Malik and Perona model, specifically aimed at segmentation by orientation differences. We sketch their model in Fig. 6.31, because it incorporates several additional mechanisms of wider interest in early vision. Its first stages—squaring and smoothing of oriented filter outputs—were discussed earlier. The next stage is *opponency*, taking the difference between energies in channels tuned to orthogonal orientations (horizontal-vertical, H-V; left-right oblique, L-R). Opponency is a well-known feature of colour vision (see Ch.2, p.35), and probably motion analysis too (Adelson & Bergen, 1985; see Chapter 8). Its function in the model is to improve the orientation-coding properties of the system by enhancing the difference in energies between regions. There is physiological evidence for opponency between orientations, as V1 cells can be inhibited by stimuli oriented at right angles to their preferred one (Morrone, Burr, & Maffei, 1982) and corresponding effects are observed in the electrical response of the human visual cortex (Morrone & Burr, 1986). This supports the proposed orientation opponency in Bergen and Landy's model. At the contrast-gain control stage (see Bonds, 1991; Heeger, 1992a), the outputs of the opponent mechanisms, (H-V) and (L-R), are each divided by the sum (H + V + L + R) of the response energies from the four oriented channels. This normalises the response values, making them independent of the overall luminance contrast of

**FIGURE 6.31**




Multi-stage energy model for finding texture boundaries defined by a change in local orientation. Adapted from Bergen and Landy (1991). See text for details.

the stimulus pattern. It also suppresses weak responses in one channel when responses are strong in the other channel. The full model was able to give a good account of psychophysical data on shape discrimination for textured regions defined by orientation difference (Bergen & Landy, 1991). We shall see many of its features reappear in our discussion of motion analysis (Chapter 8).

In summary, spatial contrast in the magnitude or "energy" of filter responses may be adequate to account for much of texture segregation. The spatial selectivity of the filters achieves a grouping by similarity of orientation or size, without having to represent the texture elements individually. This makes good sense for natural surfaces such as bark (Fig. 6.19) or skin, which have readily identifiable texture, but where it is not clear that individual "elements" even exist. Another Gestalt "law"—proximity—can be seen in the same light, as more closely packed elements increase the intensity and uniformity of the filter outputs after the nonlinear stage. For example, in the classic "rows and columns" demonstration (Fig. 6.8), closer packing in the columns will activate vertical filters more strongly than horizontal ones, and vice-versa for rows. When vertical and horizontal filters are equally activated (Fig. 6.8c), the organisation is ambiguous. The grouping and linking of explicitly derived local features—in the Marr and Julesz traditions—is therefore not the only basis for more global levels of perceptual organisation.

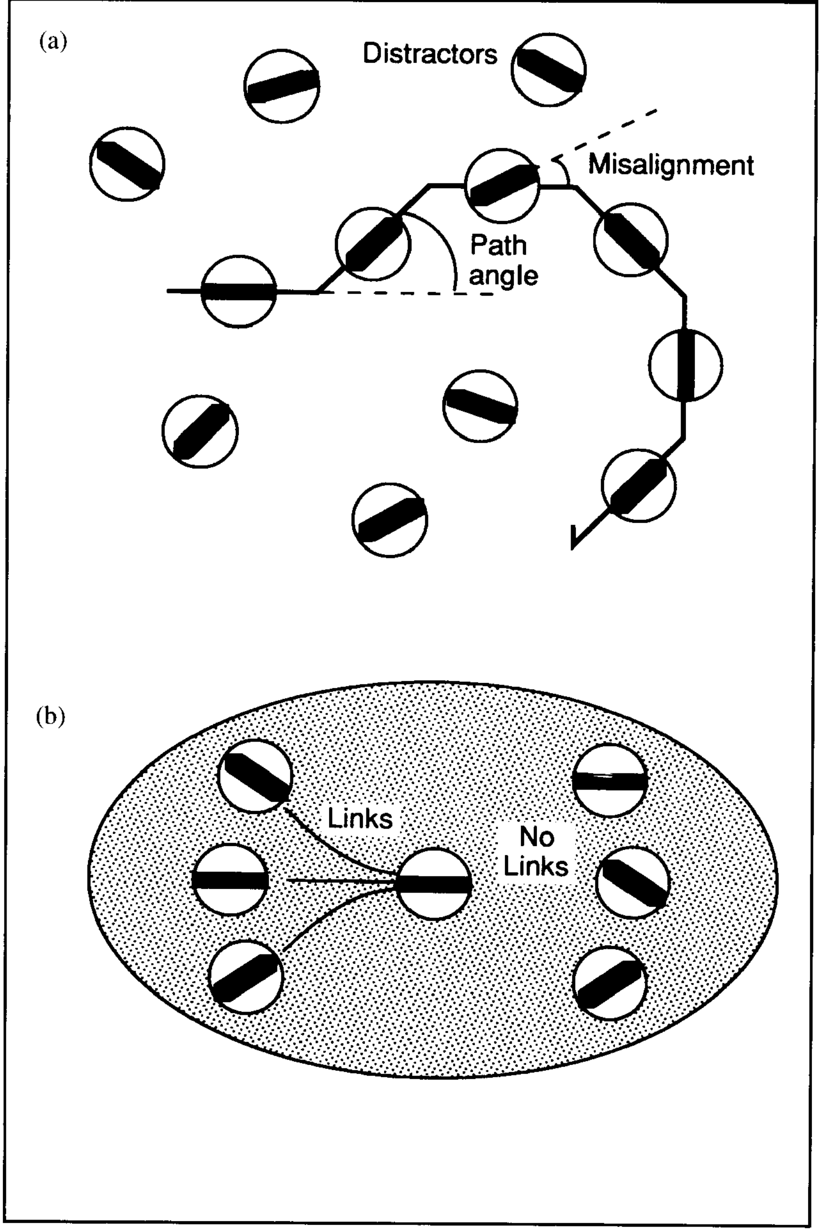
## BEYOND FILTERS: TO CONTOURS AND SURFACES

It would be wrong to conclude from these texture studies, however, that fairly simple post-processing of filter outputs provides a complete account of perceptual organisation. The energy models are relevant to the effects of proximity and similarity, but give little information about structural factors such as symmetry, closure, or "good form". We round off our discussion with several lines of recent experimental evidence pointing to further organising processes: first, that the Gestaltists' "good continuation", or equivalently Marr's notion of "curvilinear aggregation", is implemented by cooperative processes across the image; second, that elongated contours are segmented into parts that correspond to the parts of objects; and finally, that textural grouping and segmentation may depend on the representation of surfaces in depth, and not just on image-processing operations of the kind sketched in Fig. 6.31.

### Perceptual linking along straight and curved paths

Simple demonstrations of "good continuation", such as Figs. 6.10 and 6.11, are amenable to a spatial filter approach, because oriented receptive fields that spanned several elements could automatically pick up the dominant orientation along the curve, and at the intersection of two curves. The experiments of Field, Hayes, and Hess (1993), however, cannot be explained so easily, and point to a more active linking process. Their observers were presented with large arrays of 256 small, randomly scattered striped patches (Gabor patches), and within this large set of randomly oriented distractors a sub-set of just 12 patches was arranged along an undulating, snake-like path. The observer's task was to determine which of two such arrays contained the "snake", as a function of various parameters such as the "wiggliness" of the snake (path angle), and the degree to which the path elements were aligned or misaligned along the direction of the path (Fig. 6.32a). Straight paths, with elements aligned along the path, were most easily detected. Much more surprising was the finding that very jagged, undulating paths, which changed direction by up to 40–60° with every step, were still highly detectable, provided the elements were *aligned* along the path. As the path contains elements in all orientations, it cannot be distinguished from the distractors by applying any single oriented filter. Instead it was the alignment of orientations along the curve that was important. Misalignment by +/–15° reduced performance, and misalignment by +/–30° made the path almost undetectable. In contrast, alignment or misalignment of spatial phase had no effect. Field et al. (1993) interpret these findings in terms of an "association field", suggesting "a localised linking

FIGURE 6.32




(a) Display used for research on "good continuation". Subjects had to detect the presence of a continuous path of elements embedded in a field of randomly oriented distractors. The path changed direction by a given angle from step to step, and elements might be misaligned from the local direction of the path. Lines joining the path elements were not present.
(b) Association field proposed to account for the finding that alignment of elements along the path was the crucial factor in detecting its presence. Adjacent elements would be "linked" if a smooth curve could be drawn between them. After Field, Hayes, and Hess (1993).

process or association between the responses to the elements in the path according to a specific set of rules" (p.185). Their scheme for this linking process is shown in Fig. 6.32b, indicating that links will be made between two adjacent stimulus elements if both their orientation and position are such that they would lie on a simple smooth curve passing through both elements. This is more specific than the idea that elements are grouped if they are close and/or similarly oriented. The association field could serve to build a chain of linked elements representing a continuous contour, or a "flow" of texture in the image (Field et al., 1993).

Further evidence for spatially extended linking of local responses comes from experiments of Polat and Sagi (1993, 1994). Their experiments tested a very basic visual task—the forced-choice detection of very low-contrast patches of sinusoidal grating—with and without the presence of two adjacent "masking" patches of higher contrast. One might guess that contextual and grouping effects would be minimal in this kind of task, but it was not so. When the masking patches were superimposed on the test patch they made it harder to see. This is conventional masking, discussed in Chapter 5 (p.85). However, when the masking patches were adjacent to the test patch they tended to improve its detection; contrast thresholds were lower than in the baseline condition without a mask. This is *facilitation*, and the forced-choice method ensures that it is a genuine improvement in visual performance, and not an artefact of guessing or reporting bias. Facilitation was greatest when the masking patches had the same spatial frequency and orientation as the test (Polat & Sagi, 1993), and were co-axially aligned with it (Polat & Sagi, 1994). The spatial range of facilitation was extensive: it remained quite strong even when the masks were displaced from the test patch by as much as six spatial periods of the test grating. Unlike the "snake" detection experiments discussed earlier, facilitation did not seem to extend along smoothly curved paths, but was greatest for collinear alignment (Polat & Sagi, 1994). Even so, both types of study suggest that "local" spatial filters in vision actually interact across fairly long distances, in ways that serve the detection and representation of elongated contours. Facilitatory interactions across space have been found both psychophysically (e.g. Cannon & Fullenkamp, 1993) and physiologically (Nelson & Frost, 1985). Anatomical studies have revealed long-range connections across the visual cortex, linking cortical columns with a common orientation preference that may mediate these interactions (Gilbert, 1995), and the process of linking may well be associated with the selective synchronisation of cells across the cortex, discussed in Chapter 3 (pp.63–64).

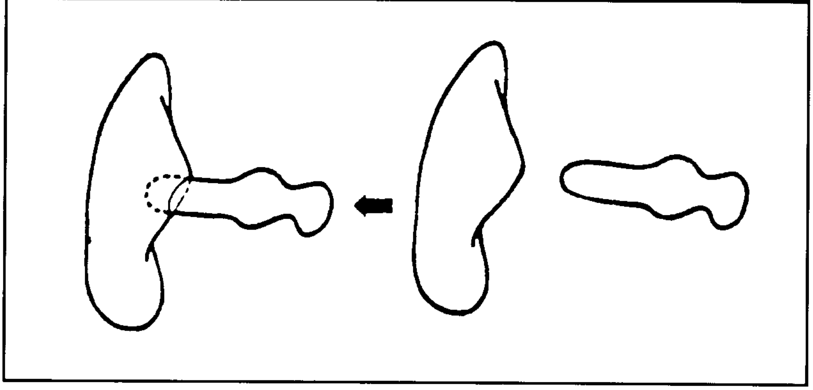
## Segmenting parts for recognition

Marr's (1976) program showed how an object's occluding contour and internal markings could be assembled from the collection of more primitive descriptions comprising the raw primal sketch, on the basis of assumptions that generally hold true in the world of natural objects. More recently, other researchers have shown how similar general assumptions can be used to segment a complex occluding contour into different "part" components—a problem reminiscent of that originally tackled by Guzman and others with artificial objects.

For example, Hoffman and Richards (1984) provided a formal computational analysis of the role played by concavities in contour segmentation. They discussed *the transversality regularity*: distinct parts of objects intersect in a contour of concave discontinuity of their tangent planes. At any point around this intersection, a tangent to the surface of one part forms a concave cusp with the tangent to the surface of the other part (concave means it points into the object rather than into the background; see Fig. 6.33). This transversality regularity means that in an image of a complex shape, "concavities" mark the divisions between the contours of distinct parts. Concavities can be detected in contours of smooth shapes by seeking places where there is greatest negative curvature (see Fig. 6.35).

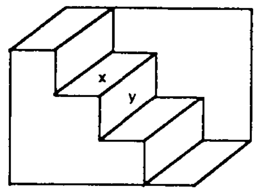
Hoffman and Richards provided some compelling demonstrations as evidence for the importance of these concavities in our segmentation of shapes. They examined a number of classic ambiguous "reversing" figures, such as the Schröder staircase (Fig. 6.34) and the faces–vase figure (Fig. 6.2), and showed how reversals of these figures are related to the possible alternative part segmentations. In the Schröder staircase (Fig. 6.34), for example, according to the partitioning scheme, "parts" of the figure must be "steps", as each of the steps is bounded by two lines of concave discontinuity. When the staircase is seen in such a way that the plane marked "x" in Fig. 6.34 appears to face upwards, then the steps that are defined by concave discontinuities pointing into the staircase are such that planes "x" and "y" are seen to belong together as faces of the same

FIGURE 6.33



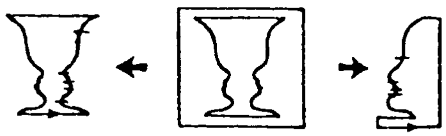
The "transversality regularity": When two surfaces interpenetrate they always meet in concave discontinuities. Reprinted from Hoffman and Richards (1984) with kind permission of Elsevier Science, The Netherlands.

FIGURE 6.34




The Schröder staircase shows how part boundaries change as the figure and ground reverse. Reprinted from Hoffman and Richards (1984) with kind permission of Elsevier Science, The Netherlands.

FIGURE 6.35



The faces–vase figure. When the vase region is taken as figure, then the concavities (minima of curvature) divide the vase into a base, stem, etc. When the faces regions are taken as figure, the concavities reveal parts corresponding to forehead, nose, etc. Reprinted from Hoffman and Richards (1984) with kind permission of Elsevier Science, The Netherlands.

step. But when the figure reverses, so that the staircase now lies in the upper right of the picture, and plane "x" appears to face downwards, the concavities pointing into the body of the staircase define a different set of steps; planes "x" and "y" now form faces of different, adjacent steps. In the faces–vase figure (Fig. 6.35), when the figure is seen as the vase, then concavities pointing into it define its parts as the base, stem, and bowl. When the figure is seen as a pair of faces, then the concavities pointing into them define parts as forehead, nose, and so forth. This demonstration shows how the same contour can be "recognised" as two distinct objects: what matters is the way in which the contour is partitioned prior to recognition, and this in turn seems to involve a simple search for concavities referred to the centre of whichever region is seen as the figure.

Hoffman and Richards (1984) have shown how the kind of occluding contour that might result from the application of the Gestalt grouping principles may be resegmented into its parts. As we will see in Chapter 9, such segmentation forms an essential stage in Marr's theory of the analysis and recognition of occluding contours, and at that point we will return to the part structure of objects.

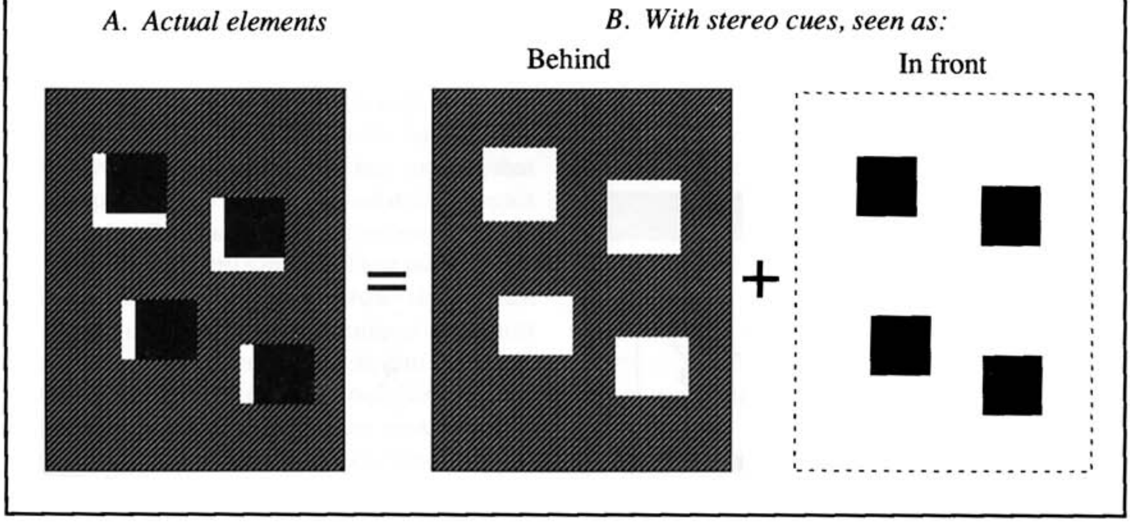
## Perceptual grouping and the representation of surfaces

Marr (1982) did not regard his early visual processing program as solving the "figure-ground" or "segmentation" problem as traditionally conceived. The goal of early visual processing is not to recover the "objects" present within a scene—for the division of a scene into component objects is an arbitrary and ambiguous affair. Which should we regard as the "objects" to be recovered—a crowd of people, each individual person, or the eyes, ears, and nose of each? Such consideration depends on the use to which the information is to be put. The recovery of the full primal sketch, in which some potentially significant structures such as occluding edges may be found, is only one aspect of early visual processing. Marr saw the goal of early visual processing as describing the surfaces present in the image. In recent years, Nakayama and colleagues have marshalled experimental evidence supporting the view that many early vision operations, such as texture segmentation and motion correspondence, take place at the level of surface representation, and not in the 2-D image domain.

Of particular relevance to this chapter are the experiments of He and Nakayama (1994) on texture segmentation. The task was to discriminate a target region containing white L-shaped elements from a background region containing white vertical bars, or vice versa (Fig. 6.36A). Thus target and background regions are distinguished by the presence or absence of horizontal bars. This relatively simple task could therefore be done by grouping features in the primal sketch, or by a filter-based scheme of the kind discussed earlier. Next, by manipulating stereoscopic cues (see Chapter 7), He and Nakayama (1994) made the "L" and "bar" elements appear to be an array of white rectangles occluded by nearer, black rectangles (Fig. 6.36B). In these circumstances, perceived texture segmentation was weak, even though the same low-level cues were present as in control conditions. In the control conditions, where segmentation was strong, occlusion of surfaces was not perceived, and the elements were seen simply as "L"s and "bars".

These results certainly present a strong challenge to both the filter-based and

FIGURE 6.36




Schematic illustration of the displays used to show that an easy texture discrimination could be eliminated by creating the appearance of two homogeneous surfaces, one occluding the other. Segmentation of regions containing white bars and Ls was initially strong (the experiments used many more elements than shown here). But when the white elements appeared to be a uniform field of white squares, partially occluded by nearer black squares, segmentation was weak. After He and Nakayama (1994).

feature-grouping accounts. In the first case, they perhaps imply that the secondary stages of smoothing and edge detection in models like Fig. 6.31 occur much later than has been supposed. In the second case, they suggest either that feature-grouping must operate at a relatively high level, at which 3-D surfaces and the occlusions between surfaces are represented, or that feature-grouping can be controlled by those higher levels of representation. It is to the representation of surfaces in depth that we turn in Chapter 7.

## CONCLUSIONS

The Gestalt psychologists, through the study of perception of simple patterns, gave us insights into the organisational principles that may apply to the perception of the world. The study of natural camouflage and concealment shows that these principles fare well in describing the utility of various surface markings of animals. Marr showed how such principles might be incorporated within a processing model that reveals the structures hidden in the messy data obtained from natural images. Recent research has begun to uncover some of the basic mechanisms underlying this recovery of "global" structure by human and animal visual systems. At a relatively low level, nonlinear transformations of the outputs of spatial filters can convert higher-order structures, such as texture boundaries, into simpler intensity edges whose detection is relatively well understood. But in addition, experiments confirm the existence of active processes linking elements across visual space, and the work of He and Nakayama implies that at least some of these processes take place at a fairly high level where a representation of surfaces in depth has been established.