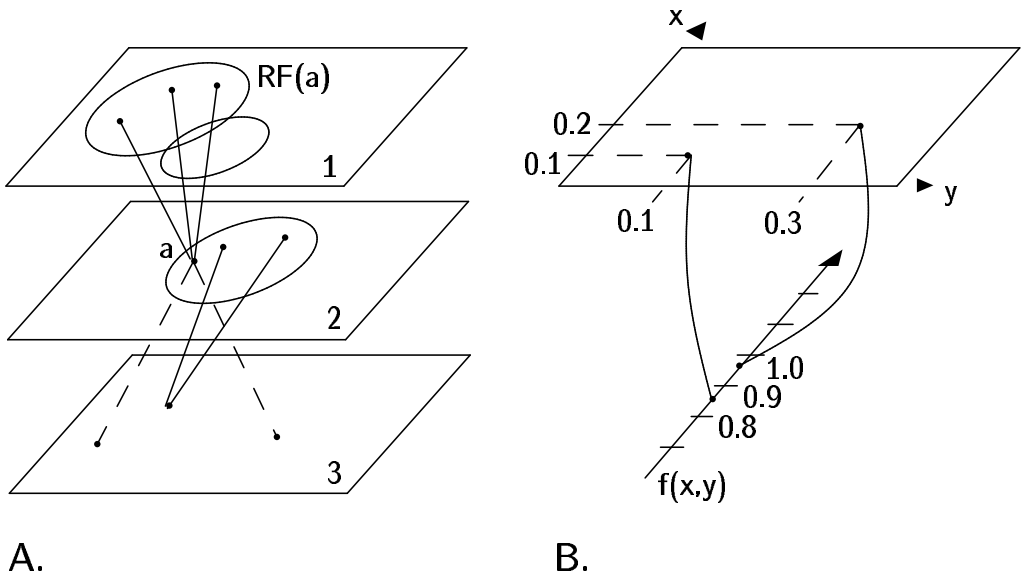


Figure 4: **A.** A highly schematic illustration of the notion of a receptive field: unit $a$ in area 2 receives input from the region marked as RF(a) in area 1, and, in turn, projects to many units in area 3. **B.** Computing the function $f(x) = e^{-(x^2+y^2)} \cos y$ with a 2D to 1D map of connections, or lookup table. Note that to find the value of the function at a point for which there in no entry in the lookup table one must resort to interpolation (in this connection, see Poggio and Girosi, 1990).

(the Dalmatian from Marr's book to be pasted here)

Figure 3: Past experience may influence the emergence of a percept where bottom-up organization and grouping alone fails (photography by R. C. James).

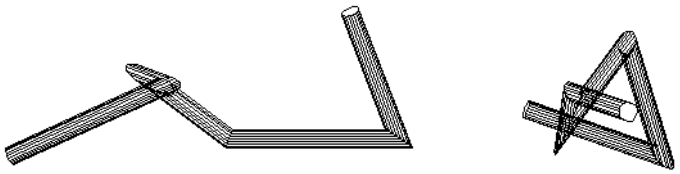

Figure 2: The appearance of a 3D object can depend strongly on the viewpoint. The image on the right is of the same object as the image on the left, rotated in depth by $90^o$. People find it difficult to recognize such objects from a radically unfamiliar viewpoint, even when stereo information (Rock and DiVita, 1987) or both stereo and motion cues (Edelman and Bülthoff, 1990) to the 3D shape of the object are available.

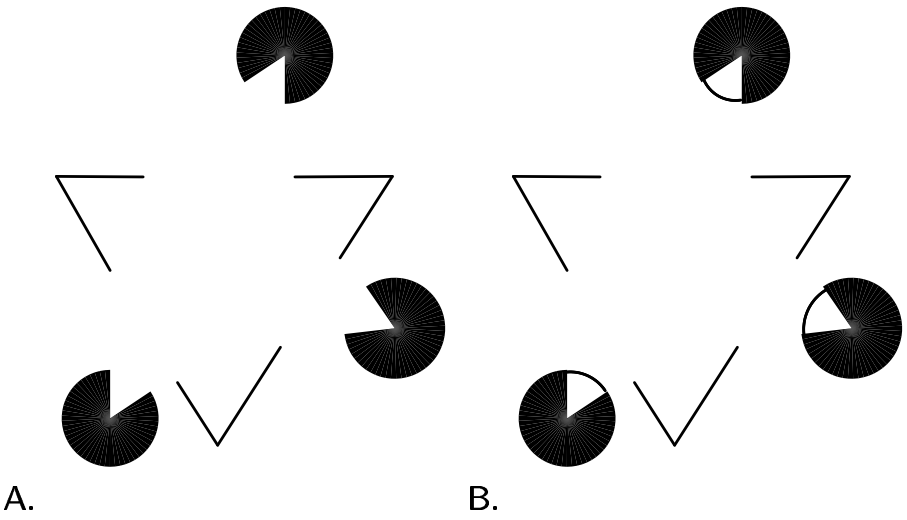


Figure 1: **A.** Illusory contours, as well as some abnormal depth and lightness percepts, arise in this version of Kanizsa's triangle. **B.** A modified version, in which the illusory percepts are much weaker.

[70] N. Weisstein and C. S. Harris. Visual detection of line segments: an object-superiority effect. *Science*, 186:752–755, 1974.

[71] S. Zeki and S. Shipp. The functional logic of cortical connections. *Nature*, 335:311–317, 1988.

[59] J. T. Todd and E. Mingolla. Perception of surface curvature and direction of illumination from patterns of shading. *J. Exp. Psychol.: HPP*, 9:583–595, 1983.

[60] A. Treisman and G. Gelade. A feature integration theory of attention. *Cognitive Psychology*, 12:97–136, 1980.

[61] S. Ullman. Filling in the gaps: the shape of subjective contours and a model for their generation. *Biological Cybernetics*, 25:1–6, 1976.

[62] S. Ullman. *The interpretation of visual motion.* MIT Press, Cambridge, MA, 1979.

[63] S. Ullman. Aligning pictorial descriptions: an approach to object recognition. *Cognition*, 32:193–254, 1989.

[64] S. Ullman and R. Basri. Recognition by linear combinations of models. A.I. Memo No. 1152, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1990.

[65] R. von der Heydt, E. Peterhans, and G. Baumgartner. Illusory contours and cortical neurons' responses. *Science*, 224:1260–1262, 1984.

[66] H. von Helmholtz. Unconscious conclusions. In W. N. Dember, editor, *Visual perception: the nineteenth century*, pages 163–170. Wiley, 1856/1964.

[67] R. D. Walk. Perceptual learning. In E. C. Carterette and M. P. Friedman, editors, *Handbook of Perception*, volume IX, pages 257–298. Academic Press, New York, NY, 1978.

[68] H. Wallach and D. N. O'Connell. The kinetic depth effect. *J. Exp. Psychol.*, 45:205–217, 1953.

[69] D. Weinshall. Perception of multiple transparent planes in stereo vision. *Nature*, 341:737–739, 1989.

[48] I. Rock and J. DiVita. A case of viewer-centered object perception. *Cognitive Psychology*, 19:280–293, 1987.

[49] A. Rojer and E. L. Schwartz. A multiple-map model for pattern classification. *Neural Computation*, 1:104–115, 1989.

[50] E. Rosch, C. B. Mervis, W. D. Gray, D. M. Johnson, and P. Boyes-Braem. Basic objects in natural categories. *Cognitive Psychology*, 8:382–439, 1976.

[51] A. Rosenfeld. Recognizing unexpected objects: a proposed approach. *Int. J. of Pattern Recognition and Artificial Intelligence*, 1:71–84, 1987.

[52] R. N. Shepard. Toward a universal law of generalization for psychological science. *Science*, 237:1317–1323, 1987.

[53] R. N. Shepard and L. A. Cooper. *Mental images and their transformations.* MIT Press, Cambridge, MA, 1982.

[54] A. Sloman. What are the purposes of vision? CSRP 066, University of Sussex, 1987.

[55] E. S. Spelke. Origins of visual knowledge. In D. N. Osherson, S. M. Kosslyn, and J. M. Hollerbach, editors, *Visual cognition and action*, volume 2, pages 99–128. MIT Press, Cambridge, MA, 1990.

[56] G. R. Stoner, T. D. Albright, and V. S. Ramachandran. Transparency and coherence in human motion perception. *Nature*, 344:153–155, 1990.

[57] G. Stratton. Vision without inversion of the retinal image. In W. N. Dember, editor, *Visual perception: the nineteenth century*, pages 143–154. Wiley, 1897/1964.

[58] M. Tarr and S. Pinker. Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, 21:233–282, 1989.

[36] S. E. Palmer. The psychology of perceptual organization: a transformational approach. In J. Beck, B. Hope, and A. Rosenfeld, editors, *Human and machine vision*, pages 269–340. Academic Press, New York, 1983.

[37] S. E. Palmer, E. Rosch, and P. Chase. Canonical perspective and the perception of objects. In J. Long and A. Baddeley, editors, *Attention and Performance IX*, pages 135–151. Erlbaum, Hillsdale, NJ, 1981.

[38] G. F. Poggio and T. Poggio. The analysis of stereopsis. *Ann. Rev. Neurosci.*, 7:379–412, 1984.

[39] T. Poggio and S. Edelman. A network that learns to recognize three-dimensional objects. *Nature*, 343:263–266, 1990.

[40] T. Poggio and F. Girosi. Regularization algorithms for learning that are equivalent to multilayer networks. *Science*, 247:978–982, 1990.

[41] T. Poggio and W. Reichardt. Visual control of orientation behavior in the fly (parts i and ii). *Quart. Rev. Biophys.*, 3:311–439, 1976.

[42] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317:314–319, 1985.

[43] M. Potter. Meaning in visual search. *Science*, 187:965–966, 1975.

[44] Z. Pylyshyn. *Computation and cognition.* MIT Press, Cambridge, MA, 1985.

[45] P. Quinlan. Visual object recognition reconsidered, 1989. submitted for publication.

[46] V. S. Ramachandran. Perception of shape from shading. *Nature*, 331:163–166, 1988.

[47] I. Rock. *Perception.* Scientific American Books, New York, 1984.

[24] J. H. Kaas. Why does the brain have so many visual areas? *Journal of Cognitive Neuroscience*, 1:121–135, 1989.

[25] E. R. Kandel and J. H. Schwartz. *Principles of neural science.* Elsevier, New York, 1985.

[26] S. W. Keele and W. T. Neill. Mechanisms of attention. In E. C. Carterette and M. P. Friedman, editors, *Handbook of Perception*, volume IX, pages 3–47. Academic Press, New York, NY, 1978.

[27] J. J. Koenderink. What does the occluding contour tell us about solid shape? *Perception*, 13:321–330, 1984.

[28] W. Köhler. *Gestalt psychology.* Liveright, New York, 1947.

[29] S. W. Kuffler and J. G. Nicholls. *From neuron to brain.* Sinauer, Sunderland, MA, 1976.

[30] E. H. Land and J. J. McCann. Lightness and retinex theory. *Journal of the Optical Society of America*, 61:1–11, 1971.

[31] R. D. Luce. *Response times: their role in inferring elementary mental organization.* Oxford University Press, Oxford, 1986.

[32] H. A. Mallot, W. von Seelen, and F. Giannakopoulos. Neural mapping and space-variant image processing. *Neural Networks*, 3:??–??, 1990.

[33] D. Marr. *Vision.* W. H. Freeman, San Francisco, CA, 1982.

[34] W. S. McCulloch. *Embodiments of mind.* MIT Press, Cambridge, MA, 1965.

[35] G. J. Mitchison and S. P. McKee. Interpolation in stereoscopic matching. *Nature*, 315:402–404, 1985.

[13] S. Edelman and D. Weinshall. A self-organizing multiple-view representation of 3D objects. A.I. Memo No. 1146, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, August 1989.

[14] J. P. Frisby. *Seeing.* Oxford University Press, Oxford, 1979.

[15] J. J. Gibson. *The perception of the visual world.* Houghton Mifflin, Boston, MA, 1950.

[16] J. J. Gibson. *The ecological approach to visual perception.* Houghton Mifflin, Boston, MA, 1979.

[17] C. D. Gilbert. Neuronal and synaptic organization in the cortex. In P. Rakic and W. Singer, editors, *Neurobiology of Neocortex*, pages 219–240. Wiley, New York, NY, 1988.

[18] A. L. Gilchrist. Perceived lightness depends on perceived spatial arrangement. *Science*, 195:185–187, 1977.

[19] R. L. Gregory. Illusions and hallucinations. In E. C. Carterette and M. P. Friedman, editors, *Handbook of Perception*, volume IX, pages 337–357. Academic Press, New York, NY, 1978.

[20] J. Jastrow. *The time relations of mental phenomena.* Hodges, New York, 1890.

[21] G. Johansson. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14:201–211, 1973.

[22] B. Julesz. *Foundations of Cyclopean perception.* University of Chicago Press, Chicago, IL, 1971.

[23] B. Julesz. A brief outline of the texton theory of human vision. *Trends in Neurosciences*, 7:41–45, 1984.

[2] I. Biederman. Human image understanding: Recent research and a theory. *Computer Vision, Graphics, and Image Processing*, 32:29–73, 1985.

[3] I. Biederman, J. C. Rabinowitz, A. L. Glass, and E. W. Stacy. On the information extracted from a glance at a scene. *Journal of Exp. Psychol*, 103:597–600, 1974.

[4] A. Blake and H. H. Bülthoff. Does the brain know the physics of specular reflection? *Nature*, 343:165–168, 1990.

[5] K. R. Boff, L. Kaufman, and J. P. Thomas, editors. *Handbook of perception and human performance.* Wiley, New York, 1986.

[6] R. M. Boynton. Color in contour and object perception. In E. C. Carterette and M. P. Friedman, editors, *Handbook of Perception*, volume VIII, pages 173–199. Academic Press, New York, NY, 1978.

[7] H. H. Bülthoff and H. A. Mallot. Interaction of depth modules: stereo and shading. *Journal of the Optical Society of America*, 5:1749–1758, 1988.

[8] P. Cavanagh. Reconstructing the third dimension: interactions between color, texture, motion, binocular disparity and shape. *Computer Vision, Graphics, and Image Processing*, 37:171–195, 1987.

[9] A. R. Damasio. The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation*, 1:123–132, 1989.

[10] D. Davidson. *Essays on actions and events.* Clarendon Press, Oxford, 1980.

[11] F. Dretske. Seeing, believing, and knowing. In D. N. Osherson, S. M. Kosslyn, and J. M. Hollerbach, editors, *Visual cognition and action*, volume 2, pages 129–148. MIT Press, Cambridge, MA, 1990.

[12] S. Edelman and H. H. Bülthoff. Canonical views and the representation of novel 3D objects, 1990. submitted to Perception.

that as a whole this structure is as heterachical as the large-scale arrangement of cortical areas mentioned above (Figure 4a).

Figure 4 here

The second basic notion in the architecture of vision, cortical mapping, pertains to the relationship among different receptive fields in the same visual cortical area. It turns out that many areas are interconnected by locally smooth maps (in fact, many areas are retinotopic, that is, their topology, but not necessarily their metric structure, conforms to that of the retina). Recent theoretical developments indicate that computing with maps or, equivalently, connections (Figure 4b) is a powerful information-processing paradigm (see, e.g., Ballard, 1986, for an integrative discussion). It has been suggested (Rojer and Schwartz, 1989; Mallot et al., 1990; Damasio, 1989) that cortical mapping is the basic mechanism of the visual function of the brain. It remains to be seen whether this concept can be extended to encompass perception (and intelligence) in general.

# 6 Conclusion

The study of visual perception is bound to inspire awe, because of the recognition of the formidable problems posed by vision, and marvel, because of the appreciation of solutions developed by the brain to address these problems. Borrowing a phrase from Warren McCulloch, one can describe this study as discovering what a thing is, that a man may see it, and a man, that he may see things (McCulloch, 1965, p.2). While looking for an answer to this question, it is worthwhile to remember that in perception, if not in intelligence, man *is* the measure of all things.

# References

[1] D. H. Ballard. Cortical connections and parallel processing: structure and function. *Behavioral and Brain Sciences*, 9:67–120, 1986.

### 5.0.7 Vision is Massively Parallel

The parallelism of visual information processing becomes apparent already at the level of the retina, where a separation occurs between several pathways each of which is functionally specialized to support certain aspects of the input, such as form, motion and color. Since eventually these have to be integrated into a coherent percept, the cortical areas fed by the different pathways are interconnected in an orderly fashion (see recent reviews in Zeki and Shipp, 1988; Kaas, 1989), so that on the whole the architecture is a heterarchy in which lateral connections and shortcuts abound, rather than a hierarchy envisaged by early visual scientists. Remarkably, in all the visual areas, as in the entire neocortex in general, information is processed by the same variety of cells, arranged in the same columnar structure (Gilbert et al., 1988). Thus, any comprehensive theory of brain function would have to specify, in information-processing terms, what the basic operation is that can be supported by the cortical architecture and is at the same time powerful enough to address the entire range of perceptual, cognitive, and motor tasks.

### 5.0.8 Maps and Receptive Fields

Although a viable and widely accepted theory of such scope has yet to be proposed, two well-established findings in the neurobiology of perception, and in particular of vision, provide an inkling as to the basic mode of information processing in the cortex. The first of these is concerned with the notion of the receptive field of a neuron. In vision, it is defined as that region of the visual field whose stimulation affects the activity of the neuron (see, e.g., Kuffler and Nicholls, 1976). Without going into details of the taxonomy and structure of receptive fields found in the visual cortex, one may describe their function schematically as the integration of information over a finite area of the visual field and its concentration at a single point — the axon of the neuron in question, or its output terminal. Note that the axon may, in turn, connect to many (potentially, tens of thousands) other neurons at a higher level, so

experience (Spelke, 1990). Significantly, the mechanisms of perceptual organization used by infants in learning how to see seem to persist through adulthood. Thus, the study of the ontogeny of visual perception may help clarify the nature of the long-term memory representations of objects and scenes.

A similar angle on the problem of representation is provided by studies in which the subjects' perceptual performance in three-dimensional object recognition is modified merely as a result of practice or exposure to the stimuli, without any feedback from the experimenter (after all, infants acquire vision — and language — without being instructed). Normally, the subjects' response time in recognition depends monotonically on the misorientation of the stimulus with respect to some canonical attitude (Palmer et al., 1981; cf. the discussion of mental rotation in the section on Methods). Perceptual learning under such circumstances, inferred from the observed changes in the pattern of performance (specifically, from increasingly uniform response times for different aspects of the stimulus), can be attributed to a shift towards a more memory-intensive and less time-consuming recognition strategy (Tarr and Pinker, 1989; Edelman and Bülthoff, 1990). Indeed, such a strategy appears to be the most suitable one for a system in which memory is cheap, but time is expensive.

# 5 The neural substrate

The architecture of the human visual system reflects the major functional constraint imposed on it, namely, the requirement of being able to recognize or classify in a few hundreds of milliseconds any object from a potentially unlimited repertoire, while taking into account a variety of visual clues (Biederman et al., 1974; cf. Rosenfeld, 1987). The following section contains several hints as to what the basic functional elements that constitute this architecture may be.

mere recollection of previously encountered stimulus exemplars. Experiments carried out by Rosch and her collaborators as part of a wider study of the structure of categories (see, e.g., Rosch et al., 1976) showed that people tend to perceive and describe objects at a certain level of detail. Importantly, this basic category level can be independently defined in terms of visual perception, language, and general cognitive development.

Figure 3 here

A less direct, but not less compelling, demonstration of the top-down principle is shown in Figure 3. Rock (1984, p.130) remarks that we would be unlikely to see a Dalmatian in this figure, unless past experience affected our perception of it. Some of the features of the processes involved in perceptual learning and memory are mentioned in the next section.

### 4.4.6 Perceptual Learning

Perceptual learning, or the adjustment of perception to the stimulus aspects of the environment, is sometimes distinguished from cognitive learning, the latter term being reserved for the modification of problem-solving behavior (Walk, 1978). In early vision, learning occurs in processes such as adaptation, mentioned above. The famous experiment first made by Stratton (1897/1964) , in which a subject wearing inverting prisms gradually adapts to this condition, is a forceful reminder that the degree of plasticity at lower levels of the visual system should not be underestimated (see Rock, 1984, for a discussion of the inversion experiments).

Exactly what is learned and what is innate in vision (and in cognition in general) has been the subject of intense philosophical debate since Plato's time (see Dretske, 1990, for an overview). A century of research in visual psychophysics and neurobiology of vision shows that the basic perceptual abilities of the human visual system (such as the ability to perceive luminance contrast) are largely innate, while others (such as some varieties of object constancy) are acquired and depend on the visual

(see Figure 1). The first of these is contour completion: the visual system prefers to perceive a nonexistent contour forming the illusory triangle rather than see the three dented discs as unrelated to each other (in which case the missing portions of the real triangle would remain unaccounted for). It turns out that the perceived shape of the completed contour can be produced by a process that minimizes a measure of its curvature (Ullman, 1976). The second phenomenon that affects the perception of figures as wholes is that of area filling-in, which may be considered a two-dimensional analog of contour completion. In Figure 1, filling-in is apparent in the increased subjective lightness of the illusory triangle. Another manifestation of filling-in is motion capture: our tendency to perceive stationary features that happen to fall within a moving contour as drifting along with the contour. Surface interpolation (see the section on Stereopsis) is an instance of filling-in that, analogously to contour completion, can be formulated as a process of optimization (see Poggio et al., 1985).

### 4.4.5 Categorization

Several illustrations of our disposition to see the environment as structured instead of chaotic were given in the previous sections, when I discussed the phenomena of object constancy and motion perception. The imposition of this kind of high-level structure on the visual world is an apotheosis of the processes of perceptual organization, linking vision to general cognition and language. Experimental evidence suggests that this connection is bilateral, and that the cognitive level can influence visual perception. The most directly relevant experiments are those in which subjects exhibit object superiority effects, for example, the facilitation of the perception of a low-level feature, such as a line segment, by virtue of its appearance as a part of the projection of a three-dimensional object (Weisstein and Harris, 1974).

In many top-down effects (including object superiority), the perceptual phenomenon is better characterized as categorization than recognition. The manner of cognitive involvement in perception is thus more flexible and more general than

be quite expensive. In these cases, attention provides an acceptable trade-off between resources and time.

### 4.4.3 Perceptual Organization

Perceptual organization is a collective term for a diverse set of processes that contribute to the emergence of order in the visual input. Some of the phenomena I have already mentioned can be considered particular instances of perceptual organization. Two examples which are, really, opposite sides of the same coin are shape constancy and visual motion: the perception of a deforming two-dimensional retinal stimulus pattern as a three-dimensional object in motion amounts to organizing the visual input so that it can be described in a simple and stable fashion. The study of perceptual organization has a long history in psychology. Palmer (1983), following Cassirer, Pitts and McCulloch, Gibson, and others, outlined a uniform framework for the study of organizational phenomena, based on the mathematical notion of invariance under transformation groups. The transformational approach allows problems of constancy and motion to be addressed in the same language as the classical issues of perceptual organization: figural goodness, grouping and frames of reference. A major attempt to understand these issues, motivated by a conviction that such understanding would shed light on perception in general, led to the formation of the Gestalt school in psychology (Köhler, 1947). Many of the laws of organization proposed by Gestalt theorists, as well as the concept of perceptual goodness ("Praegnanz") they have introduced to account for a variety of perceptual phenomena, have been incorporated in a more rigorous formulation into the currently prevailing paradigm of visual perception (see the discussion in Marr, 1982, p.187).

### 4.4.4 Completion and Filling-In

The group of phenomena that can be characterized by a tendency to optimize figural goodness includes two that have been mentioned in the section on Visual Performance

rest is supported by physiological processes that operate at the photoreceptors and in the higher levels of the retina. Presumably, some of the visual aftereffects that may be classified as adaptation phenomena happen at higher levels of the visual system. A well-known example of this type is the motion aftereffect: we tend to perceive the stationary parts of a waterfall scene as moving upwards after having concentrated for sufficient time on the downward motion of the falling water. One possible explanation of the aftereffects is that the visual system represents qualities such as planarity and immobility as a dynamic balance between representations of opposites such as convexity/concavity and upward/downward motion. A fatigue-like reduction in the activity level of the representation substrate of downward motion, for example, will then cause a perception of upward motion in a stimulus that is, in fact, static. Note that this account relies on another common property of perception related to adaptation: the preferential response to temporal change and spatial contrast, as opposed to status-quo and uniformity.

### 4.4.2 Attention and Search

The visual system allocates different amounts of processing resources to different portions of the visual field. Most of this non-uniformity is architectural and is due to the gradual decrease in the photoreceptor density (which entails a decrease in acuity) between the fovea and the periphery. The other part is more flexible and can be manipulated at will. The phenomenon of diverting processing resources to specified locations in the visual field or to a specified submodality is called visual attention (e.g., Keele and Trammer Neill, 1978). Attention can be shifted to a new location overtly, by executing an appropriate saccadic eye movement, or covertly, by a mechanism whose precise nature is not yet known. The benefit of being able to shift attention is in the economy of processing resources required for adequate functioning. For some visual operations, such as search for compound stimuli (Treisman and Gelade, 1980), maintaining uniform processing capability over the entire field may

identify and describe its source. In other cases the knowledge source is not so readily apparent to the observer. For example, the visual system seems to take into account the physics of specular reflection in the perception of three-dimensional shape in shaded images (Blake and Bülthoff, 1990). In another example, motion perception appears to involve implicit familiarity with the physics of transparency (Stoner et al., 1990).

## 4.3 Illusions

The unconscious rules of thumb that are in part responsible for the incredibly fast performance of the visual system in a wide variety of vital tasks are, necessarily, limited in scope. Certain visual stimuli produce illusory or non-veridical percepts by causing the breakdown of those perceptual processes that rely on unconscious rules. Another source of illusions is in the inherent anatomical and physiological properties and limitations of the visual system. The illusory contours of Figure 1a for which a tentative physiological mechanism has been identified (von der Heydt et al., 1984) are an example of this kind (see also Gregory, 1978).

## 4.4 Processes

It is remarkable that a number of central parts of the process of visual perception can be described functionally in a manner that is largely independent of the particular goals they serve. What follows here is an attempt to briefly characterize these sub-processes.

### 4.4.1 Adaptation

The most common example of adaptation is the adjustment in the light sensitivity of the eye that follows any change in the ambient illumination. The dynamic range of light adaptation is very wide (at least five orders of magnitude). Only a small part of this range (a factor of sixteen or so) is attributable to the changes in pupil size; the

### 4.1.6 Explanations of Constancy

The constancies of visual perception constitute an essential part of the visual experience as we know it. Imagine what a bewildering world it would be in which red tomatoes turned yellow when seen in the kitchen under incandescent light, coins looked elliptical unless viewed from the proper vantage point, and the slightest movement of the head sent the entire surrounding scene careening about. How does the visual system achieve perceptual constancy? The stimulus-relation explanation, favored by the direct perception school, sees the stimulus itself as the sole cause of constancy. According to this account, the context in which the stimulus appears affects the way it is perceived. For example, the apparent size of an unfamiliar object may be affected by the presentation next to it of another object of known size. In many situations, however, context information is unavailable, yet size constancy still holds. To continue the above example, the size of a luminous object in a darkened room may be correctly judged as long as distance information is available, e.g., from accomodation and convergence. If, however, the object is seen through a narrow aperture that eliminates distance cues, the constancy breaks down and the object appears to be of indeterminate size (Rock, 1984). This phenomenon prompts an alternative explanation of constancy which proposes that viewers take unconsciously into account independent or prior knowledge relevant to a given situation. As we shall see in the next section, the involvement of prior knowledge can account also for other features of vision besides constancy.

## 4.2 Implicit Knowledge

An example of the facilitation of perception through the use of information not available in the immediate sensory input is the visual system's superior performance in recognizing and remembering objects and scenes that make sense, as opposed to those that do not (Biederman et al., 1974; Potter, 1975). Although the implicit knowledge in this case is used unconsciously, the perceiver would normally be able to

(1984).

### 4.1.1 Lightness Constancy

The perception of the shade of a surface's lightness varies in general with the true albedo of the surface rather than with its luminance (the intensity of the light reflected by the surface, which changes, for example, because of varying illumination and shadows).

### 4.1.2 Color Constancy

When the spectral content of the illuminant undergoes a radical change, a surface will no longer reflect light that corresponds to its true color (i.e., the color it reflects when illuminated with white light). Nevertheless, its perceived color will in general appear close to the true one.

### 4.1.3 Size Constancy

A given object in the world appears to be about the same size, irrespective of the variation of the size of its retinal projection (due, e.g., to its varying distance from the eye).

### 4.1.4 Shape Constancy

The perceived shape of an object remains constant despite changes in the shape of its retinal projection caused by the movement of the object relative to the observer.

### 4.1.5 Space Constancy

The visual world appears to us as stable and unmoving despite continuing movement of the retinal image, caused by the movement of the eyes, as in visual tracking and saccades (see the section on Attention below), the head and the entire body (as in locomotion).

view amounts to much more than a theory of object recognition: it dictates the interpretation of processes of early vision in terms of the reconstruction of a replica of the visual world. Recently, this view has been disputed on philosophical, empirical and computational grounds (see Sloman, 1987, Quinlan, 1989, Edelman and Weinshall, 1989, Edelman and Bülthoff, 1990).

Three-dimensional object-centered models, envisaged by Marr, and other 3D structural representations (e.g., Biederman, 1985) are but one of a number of competing theories in the field of recognition. Major alternatives (see Ullman, 1989, for a review) are template matching, description by invariant features and shape normalization. Although it is clear that structural descriptions of the type suggested by Marr, Biederman and others must be invoked to explain some perceptual phenomena, the emergence of recognition models based on interpolation among prototypical two-dimensional views (e.g., Ullman and Basri, 1990; Poggio and Edelman, 1990) indicates that memory for specific instances may be more important for recognition than previously thought.

# 4 Cross-modal Characteristics and Processes

This section lists several characteristics of perception whose common denominator is generality and pervasiveness in vision. These are grouped into two classes. The first class includes *phenomena* that are common to more than one of the building blocks mentioned earlier. The other class comprises dynamic *processes* whose scope spans several visual submodalities.

## 4.1 Constancy

Perceptual constancy is our tendency to see properties of objects as invariant despite perpetually changing retinal stimuli. Following is a list of the most prominent constancy phenomena (some of these have already been mentioned in the preceding sections). Concrete examples of each phenomenon can be found, e.g., in Rock,

main aspects of the integration problem are the nature of the output representation and the relative weight given to each cue. Situations in which the cues are conflicting can be especially interesting. For example, when the contents of Figure 1a are shown stereoscopically in such a manner that the vertices of the illusory triangle appear, in conflict with the (imaginary) interposition cue, behind those of the real one, the illusion becomes weaker (Gregory, 1978).

## 3.7 Object Recognition

Mechanisms that support object recognition in human vision are the subject of considerable controversy among psychologists. Although most would agree that recognition involves comparison between the stimulus and an internal model or representation kept in memory, no consensus exists as to the nature of that representation. Consider the problems encountered by the visual system that attempts to identify an object present in the field of view. Assuming that the candidate object has already been detected and its approximate location estimated, the system must segment the object from the environment and factor out variations in its appearance due to changing illumination, changing viewpoint (see Figure 2) and, possibly, changing shape of the object (e.g., as in the recognition of a moving animal; see the section on Motion above).

Figure 2 here

Does the visual system represent objects in a straightforward fashion, e.g., by remembering sets of 2D "snapshots" taken from different vantage points, or are the object models, geometrically, 3D analogs of the objects they represent? Arguments based, among other phenomena, on our ability to perceive and describe the three-dimensional shape of novel objects led many researchers (notably, Marr, 1982) to postulate the formation of 3D object-centered (i.e., viewpoint-invariant) representations of the environment to be the ultimate goal and the final product of vision. This

ability to perceive 3D structure in moving 2D patterns, such as those that appear on a television screen, and even in random-dot kinematograms (Ullman, 1979). The contribution of motion to our overall impression of the world can be appreciated by anyone who has watched a film taken from the vantage point of a roller-coaster rider: the somatic illusions evoked by such stimuli are strong enough to override vestibular and somatosensory cues.

As Rock (1984) has pointed out, the presence of retinal displacement of objects is neither sufficient nor necessary for the perception of motion, despite the indisputable fact that such displacement is the starting point of neural motion processing. On the one hand, moving objects that are fixated and tracked are effectively stationary with respect to the retina, but are still perceived as moving. On the other hand, we perceive the visual world as immobile when our eyes move between fixations (but not when the eyes are moved externally, e.g., by gently pressing on the eyeball from the side).

## 3.6 Three-dimensional Space and Shape

The perception of the three-dimensional layout of visual space and of solid objects embedded in it relies on a combined action of all the basic modules mentioned above, as well as on a variety of perceptual "rules of thumb" that cannot be readily said to belong to one of those modules (Ramachandran, 1988). The role of shading, textural cues, retinal disparity and motion information in seeing depth has been outlined above. Another class of depth cues is provided by the oculomotor system: depth can be estimated from convergence of the eyes and from the accomodation status of the lens. Among the pictorial cues are interposition (inferred from occlusion of some objects by others), shadows, perspective, and familiar size information (see Rock, 1984). The problem of understanding the integration process that brings the depth cues together, traditionally neglected in favor of the study of individual visual modules, has recently received increasing attention (Bülthoff and Mallot, 1988). The

stereograms (image pairs consisting of random dots, some of which are displaced in one image with respect to the other to form a stimulus that can be perceived only through stereo vision; see Julesz, 1971). In a random-dot stereogram, each dot in one image can match potentially any dot in the other image. Although matching is usually selective enough to disambiguate such situations, in some cases people may perceive simultaneously several surfaces corresponding to multiple matchings between elements in the two images (Weinshall, 1989). Moreover, the relationship between the outcome of the matching and the perceived depth is sometimes not unequivocal. For example, the perceived depth may correspond to an average disparity rather than to one of the actual disparities derived from a possible matching (Mitchison and McKee, 1985).

The second process involved in stereopsis is surface interpolation, which fills-in the gaps between those locations in the image where exact disparities are found through matching. Similar to matching, the surface interpolation subsystem possesses several distinct features that are not well-understood and are not reproduced by machine vision algorithms. Two of these are simultaneous perception of multiple transparent surfaces and the integration of different depth cues in surface perception (see the section on Integration below).

## 3.5 Motion

Visual motion contains many important cues about the outside world. Moving patterns of light projected onto the retina provide information that can be used to segment the surrounding scene into objects according to their motion and to navigate in the environment while avoiding collisions with both stationary and moving obstacles. Visual motion can also be interpreted to yield the three-dimensional structure of objects (Wallach and O'Connell, 1953), even when the objects themselves are allowed to deform while moving, a common phenomenon in the motion of living things (see Johansson, 1973). The autonomy of visual motion perception is demonstrated by our

field. The identification of such features provides important clues to the structure of early visual processes. An illustration of this approach may be found in recent work by Fahle (1990), who found that vernier offset stimuli in the hyperacuity range (pairs of abutting line segments displaced by an amount that is smaller than the spacing of the photoreceptors in the retina) can be detected in parallel.

## 3.4 Stereopsis

Since the retinal image is a projection of the three-dimensional world onto a two-dimensional surface, the information on the third dimension, depth, is already lost at the very first stage of vision. The perception of depth, or stereopsis, can, however, still be attained by combining information from the two eyes (other — monocular — depth cues are mentioned below in the section on space perception). Stereopsis works because the separation between the retinal images of objects (retinal disparity) is different in the left and right eyes, depending on the separation of the objects in depth. Binocular stereo resolution is extremely fine: a difference in depth of one millimeter can be perceived at a distance of one meter. The disparity between the two eyes' views under such conditions is many times smaller than the size of a single retinal photoreceptor.

Stereopsis has received much attention in the study of vision (Julesz, 1971; Poggio and Poggio, 1984; see also the article on Stereo Vision in this volume). The behavioral importance of depth perception becomes apparent if one attempts to thread a needle, or catch a fly, with one eye closed. Stereopsis is an acquired ability: newborn babies do not perceive binocular depth until the age of three or four months. Disorders such as strabismus (the inability to fixate the same object simultaneously with the two eyes) present during this period of plasticity can cause permanent stereoblindness by hampering the development of one of the two main processes underlying stereopsis: matching the two retinal images to produce the disparity field. The power of the matching process is illustrated by our ability to perceive depth in random-dot

## 3.2 Color

In the perception of color, as in the perception of lightness, the human visual system exhibits an impressive disregard for irrelevant variables. In this case, the intensity and the spectral content of the illuminant must be factored out if a reasonable approximation to the color of the viewed surface is to be inferred (see the review in Boynton, 1978). The mechanisms responsible for this function appear to be similar for lightness and for color in that they depend on local contrast while ignoring slow and gradual changes in image intensity and spectrum, which in many cases can be safely attributed to the influence of the illuminant (Land and McCann, 1971). Observations that are not easily accounted for by an application of such simple fixed rules were made by Gilchrist (1977), who found that global (and cognitive) factors such as the knowledge of the spatial arrangement of surfaces may affect their perceived lightness.

## 3.3 Texture

In the natural world, texture, along with color, is an important cue to the physical composition of visible surfaces and can be used to segment complex scenes into surface patches that have distinct origins (e.g., belong to different objects). Also, texture gradients can be readily interpreted in terms of the orientation of the underlying surfaces and thus contain cues to the three-dimensional structure of the visual space (Gibson, 1950). Texture is a mass phenomenon: a surface must bear more than a few markings to be perceived as textured. The problem of texture perception can be formulated either in statistical terms, or in terms of the detection of the underlying texture elements or textons (see Julesz, 1984). An issue that was originally raised in the context of texture perception and has since been intensively studied is that of preattentive discrimination (i.e., distinguishing between different stimuli without the involvement of attention or scrutiny). Presumably, features that combine into preattentively discriminable textures are processed in parallel over the entire visual

ject identity between two simultaneously shown images of three-dimensional objects, and the response time was found to depend linearly on the relative misorientation of the objects with respect to each other.)

# 3 Building blocks

Reductionistic methods that investigate the structure of the perceptual system encourage the dissection of vision into submodalities. Some of these building blocks of perception – lightness, hue, texture, stereopsis, visual motion, the perception of space, and object recognition – are briefly described below. Within the scope of this article, little more than a hint can be given as to the perceptual problems solved by each module.

## 3.1 Lightness and Shading

Confronted with a gradient of illumination across the viewed surface, the visual system must separate the effect of illumination from the effects of surface albedo, orientation and shape. Disentangling all the factors that contribute to the intensity of the image at a given point on the retina is probably the most complicated computational problem in vision. It is not surprising then that some of the more compelling illusions, such as the Mach bands and the Craik-Cornsweet-O'Brien effect (see, e.g., Frisby, 1979, for an overview), are caused by peculiarities of lightness perception mechanisms (one such peculiarity is illustrated in Figure 1a, where the illusory triangle is perceived to be brighter than the background). In addressing the lightness problem, the human visual system appears to have settled for qualitative rather than quantitative solutions (Todd and Mingolla, 1983). Moreover, these solutions often seem to be based on high-level heuristics (Ramachandran, 1988) and are easily downplayed if the relevant information is available from other sources, such as the shape of the occluding contours of the surface (Koenderink, 1984), or stereopsis (Bülthoff and Mallot, 1988).

notion of a general psychological space may apply to stimuli that themselves possess a well-defined metric structure, the chances are meager that the more cognitive perceptual processes such as object recognition would admit a similar universal law-like or nomological description. If those philosophers are right who claim that mental processes are anomalous instead of nomological (Davidson, 1980), accounts of cognition in terms of prototypes and narrow-scope or local rules should be more fitting than invocations of universal laws.

## 2.2 Methods

There are many ways to collect the data necessary for building a theory of perception. One may distinguish between psychological approaches, which concentrate on the perceptual capacities and experiential aspects of perception, and biological approaches, which focus on the anatomy and physiology of the sensory nervous system. I discuss here only one approach — that of experimental psychology.

Experiments which quantify and measure the psychometric function – the response of a subject to a controlled stimulus – have traditionally been the principal method of the psychological study of perception. This experimental paradigm, called psychophysical because it relates the magnitude of a psychological response variable such as response time to some physical quality of the stimulus, dates back to the previous century. A clear formulation, due to Jastrow (1890), states that if the process of perception is indeed structured, then different paths through this structure will yield different response times. If this is true, one may hope to infer the structure of perception from the patterns of response times obtained under different experimental conditions. Mental rotation (see Shepard and Cooper, 1982, for an overview, and Pylyshyn, 1985, for a critique) provides an outstanding example of a phenomenon in which the dependency of response time on a characteristic of the stimulus has triggered hundreds of experiments and was incorporated into the foundations of a theory of visual representation. (In this case, the task called for a judgement of ob-

# 2 The Study of Perception

## 2.1 Goals

One way to begin investigating the basic questions mentioned above is to decide upon the form into which the answers should be cast. As in other natural sciences, this calls for the formation of a mathematical framework that would encompass the existing body of data on perceptual performance and would admit generalization by successfully predicting performance under novel conditions. In relatively primitive organisms the process of perception appears indeed to be amenable to precise mathematical treatment. As Poggio and Reichardt (1976) have shown, in the fly the transfer function of the module that supplies visual motion information for the purpose of flight control can be specified by a simple expression with only a few terms. The situation in human perception is rather more complicated. Among other reasons, this is because of the sheer diversity and complexity of the human visual system, because the output representations of visual modules are as yet unknown, and because in many cases top-down influences tend to interfere with "pure" perception (see the Phenomena section). A typical example of the classical mathematics of perception is Luce (1986), where one can find an exhaustive survey of the response time paradigm (see the Methods section below). The main mathematical tools of this paradigm are descriptive statistics and statistical models borrowed from signal detection theory (Green and Swets, 1966).

Shepard's proposal of a "universal law of generalization for psychological science" (Shepard, 1987) represents a more ambitious attempt at the mathematization of perception. The article, written for the occasion of the tercentenary of the publication of Newton's *Principia Mathematica*, suggests that the generalization of response from a learned to a novel stimulus depends on the distance between the stimuli in an abstract space that has the same metric structure for a wide variety of tasks, ranging from shape, size and color judgements to auditory signal perception. While the

resentations (Marr, 1982). The notion of perception as unconscious inference is usually traced back to Helmholtz (1856/1964). In his treatise on physiological optics Helmholtz argued that most often the information present in the visual world as projected on the eye's retina is too incomplete to support the richness of perception, which perforce must largely rely on previously or independently available data and on unconscious reasoning. According to Marr, an important source of this additional information is a knowledge of physics of the outside world and of image formation.

Consider again the example of Figure 1. A computational account of the perceived illusory triangle would have it that (i) all contours, including the illusory ones, are explicitly represented at some stage of the visual system, and (ii) the transformation between the stimulus and the explicit representation can be computationally specified (and implemented in the hardware of the brain). A distinct advantage of computational theories of perception is that, unlike "direct" accounts, they can be made to generate concrete predictions, which in turn can be experimentally upheld (or refuted). Computational, psychological and physiological investigations of illusory contours are reported in Ullman (1976) and Cavanagh (1987), and von der Heydt et al. (1984), respectively.

To summarize, the interesting questions that can be asked about visual perception are those of performance (what conditions affect the way we see?) and of competence (what does it mean to see? how do we see?). The computational/representational paradigm proved useful in addressing all these questions. At present this approach dominates the psychology of vision, consequently, in the remainder of the article I shall employ its assumptions and terminology without qualification.

an answer by trying to influence the percept of the illusory triangle by manipulating the stimulus (Figure 1b). As a by-product, this process of perceptual experimentation frequently comes up with notions that are useful for understanding other visual phenomena (in this specific example, the important notions are those of completion and filling-in, to be discussed later), including the perception of natural scenes.

## 1.2 Visual Competence

A different angle on the problem of perception is provided by the opening lines of Marr's book on computational vision (1982). Marr asks: what does it mean to see? This question, which (using Chomsky's terminology) is about competence rather than performance, cannot be elaborated upon much further without making a commitment to an important part of the answer. One of the two main approaches to this problem is through the notion of direct perception (Gibson, 1979). According to Gibson, perception is direct because the organism picks up relevant perceptual invariants, such as the three-dimensional shape of objects, from the visual world without intervening processing or representation (for a long time, Gibson went so far as to deny the existence of retinal images and the relevance to perception of geometrical optics involved in their formation). In its rejection of internal representations and processes that can be described as unconscious inference, the direct perception stance is related to behaviorism. The answer of Gibson's school to the question "what does it mean to see" is: to be attuned to certain invariant qualities present in the optic array, and, potentially, to become disposed to act in certain ways, given the appropriate stimulation. Under this view, phenomena such as that illustrated in Figure 1 tend to be dismissed as unnatural and "ecologically invalid".

The other approach, called *representational* by Marr (1982), postulates that the goal of any visual system, including biological vision, is to produce representations of the environment, and that various specific visual tasks, such as recognition and navigation, are solved through inference, or computation, defined over these rep-

on perceptual performance). Computational accounts of vision (Marr, 1982), also offered elsewhere in the encyclopedia, will be pointed to briefly. Finally, most of the topics that have to do with the biological substrate of vision, except some of the most pervasive notions (such as receptive fields and cortical maps), will be left out altogether. For information on the anatomy and the physiology of vision the reader is referred to other sources (e.g., Kandel and Schwartz, 1985).

## 1.1 Visual Performance

Gibson (1950) starts his book on visual perception with a question: why do things look as they do? An obvious (but not very informative) answer to this would be: because of the way our visual system is built. As a matter of fact, the recounting of conditions that must be fulfilled before anyone can see, which comprises the first paragraph of Gibson's book, begs for a more constructive formulation of the basic empirical question of visual perception: how does the visual system perform under different conditions, or, what are the factors that affect the way things look? Some of the factors that influence perceptual performance, such as the basic architecture of the visual system, are internal to the observer and belong to the domain of visual neuroscience. Other factors — the physical characteristics of the stimulus and of the manner of its presentation — can be externally manipulated. Experimental psychological study of visual perception is aimed at understanding the outcome of the perceptual process, its building blocks, and its neural substrate, through controlled manipulation of the external stimulus.

Figure 1 here

Consider as an example Figure 1a. When asked to describe its contents, average observers usually state that it shows two overlapping triangles, of which one has its vertices at the centers of the black disks and consists predominantly of illusory contours. Why do we see this second triangle? One can make an initial step towards

# Visual Perception

**Shimon Edelman**

Center for Biological Information Processing
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139, USA

## 1 Questions

The veridicality of visual perception — the correspondence between the percept of the outside world and its physical features that can be verified, e.g., through the sense of touch — is probably the most striking quality of vision. Despite its numerous shortcomings, catalogued and studied by psychologists as visual illusions, the visual system virtually never fails to provide information about the outside world that is of genuine behavioral importance. The three-dimensional layout of surfaces in the vicinity of the observer, their motion, their compositions, and their grouping into well-defined objects are representative examples. The resulting impression of the visual world is complete: the visual experience of the surrounding space has no gaps in it, even in those portions of it for which no input information is available (such as that part of the space projected onto the blind spot of the eye). At the same time, through fixation, attention, and scrutiny we can perceive many fine details of the environment.

Historically, reflections of the sort outlined above have led to many questions pertaining to the phenomenon of perception. These questions can be divided into distinct classes, according to the aspects of vision they address. This essay will deal with only a few of these aspects. It will survey some common features of the human perceptual performance, the processes involved in that performance, and how these are studied experimentally (Boff et al., 1986, is a recent comprehensive handbook