11

FORM PERCEPTION AND ATTENTION
Striate Cortex and Beyond

Anne Treisman, Patrick Cavanagh, Burkhart Fischer,
V. S. Ramachandran, and Rüdiger von der Heydt

I. INTRODUCTION

Consider a "snowman" seen against a blanket of snow. How does the visual system demarcate its boundaries, and how do we recognize it as a "man" even though its surface is made of snow and the nose is a carrot? We do this almost effortlessly, even though snowmen may be perceived from different angles, at different distances, and with varied shapes, sizes, and features. The visual system has a remarkable capacity to determine the spatial relationships between the features of an object before categorizing it in the appropriate semantic category.

What kind of neural circuitry could perform this deceptively simple task? The answer is far from known. The great difficulties encountered in developing computer vision systems, even on the most advanced computers, have helped us appreciate the power and sophistication of our own visual processes (Feldman, 1986; see Chapter 15, this volume). The brain solves the problem of vision by dedicating massive amounts of cortex to it: Areas involved in vision account for over 30% of the neocortex in macaque monkeys, a species whose visual system closely resembles our own. Trying to understand the workings of such a complex system is a formidable task, but it may be made easier if parts of the system can be studied in isolation.

A widely held view of object perception divides the visual process into a number of distinct functional stages. This may be misleading for a highly interactive system like the brain, but as a first approximation it can be used to guide research and to provide a framework for results. An early stage within this hierarchy is said to extract different elementary image features and to group them on the basis of similarity and spatial proximity. The areas defined by these grouping processes become the candidates for further visual processing, culminating in identification. The next stage delineates the boundaries and surfaces present, forming a three-dimensional, viewer-centered

Visual Perception: The Neurophysiological Foundations.

representation of objects and their backgrounds. Finally, the representations are matched to a thesaurus of stored descriptions and identified, if possible, as familiar individuals or as instances of familiar categories.

Many questions arise within this framework: What defines the inventory of elementary features? How are they detected and grouped? How are object boundaries located? At what stage or stages are two-dimensional retinal image features mapped into the properties of real objects from which they originate? What kind of object representation is formed, and what constitutes a match to objects previously experienced? How are objects recognized independently of their locations, distances, and sometimes of their orientations, while at the same time information about these spatial properties is preserved and used? For all these stages, what neural mechanisms might implement the functions we hypothesize? In dealing with processes as complex as object perception, we are unlikely to find answers by looking only at the anatomy and physiology of the visual system. Although single cells show considerable specialization in the stimuli to which they respond best (whether oriented gratings, illusory edges, or faces), it seems highly unlikely that signals from any one cell can convey meaning in isolation. We must specify the functions underlying subject performance in a given task and then see at what levels they can be linked to the pathways identified by neuroanatomical studies. Visual phenomena are used to infer the existence of separate visual mechanisms, and these in turn are checked both by converging evidence from other behavioral tasks and by possible links to physiological processes.

It is impossible in a single chapter to review all the relevant evidence. We select examples of research which seem to us to probe central issues at the different stages of object identification and to throw light on the mechanisms which may be involved. Our understanding of these complex issues is still rudimentary; the hypotheses proposed to link behavior to physiology are tentative at best.

II. EXTRACTION OF ELEMENTARY IMAGE FEATURES

The visual system encodes a number of different classes of information: It must extract features that specify textures, surfaces, and their spatial layout, features that specify events (movement, change), and finally features that define the shapes and structures of objects. These properties of the world may each be inferred from a variety of physical parameters of the retinal image. There is no simple one-to-one mapping between the two. Thus, motion across the retina may specify object shape (structure from motion) and distance (motion parallax) as well as motion of objects or of the observer. If the goal of perception is to specify the external environment, we need not expect to find anatomical specialization reflecting simple physical parameters of the retinal stimulus; stimulus cues may be functionally grouped in terms of the real properties that they help to specify. (For more detailed reviews of psychological research on feature extraction, see Anstis, 1975; Cowey, 1979; Graham, 1985; Treisman, 1986a,b).

The first class of information, specifying the spatial layout of the scene is likely to be encoded at an early stage, in order to group the parts of potential objects and to segregate them by their surface properties from their backgrounds. The second class of properties specifies "events" such as onsets, offsets, or changes of illumination (see Chapter 7, this volume) or codes properties of motion such as direction, velocity, acceleration, and looming (Regan, 1986). Finally, there may be an inventory of parts or properties used to characterize and to identify objects. Examples might be the volumetric units or "geons" proposed by Biederman (1987). These various kinds of visual information have often been treated together in accounts of early feature analysis, but this may lead to confusion. They serve different perceptual functions and are likely to reveal their effects in different behavioral tests (see Julesz, 1984; Treisman & Gormican, 1988). As a result, disagreements have arisen about the candidate members for an inventory of basic visual features (so-called primitives) from which more complex percepts are constructed, and discrepancies between inferences from physiology and psychophysics may occur. In this section, we discuss evidence relating to the early texture and surface-segregating features. In addition, we outline behavioral criteria and neurophysiological evidence that might be relevant to this "initial parsing of the visual field."

A. Visual Grouping and Texture Segregation

A prerequisite for features involved in the initial segregation of objects and background surfaces is that they should be detected in parallel across space. That is, they should allow easy, "preattentive" grouping of areas that contain them, and set up salient boundaries with areas that do not. Figure 1A shows examples where such segregation spontaneously takes place. The Gestalt psychologists (e.g., Wertheimer, 1923, 1938; Metzger, 1953, 1975) were the first to demonstrate good phenomenal grouping based on spatial proximity of local elements, on common direction of motion, and on "similarity." An historical account of this research is found in Pastore (1971). Other attempts to specify the effective forms of "similarity" found easy detection of groups and boundaries based on the simple features of color, brightness, and line orientation (Beck, 1966, 1967, 1982; Olson & Attneave, 1970; Attneave, 1971; Julesz, 1975, 1984; Treisman & Gelade, 1980). Examples are given in Fig. 1A. Julesz (1964) added binocular disparity, line ends (terminators), and intersections, and Treisman and Paterson (1984) reported evidence that closure (a wholly or partly enclosed area) also functions as a feature whose presence can segregate one area from another. The cues which support this operation of grouping and boundary formation are likely to be detected and processed early within the visual pathways.

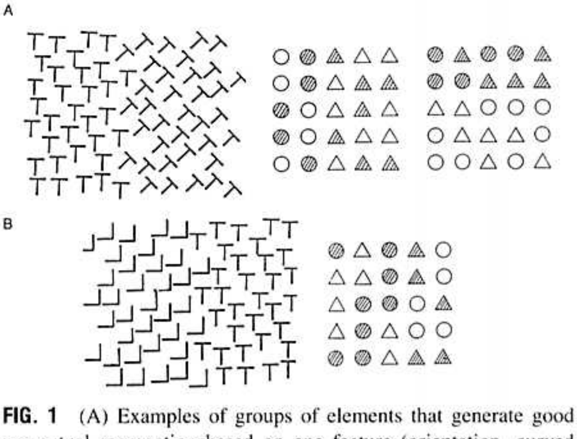

FIG. 1 (A) Examples of groups of elements that generate good perceptual segregation based on one feature (orientation, curved circles versus straight-edged triangles, shape, and color; in the experiment open shapes were red and filled shapes blue). (B) Examples of groups of elements differing only in the spatial arrangement or conjunctions of their features. No clear division is apparent between the left and right sides of the figures. From Beck (1966) and Treisman (1985).

Spatial arrangements of parts and conjunctions of features do not define bounded areas with the same salience and clarity (Beck, 1966; Treisman & Gelade, 1980). Figure 1B shows examples where the separate areas can be found only with focused attention, and with much longer reaction time. Perceptual grouping seems, then, to be determined by separate features and not by the integrated wholes that they characterize.

B. Parallel Processing in Visual Search

Another visual performance test for early parallel detection of features uses visual search tasks to discover targets that "pop out" against a background of distractors (nontarget items). Such targets are detected equally quickly regardless of the number of distractors that are present. Fast and spatially parallel detection is taken as evidence that the features in question are coded early in the visual process. Targets that pop out when displays are searched usually differ from the distractors in one or more of the same set of elementary features that mediate grouping and segregation. They may possess a unique color or brightness, a unique orientation, disparity, or direction of movement, or they may be the only elements with terminators or closure (see, e.g., Neisser, 1963; Egeth, Jonides & Wall, 1972; Treisman & Gelade, 1980; Bergen & Julesz, 1983; Treisman & Gormican, 1988). Results of visual search studies suggest two further inferences: First, they are consistent with the idea that different features are processed in separate subsystems or channels, and, second, they reveal striking asymmetries in feature coding between target and distractor. These inferences are now described in greater detail.

C. Modularity of Feature Analysis in Early Vision

A module (Marr, 1982; Fodor, 1983; see also Chapter 10, this volume) is a specialized subsystem coding a particular class of properties independently of other properties. For example, stereoscopic depth might be coded independently of motion or of color. Search tasks provide evidence that early visual analysis may be modular. If a target-defining feature is coded independently of others, variation on irrelevant dimensions should have no effect on the speed of search. If, on the other hand, search depends on a more global process, heterogeneity of background elements should seriously disrupt detection of a target item, whether the variation is on the target-defining dimension or on other, irrelevant dimensions. Treisman (1988) found no interference with detection of a target (e.g., a blue bar among green bars, or a horizontal bar among vertical bars), when the distractors varied on other dimensions, although search was slower when the distractors varied on the relevant di-

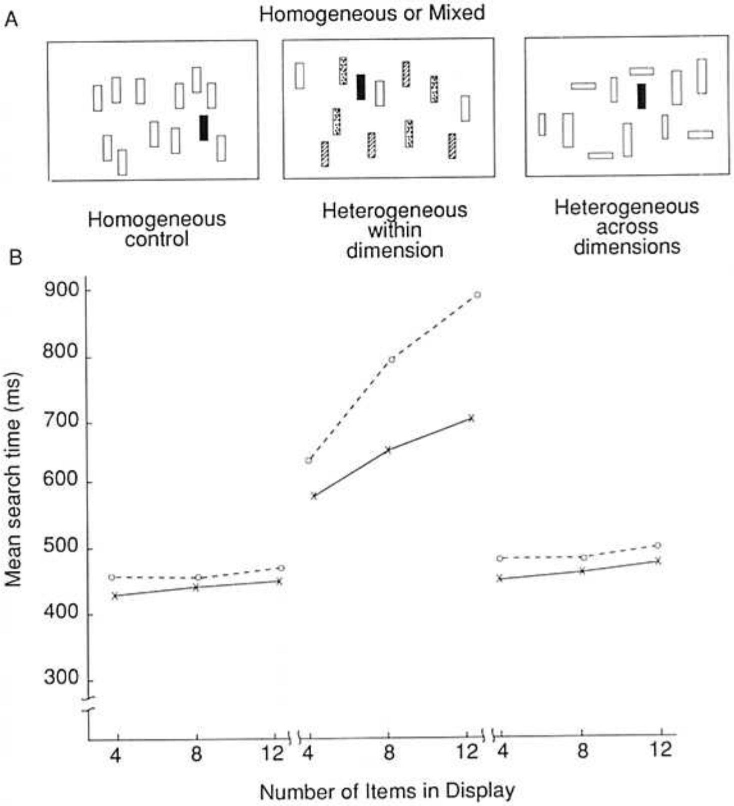

FIG. 2 (A) Examples of displays testing the effects of distractor heterogeneity, within and between dimensions. In the displays illustrated, the target is a blue bar in each case. In conditions not illustrated here, targets defined by orientation (horizontal) were presented among distractors that were either all vertical (homogeneous condition) or vertical, left diagonal, and right diagonal (heterogeneous within dimension), or vertical in three sizes and three colors (heterogeneous on irrelevant dimensions). (B) Mean search times with such displays. The results are the means for targets defined by color (blue) and by orientation (horizontal). Display size indicates the number of elements in each display, whether homogeneous or heterogeneous. Crosses indicate that a target was present in the display; circles indicate that it was not. From Treisman (1988).

mension (see Fig. 2). This finding is consistent with the idea that separate feature maps are formed in early vision, with salience defined only *within* dimensions. It conflicts with the alternative account that the locations of feature differences are detected before their nature is determined (Sagi & Julesz, 1987). On the other hand, when the target is unknown, "the odd one out" is detected more slowly than when the target is specified in advance, suggesting that it takes longer to check for its presence in several different modules than to check a single known module (Treisman, 1988).

D. Asymmetries in the Coding of Features

The failure of distractors to impair performance when variation occurs on irrelevant stimulus dimensions suggests independence of coding *across* different stimulus dimensions. Visual search results can also suggest which values *within* a dimension are coded as separable features. A surprising symmetry between different values within dimensions has been discovered, using the visual search task (Treisman & Souther, 1985; Treisman & Gormican, 1988). Given the same physical discrimination, for example, between a curved and a straight line, search latencies may be drastically different depending on which of the two features defines the target and which the distractors. A curved line can be found quickly in a display of straight lines, whereas a straight line appears to require much slower, serial search when presented in a display of curved lines. Similar asymmetries are found with many other dimensions: a tilted line pops out of a display of vertical lines, whereas a vertical (or frame-aligned) line is harder to find among tilted lines. A pair of converging lines is easier to find among pairs of parallel lines than the reverse. A circle with a gap pops out of complete circles but not the reverse. An ellipse is found more quickly among circles than a circle among ellipses. Figure 3 gives examples of the displays and of the search functions obtained in these different tasks.

The following paradigm cases may help us to interpret the others: (1) A circle with an added intersecting line pops out of circles without lines, but a circle without a line is found only by serial search through a display of circles with intersecting lines (see Fig. 4). Thus the *presence* of an added feature seems to be detected immediately and with no effect of the number of distractors, but its *absence* does not. (2) On quantitative dimensions (such as line length, number, or contrast) search for a high value among distractors with low values is also faster than search for a low value among distractors with high values. It seems that adding a new feature or increasing the value on a quantitative dimension makes a target item more easily detectable. On the other hand, removing or decreasing the feature which distinguishes the target from the distractors makes it preattentively "invisible." By analogy, certain qualitative dimensions may also be asymmetrically coded, with one direction of change on the dimension coded as "adding a feature," or giving an increase in neural activity. This change would make a target easy to find, whereas a change in the opposite direction would make it hard. A curved line, by this analogy, appears to have an added property (curvature) that a straight line lacks; a tilted line has an added property (tilt) that a vertical line lacks.

The general rule, extracted from a number of experiments showing search asymmetries, seems to be that a stimulus that departs from a standard or reference value is coded as having an added or positive feature that the standard value lacks. An ellipse has an additional feature (elongation) relative to a circle; magenta has some additional blue relative to a standard red; converging lines have an extra property (convergence) relative to parallel lines; curved lines add curvature to the reference value of straightness. All appear to signal the *presence* of an additional unique feature besides the standard or reference value for the dimension in question. If this interpretation is correct, the direction of search asymmetry can be used to discover which stimulus in any new pair is treated as the reference value in early visual coding. For example, it appears that a dot outside a closed shape is treated as the deviation and a dot inside

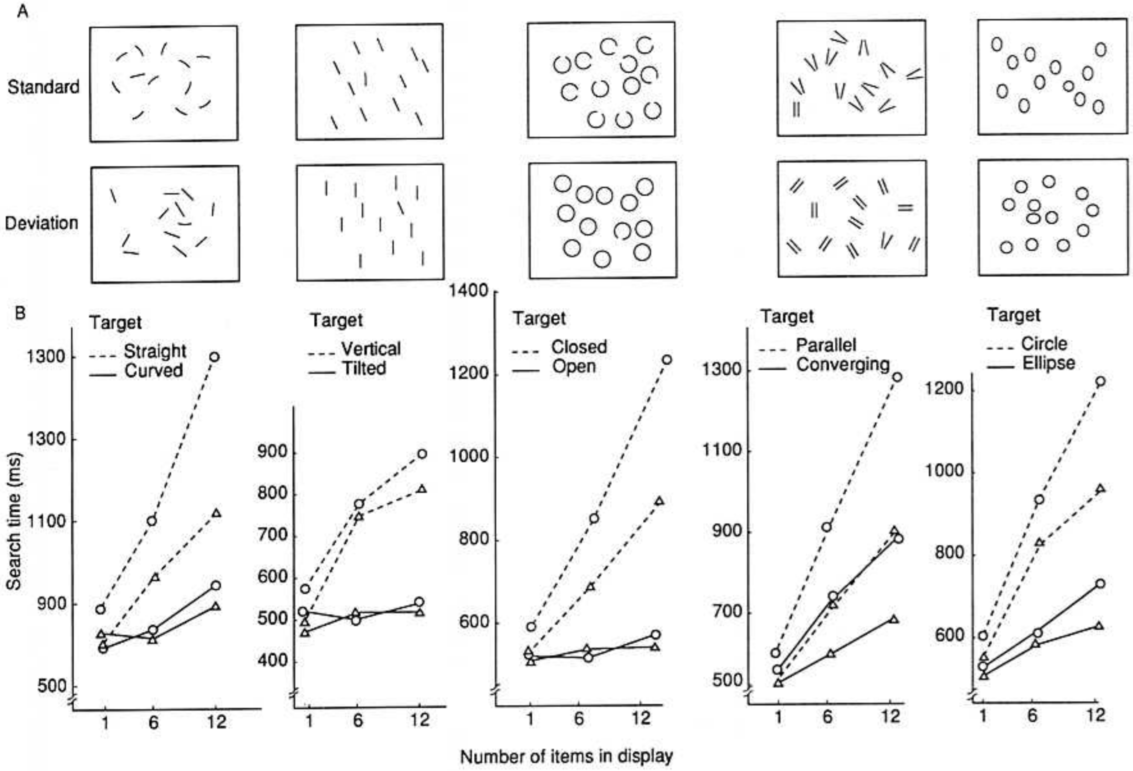

FIG. 3 (A) Examples of five pairs of displays that generate search asymmetries between target and distractor. (B) Mean search times through such displays. In each case, the dotted lines show the times for the "standard" target and the solid lines for the "deviation" target. Triangles indicate that the target was present in the display, and circles signify that it was not. Adapted from Treisman & Gormican (1988).

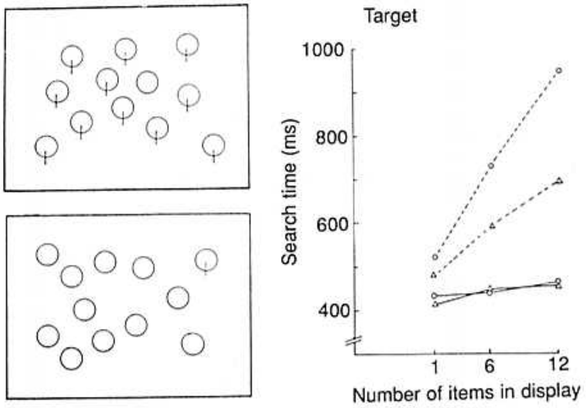

FIG. 4 Displays with a target defined by either the presence or the absence of a feature (added line), and search times for such displays. From Treisman and Souther (1985). Copyright (1985) by the American Psychological Association.

the shape as the standard. So far, no physiological correlate has been found for any of these coding asymmetries.

E. Parallel Coding versus Parallel Access to Previously Coded Features

One of the criteria to identify which features are coded as the basic elements in early visual processing is that they should be detectable preattentively. They are characterized by flat functions (near zero slope) relating search times to the number of distractor items. In addition, the mean reaction times in both search and boundary detection are typically quite short; response times to detect a target or a texture boundary average 400 to 600 msec. Some perceptual grouping phenomena, however, appear to reveal spatially parallel access after a much longer initial delay. For example, the emergence of stereoscopic depth in random dot stereograms may take several seconds (Julesz, 1971), yet the perceived depth is clearly apparent across the field as a whole, once the three-dimensional organization has emerged. (On repeated presentations, shape from depth is discerned more quickly.)

There may be other forms of perceptual grouping that share this character of slow emergence, with parallel (global) awareness of all the elements prevailing, once the organization of the field has become apparent. They contrast with other displays whose texture boundaries are found equally slowly (e.g., displays in which different areas are defined only by conjunctions of features) but which do not afford the same continued awareness of the boundaries after they have been located. For example, Ramachandran (1988a,b) has found that three-dimensional shape from shading can provide cues for perceptual grouping of this slow emerging, but global type. Thus, in Fig. 5A there is a strong tendency to perceive one set of objects (the ones that are white on top) as convex and all the others as concave. In this display, it is possible to group perceptually all the convex objects together, so that they form a separate depth plane standing out clearly from the background of concave objects. This is a surprising result, since it is generally assumed that only elementary image features like brightness and contrast, color, stereodepth, orientation, or motion can serve as a basis for perceptual grouping. The results imply that even three-dimensional shapes defined exclusively by shading can provide perceptual features that allow phenomenal grouping and segregation.

To ensure that this phenomenal grouping was not simply due to a more primitive image feature (such as luminance polarity), Ramachandran produced a control display (Fig. 5B) that was similar to Fig. 5A in terms of luminance polarity but did not convey any depth. Grouping was very difficult to achieve in this display, suggesting that the effects observed in Fig. 5A must have been based on three-dimensional shape from shading. It may take as long as 30–60 sec for depth to emerge; however, there is again a tendency for this time to decrease with repeated trials.

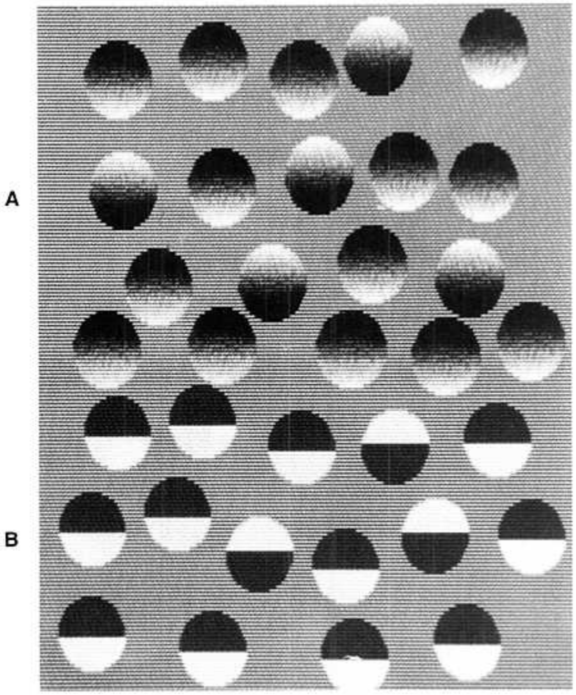

FIG. 5 Random mixture of shaded objects that have opposite luminance polarities. The objects which are light on top are usually perceived as spheres, and they can be visually grouped together and segregated from the background of concave objects. Hence, one may conclude that three-dimensional shapes defined by shading can provide tokens for perceptual grouping and segregation. (B) In the "control" no three-dimensional shapes are seen, and it is also impossible to perceptually segregate the objects on the basis of luminance polarity. From Ramachandran (1988a,b). Copyright © 1988 Macmillan Magazines Ltd.

These findings are particularly revealing because shape from shading relies on a fairly complex "interpretation" of the display in terms of an assumed source of illumination. The derivation incorporates the constraint that there is only *one* light source in the entire image (or a large portion of it), perhaps because our brains evolved in a solar system that has only one sun.

The strong sense of depth observed in Fig. 6 depends exclusively on shading. The sign of perceived depth, however, is ambiguous, for the brain has no

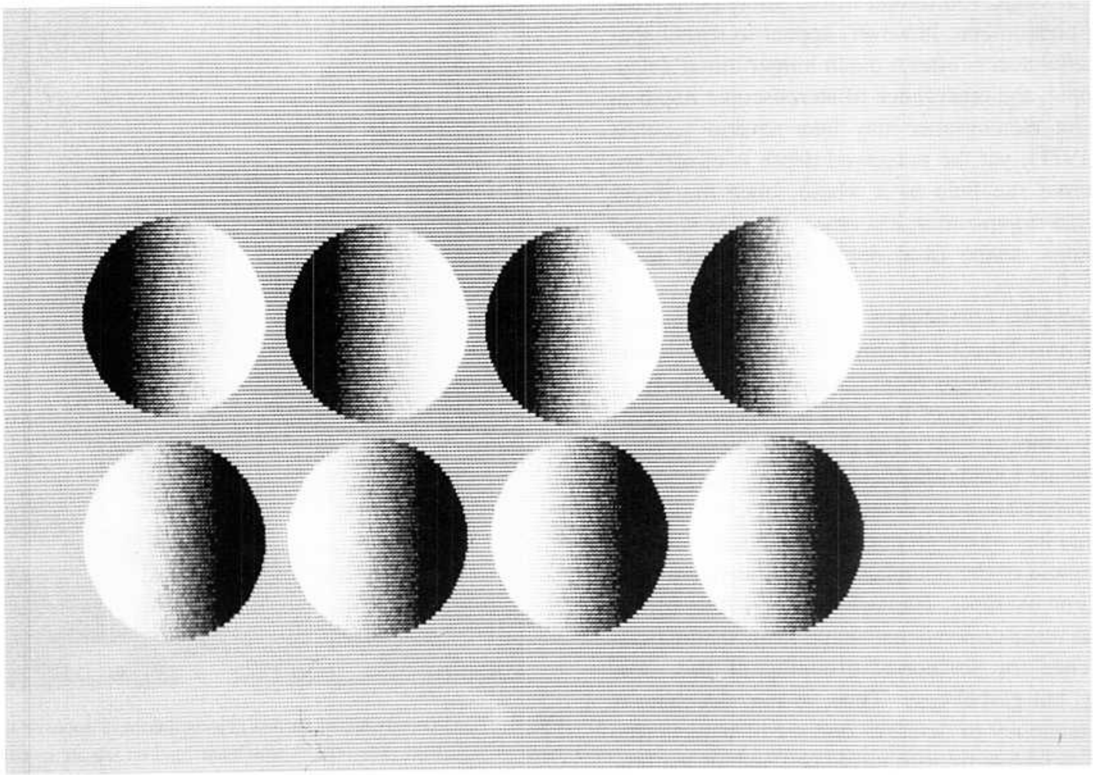

FIG. 6 Bistable display in which shapes can be perceived either as convex or concave by mentally shifting the assumed light source from left to right. When the top row is seen as convex, the bottom row usually appears concave (or flat), whereas when the bottom row is seen as convex, the top row appears concave. The effect may take some time to emerge and stabilize. The display illustrates the point that the visual system has a built-in "assumption" that there is only one light source illuminating the entire image. From Ramachandran (1987a,b). Copyright © 1987 Macmillan Magazines Ltd.

way of knowing where the light source is. It seems that when one row of objects is convex, then the mirror images are always concave (and vice versa). In fact, it is impossible to see all the objects as being simultaneously convex or concave. This observation suggests that the brain prefers a "common-light-source" assumption to a "common-depth" assumption. On the other hand, the shapes that are white on top are almost always seen as convex, and it is very difficult to generate voluntary perceptual reversals. This suggests that in addition to the single-light-source constraint, there is also a tendency to assume that the light source is at the top; an effect that is well known to artists.

While viewing these displays, one has the impression that the visual system goes through several operations before the final percept emerges. In the earliest stage the computation required for defining the three-dimensional shapes (i.e., depth) is performed. This can take several seconds. Next, the objects that differ from the others pop out (i.e., segregate) as in Treisman's displays. Once the targets have emerged, however, one has the distinct impression of being able to "hold on" to them indefinitely (hysteresis) in order to group them with other similar items in the display (grouping). Finally, after grouping has occurred, the objects are clearly segregated from irrelevant items cluttering the background (figure–ground separation). Thus, what is usually described as a one-step operation—the extraction and grouping of elementary features—may, in fact, involve several distinct perceptual capacities which contribute toward the ultimate goal of delineating figure and ground.

There is, unfortunately, no simple way to disentangle and measure these effects. One promising approach involves the use of apparent motion to determine the extent of perceptual segregation (Ramachandran, Rao & Vidyasagar, 1973a). Apparent motion of a group of items as a whole may provide converging evidence for parallel access, since all the elements must be matched to the corresponding elements in the preceding display in order to determine the extent and direction of motion. If the global motion of the group is immediately apparent, one can infer that all the items are capable of being matched in parallel. This is shown by the following experiment. Consider two frames (similar to Fig. 5A) of an apparent motion sequence. The first frame has a cluster of three convex objects on the left, whereas in the second frame the cluster as a whole is shifted to the right. The positions of the remaining elements themselves are random and uncorrelated in successive frames. The results are quite striking. As long as no depth is perceived, the elements in the display simply appear to flicker. Once depth emerges, however, the triangular cluster of convex objects jumps vividly between the two locations as the frames are alternated. (Optimum apparent motion is seen when the stimulus onset asynchrony is about 200 msec.) These results suggest that shapes defined exclusively by shading can provide cues for grouping and long-range apparent motion correspondence (see Chapter 9, this volume).

A different case of apparent motion is illustrated in Fig. 7. In this case, there is no correlation at all between the elements of the two textures in the successive displays; only the shape of the boundaries between the different areas in the two frames is correlated. Yet, when the frames are alternated, one observes vivid apparent motion between the two inner "squares." This result suggests that the long-range apparent motion system can accept previously abstracted texture borders as an input, although the individual elements constituting the apparent shape are not matched. Both the boundaries and the shape from shading, once computed by the brain, can remain perceptually available as a basis for further visual computations. Given the very long latency with which the shape from shading becomes apparent, it is possible that serial processing is required to set up the initial representation of the individual elements. Once computed, however, the results can be "held on to" and used as cues for apparent motion, global shape perception, and other visual analyses.

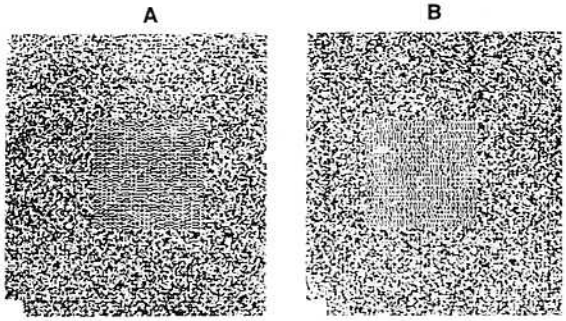

FIG. 7 Two frames of a movie (A and B), presented side by side for clarity. The square in Frame B is shifted horizontally in relation to Frame A, but the elements constituting the squares are completely uncorrelated in successive frames. When A and B are optically superimposed and alternated, the square is seen to move back and forth horizontally, suggesting that the motion system accepts texture borders as an input even in the absence of point-to-point luminance correlation. From Ramachandran, Rao, and Vidyasagar (1973a,b). Copyright © 1973 Macmillan Magazines Ltd.

In contrast, not only do stimuli defined only by arbitrary combinations of properties (e.g., red and vertical) require serial processing in most cases, but they also fail the apparent motion test. A. Treisman (unpublished) repeated Ramachandran et al.'s (1973a) experiment using conjunctions of color with orientation, color with size, and orientation with size to define sets of coherently moved elements. For example, a set of six pink bars tilted 45° to the left appeared in two different locations in alternating displays, either preserving their relative positions within the set (coherent motion) or with changes in their relative positions (incoherent motion). These six elements were embedded in a larger display of pink bars tilted 45° to the right and green bars tilted 45° to the left. Each of these background elements was slightly displaced in a random direction between the two displays, so that the targets could not be distinguished simply by contrast with a continuous background. The task was to discriminate coherent global motion of the target group from incoherent motion of its separate elements. Subjects were, in general, quite unable to do this when the target group was defined only by conjunctions of features, and, in fact, no global motion was seen. The target group did not become phenomenally salient as a whole, even once its elements had been located, and it failed to emerge in apparent motion across successive displays, even when subjects were given unlimited time to observe the alternation. The only case in which performance was above chance was when one of the features defining the target conjunction group was the larger size. In contrast, if the target group differed in a single feature rather than a conjunction, apparent motion was usually seen. For example, grey vertical bars embedded in the same background of pink and green bars tilted right and left gave significantly better performance. Thus, a difference in color, size, or orientation presumably gave access to the target group as a whole, allowing a match to be made between the displays.

F. Conclusions

These tests of perceptual grouping and segregation, search, and apparent motion suggest the existence of three classes of perceptual units: (1) simple features, which mediate rapid segregation, which pop out in search, and which create texture boundaries that can be abstracted from different successive displays to mediate apparent motion of the areas they contain; (2) more complex features of surfaces in depth, such as stereoscopic planes and shape from shading, which may be processed slowly, in some cases serially, by visual routines (Ullman, 1984) but which remain accessible in parallel once the scene has been perceptually organized; and (3) conjunctions of features that do not create new emergent properties. These are visible without difficulty when presented or attended singly. However, they appear neither to be processed in parallel nor to remain available as parallel inputs to further visual processing once they have been identified. Phenomenologically, it is difficult to maintain more than two or three of them as a salient group against the background elements. They are unlikely to play any role in segregating objects from their background. They may, however, be critical in identifying objects once they have been located.

Treisman's paradigm provides a useful method for deciding whether a given feature is "elementary" or not. For example, it is well known that orientation in two-dimensional space is an elementary feature. What about orientation in three-dimensional space? In experiments to address this issue (Ramachandran & Rogers-Ramachandran, 1988), subjects were presented with a stereogram consisting of a random array of needles tilted in three-dimensional space. For all the needles, except one, the upper ends were tilted toward the observer. (The centers of all the needles were at zero disparity.) Subjects could preattentively detect the single anomalous needle whose upper end was tilted away from them, that is, the reaction time for detecting this anomalous needle did not vary with the number of distractors.

At what stage in visual processing does popping out of elementary features occur? For example, we know that depth can be conveyed by a variety of parallel cues such as stereo disparity, shading, occlusion, and motion parallax. At least two of these (shading and disparity) constitute elementary features in the sense described by Treisman (1986a). But where does this segregation occur? Does it occur at a very early stage where there are (presumably) separate feature maps for different depth cues or at a later stage where different sources of information are combined into a common "depth map"? To answer this question Ramachandran and Rogers-Ramachandran (1988) used targets that convey conflicting depth cues. They designed a stereogram composed of small squares whose corners partially overlapped. Disparities were introduced between the two images so that for all but one square, the two cues (occlusion and disparity) were *consistent* with each other, that is, they conveyed the same depth. For the single anomalous target, on the other hand, the occlusion and disparity were of opposite sign so that they tended to cancel each other. The magnitude of depth seen in this target was considerably less than in the distractors, yet it did not seem to pop out preattentively. This result suggests that grouping must occur before the conflicting cues are combined.

III. NEUROPHYSIOLOGICAL AND ANATOMICAL EVIDENCE ON EARLY STAGES OF VISUAL CODING

A. Single Cell Recordings and Separate Neural Channels

It seems likely that the grouping and parsing phenomena attributed to early vision depend on processing in the various areas of the visual cortex in which cells respond to simple visual stimuli; these include Areas V1, V2, V3, V4, and MT. In this section, the main findings from neurophysiological recordings in these areas are outlined. As yet, there has been little attempt to relate the single-cell data directly to those from behavioral studies. One important reason is that there need not be a direct correlation between the responses of single units and the response of the system as a whole. If such a correlation is found, as it is in some specific cases, it may be very informative. In other cases, however, the correlation could be with patterns of activity across populations of individual cells; it will then be much more difficult to detect and interpret.

Most of the neurophysiological data about the striate cortex and the extrastriate visual areas come from experiments with isolated spots, edges, bars, or gratings. The main variables that have been studied are stimulus position, orientation, temporal modulation, direction and velocity of movement, side of monocular stimulation, binocular disparity, wavelength, width and length of bars, spatial frequency, and contrast. The proportion of cells that are sensitive to each of these parameters as well as their degree of sensitivity in different areas and subdivisions of the visual cortex offer important clues to their function (see Van Essen, 1985).

Any visual system must make a compromise between temporal resolution and spatial resolution. The mammalian visual system appears to resolve this problem early in the retina, where α- and β-type ganglion cells have evolved specialized functions for sharpness in time and space, respectively. Thus, when information reaches the cortex, via the separate magnocellular (originating in α cells) and parvocellular (originating in β cells) layers of the lateral geniculate nucleus (LGN), a division has occurred between the processing of pattern–color, on the one hand, and motion, depth, and spatial orientation, on the other. This division of information flow through the brain continues from visual cortex through temporal and parietal cortex. How, then, do the earliest areas (V1, V2, V3, MT) sort out the cues just mentioned for further elaboration by higher cortical modules?

A functional differentiation appears even at the beginning of cortical processing, between three parallel subsystems that are anatomically interlaced in Areas V1 and V2 as revealed by cytochrome oxidase staining (see Hendrickson, 1985, for a review) and by tracing the connections within and between these areas (Livingstone & Hubel, 1984; 1987a,b). Figure 8 provides a summary of these pathways (see also Chapters 7 and 8, this volume). The first subdivision is formed by layer 4B of Area V1 and the "thick stripes" of V2, the second by the "blobs" of V1 and the "thin stripes" of V2, and the third by the "interblobs" of V1 and the "interstripes" or "pale" stripes of V2. The first has been related to the perception of depth and motion since, in the macaque, cells in the thick stripes are often sensitive to binocular disparity and direction of movement but not to color; the second has been related to color, since most cells in the blobs and the thin stripes are unselective for orientation and direction of movement but many are wavelength selective or show color opponency; the third has been related to form perception since cells in the interblobs and the interstripes were often found to be

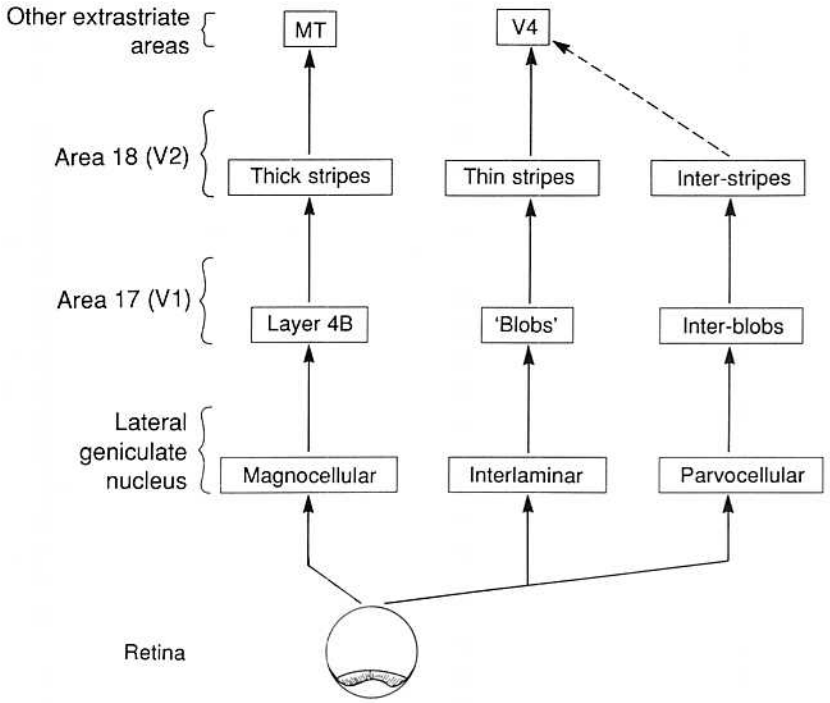

FIG. 8 Three parallel channels in the primate geniculocortical pathway. The diagram summarizes information from several sources including Hubel and Livingstone (1985) and Van Essen (1985). For clarity, the relay through layer 4C has been left out. Also, the interlaminar projection to the blobs is still somewhat controversial. After Ramachandran (1987a).

orientation selective, but not color selective (Hubel & Livingstone, 1985, 1987; Shipp & Zeki, 1985; Livingstone & Hubel, 1984b).

Further evidence comes from the projections from V2 to other areas. The thick stripes were found to project mainly to the middle temporal area (MT), the thin stripes and interstripes to Area V4 (DeYoe & Van Essen, 1985; Shipp & Zeki, 1985), which themselves are specialized, as described below. By means of a double labeling technique it was shown that individual V2 cells project either to V4 or to MT but not to both (DeYoe & Van Essen, 1985). Estimates of the numbers of specific cells in the subsystems of Area V2 do not all agree across these studies. For example, according to Hubel and Livingstone (1985, 1987) and Shipp and Zeki (1985), the thin stripes and interstripes differ markedly in the frequency of orientation-selective cells, whereas DeYoe and Van Essen (1985) find only subtle differences, if any, between these stripes.

The early anatomical segregation within Area V1 is maintained in the later stages of visual processing. There is some evidence that motion and color are analyzed by specialized cortical areas. The dichotomy into a "motion system" and a "color and form system" (Maunsell & Newsome, 1987) appears embedded in the more general distinction between two cortical streams (see Fig. 8): the occipitotemporal pathway including V4, devoted to object perception and recognition, and the occipitoparietal pathway including Area MT, concerned with perceptual localization (Mishkin, 1972; Pohl, 1973; Ungerleider & Mishkin, 1982; Desimone, Schein, Moran & Ungerleider, 1985). Thus, direction of motion is emphasized in Area MT (Dubner & Zeki, 1971; Zeki, 1974; Maunsell & Van Essen, 1983a,b), whereas processing of color is emphasized in V4 (Zeki, 1973, 1983c). Cells in Area V4 are also selective for orientation, the shape of bars, and spatial frequency, suggesting that form processing occurs here (Desimone & Schein, 1987). It is worth noting, however, that the orientation of texture elements can also be important in the segregation of surfaces by their texture. It is not always clear whether cells that are tuned to orientation are specialized to code texture elements or the orientation of global boundaries between surfaces, or both. Hammond (1978) and Nothdurft & Li (1985), for example, found different directional tuning of the same complex cells for texture motion and for motion of oriented bars. This suggests that oriented and textured stimuli may be coded by functionally different cortical pathways, which nevertheless share some of the same complex cells (Hammond, 1985).

B. Evidence from Neuropsychology and Psychophysics

Visual deficits following brain lesions in human patients have shown independent losses of at least three different attributes. Zihl, von Cramon, and Mai (1983) described a patient having difficulty perceiving moving stimuli. The patient reported that objects seemed to change position but gave no impression of motion. Botez (1975) has reported the converse, a patient who can identify objects only when they move. Cortical color blindness is an occasional consequence of brain trauma (Damasio, Yamada, Damasio, Corbett & McKee, 1980), while the reverse (luminance blindness) has also been reported (Rovamo, Hyvärinen & Hari, 1982). In the latter case, the patient complained of not being able to see anything on black and white television, even though she could see the pictures on color television without any problem. These clinical cases argue for some physical separation in the cortical representations or projections of color, luminance, and motion in the visual system.

The anatomical segregation can also be linked to psychophysical correlates. Since the magnocellular stream is "color-blind" (see Chapters 6 and 8, this volume), one might expect human motion perception, mediated by this stream, to be color-blind as well. This prediction was tested by Ramachandran and Gregory (1978) using a random dot kinematogram. A central square area in a random black and white field of dots was shifted in one display relative to another. When the two displays were alternated at the correct intervals, the square emerged from its random dot background and was seen to oscillate in apparent motion. However, when the black and white dots in the kinematogram were replaced with red and green dots at equiluminance, the appearance of the oscillating central square was lost completely. This suggested that the human motion system cannot accept purely chromatic borders as an input.

The separation does not, however, appear to be complete. Information from the LGN arrives in Area V1 through opponent-color and nonopponent or weakly opponent (luminance) pathways (DeValois, Abramov & Jacobs, 1966; Derrington, Krauskopf & Lennie, 1984). Since there is a higher area, V4, specialized for the analysis of color (Zeki, 1978a,b), it is often assumed that color information contributes only to the pathway leading to this area and not at all to the pathways involved in the analysis of motion and binocular disparity. However, several psychophysical studies have shown that color can in some conditions contribute to the perception of motion (Cavanagh, 1988) and to stereopsis (DeWeert & Sadza, 1983; Grinberg & Williams, 1985). Indeed, Ramachandran and Gregory (1978) found that although motion was lost at equiluminance in random dot kinematograms, motion could still be seen if simple line targets were used. Based on this observation, they concluded that the early motion system was color blind but that the long-range motion system can use any type of contour, including chromatic contours or even equiluminous texture borders. Other studies (Cavanagh, 1988) suggest that the contribution of color to motion passes through the opponent-color (parvocellular) pathway from the retina to the striate cortex rather than "leaking" into the magnocellular, nonopponent pathway that carries the luminance contribution to motion (see Maunsell & Newsome, 1987). These separate routes for color and luminance then converge to form a common motion pathway and a common site for motion aftereffects (Cavanagh, 1988).

Although the physiological evidence for separate streams leading to Areas V4 (form/color) and MT (motion) is strong, the route for analysis of other perceptual dimensions such as texture and stereopsis is less clear. Texture, for example, may require a secondary analysis performed by a subset of cells in the luminance pathway. Similarly, the analysis of binocular disparity may occur in the same region, MT, as the analysis of motion (Maunsell & Newsome, 1987). This could explain the close coupling between stereopsis and motion observed by Ramachandran and Anstis (1986a,b). They presented two vertically aligned spots in Frame 1 of an apparent motion sequence followed by two spots shifted horizontally (Fig. 9A). The spots were always seen to "jump" horizontally and never crossed paths (see also Kolers, 1972; Ullman, 1979a,b), even if diagonally opposite pairs (in Frame 2) were made similar in color or "form." However, if two diagonally opposite corner spots were presented in a separate stereoscopic plane from the other two (Fig. 9B), the spots did cross paths. This was especially true if several such displays were viewed simultaneously. Thus, stereoscopic depth can change the rules of (apparent) motion processing. Motion and stereo disparity processing are closely linked in the brain. The different channels coding object shape are discussed in more detail in Section IV, together with other psychological findings concerning their interrelations.

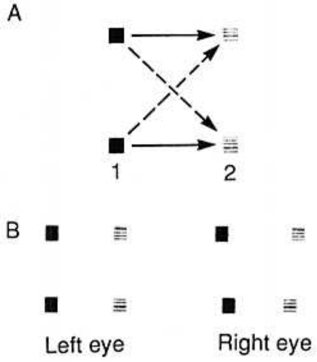

FIG. 9 (A) Two vertically aligned spots are flashed in Frame 1 (black squares) and are followed by two spots shifted horizontally in Frame 2 (gray squares). The spots never appear to cross each other's paths (dashed lines), but always seem to move horizontally (solid lines). Even if diagonally opposite pairs are made similar in color or "form," crossing is never perceived. (B) When diagonally opposite pairs occupy slightly different depth planes as in a stereogram, however, the spots cross each other's trajectories. Thus, there is a close perceptual coupling between motion and stereoscopic depth.

C. Specificity of Cortical Areas

The use of quantitative methods for assessing physiological and anatomical specificity (e.g., Schiller, Finlay & Volman, 1976a,b; Baker, Petersen, Newsome & Ullman, 1981) is obviously important in determining how strictly the different visual areas are segregated in the brain (see Felleman & Van Essen, 1987, for a synopsis of the specificity estimates of various studies of six cortical areas of the macaque). Often the differences between areas or subdivisions of areas seem to be only relative. A number of examples show this. (1) Orientation selectivity and binocular interaction are found in the thin stripes of Area V2, which are part of the "color system," as well as more frequently in the thick stripes (DeYoe & Van Essen, 1985). (2) Cells in the "color area" V4 can also be selective for orientation, spatial frequency, and even direction of motion (Schein, Marrocco & De Monasterio, 1982; Desimone & Schein, 1987). (3) Many cells in the "motion area" MT are sensitive to bincular disparity (Maunsell & Van Essen, 1983a,b). (4) Substantial numbers of wavelength-selective cells have been found in Area V3 (Felleman & Van Essen, 1987), which was previously thought to lack such cells (Zeki, 1978a,b, 1983c). Thus, even though the existence of multiple representations in extrastriate cortex is widely accepted, the routing of the pathways through these areas and the stimulus and/or response attributes processed by each are not yet clearly established.

It is sometimes assumed that successive stages of cortical processing serve to improve selectivity or "sharpen the tuning." However, there is little evidence that dimensional selectivity always increases in higher areas. For example, the width of orientation tuning is not narrower in Area V2 than in V1 (compare Burkhalter & Van Essen, 1986, versus Schiller et al., 1976a, and De Valois, Albrecht & Thorell, 1982). Cells in Area V4 are no more selective for wavelength than are retinal ganglion cells (DeMonasterio & Schein, 1982; but see Zeki, 1980). Tuning to spatial frequency appears to be broader in Area V2 (Foster, Gaska, Nagler & Pollen, 1985) and V4 (Desimone & Schein, 1987) than in V1 (Schiller et al., 1976b; DeValois, Yund & Hepler, 1982). Disparity selectivity does not differ much between Areas V1 and V2 (Poggio & Fischer, 1977; Poggio, 1984). A notable exception is velocity of movement, where cells in Area MT are more selective than cells in V1 (Maunsell & Van Essen, 1983a,b). Thus, looking at the proportions of cells devoted to one or the other parameter and their degree of selectivity, there does not appear to be any regular increase in specificity from level to level in visual cortical processing.

Perhaps a better way of studying visual cortical specificity is to try to relate physiology to behavioral function, and to ask what computational problems might be solved at a given stage (Marr, 1982). As noted earlier, the perceptual qualities of color, contour, motion, and depth depend on stimulation in a much more complicated way than the stimulus properties to which they seem to correspond, such as wavelength composition, luminance gradient, velocity, and binocular disparity. Although perception depends on these properties, the relationship is not simply one-to-one, nor is it many-to-one. For example, many different wavelength compositions can produce the same perceived hue, while a fixed wavelength can appear in many different colors while reflecting the same composition of light into the eyes. Thus, in order to see how visual information is processed in the cortex, we should analyze the stimuli in terms of the properties of the world that they reflect and that the visual system has presumably evolved to detect. If there are modules coding different properties, their function is probably to transpose image features into properties of the three-dimensional world.

Relatively few studies have explicitly used this approach (von der Heydt, 1987). Among the exceptions are studies by Maffei, Morrone, Pirchio, and Sandini (1979) and by Albrecht and DeValois (1981) using square-wave gratings with and without the fundamental component. These gratings appear similar perceptually in that the same periodicity is perceived; however, neurons in striate cortex (cat, monkey) fail to signal the periodicity of the missing fundamental grating. Zeki (1983a,b) has demonstrated that neuronal responses in Area V4 of the monkey may correlate with the phenomenology of color perception when multicolored displays ("Mondrians"; Land 1983) are used, whereas responses of single units in Area V1 are generally determined just by the spectral composition of the light falling on one small patch of retina. Movshon, Adelson, Gizzi, and Newsome (1985) have used two crossed moving gratings to show that responses in Area MT may signal the global direction of movement that is perceived in such a stimulus, whereas those in V1 signal only the directions of the grating components.

Consideration of the functional goal of retrieving real-world properties from the retinal image may help to explain many visual phenomena showing interaction between stimulus parameters that are inexplicable assuming strict anatomical segregation. Examples of such phenomena are the effect of stereo disparity on the perception of illusory contours (Harris & Gregory, 1973; Ramachandran, 1986a), the perception of shape from motion (Wertheimer, 1923, 1938; Wallach & O'Connell, 1953), and the brightness and color effects associated with illusory contours (Ehrenstein, 1941; Van Tuijl, 1975; Ramachandran, 1987a). Indeed, studies of neuronal signals in Area MT with focal and large-field motion stimuli indicate a role for this area in figure–ground discrimination (Allman, Miezin & McGuinness, 1985; Tanaka, Hikosaka, Saito, Yukie, Fukuda & Iwai, 1986). Monkeys with lesions of Area MT were impaired in the perception of motion (Newsome & Pare, 1988) as well as the perception of three-dimensional shape in motion displays (Siegel & Andersen, 1988). On the other hand, several researchers have found cells in the temporal cortex that were highly selective for stimulus form, having receptive fields that were selective for head and body parts (faces) and body movements (Gross, Rocha-Miranda & Bender, 1972; Perrett, Rolls & Caan, 1982; Schwartz, 1984; Perrett, Harries, Mistlin & Chitty, 1989).

IV. CODING OF BOUNDARIES AND CONTOURS

The visual features described so far define the groups of elements that provide the input for later identification of objects. An essential prerequisite is to locate and delineate the boundaries between the groups, since these are likely to correspond to the contours of three-dimensional objects.

If one tries to extract the contours from an image by computer, one realizes that contour is a sophisticated abstraction that involves far more than just intensity gradient detection. Mathematically, contours (specifically "occluding contours") are defined in images of three-dimensional scenes as the lines of discontinuity of the third dimension (Marr, 1977; Koenderink, 1984). At first glance, the perception of contours appears as simple as this definition; we can generally distinguish the various objects in a scene with ease, even if the objects are unfamiliar and we are not able to recognize them. This leads to the impression that the objects are delineated in the image by luminance edges or lines. Therefore, the determination of contours has often been equated with edge and line detection. Since in most cases, neither an object nor its background is homogeneous in luminance and chromaticity, the gradient across contours is variable; it may even vanish at some points and be hard to detect at others. Contours defined by intensity gradients frequently merge with background structures, or with structures on the surface of the object. Moreover, the conditions change continuously when the projection of an object moves relative to the background, as in the case of motion parallax.

A. Perception of Contour

Perception of contour is not confined to intensity or color edges. Contours can be defined by discontinuities in several other attributes, including texture, binocular disparity, and relative motion. For all the extrastriate areas up to the level of Areas MT and V4, the representation of space is fairly retinotopic (see Maunsell & Newsome, 1987). Therefore, the two-dimensional shape of a stimulus is coded on a global level by the spatial arrangement of image information for each of these attributes within one or more of these extrastriate areas. On the other hand, the nature of the local sampling of each cell, within its receptive field, varies from pathway to pathway (see Chapter 10, this volume). The structure of the receptive field for cells in a given area reveals the shape dimensions that can be assumed to be important for further processing in the visual system. In the retina and LGN, for example, information is coded by receptive fields having an antagonistic center–surround organization, and these contribute, in the striate cortex, Area V1, to the formation of receptive fields that are orientation and size selective (Hubel & Wiesel, 1968) as well as possibly curvature selective (Dobbins, Zucker & Cynader, 1987).

Information on stimulus orientation and size therefore appears to be available in Area V1 for stimuli defined by luminance. However, the explicit coding of boundaries between regions defined by other attributes may not be available until the information reaches the extrastriate visual areas. For stimuli defined by wavelength, Gouras (1974), Michael (1978), and Thorell, DeValois, and Albrecht (1984) reported size- and orientation-tuned cells selective for color in Area V1. Color-selective cells in Area V4 appear to have both oriented and nonoriented receptive fields (Zeki, 1978a,b). Many cells in Area MT (Movshon, Adelson, Gizzi & Newsome, 1986), although directionally selective, do not appear to be orientation selective, at least not for the orientation of a moving bar. Whether they are selective for the orientation of bars defined by relative motion has not been determined. Cells that respond to random dot stereograms in Area V1 do not appear to be selective for the orientation of bars presented as random dot stereograms either (Poggio, Motter, Squatrito & Trotter, 1985). Nothdurft and Li (1985) did not find any cells in Area V1 that responded to texture bars. The explicit analysis of the orientation and size of regions defined by color, motion, texture, and binocular disparity therefore probably occurs beyond area Area V1. At these higher levels the shape primitives available for each stimulus attribute limit and, in a way, identify the algorithms the visual system can use to represent shapes, to derive position-, size-, and orientation-invariant descriptions that form the basis for memory and recognition (Cavanagh, 1984, 1985).

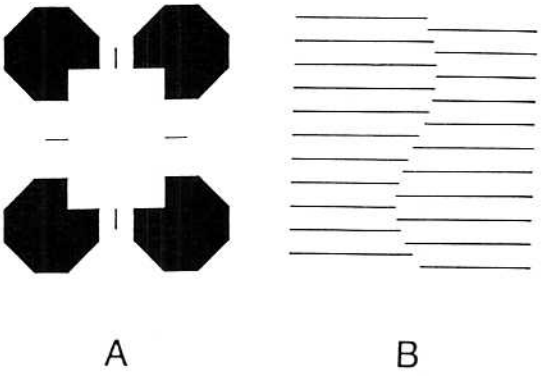

FIG. 10 Examples of subjective contours. (A) Brightness enhancement within a rectangular area delineated by subjective contours. (B) Subjective contour without brightness enhancement. From Kanizsa (1979).

B. Neural Coding of Subjective Contours

This section discusses the neural coding of subjective contours, as described by von der Heydt and colleagues (Peterhans & von der Heydt, 1989; von der Heydt & Peterhans, 1989a; von der Heydt, Peterhans & Baumgartner, 1984). The link between neurophysiological data and psychophysical observation is particularly close in this area, while subjective contours themselves reveal the algorithms employed by the brain in attempting to represent the visual world.

Contours can be perceived in the absence of any real edges or lines (Fig. 10). Such contours have been called "virtual," "anomalous," "illusory," "quasi-perceptual," "subjective," and "cognitive" (Schumann, 1900; Kanizsa, 1955, 1979; Lawson & Gulick, 1967; Coren, 1972; Gregory, 1972). The term "subjective" contours is preferred here because they have no direct physical analog in the stimulus, that is, a brightness step is perceived although there is no step in luminance. The ability to perceive such contours is likely to be useful under normal conditions of vision. Illusory figures, such as those in Fig. 10, demonstrate that the system has the means to reconstruct contours from cues other than luminance differences.

One difficulty in investigating subjective contours by recording from single cells is the interpretation of the response. Since these contours can only be generated by arrangements of ordinary lines or edges, how can we distinguish whether a neuronal response signals the subjective contour or the stimulus elements generating it? One possible approach is to use two abutting gratings as in Fig. 10B. In this way a vertical illusory contour can be produced by merely presenting horizontal lines. If a cortical neuron that is selective for vertical orientation responds to such a stimulus of purely horizontal lines, its responses can be presumed to relate to the perception of the subjective contour.

Figure 11 shows an example of responses of a single cell recorded in Area V2 of a rhesus monkey (von der Heydt & Peterhans, 1989a). The animal was trained to fixate a small target while the stimulus was presented near the fixation point. The cell responded to the abutting gratings (Fig. 11B) at the same orientation of the subjective contour at which the bar (Fig. 11A) was also effective. A continuous grating presented as a control elicited no response (Fig. 11C). Thus, one may assume that this kind of cell signals an illusory line and may actually mediate its perception although it is not physically present. Similar responses were found in about one-third of the

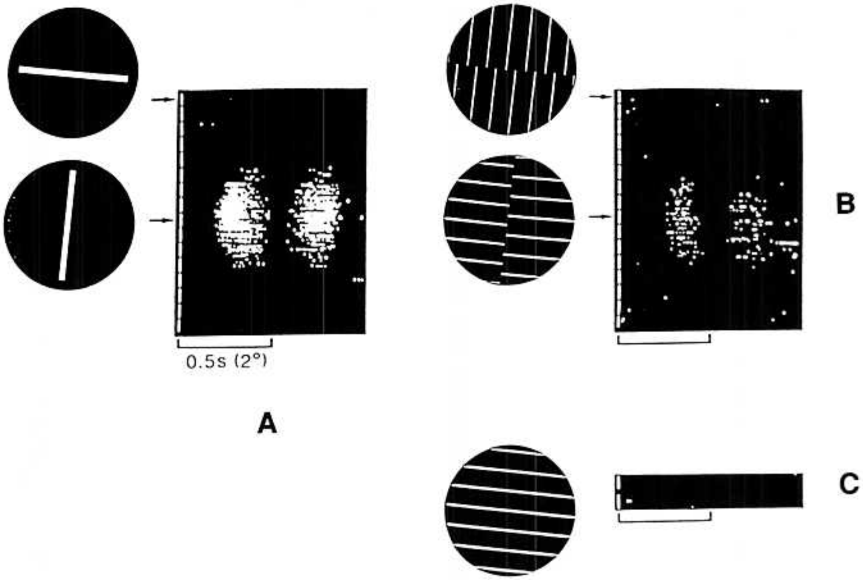

FIG. 11 Representation of subjective contours at the neuronal level. Responses of a cell in cortical Area V2 of an alert monkey. While the monkey was looking at a stationary fixation target, a light bar, or a subjective contour produced by abutting gratings, was moved back and forth across the receptive field at 16 different orientations covering 180° (only two are illustrated). The rows of dots represent the responses. (A) Bars elicited responses selectively at orientations near the vertical. (B) With abutting gratings, the cell responded according to the orientation of the subjective contour. It did not respond to the grating of switching phase (C). From von der Heydt and Peterhans (1989a).

cells in Area V2. Additional tests showed that the line ends by themselves were not effective stimuli; see Fig. 12.)

In a natural scene, when a number of lines terminate at a subjective border forming a straight line or a smooth curve, the border is usually perceived as the contour of an object. A possible inference is that the lines terminate because they are partially occluded by the object; the greater the number of aligned terminators the more likely a contour is present. The perception of subjective contours will therefore usually match the real contours of an object. The perceived contours become more salient as the number of terminators is increased (Fig. 12, top). No contour is perceived at the end of just one line; however, a weak contour begins to appear with three or four lines and becomes even more distinct with greater numbers. Apparently the perceived strength of the contour increases gradually with the probability of an occluding contour. The responses of a cell in Area V2 to such a stimulus is shown in Fig. 12. The increase in response parallels the perceptual salience of the contour.

Neurons were also tested with other illusory figures, such as the well-known Schumann pattern (Fig. 13B), in which the subjective contours have orientations in common with the inducing elements. Here, it is necessary to distinguish between a neural representation of the subjective contours and the response to the edges inducing them. A possible solu-

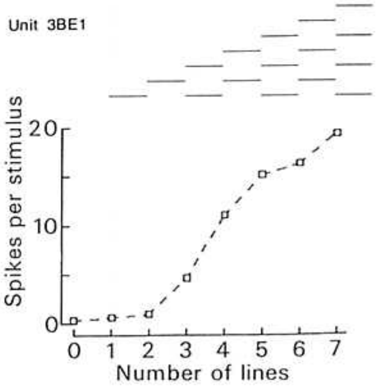

FIG. 12 The dependence of subjective contour and neuronal response rate of a cell in Area V2 of the monkey on the number of lines inducing the contour. In the experiment, the stimuli were centered over the cells receptive field. From von der Heydt et al. (1984). Copyright 1984 by the AAAS.

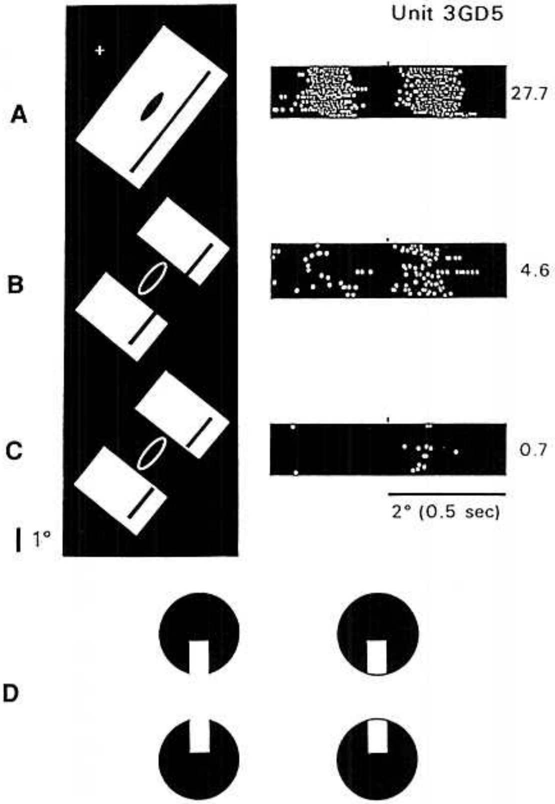

FIG. 13 Figure completion and the effect of closure in a cell of monkey Area V2. Ellipses indicate the cell's response field, i.e., the region outside which a bar evoked no responses, a cross marks the monkey's fixation point. (A) The responses to a moving light bar in a stationary dark rectangle, at the cell's preferred orientation. (B) When the central part, including the response field, was blanked out, the cell still responded regularly, although weakly, just as if a bar of low contrast were passing its response field. A moving illusory bar could be perceived crossing the gap. (C) When the ends of the bar were closed by thin lines, the responses were abolished just as the illusory bar disappeared (C). (D) Stationary figures that demonstrate a similar perceptual effect. The gap was 2° wide, the closing lines 2 minutes of arc. From von der Heydt, Peterhans and Baumgartner (1984). Copyright 1984 by the AAAS.

tion is to use a slightly modified figure which consists of similar elements but produces weaker or no subjective contours. Such a control figure is obtained by sealing off the ends of the upper and lower halves of a real bar with thin lines (Fig. 13C). Such closure abolishes the neuronal response and also eliminates the perception of the illusory bar (Fig. 13D). This result is particularly interesting, since it shows that very small changes in luminous flux made in the appropriate location can alter the neural response drastically. About one-third of the cells in Area V2 were found to signal subjective contours, whereas cells in Area V1 seemed to be "blind" to such contours and did not respond.

Many cells of V1 responded, of course, to the stimuli that produced subjective contours, but only if real edges or lines passed over the cell's response field at the preferred orientation. Thus, we can say that responses in Area V1 correlate with the perception of physically present contour, whereas many cells in Area V2 show an ability to construct or extrapolate contours in plausible ways, an ability which some have even deemed "cognitive" (Gregory, 1972; Coren, 1972).

Subjective contours have also been investigated in the cat visual cortex (Redies, Crook & Creutzfeldt, 1986). A border between abutting gratings similar to that of Fig. 10B was used. It was found to generate responses in complex cells of Area 17 as well as 18, but not in simple cells. The criterion was the continuity of the representation in a two-dimensional scanning plot of the responses. Whereas simple cells and end-stopped cells (if they responded at all) responded only to the line ends, complex cells brought out a continuous line. These results from cat Area 17 contrast with those from Area V1 in the monkey where orientation-selective cells in general did not respond at all to lines or patterns of lines orthogonal to the cell's preferred orientation (von der Heydt, Peterhans & Baumgartner, 1984). Nothdurft and Li (1985), using the same technique of analysis as Redies et al., found that complex cells in cat Area 17 do not make isoluminant texture borders explicit. With textures of short parallel lines, for example, the border between regions of different line orientation was not signaled as an edge or line would have been signaled.

If one can draw conclusions about cortical processing in humans from experiments in monkeys, the results described above cast new light on old questions of contour perception. First, the finding that cells in the primary visual cortex (V1) generally fail to signal subjective contours is contrary to the common assumption that a peripheral mechanism, such as lateral inhibition, plays a significant role in the perception of subjective contours (Brigner & Gallagher, 1974; Frisby & Clatworthy, 1975). It also makes it unlikely that subjective contours can be accounted for in terms of filtering out certain spatial Fourier components of the stimulus, as has been suggested by Ginsburg (1975) and Becker and Knopp (1978). Although the simple cells of Area V1 can be thought of as oriented spatial filters, they show no response to subjective contours. Furthermore, the neurons of Area V2 that do signal subjective contours are very sensitive to tiny changes in configuration (Fig. 13). They also respond vigorously to the border between gratings that is not represented in the Fourier spectrum at all. [Integrating the luminance along lines parallel to the border gives the same, constant value on either side of the contour; hence convolution with sinusoidal gratings of this orientation yields zero (see also Kelly, 1976; Parks & Pendergrass, 1982).] These findings are again hard to explain by the filter hypothesis.

Theories of subjective contours are often based on high-level inferences of occluding surfaces. Gregory (1972) and Rock and Ansom (1979) suggest that an occluding surface is hypothesized to simplify the interpretation of the stimulus. Such theories belie the fact that neuronal signals which might give rise to the perception of subjective contours are found as low as in Area V2, indicating that their perception need not be the result of higher level processing. The neurons in Area V2 are still mainly stimulus dependent, and the responses are stereotyped and reproducible, even in an alert animal. If the perception of subjective contours resulted from an attempt to reconcile the available stimulus with our perceptual postulate of it (Gregory, 1972), the observation that the strength of the illusion gradually increases with an increasing number of elements would be difficult to explain. Perceptual postulates cannot be gradual. On the other hand, neuronal mechanisms involving cooperative networks (Grossberg & Mingolla, 1985a,b; see Chapter 16, this volume) provide a plausible explanation for these phenomena. The finding that behaving cats might perceive illusory figures similar to the Kanizsa triangle (Bravo, Blake & Morrison, 1988) may also be interpreted as showing that cognitive processes need not be involved. Nonetheless, the possibility of top-down influences from higher areas cannot be ruled out completely.

To explain subjective contours in illusory figures, Peterhans, von der Heydt, and Baumgartner (1986) suggest a simple feed-forward mechanism that is compatible with the neuronal data and requires few assumptions. Their model assumes a convergence at the level of Area V2 of two parallel paths of computation, one for edge detection and one for detecting configurations formed by a series of terminators (e.g., line ends and corners). In this model, simple and complex cells constitute the edge-detecting path, while end-stopped (hypercomplex) cells contribute to the grouping path. A set of such receptive fields is lined up in a row according to the orientation signaled by the target cell, while the single fields are predominantly orthogonal to it (Fig. 14). This scheme explains, for example, the dependence of subjective contour on the number of lines (by summation in the grouping path) and the effect of closure (by the stimulation of inhibitory end zones). It also interprets the neurons with end-stopped receptive fields in a new way: their primary function would not be the detection of object features, but the evaluation of partially occluded patterns resulting from interposition (see also Redies, Spillmann & Kunz, 1984).

Several predictions follow from this model. Since the end-stopped fields of the grouping path are themselves orientation sensitive, the orientation of the inducing elements in subjective contour figures should influence the neuronal response. Subjective contours should be strongest when induced by orthogonal lines and less strong with oblique lines. This is indeed so (Fig. 15B; Kennedy, 1978). The perceived orientation of the subjective contours should also be biased toward an orientation orthogonal to the inducing lines. This is what one finds. The depth or overlay which is frequently perceived in subjective contour figures (Coren, 1972; Gregory & Harris, 1974; Coren & Porac, 1983) and the brightness enhancement in the

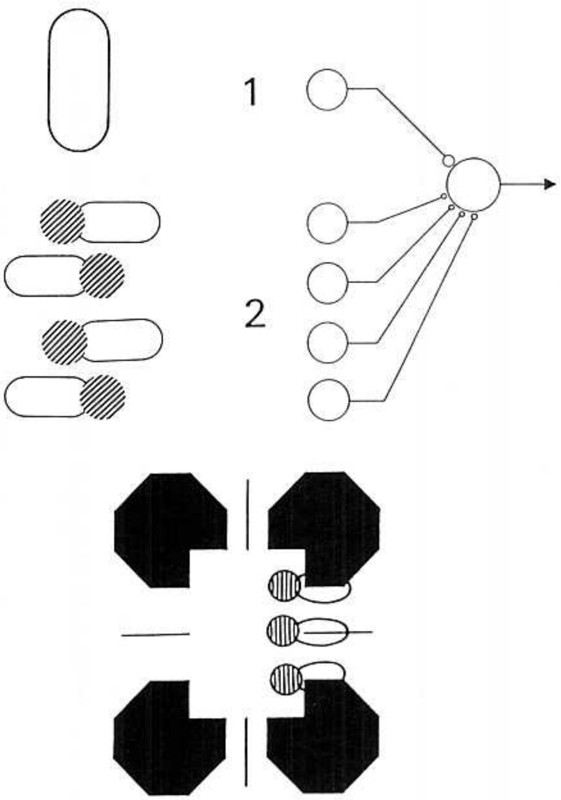

FIG. 14 A schematic representation of a neuronal mechanism that might underly the contour-related signals found in V2. A single cell of V2 is assumed to have input via two parallel paths, (1) from simple or complex fields ("edge detecting path"), and (2) from a set of end-stopped fields ("grouping path"). Both sets of fields overlap on the same patch of retina. Pairs of end-stopped fields straddling the center are connected by multiplication (×) with the effect that at least two of the end-stopped fields have to be excited simultaneously to produce a signal at the output, and the results are summed together with the edge signal. The figure at the bottom shows how the grouping path may be activated by a figure eliciting the perception of a subjective contour. After Peterhans, von der Heydt and Baumgartner (1986).

Kanizsa and Ehrenstein illusions (Fig. 10A and 15A) can similarly be explained by this hypothesis (von der Heydt & Peterhans, 1989b).

A neuronal mechanism derived from single-unit experiments with subjective contours can be interpreted as the computational solution to the problem of defining the contours in images of three-dimensional

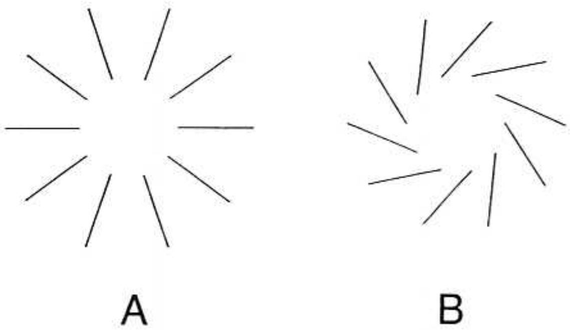

FIG. 15 (A) Subjective contours and brightness enhancement induced by line ends are most vivid when the contour runs at right angles to the lines; (B) with the inducing lines oriented tangentially there is no effect (Ehrenstein, 1941; Kennedy, 1978).

scenes without prior knowledge about the objects in the scene, a problem that can be solved only by exploiting the statistical properties of occluding contours. This hypothesis explains not only the perception of subjective contours, but also the brightness and depth illusions associated with them.

What does the visual system do with subjective contours once they have been extracted? There is a wealth of psychophysical evidence to suggest that they constrain many other aspects of perception such as stereopsis (Ramachandran, 1986a), apparent motion (Petersik, Hicks & Pantle, 1978; Ramachandran, 1985), and color (Ramachandran, 1987a,b). One wonders whether the many extrastriate visual areas (Van Essen, 1985) are somehow involved in mediating these interactions. Understanding the manner in which different visual "modules" interact is one of the great challenges for the future.

V. REPRESENTATION OF SHAPES

Once the visual scene has been parsed into the global areas that share particular features and that are likely to belong together as part of the same real objects, the three-dimensional shape of each area must be determined. Perceptual groups and areas can be defined by a number of stimulus attributes including luminance, color, motion, texture, and stereopsis, and the coding of shape might occur independently for each of these attributes. Using a variety of behavioral tasks, Cavanagh (1987, 1988) and colleagues explored the possibility that the analysis of each of these five attributes constitutes an independent "perceptual pathway." Cavanagh started from a larger set of potential pathways than is suggested by either physiological or neuropsychological data. Physiological studies show clear evidence for two independent streams, color/form and motion (Maunsell & Newsome, 1987) while neuropsychological reports of brain-damaged patients suggest only three physically separate and dissociable analyses, luminance (Rovamo et al., 1982), color (Damasio et al., 1980), and motion (Zihl et al., 1983). The five attributes that Cavanagh examined are meant to be exhaustive in that all other properties of the visual image (size, orientation, form, etc.) are properties of the spatial representations of one or more of these five attributes. Given that these attributes are only potential candidates for independent analysis, each of them was tested in several experiments for evidence of actual independence.

A. Evidence for Separate Pathways in Shape Perception

The first research on the perceptual capacities of individual pathways was the pioneering work by Julesz (1971) on images defined only by binocular disparity. [Prestriate Areas V3 and V3A may be involved in stereopsis (Zeki, 1979), possibly along with Area MT (Van Essen, 1985).] Julesz (1971) not only studied the sufficient conditions for the perception of depth in random dot stereograms, but, more importantly, he examined what qualities could be seen with images defined in this manner when no monocular cues were present. From Section II we know that the analysis of binocular disparity leads not just to the extraction of depth but also to the representation of the shapes of regions defined by their different depths. Julesz examined whether such shapes could produce classical visual illusions, identifiable letters, and various other perceptual phenomena. This approach has been extended to additional pathways, and comparisons have been made across these pathways (Cavanagh & Leclerc, 1989; Cavanagh, 1987).

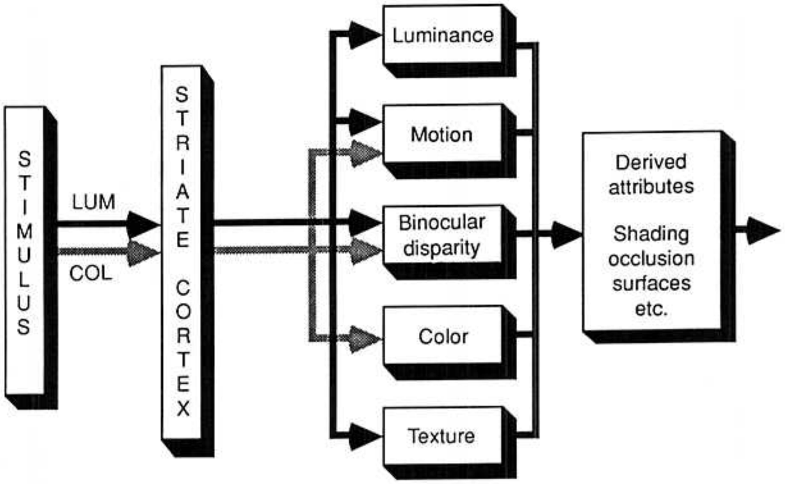

FIG. 16 Perceptual pathways in the visual system. Luminance (nonopponent) and color (opponent) pathways carry information from the retinal ganglion cells to the striate cortex where multifunction cells begin the analysis of orientation, motion, and binocular disparity. Beyond the striate cortex, information is routed to areas performing specialized analyses of various attributes such as color (Area V4; Zeki, 1978a,b) and motion (Area MT; Van Essen, 1985). Luminance, binocular disparity, and texture are other stimulus attributes that may receive specialized analyses in separate areas of extrastriate cortex. Each of these specialized areas generates a two-dimensional representation of the attribute being analyzed, contributing to an overall representation of stimulus shape. These independent representations may be followed by one or more high-level areas that recombine information from all attributes.

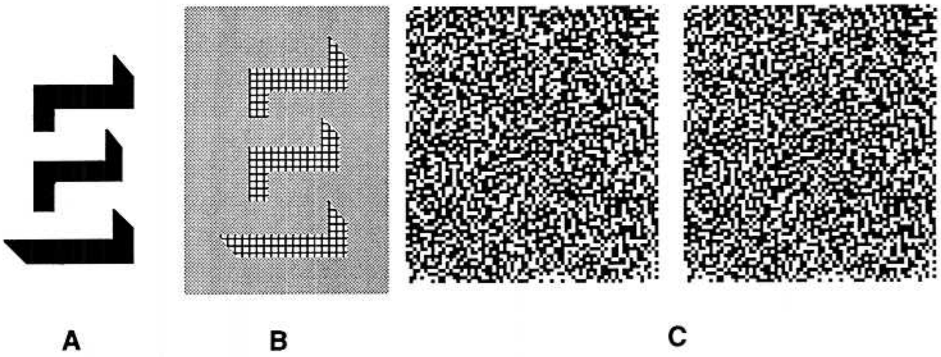

FIG. 17 Example of a shadowed block letter, E, represented by three different attributes: (A) luminance, (B) texture, and (C) binocular disparity. Most observers report the three-dimensional interpretation of the figure only for the leftmost stimulus defined by luminance. From Cavanagh (1988).

The studies described here examine perceptual pathways for five stimulus attributes (Fig. 16): color, luminance, texture, binocular disparity, and relative motion. To construct images defined by a single attribute, a black and white figure (Fig. 17A) was modified by replacing the black areas with, for example, a random texture moving in one direction, and the white areas with a similar texture moving in the opposite direction. Figure 17B,C demonstrates this attribute replacement for texture and binocular disparity.

In addition to the specialized analysis that is characteristic of each particular pathway, each possesses some degree of retinotopy in the underlying physiological structure (Maunsell & Newsome, 1987), and therefore each is inherently capable of representing two-dimensional contour. It is this common ground of two-dimensional shape representation that permits comparisons of coding and image analysis to be made across the pathways.

The first question addressed by these experiments is the type of shape coding that occurs at this low level. To segment image areas defined by a particular attribute, the visual system only needs to distinguish one region from another, which it could do very well by simply coding the value of the attribute (say, color) at each point in a retinotopic map (Fig. 18). We know, however, that in the case of luminance the visual sys-

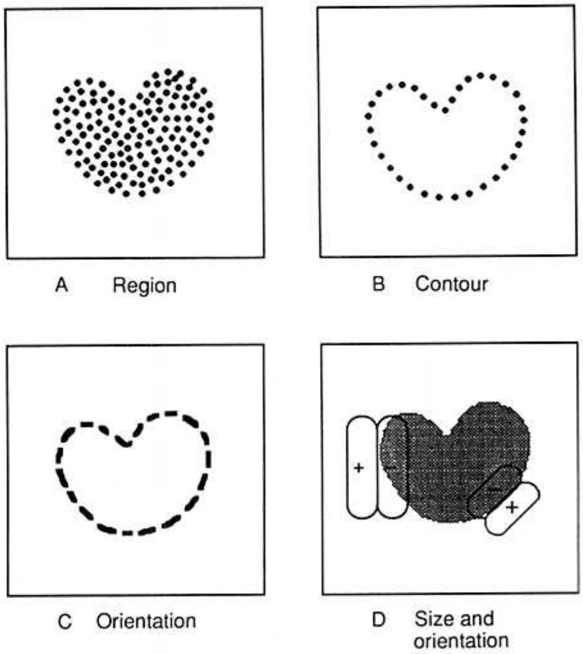

FIG. 18 Examples of the use of shape primitives to encode a two-dimensional region. (A) Point-by-point encoding of the attribute value preserves the region's shape. (B) Coding of only the region's contour preserves the shape and reduces the number of units needed to represent the shape. (C) Coding the contour with oriented tangents can improve the estimate of the region's outline if it is in a noisy presentation. (D) Overall shape can be decomposed into size- and orientation-specific units.

tem goes beyond this simple coding and extracts local structure (shape primitives) directly with cells that are selective to orientation and size (e.g., Hubel & Wiesel, 1968) and perhaps curvature (Dobbins et al., 1987). The visual system may use similar shape primitives for all attributes.

The second issue addressed by the experiments is the ability to use the two-dimensional shapes of regions defined by different surface attributes to infer three-dimensional object structure through cues such as shadows, occlusion, perspective, and contour junctions (Fig. 19). If this recovery of three-dimensional structure (three-dimensional shape from two-dimensional shape) is a higher level process that operates equally well on shape descriptions from any pathway, then the "shape from shape" approaches should work for two-dimensional shape defined by any surface attribute that the visual system can distinguish. For example, we should be able to interpret an ellipse as a tilted circle whether the ellipse is defined by color (red on green of equal luminance), by a random dot stereogram, or by luminance (black on white).

Finally, composite images were used to demonstrate the extent of cooperation between the pathways in producing geometrical illusions, stereoscopic depth perception, and apparent motion. These studies exploited the multiple pathways to localize various processes within the visual system. For example, if we can read a letter E whose horizontal bars are defined only by color and whose vertical bar is defined only by relative motion, then the identification process must have access to the combined information of the separate pathways.

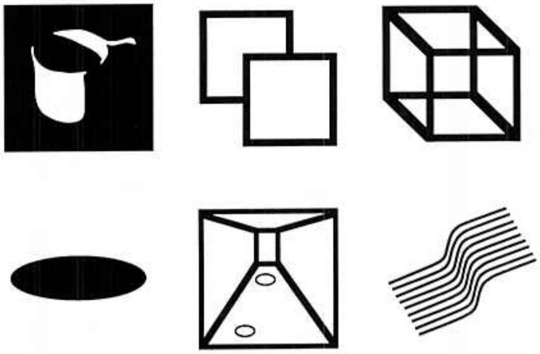

FIG. 19 Examples of stimuli whose three-dimensional interpretation is derived solely from two-dimensional shape cues. The examples do not have any gradient cues (shading, optical flow, or texture gradients) to convey three-dimensional information, but instead rely on shadows, occlusion, and perspective. Adapted from Cavanagh (1987) and Cavanagh (1988).

B. Use of Aftereffects to Infer Coding of Shape Primitives

When a stimulus sharing a particular property is viewed for some period of time, it induces an aftereffect which alters the threshold or the appearance of test stimuli viewed after the adaptation period. Figure 20 illustrates how adaptation to a fine grating makes a test grating appear coarser and vice versa. Similarly, adaptation to a grating tilted to the right from vertical makes a vertical stimulus appear to tilt in the opposite (to the left) direction and vice versa. Size (Blakemore & Sutton, 1969) and tilt aftereffects (Campbell & Maffei, 1971), as well as simultaneous induction paradigms (Georgeson, 1973; Klein, Stromeyer & Ganz, 1974), have been used to infer the existence of size and orientation coding in the luminance domain. The same procedures can be applied to stimuli defined by other attributes. Favreau and Cavanagh (1981) demonstrated that the color pathway has size-specific coding; Elsner (1978) showed a tilt aftereffect for equiluminous, colored stimuli; and Tyler (1975c) reported size and tilt aftereffects for random dot stereograms.

Using the size aftereffect paradigm, Favreau and Cavanagh (1981) induced simultaneous and opposite size aftereffects for color and luminance stimuli, suggesting these channels are separate. More recently they were able to make the same demonstration for the tilt aftereffect. It makes sense for the visual system to provide the color pathway with a rich set of shape primitives since shape analysis based on color information is frequently superior to luminance analy-

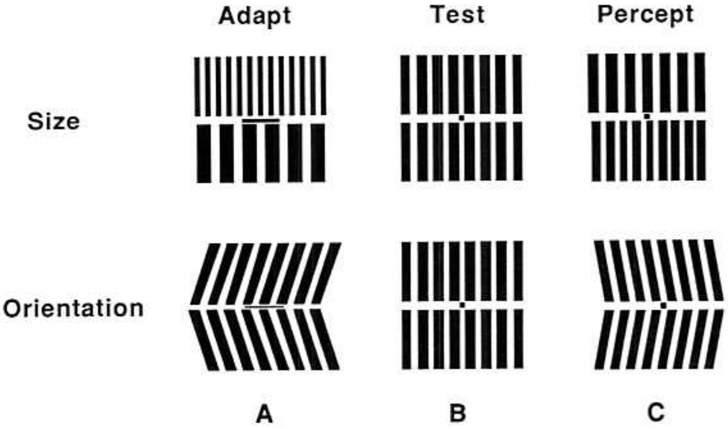

FIG. 20 Size and tilt aftereffects have been used to infer size and orientation coding in the visual system. Observers adapt to stimulus A for about 30 sec while scanning the central fixation bar. They are then asked to judge the relative size or tilt of stimulus B. Observers typically report that the central figures appear to have changed size (top) or tilt (bottom) as shown in C. From Cavanagh (1988).

sis. Color edges reliably move with the objects, while luminance edges are often confounded by the clutter of shadows unrelated to the objects across which they fall.

Tilt aftereffects were measured in each pathway using a standard induction and test procedure. Observers were first exposed to adapting gratings whose bars were tilted 15° from vertical and defined by relative motion. They were then presented with a vertical test also defined by relative motion (Fig. 20). When the test was presented, the observer matched its apparent tilt by adjusting a comparison stimulus presented in an unadapted region of the visual field. A tilt aftereffect of about the same magnitude was found for each type of stimulus. These data suggest that each pathway may perform a similar analysis of orientation.

The tilt or Zöllner illusion was also used to evaluate orientation coding in each pathway. In this illusion, parallel lines appear tilted away from crossing lines depending on the angle of intersection. This interaction has been claimed to result from inhibition between the orientation signals of the lines. If orientation coding exists for attributes such as color, texture, binocular disparity, and relative motion, then the illusion should persist when the stimulus is defined by each of these attributes. Indeed, the induced tilt was evident for all presentations and had approximately the same value, supporting the notion that orientation processing is ubiquitous in all pathways.

C. Visual Search in Separate Pathways

Evidence from visual search supports the inferences drawn from experiments using adaptation and aftereffects. Cavanagh, Arguin and Treisman (1990) used the search paradigm described earlier to determine whether size and orientation are privileged features, not only when defined by luminance differences, but also when defined in other pathways such as motion, color, and texture. Stimuli consisted of a large disk among small disks (size task) and a tilted bar among vertical bars (orientation task). They were defined by each of the five stimulus attributes. For the orientation task reaction time required to detect the presence of the target was independent of the number of distractors for all stimulus types except binocular disparity, indicating again that orientation may be a basic shape primitive in at least four of the five pathways tested. For the size task, reaction times were unaffected by the number of distractors for stimuli defined by luminance, color, texture, and relative motion but did show a significant increase for stimuli defined by binocular disparity (see also Nothdurft, 1985a,b). These data suggest that size coding may occur for four of the five pathways; however, further experiments may extend this to all five. In particular, experiments on size aftereffects by Tyler (1975c) do show adaptation for stimuli defined by binocular disparity, indicating that size is coded for this attribute. In summary, orientation and size coding appear to be very common in the visual system and may be used as shape primitives for each of the five attributes tested.

D. Shape Imaging Capacities of Individual Pathways

What are the elementary processes that underlie "shape-from-shape" inferences? Cavanagh (1987) has attempted to identify these processes by discovering the common aspects of a range of tests that *fail* in one pathway but not another. Tests are made with stimuli such as the Necker cube, perspective drawings, geometrical illusions, and subjective contours with each figure defined by a single attribute. Since the pathways other than luminance have poor spatial resolution (see Chapter 6, Fig. 7), it is necessary to use highly visible figures with no fine detail. Otherwise the inability to perform a particular task may be due simply to the reduced resolution or contrast inherent in a particular representation.

1. Figures with Explicit Contours

Information signaled by stimuli having explicit contours, for example, T-junctions indicating occlusion, was effective no matter which visual pathway was used to present them (Fig. 21). Simple, two-dimensional letter shapes could be easily identified, and objects defined by complete contours in line drawings involving occlusion and perspective gave rise to similar three-dimensional interpretations whether the line drawings were represented by luminance, color, or texture. When relative motion or random dot stereograms were used to present these same drawings, the depth suggested by two-dimensional occlusion and perspective sometimes conflicted with the depth indicated by the relative motion or binocular disparity used to present the figure. Many observers could see the depth implied by the drawing despite the conflict, although others could not. In cases where there was no conflict between the depth inferences in the picture and the depth used to present the picture (based on either relative motion or binocular disparity), the pictures were interpreted in the same manner as for luminance, color, or texture presentations. Shape information involving explicit object contours therefore appeared to be represented equally well in all of the five pathways. The depth and surface inferences based on these shape representations probably occur after the level of separate pathways and accept shape descriptions from any pathway. There was no indication that luminance information played any privileged role in these images.

2. Figures with Implicit Contours

The results for stimuli involving implicit contours were strikingly different. Figure 22 shows the two stimuli studied: subjective contours (Gregory, 1977) and shadows (Cavanagh & Leclerc, 1989). In both cases, a luminance difference was necessary between figure and ground for three-dimensional shape to be recovered. If the parts of the stimuli were presented without a luminance difference, they were interpreted as separate, unconnected islands of color or texture (see Liebmann, 1927). If a luminance difference was then introduced, the overall global organization of the stimulus became visible. Thus, the luminance pathway is essential for shadows and subjective contours, but it appears that it is the edges that are signaled at this low level and not the entire shadow region or subjective surface.

It might seem self-evident that shadows would require luminance information to be properly interpreted: a real shadow is always darker than the adjacent nonshaded region. Shadow analysis may therefore be part of the specialized processing in the luminance pathway. However, ignoring shape information in other pathways, the visual system must give up opportunities to reject areas as shadows because of "impossible" colors or inappropriate depths, motions, or textures. Indeed, observers saw depth in

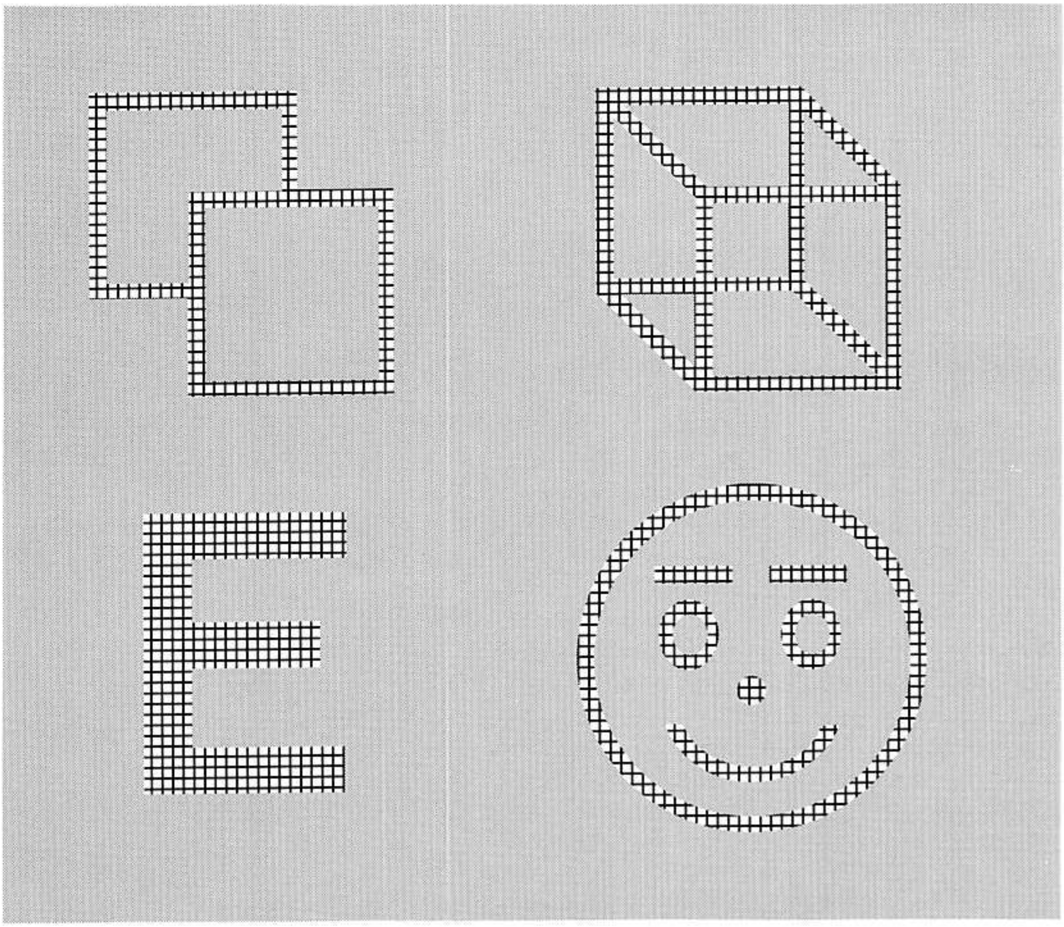

FIG. 21 Stimuli with explicit contours. The stimuli are defined only by a difference in texture and should be invisible if you squint your eyes. In a Necker cube perceived as a wire frame the background can be seen through the areas between the cube contours; in an occlusion figure consisting of one square covering one corner of a similar square, the areas between the contours hide the background. In general, the same perceptions are reported for these stimuli whether defined by luminance, texture, color, relative motion, or binocular disparity. However, large individual differences are sometimes found in the latter two cases. These occur when there are conflicts between depth suggested by the picture and depth defined by the relative motion or binocular disparity. From Cavanagh (1988).

shadow images even when they violated the color, depth, motion, and texture constraints of natural shadows (Cavanagh & Leclerc, 1989).

The coding of subjective contours by pathways other than luminance is doubtful. Brussell, Stober, and Bodinger (1977), Gregory (1977), and Prazdny (1985a) all report that subjective contours were visible only when there was a luminance difference between the regions defining the shapes. However, Kellman and Loukides (1987) and Prazdny (1987) were able to create subjective contours in the motion pathway using kinetic stimuli without any luminance contrast.

3. Illusions

Gregory (1977) evaluated several illusions for figures defined by color and Julesz (1971) for figures defined

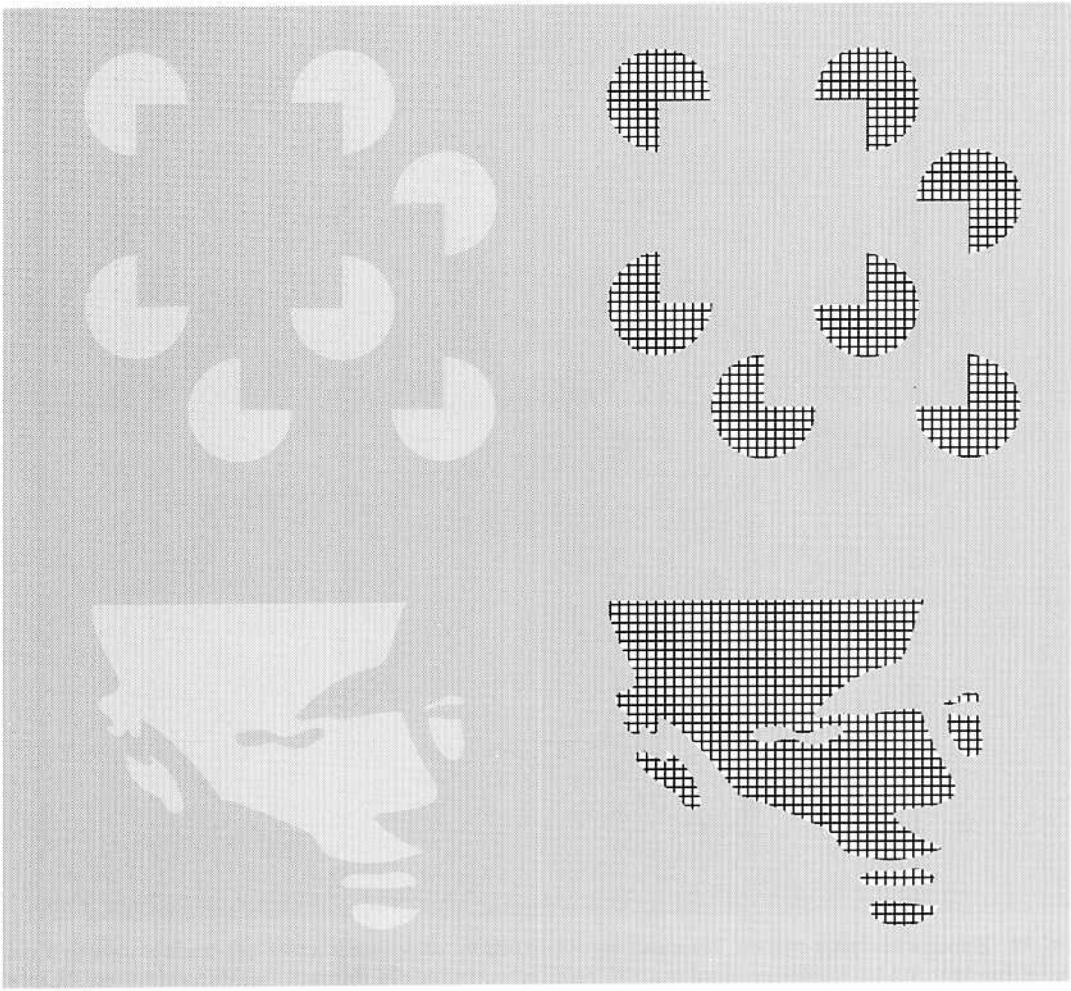

FIG. 22 Stimuli with implicit contours. Top row: subjective contours. When there is a luminance difference between the inducing elements and the background, subjective contours can be seen (left). When the same figure is presented without a luminance difference (right), subjective contours are weak or absent. Bottom row: shadows. Many of the contours in the figure (left) are shadow contours, not object contours, and many of the object contours, both external and internal self-occlusions, are implicit. The interpretation of the figure as a face changes when no luminance difference is present (right). The interpretation of surface relief due to shadows occurs for stimuli defined by luminance and not for stimuli defined by texture (shown here), color, relative motion, or binocular disparity.

by binocular disparity. Both report that the illusions remain undiminished with two exceptions, the café wall illusion and Fraser spiral, which Gregory discusses in detail. The strength of two additional illusions that Cavanagh tested also remained undiminished. In the Ponzo illusion a horizontal bar which appears closer also appears shorter than an equally long bar which appears to lie farther in the distance and in the horizontal–vertical illusion, a vertical line appears longer than an equally long horizontal line. Both illusion figures were presented in each of five stimulus dimensions (luminance, color, motion, binocular disparity, and texture), and both illusions showed the same strength for all presentations. The perspective and spatial scaling processes involved therefore appear to operate on shape codes that are available in each pathway.

E. Interpathway Cooperation and Localization

Interocular transfer and dichoptic viewing are useful techniques for localizing processes in vision, identifying whether the locus of origin of a given effect precedes or follows the emergence of binocular units. The multiple pathways of the extrastriate cortex may provide a similar opportunity. If certain analyses of shape operate only within independent pathways whereas other higher level analyses can operate on information from any pathway, the identity and locus of visual processes can be revealed using "composite" stimuli where different components of the stimulus (adapt and test, or inducing and test elements) are defined by different attributes. If, for example, the orientation analysis that underlies the tilt illusion occurs only within each pathway, adaptation to tilted lines defined by texture should induce an apparent tilt in a test line also defined by texture, but not in a line defined by relative motion (Fig. 23). On the other hand, if depth from perspective is a high level process, the impression of perspective should be produced even by the relative angle between two lines defined by different attributes. Experiments of this type can provide guidelines for the processing hierarchy of the visual system. Specifically, a perceptual effect that is preserved in a composite stimulus is necessarily located at some stage beyond the independent pathways. They may be localized beyond the prestriate cortex where information from all the pathways is accessible and integrated (Fig. 16). An effect that does not survive in a composite image must therefore be located in one or more, but not all, of the independent pathways.

1. Illusions

Three illusions, horizontal–vertical, perspective (Ponzo), and tilt (Zöllner), were tested with composite figures. With two exceptions (color and stereo could not be combined because red–green anaglyphs were used to present the random dot stereograms), all possible combinations of five attributes were applied to these figures. There were 23 different versions, 5 "within" figures with both the inducing and the test portion of the illusion figure defined by the same attribute, and 18 "between" figures in which the two components were defined by different attributes. A "between/within" ratio was computed by dividing the average strength of each illusion for all "between" figures by the average strength for all "within figures". A value of 1.0 for this ratio indicates that the processes responsible for the illusion had access to a combined shape representation, whereas a value less than 1.0 indicates that the processes responsible must, at least in part, be located in the segregated pathways so that the integrated figure was unavailable to them.

Both the horizontal–vertical and Ponzo illusions had between/within ratios very close to 1.0, indicating that the processes underlying the illusions had access to an integrated image. Processes involved in spatial scaling of horizontal and vertical dimensions and scaling size with distance (the perspective cue in the Ponzo illusion) should be expected to operate on an integrated image since they must deal with scenes made up of objects defined by an arbitrary assortment of attributes. For the Zöllner illusion, on the other hand, the ratio was significantly less than one, implying that the orientation coding underlying this tilt illusion occurs independently in each pathway: a composite figure would therefore produce a weaker illusion than one defined by a single attribute (Fig. 23). The ratio was, nevertheless, also greater than zero, indicating that there is some interpathway interaction for orientation coding or alternately some orientation coding in a common high level representation.

2. Motion

Apparent motion can be perceived between two shapes defined by different stimulus attributes (Cavanagh, Arguin & von Grünau, 1989). Ramachandran, Rao, and Vidyasagar (1973a,b) reported earlier that motion can be seen between two patches defined only

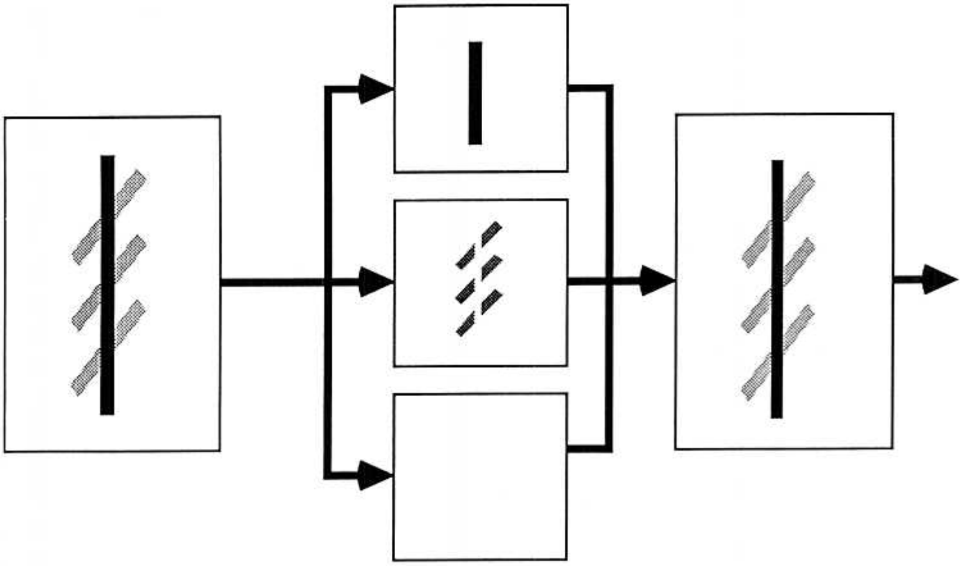

FIG. 23 Composite stimulus for inducing the tilt effect. Inducing bars are defined by one attribute and the test bar by another. On the left the stimulus is represented as a whole in the first cortical area, V1. However, at this level, no cells code for the orientation of bars defined by texture, relative motion, or binocular disparity (as a random dot stereogram). Therefore, no inhibition between orientations could be expected to produce the apparent tilt at this level when these attributes are involved. Similarly, in the center, the inducing and the test lines are represented in separate cortical areas (pathways) and thus have little opportunity to interact. Finally, on the right, in a representation where the individual attributes are recombined, a tilt illusion could occur if orientation is coded there.

by texture. Similar demonstrations have been made for stimuli defined by color (Ramachandran & Gregory, 1978) and binocular disparity (Julesz, 1971; Prazdny, 1986a,b). The demonstration that motion can be seen equally between patches defined by *different* attributes indicate that this motion process must have access to an integrated image. The strength of the motion, as measured by the maximum displacement over which the motion could be seen, was similar whether the two patches were defined by the same attribute or by different attributes.

3. Stereo

Similar studies of stereograms, where the figure presented to each eye could be defined by the same or different attributes, showed different and sometimes conflicting results. Binocular disparity between the monocularly visible patches seen by both eyes did produce a vivid impression of depth if the image in each eye was defined by the same attribute whether it was color, luminance, texture, or relative motion. Similar data were reported by Ramachandran, Rao, and Vidyasagar (1973a,b) and by Lee (1979). However, when different attributes were used in each eye, Cavanagh found that no depth was visible. This is not simply a matter of binocular rivalry; the images of the two eyes are also rivalrous in their luminance detail when both are defined by, say, relative motion, and yet depth is perceived in that case. Cavanagh concluded that, unlike the motion process, binocular disparity cannot operate on composite stimuli combining attributes across the eyes.

On the other hand, Ramachandran et al. (1973b) reported that one can obtain stereopsis by fusing a luminance border from one eye with a disparate texture border or chromatic border from the other eye, implying that different types of contours or edges may eventually be processed by a single mechanism in the brain that is indifferent to the manner in which the contour is constituted. Ramachandran et al. suggest that the purpose of such a mechanism would be to signal the presence of occlusion borders in the visual image and that there could even be a "channel" somewhere in the brain that is specialized exclusively for this role.

F. Conclusions on Separate Pathways Coding Shape

The studies reviewed here show that perceptual pathways can be probed with stimuli defined by luminance, color, relative motion, texture, and depth. Aftereffect, illusion, and visual search paradigms demonstrate that orientation coding is used for all the attributes. For the four attributes where evidence was available, size coding was also used. Simultaneous and opposing aftereffects showed that there are functionally independent streams of processing for luminance and color and that both perform similar analysis of size and orientation. The shape information available in each pathway was sufficient to support three-dimensional interpretations of images involving explicit contours but not for images involving implicit contours, namely, subjective contours and shadows. These images therefore appear to depend on a particular process or shape code available only in the luminance pathway. Composite images showed that the processes underlying perspective, the scaling of horizontal versus vertical dimensions, and form-based motion have access to high level representations where information from all the attributes is recombined. In contrast, orientation coding underlying the tilt illusion appears to be based in the individual pathways although some orientation coding may also occur in a high level representation. Finally, although depth could be perceived from form-based binocular disparity for each attribute, there is some doubt whether it could be achieved when the same form was defined by different attributes in each eye.

VI. PERCEPTION AND IDENTIFICATION OF OBJECTS

Once objects have been delineated by early processes of feature-based segregation and boundary formation, they must be *identified* as known individuals or instances of a familiar category. Identification seems to involve the use of two different strategies. The first is based on the use of particular, salient cues that directly label the object rather than on elaborate processing of all its parts. The second strategy allows us to extract spatial *relationships* between features (i.e., what is above, what is below). It is this second process that enables us to recognize pumpkin faces and snowmen in spite of locally misleading salient features (such as a carrot used for a nose). There is evidence that the two processes can be dissociated in visual agnosia (Humphreys & Riddoch, 1987) and in certain visual illusions such as the Thatcher illusion (Thompson, 1980). A frowning upside-down face is seen as a face because of the spatial relationships between features but it is seen as smiling because the features themselves are processed separately and can act as inborn ethological releasers.

A. Feature Integration

In most cases, object identification depends on comparing a complex conjunction of features to a stored representation. It is of interest to investigate how the visual system codes conjunctions, given the evidence for modularity of feature analysis described in Section II.

One important observation suggests that conscious experience and voluntary behavior necessarily depend on these later stages of perceptual representation. Although there is evidence that the analysis of color, motion, orientation, and other properties is modular at the early stages of coding (see Sections II, C and V), it is difficult for subjects to access separately the output of one module in order to control behavioral responses, while suppressing the outputs of other modules coding different properties of the same perceptual object. The Stroop test (Stroop, 1935) is a well-known example of such interference: In naming the colors of the ink in which words are printed, subjects are considerably slowed down if the words themselves are the names of other colors. The natural process of perception unites the outputs of the color and shape modules before they become available to introspection. In a task such as the Stroop test, in which the two modules evoke conflicting responses, a clear failure of selective attention is observed. We can select one object rather than another, but not one property of an object while excluding another property of the same objects (see also Kahneman & Chajzcik, 1983; Treisman, Kahneman & Burkell, 1983).

On the other hand, in tasks designed to tap the earlier stages of visual coding, the evidence suggests dimensional separation and modularity. In section II, we reported a number of results suggesting failures to conjoin or integrate different features: areas containing elements differing only in conjunctions of features do not spontaneously segregate or form perceptual boundaries, and targets in visual search are detected in parallel only when they have some unique defining feature not shared by the distractors. In this section the conjunction process is examined in more detail, together with the conditions under which features can be interrelated and structured to define perceptual objects.

Experiments on texture segregation suggested that elements are initially grouped within perceptual dimensions, forming, for example, a color map, an orientation map, a motion map, but not across dimensions. Thus, an element located between two groups and sharing a different feature with each group is less conspicuous and more often overlooked than an element with a unique feature (Treisman, 1982). A red O between a group of blue O's and red X's would be grouped with the red X's in the color map and with the blue O's in the shape map, and it would not appear at this level of processing to have an identity of its own. Targets in visual search tasks that are defined only by conjunctions of features (like the red O in the above example) are usually found through a serial process of checking and rejecting the nontarget items or distractors (Treisman & Gelade, 1980). The search time increases linearly with the number of distractors in the display, suggesting that attention must be focused to each item in turn in order to conjoin features correctly. The serial process could not be attributed to the heterogeneity of the distractors, because a target with either of two unique features (e.g., green or S) among the same heterogeneous distractors was detected in parallel. If the distractors were spatially grouped into homogeneous clusters (Fig. 24), search appeared to be serial across groups. However, the

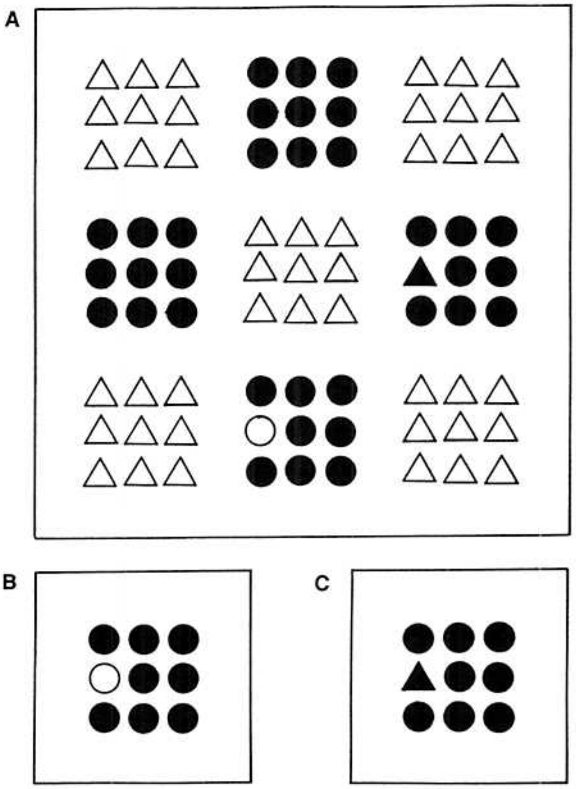

FIG. 24 (A) Display in which search for a conjunction target (e.g., a white circle or a black triangle) appears to be serial across groups but parallel within any one group. The items within each uniform group can receive divided attention, because within the group the target is defined by a single feature and will pop out, as it does in B and C. From Treisman (1985).

items within each group were checked in parallel (Treisman, 1982). In effect, the target could be distinguished from the distractors within any group on the basis of a single distinguishing feature. Similar results have been found with conjunctions of color with shape (curved versus angular as in O versus N or vertical versus diagonal components as in X versus T) and with conjunctions of parts of shapes (e.g., R's among P's and Q's).

Figure 25 shows the model proposed by Treisman (1986a,b) as a first attempt to account for the contrast between single features and conjunctions of features. The modular feature maps (as discussed in Section II) are linked to a master map representing their spatial locations. Attention is represented by a "spotlight" or a "window" (Kosslyn, 1987) selecting the features represented in a subarea of the master map. The features in this preselected location are integrated to form a unitary representation specifying their structural relations. The attention window is adjustable in size, depending on the precision required. Thus, when distractors are grouped into homogeneous clusters, the attention window can be set to accept each cluster as a whole. In a search task, attention scans the master map until all the distractors have been rejected.

The model distinguishes a representation of space (the master map of locations) from a representation of the features (e.g., color, orientation) that occupy locations in space (the feature map). As already men-

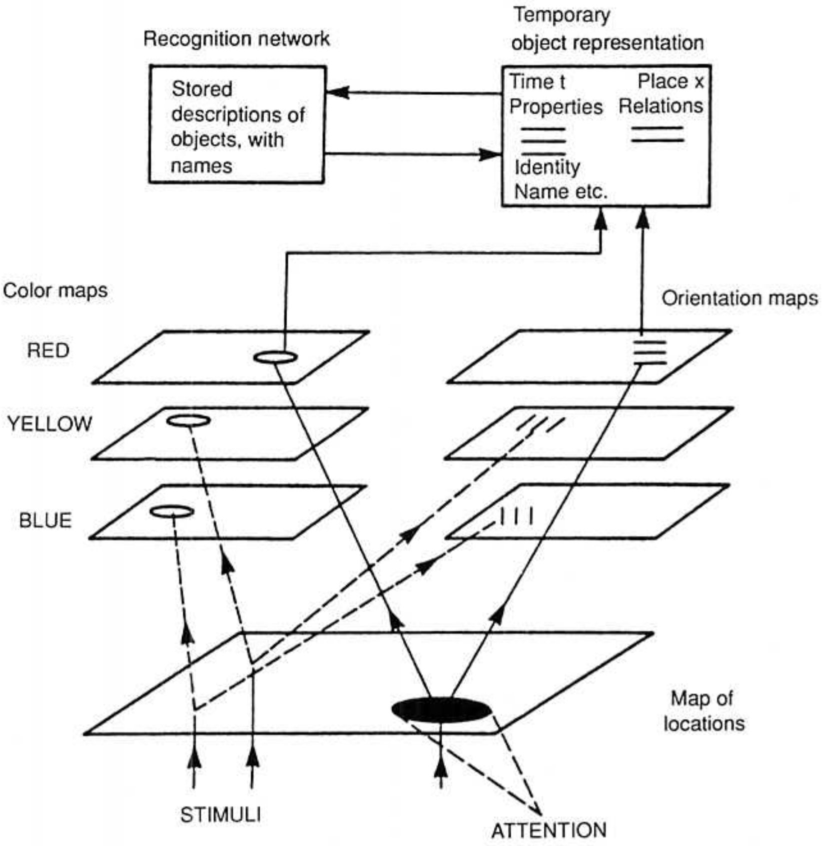

FIG. 25 General framework relating hypothesized feature maps to a master map of locations through which focused attention serially conjoins the properties of different objects within the scene. See text for details. From Treisman (1988).

tioned in Section II, a similar where-versus-what distinction has been drawn by neuroscientists (Mishkin, Ungerleider & Macko, 1983). Mishkin et al. have traced two separate pathways in the brain, one through the posterior parietal area, which appears to deal with spatial discriminations, and one leading to the inferotemporal cortex, which appears to deal with object recognition. Lesions to each pathway lead to quite different behavioral deficits. Other researchers have linked the parietal areas to spatial attention. For example, parietal lesions in human patients have been shown to lead to "neglect" of contralateral space (Jung, 1974; DeRenzi, 1982). Neurophysiologists (e.g., Robinson, Goldberg & Stanton, 1978; Mountcastle, Andersen & Motter, 1981; Wurtz, Goldberg & Robinson, 1982) have identified single units in the posterior parietal lobe of monkeys that respond only when the monkey attends to the stimulus. It is not inconceivable, then, that the parietal areas might play a role in integrating information relating to different attributes of objects, by controlling a serial scan of the locations in which they appear. A related hypothesis was proposed by Crick (1984), who attributed the selective and integrative role of the attention spotlight to facilitating bursts of activity from the pulvinar. The lateral pulvinar has strong connections to Area 7 of the parietal cortex, and cells in these areas appear to respond in similar ways to visual properties and to manipulations of attention (Robinson & Petersen, 1984).

Quite recently, however, the generality of the behavioral results with conjunction targets in visual search has been questioned. Nakayama and Silverman (1986) found that when the features defining a conjunction target are highly discriminable, the slope relating target search time to the number of distractors may become quite shallow (less than 10 msec per item) or even flat. This result suggests that conjunctions of features can sometimes be detected when attention is spread over several items at a time, or even over the whole display. The parallel detection of such stimuli is usually associated with a phenomenological impression of clear segregation between the two sets of distractors; these often appear to be in separate depth planes.

Treisman (1988) suggested that, with highly discriminable features distinguishing the two sets of distractors, attention can be directed to homogeneous subgroups of items as a whole (i.e., parallel access), even when they are spatially intermingled with other items. In terms of the framework in Fig. 25, this could be achieved by preselecting one complete set of distractor locations in the master map through inhibition (or activation) from the feature map that has this attribute. For example, if the target is green and horizontal among red horizontal and green vertical distractors, the locations of all red items might be selectively inhibited in the master map of locations, leaving activity mainly or only in the locations of the green items. These could then be checked with divided attention, since any remaining horizontal item would have to be the target.

Wolfe, Franzel, and Cave (1989) proposed a similar model and reported two new pieces of supporting evidence. (1) A target defined by a triple conjunction that differed in *two* features from each distractor was found more quickly than either a triple or a double conjunction that differed only in one feature from each distractor. For example, a large red cross gave very rapid parallel detection among large green circles, small red circles, and small green crosses. This would follow if inhibition coming from two feature maps could summate to reject each distractor location more effectively than inhibition coming from only one. (2) A target defined only by the spatial arrangement of its parts (e.g., a target T among distractor L's), could not be detected in parallel. This would follow from the hypothesis proposed, since neither feature of the distractors could be inhibited without also rejecting the target.

B. Temporary Object Representations

At the final stage of perceptual processing, objects are identified by matching them to stored descriptions of familiar objects. Many models of perception have equated "seeing" with "identifying" an object (LaBerge, 1975; Shiffrin & Schneider, 1977; Johnston & Dark, 1982). Identification is said to occur when the "nodes" standing for the object and for its properties are activated in a recognition network. The same idea appears in the physiological theories that equate perception with the activation of single "grandmother" cells (Barlow, 1972b) or assemblies of cells (Hebb, 1949). Recently discovered cells responding selectively to faces (Perrett et al., 1982) could be an example of such a perceptual model, at least for these highly significant biological stimuli.

However, in addition to the activation of nodes mediating recognition, one can argue for the existence of temporary object representations (see Fig. 25) that assemble the feature information in the correct structural relations and mediate "seeing" or perceptual experience (Treisman & Schmidt, 1982; Kahneman & Treisman, 1984). It is quite possible for us to "see" an unfamiliar object for which no prior representation has been established. We are also quite capable of seeing multiple replicas of an identical object when they are present, although each replica may take a measurable time to set up. The perceptual system must therefore form representations of "tokens" as well as of types. It must allow us rapidly to form one or many identical temporary representations of any arbitrary and unfamiliar conjunction of features in a visual field which may also contain a large number of other objects with potentially interchangeable properties. Moreover, it must be able to retain the perceptual identity and continuity of an object representation even when the object moves or changes its perceptual appearance.

Little research has been devoted to understanding these representations that presumably mediate our conscious experience. The separation of tokens (temporary object-specific representations of currently present perceptual objects) from types (the long-term stored representations abstracted from past experience and used for object recognition or labeling) has so far not found any physiological correlate.

C. Object Identification

The final issues, on which very little is known, concern the relationship between the temporary object representations and the stored descriptions to which they are matched (Fig. 25) as well as the matching operation itself. The object representations should, of course, specify three-dimensional solids in a three-dimensional spatial field rather than two-dimensional images. Their properties will therefore belong to a different vocabulary from the features of the image that reflect their real world origins only indirectly. Whether the features which are initially entered into the temporary object representations are the features of three-dimensional surfaces or whether the mapping is achieved only with focused attention once the features have been conjoined is still an open question. It seems plausible, however, that the visual system has evolved to extract information, even at the earliest stages, in a form which specifies the external world. Thus, rather than coding the properties of the retinal image as a physicist might (specifying the wavelength, the intensity, and the geometrical properties of the two-dimensional projection), the visual system may code such properties directly. Examples may be the gradients of texture and of motion that specify three-dimensional shapes and surfaces (Gibson, 1966; Braunstein, 1976) or the motion of edges in opposite directions that specifies "looming" in an approaching object (Regan & Cynader, 1979). Although these properties depend on relations across space and time, they need not be treated as conjunctions; specialized detectors may have evolved to code them directly. Such dynamic or relational properties actually appear to be detected and used earlier in infant development than static properties (Owsley, 1983; Kellman & Loukides, 1987).

Marr (1982) suggested that the representations which form the basis for perceptual recognition should be object centered rather than viewer centered. In other words, their descriptions should be independent of the viewing angle, distance, and illumination. The relations between the parts should be specified relative to the intrinsic axes of the object itself rather than of the visual field of the viewer. It is interesting to note, however, that the representation which we consciously experience as seeing is not object centered. Instead, perception is egocentric. Although the perceptual constancies of size, shape, brightness, color, and motion reduce the variations in the retinal image due to distance, three-dimensional tilt, line of sight, and illumination, they do not usually eliminate them completely. Moreover, we do not "see" the backs of objects or fill in occluded parts, although in most cases we easily infer them and are aware of their existence. Of course, our current spatial relation to the object is often as relevant to us as its identity; it determines our motor behavior in grasping, navigation, and so on. This may explain why conscious perception seems to represent both in an uneasy compromise that is neither completely viewer centered nor completely object centered. Whether the same compromise is present in the functional codes that represent the objects or whether those are fully object centered, leaving separate codes to specify the viewing conditions, is not known.

VII. PHYSIOLOGY OF VISUAL ATTENTION

Visual perception depends not only on the retinal image and its cortical representation but also on the focus of attention. We usually look at what we are interested in so that our direction of gaze coincides with our direction of attention in space. However, if needed we can also dissociate fixation from attention by attending to an event in our peripheral visual field without directly looking at it. The eye movement required to bring the fovea into register with the peripheral target is called a saccade. Saccades usually take 200–300 msec before they are initiated, a latency that is highly dependent on the stimulus conditions and state of attention.

The modulation of visual perception by attention has been studied in psychophysical as well as physiological experiments in man and monkey. The results obtained from single-cell recordings in various areas of the brain are described first. Neurophysiological studies of attention were only made possible by the pioneering work of Evarts (1966) in the motor system. He combined microelectrode techniques with psychological test procedures. Wurtz (1969) adopted this technique for the visual system and also precisely controlled the position of the eye. This was done by implanting a magnetic search coil (Robinson, 1963) in the eye or by using infrared light reflected from the cornea (Bach, Bouis & Fischer, 1983). All studies were done in the awake behaving monkey. Extensive reviews on the preparation of visually guided saccades (Fischer 1987) and on the relationship between saccades and visual attention (Fischer, 1986; Fischer & Breitmeyer, 1987) are available.

The basic idea in all studies on the effect of visual attention on single-cell activity is that the cell's response to a given stimulus can be modulated by the behavioral circumstances under which the stimulus occurs. Because it is impossible to know whether or not the animal is attending to the stimulus, one must rely on a motor reaction by the animal as a measurable response that can be related to the particular stimulus used.

Whereas in many psychophysical studies a manual response such as a key press is used, most studies in monkey have employed the oculomotor reflex, that is, the onset of a saccadic eye movement in response to a peripheral stimulus. In this way one can be sure that the stimulus that is viewed is also being attended to. The disadvantage, however, of this technique is that any modulation of a visual response in conjunction with a saccade may reflect the motor preparation of the eye movement rather than an effect of attention per se. In order to be sure that a given effect reflects attention only, one has to ascertain that the neuronal response can be obtained without a temporal change of the physical stimulus and without an accompanying motor response. We next describe a number of tasks that allow effects of visual stimuli, oculomotor occurrences, and attention on single-cell behavior to be assessed separately.

A. Experimental Tasks

1. Fixation Task

The first step in all experiments on the neurophysiology of the visual and oculomotor systems is to teach

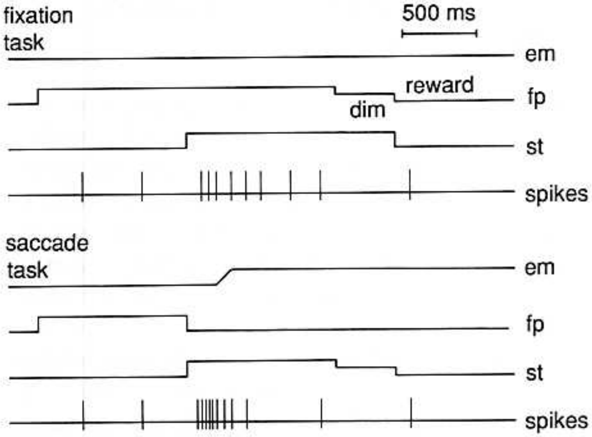

FIG. 26 Schematic illustration of the fixation task and the saccade task. The traces are identified as em (eye movement), fp (fixation point), and st (stimulus). The lowest trace (spikes) indicates the impulse activity of a single cell: in the fixation task the cell produces a moderate number of extra spikes after stimulus onset. In the saccade task, the same cell produces a burst of spikes. This response modulation is called the enhancement effect.

a monkey to fixate in a given direction. For this purpose a small fixation point is used (Wurtz, 1969). The animal quickly learns to look at the fixation point because he is rewarded each time he detects a mild dimming of the fixation light. The monkey signals the dimming by manually activating a switch. Following the detection and reward, the fixation point is extinguished, and a new cycle of fixations begins. In this way, up to 2000 fixation periods, each 2 to 8 sec long, may be obtained every day. This task is schematically illustrated in the top half of Fig. 26.

2. Saccade Task

During the fixation task, the fixation point is switched off and a new stimulus appears without delay in the monkey's peripheral field of vision. This stimulus has been positioned to project into the receptive field of the neuron from which recordings are made. The animal quickly learns to change his direction of gaze when the luminance of the new stimulus is turned down, and he will soon respond to the dimming of the peripheral light the way he responded to the fixation point before. The neural response to the onset of the peripheral stimulus (when placed in the cell's receptive field) will precede the beginning of the eye movement and may therefore reveal important cues to the processes occurring during the preparatory period. This saccade task is schematically illustrated in the bottom half of Fig. 26.

3. Delayed Saccade Task

To avoid confounding the occurrence of the saccadic target and the preparation of the eye movement, Fischer and Boch (1981b) introduced the delayed saccade task. In this task, the new stimulus appears before the fixation point is turned off so that any neural response elicited by the stimulus occurs well before the animal begins to prepare his saccade.

4. Suppressed Saccade Task

The suppressed saccade task is identical to the delayed saccade task with the only difference that the animal is required to suppress his eye movement to the new stimulus; he keeps looking straight ahead in the direction of the fixation point even after it has already disappeared (Fischer & Boch, 1985). A modification of this task is the blink task (Richmond, Wurtz & Sato, 1983). Here, the fixation point is temporarily turned off and a probe stimulus is flashed into the receptive field during this blink. Thereafter, the fixation point reappears, dims, and the animal is rewarded for pressing a bar upon the detection of the dimming.

5. Peripheral Attention Task

The peripheral attention task is similar to the suppressed saccade task except that here the fixation point remains visible while dimming occurs in the peripheral stimulus (Wurtz & Mohler, 1976a).

B. The Enhancement Effect

The enhancement effect was first reported by Goldberg and Wurtz (1972) for neurons in the superior colliculus nucleus. Figure 26 shows that during the fixation task the cell responds to the onset of a light placed in its receptive field with a moderate number of spikes distributed over the entire stimulus interval. However, during the saccade task, the same cell responds with a sudden burst as the stimulus now becomes the target for the saccade. The response is clearly enhanced. Note that the enhancement occurs already before the eyes begin to move, that is, at a time when the retinal image is still the same as in the fixation task. This effect has been attributed to the focusing of the animal's attention on the fixation point and its subsequent shift to the peripheral stimulus as the new target (Wurtz & Goldberg, 1972; Wurtz & Mohler, 1976a).

When a control task was introduced in these experiments, it became clear that the enhancement effect observed in collicular cells as well as those of other brain structures cannot always be attributed to attention. Instead, this effect is often related to the preparation of a motor response or to the transition from active to passive fixation.

C. Cortical and Subcortical Structures for Visual Attention

Possible candidates for the control of attention may be found by examining the anatomical pathways that connect the visual (afferent) system with the oculomotor (efferent) system, that is, the retina with the oculomotor nuclei. The most prominent structures are (in this order) the lateral geniculate body, superior colliculus, striate cortex (Area V1), prestriate cortex (Areas V2 and V3), prelunate cortex (Area V4), the cortex hidden in the superior temporal sulcus (Areas MT and MST), inferior temporal cortex (Area 22), posterior parietal cortex (Area 7), frontal eye fields (Area 8), and prefrontal cortex (Area 46). Figure 27 shows the location of the cortical areas in a side view of the left hemisphere of the rhesus monkey. Transections through the frontal and occipital brain along lines A and B, respectively, are also shown.

Earlier work on cells in the superior colliculus nucleus, the frontal eye fields, and the posterior parietal cortex has been summarized by Wurtz et al. (1982)

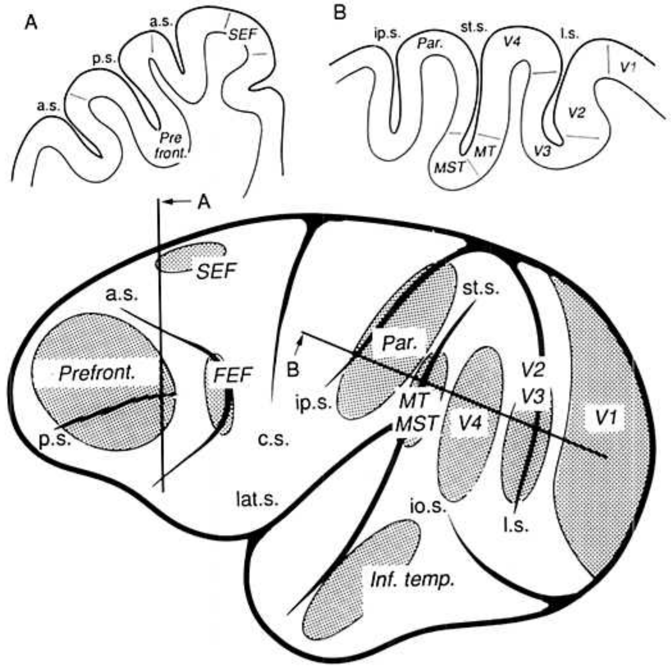

FIG. 27 Lateral view of the left hemisphere of rhesus monkey brain. Sulci are denoted by small letters as follows: ps., principal sulcus; a.s., arcuate sulcus; c.s., central sulcus; ip.s., intraparietal sulcus; lat.s, lateral sulcus; st.s., superior temporal sulcus; io.s., inferior occipital sulcus; l.s., lunate sulcus. Other abbreviations: Prefront., prefrontal cortex; SEF, supplementary eye field; FEF, frontal eye field; Par., parietal cortex; MT, middle temporal cortex; MST, medial superior temporal cortex; Inf. Temp., inferior temporal cortex. The transsections, made along lines A and B, show that many areas stippled in the lower part are, in fact, buried in the sulci and cannot be seen in a lateral view.

TABLE 1 Modulation of Neural Impulse Activity of Single Cells in Different Brain Structures as Tested with Different Tasks

	Task[a]			
Structure	Saccade	Delayed saccade	Suppressed saccade	Peripheral attention
Superior colliculus	Selective (+)	(+)	Not tested	(−)
Striate cortex, Area 17	Nonselective (+)	Selective (+)	Selective (+)	Not tested
Prestriate cortex, Area V2	Nonselective (+)	(+)	(+)	Not tested
Prelunate cortex, Area V4	Selective (+)	Selective (+)	Selective (+)	(+)
Posterior parietal cortex, Area 7	Selective (+)	Selective (+)	Not tested	(+)
Inferior temporal cortex, Areas 21 and 22	Not tested	Not tested	(+)	(−)
Frontal eye fields, Area 8	Selective (+)	Not tested	Not tested	(−)
Prefrontal cortex, Area 46	Selective (+)	Selective (+)	Selective (+)	Not tested
Pulvinar cortex	Nonselective (+)	Not tested	Not tested	Selective (+)

[a] Plus (+), enhanced visual activity; minus (−), tested with negative results.

and Goldberg (1986). Here, new results are added and briefly discussed for each brain structure in which one of the above tasks has been used. For an overview, see Table 1.

1. Superior Colliculus

The superior colliculus consists of different layers of cells having visual and oculomotor properties. The cells in question are located in the intermediate layers. They show an enhancement effect concomitant with the saccade that brings the fovea in register with the image of the peripheral stimulus. The stimulus need not be visual; it can also be auditory (Jay & Sparks, 1987). This multimodality shows that the superior colliculus is involved in a sensory-to-motor conversion rather than in attentional mechanisms per se. However, motor commands are prepared and generated by this nucleus for those stimuli that have been selected as potential targets. It may therefore be assumed that the collicular cells are controlled by an attentional signal, even if this signal occurs together only with a saccade.

2. Striate Cortex

In the saccade task, a few striate cortical cells show enhanced visual responses. However, the enhancement is not spatially selective because it also occurs with saccades to targets located outside the cell's receptive field (Wurtz & Mohler, 1976b). In contrast, Boch (1986) found cells in striate cortex that were activated in the suppressed saccade task. Therefore, if striate cortex plays a role in visual attention, its contribution may be weak and of minor importance. Note that although electrical stimulation of striate cortex does elicit saccadic eye movements, cell discharges in that same area do not lead to saccades.

3. Prestriate Cortex

Cells in Areas V2 and V3 may also show an enhanced visual response in the saccade task, but again the enhancement is not spatially selective (Robinson, Baizer & Dow, 1980). The suppressed and delayed saccade tasks were used in a few cells (Fischer & Boch, 1985) with the result that even at this early stage of visual processing one can already find clear signs of extraretinal modulations related to visual attention and/or fixation.

4. Prelunate Cortex

In contrast to Areas V1, V2, and V3, Area V4 has been studied extensively. Many V4 cells show spatially selective enhancement in the saccade task (Fischer & Boch, 1981a). They are also activated prior to a saccade by a constantly illuminated stimulus in their receptive fields (Fischer & Boch, 1981b), in the delayed saccade task (Fischer & Boch, 1982, 1983), as well as in the suppressed saccade task (Fischer & Boch, 1985). Cells in Area V4 also differentiate between attended and unattended stimuli within their receptive field (Moran & Desimone, 1985). Furthermore, stimuli to which the animal has been "cued" elicit a response that is stronger than that when the same stimulus is uncued (Haenny, Maunsell & Schiller, 1988). In this experiment, the monkey is presented with a certain stimulus, the cue, to which he must respond if it recurs in the same trial. Other stimuli that might occur instead are uncued stimuli. The question of whether certain stimulus features, such as color and orientation, affect the selectivity of Area V4 cells still needs to be answered.

5. Posterior Parietal Cortex

For a long time, the parietal cortex has been considered a major source of attentional signals (Lynch, Mountcastle, Talbot & Yin, 1977; Robinson & Petersen, 1984). Parietal cells show a spatially selective enhancement effect which is independent of the type of motor reaction, be it eye movement or hand movement (Robinson & Petersen, 1984; Goldberg & Bruce, 1985).

Interestingly, cells in parietal cortex may be strongly activated by a visual stimulus placed in a (peripheral) receptive field while the animal attentively fixates the central fixation point. The response of these cells may be attenuated or disappear altogether if the same stimulus is used during eye movement or when the animal is not fixating (Mountcastle, Motter, Steinmetz & Sestokas, 1987). No difference of this kind has been found in prelunate cells, a result that has yet to be explained.

6. Inferior Temporal Cortex

Cells in the inferior temporal cortex may also be involved in attentional mechanisms. Most cells respond vigorously in the suppressed saccade task in which attention to the stimulus is not required (Richmond et al., 1983). This indicates that the absence as well as presence of a fixation point can modulate the visual activity both in inferior temporal and prelunate cortex. It is later shown that under similar conditions the monkey fails to produce short latency saccadic eye movements.

7. Prefrontal Cortex

Boch and Goldberg (1989) extended these studies to Area 46, a region located in front of the frontal eye fields (Area 8). Here they found neurons showing an enhancement effect in the saccade task. In addition, neural activity was increased both in the delayed and suppressed saccade tasks. Earlier studies using the delayed response task had already shown that neurons in the prefrontal cortex may be related to attention (Fuster, 1973). More recent work supports this notion on the basis of the differential responses in go versus no-go trials (Bakay, Pragay, Mirksy & Nakamura, 1987). These results further substantiate findings indicating that the neural activity in the frontal eye fields is closely related to the preparation of the oculomotor response (Bushnell, Goldberg & Robinson, 1981; Goldberg & Bushnell, 1981; Bruce & Goldberg, 1985).

D. Stimulus Selectivity and Receptive Field Properties

The concept of the receptive field of a retinal cell implies that a cell's activity can be modulated by a probe stimulus only if it is positioned in a restricted area of the retina and at a specific location. The concept does not specify any features the stimulus should have except that the probe must be small as compared to the receptive field size. The receptive field may possess an antagonistic center–surround organization, thus favoring concentric stimuli to produce a maximal response.

As one proceeds from the retina and LGN to higher cortical structures such as the prestriate and parietal cortex, the spatial requirements for a stimulus eliciting a neuronal response are largely the same. However, as the receptive fields become larger and less specific for physical stimulus properties such as form and color, other aspects of the stimulus become important. These aspects must be related to the behavioral context within which a certain stimulus occurs. This relationship has been demonstrated by Haenny et al. (1988) for cells in prelunate cortex. Area V4 cells may respond better to the presentation of a grating of one orientation than to the same grating presented with another orientation. Yet, if the animal is cued to the latter orientation the response is enhanced and may be stronger than the response to the "best" uncued orientation. Another aspect for V4 cells is the large size of their receptive fields as compared to the optimal size of the stimulus. Weber and Fischer have measured the distribution of optimal sizes of V4 cells. The same animal was then trained to make saccades to a constant small stimulus which thus became a revelant stimulus. When they repeated the measurement of optimal stimulus sizes, the same cortical area appeared to exhibit a preponderance of cells specific for small stimuli. Thus, cells in Area V4 do not seem to have fixed specificities but may change their responsiveness in accordance with stimuli in which they are "interested."

E. Conclusion

Table 1 depicts the different tasks used to study cells of the different brain structures. With the exception of prelunate cortex, none of the areas has been studied using all tasks. Nevertheless, it has become clear that the superior colliculus and the frontal eye fields, even though they may receive a signal for attention, are closely related to the initiation of a saccade without being purely of motor nature. In comparison, striate cortex as well as inferior temporal cortex seem to be more involved in the analysis of sensory information. These findings imply that a voluntary change of the direction of gaze requires a release from attention. Conversely, during a state of peripheral attention, eye movements may be blocked (inhibited) because attention is fixed (engaged) to a certain part of the visual field.

The following section summarizes a number of new experiments on the initiation of saccades which show that disengagement, that is, the release from active (focused) attention, indeed enables the visual-to-oculomotor system to produce goal-directed saccades after extremely short reaction times.

F. Saccadic Eye Movements

A saccadic eye movement is a motor reaction optimally suited to study the mechanisms underlying visual attention. In the past, however, reactions of the hand or finger in response to visual stimuli have been used more frequently. This is mainly because it is easier to record manual reaction times than to accurately measure eye position and determine the latency for a saccade. The use of an infrared light-emitting diode and two photocells mounted on a simple spectacle frame provides an inexpensive and simple method for detecting horizontal saccades and for measuring saccadic reaction times with a precision of one millisecond.

1. Express Saccades

In 1967, Saslow demonstrated in human observers that the introduction of a temporal interval between the offset of a central fixation point and the onset of a peripheral saccade target (so-called gap trial) significantly reduced the saccadic reaction time (SRT) as compared to the case in which the fixation point remained on (overlap trial). The reduction consisted of an overall shift of the SRT distribution to shorter values. Fischer and Boch (1983) using gap trials in the monkey found that SRTs were reduced but, more importantly, two separate peaks occurred in the latency distribution, one at about 70 msec and the other at about 140 msec. Saccades contributing to the first peak were called express saccades (because of their extremely short reaction time) and those contributing to the second, fast regular saccades (Fischer & Boch, 1983). Slow regular saccades forming a peak at about 200 msec were only obtained in overlap trials.

The existence of express saccades in man was first reported by Fischer and Ramsperger (1984) who repeated Saslow's experiment. Saslow may have failed to find express saccades because he used large bin widths for analyzing his latency distributions. Figure 28 shows the distribution of saccadic reaction times for two human subjects. Bimodal distributions like these are easily obtained when the fixation point is turned off before the saccade target occurs.

The introduction of a temporal gap, however, is not a necessary condition for the occurrence of express saccades. In humans the instruction "Do not pay attention to the fixation point" is sufficient to allow for express saccades in overlap trials (Mayfrank, Mobashery, Kimmig & Fischer, 1986). In the monkey, regular daily practice using overlap trials also produces express saccades (Boch & Fischer, 1986). In both species, daily practice increases the percentage of express saccades and reduces their latency (Fischer, Boch & Ramsperger, 1984; Fischer & Ramsperger, 1986). The observation that saccadic reaction times decrease drastically following changes in instruction and after periods of daily practice has led to new concepts of visual attention and a better understanding of its relation to saccadic eye movements and fixation.

G. Attention and Reaction Time

The basic idea in relating attention to reaction time is that any change in the state of visual attention takes time which may or may not be included in the saccadic reaction. Posner, Cohen, and Rafal (1982) and Posner, Walker, Friedrich, and Rafal (1984) used key pressing as a motor reaction to study attentional effects in man. From their results they concluded that the attentional system can be in either of two different states: engaged or disengaged. This distinction holds also for saccades: express saccades are obtained in trials where attention is already disengaged from the

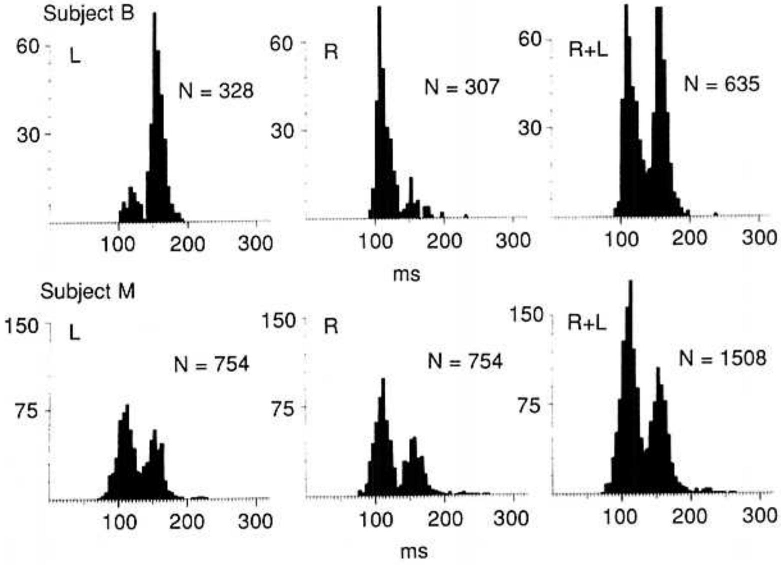

FIG. 28 Distributions of saccadic reaction times in humans. The results of two different subjects are shown. Ordinate: Number of saccades in a bin width of 10 msec; abscissa: saccadic reaction time. The saccade target occurred in random order at 4° to the left (L) or right (R) of the fixation point. The total number of saccades (*N*) is indicated in each diagram. The summed distribution for right and left directed saccades is clearly bimodal. Note the asymmetry in the response latencies of subject B for left directed versus right directed saccades. Saccades contributing to the first peak are called express saccades.

fixation point. On the other hand, reaction times are extremely long in trials where attention is still engaged. Engagement need not be in the location of the fixation point but can be anywhere in the visual field. Even if one attends to the location of the future target, SRTs will be long with express saccades virtually absent (Mayfrank et al., 1986).

It thus appears as though visual attention blocks saccades as long as it is engaged. In order to move the eye from one target to another, visual attention must be disengaged regardless of where in the visual field it is engaged. This interpretation suggests a cyclic mechanism governing vision with moving eyes under normal conditions (Fischer, 1987): As the direction of gaze changes due to saccades (i.e., every 200 to 300 msec), the state of attention also changes from engaged to disengaged. Each saccade is preceded by disengagement and followed by engagement. During periods of engagement visual information enters the afferent system, whereas during disengagement the next saccade is prepared by the efferent system.

H. Model for the Control of Saccadic Eye Movements

The fact that visual responses, saccades, and attention-related activity can be found in different cortical and subcortical areas and the multimodal distribution of saccadic reaction times have led to a concept of three loops that are involved in the control of vision and eye movements. Loop 1 would control the disengagement of visual attention; Loop 2 the decision to make a saccade; Loop 3 the computation of amplitude and direction. Figure 29 illustrates the loops and their functional interconnections. Although Fig. 29 suggests that each loop acts in parallel with the others, this is not necessarily so. To the contrary, each square may symbolize a logical AND-gate to ensure that all processes must have taken place before a saccade can be made. One may even postulate that processes involved in loops 1, 2, and 3 take place in a serial manner if one assumes that processes in loop 2 can be started only after the processes in loop 1 have been completed and that processes in loop 3 can begin only after the processes in loop 2 are finished. Such a mechanism requires that all brain structures involved must function properly to generate a correct saccade. In view of this complexity one immediately understands why saccadic reaction times are so long (about 200 msec) in one condition and extremely short (about 70 msec) in another. It also shows how fixation may be achieved. Suppose the neural processes in one of the loops, let us say loop 1, cannot be completed as long as attention is engaged to a particular part of the visual field. In this case, saccades are arrested even if the object attended to is not being fixated. On the other hand, it assumes that if attention is disengaged the eye can execute a goal-directed saccade provided that the processes in the other two loops are completed.

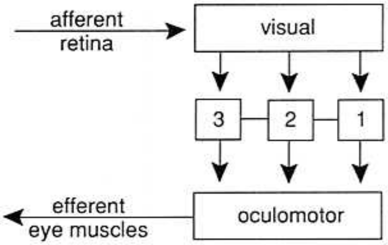

FIG. 29 Block diagram illustrating the three-loop concept for the control of eye movements by vision and attention. Loop 1 disengagement; Loop 2 decision; Loop 3 spatial coordinates.

The concept of the three loops is a functional, not an anatomical concept. Nevertheless one can attempt to identify different anatomically defined structures as contributing to one or the other loop as follows: (1) striate cortex and superior colliculus; (2) frontal cortex, in particular the frontal eye fields; (3) prestriate cortex and partietal cortex.

VIII. CONCLUSIONS

We are still in the very early stages of exploring the perception of form and relating it to the neuropsychological mechanisms of the visual system. In recent years many relevant discoveries have been made, but they chiefly emphasize the complexity of processing that must underlie the effortless experience of seeing. The broad picture seems to be one of initial analysis by a number of specialized channels or modules, and a subsequent coordination and integration of the information extracted.

Attention seems to play a central role, both in selecting the relevant object at any given time and in specifying its particular structured assembly of properties. Further issues for future research will concern the mechanisms of selection—whether attention acts by facilitating relevant stimuli or suppressing irrelevant ones; and whether it scans the visual field with continuous motion, as suggested by the frequently used spotlight and searchlight analogies; or whether it switches in discrete steps from object to object. These psychological debates may benefit from the accumulating evidence relating physiological recordings (e.g. Moran and Desimone, 1985) to behavioral tasks requiring selective attention.

The chapter opened with the example of a snowman seen against a blanket of snow. The evidence we have reviewed suggests that his features will be represented in multiple visual maps coding their color, location, orientation, and size, specifying boundaries and creating subjective contours where they are missing from the retinal image. How we reach, within a quarter of a second, our unified awareness, both of what he looks like and of what he is, remains a mystery which will challenge both psychologists and physiologists in future research.

ACKNOWLEDGMENTS

Preparation of this chapter was supported by grant number AFOSR 87-0125 to Anne Treisman from the Air Force Office of Scientific Research, Air Force Systems Command, USAF. Patrick Cavanagh received a fellowship from the Canadian Institute of Advanced Research and a grant from the Natural Science and Engineering Research Council of Canada. Rudiger von der Heydt was supported by grant number 3.939.84 from the Swiss National Foundation. Burkhart Fischer was supported by a grant from the Deutsche Forschungsgesellschaft (Sonderforschungsbereich 325, C5).