PATTERN RECOGNITION: HUMAN AND MECHANICAL

SATOSI WATANABE
University of Hawaii

A Wiley-Interscience Publication
JOHN WILEY & SONS
New York • Cichester • Brisbane • Toronto • Singapore

RS

CHAPTER 2
Pattern Recognition as Perception

2.1 OBJECT AND FEATURES

Whether we like it or not, under all the works of pattern recognition lies tacitly the Aristotelian view that the world consists of a discrete number of self-identical objects provided with, other than fleeting accidental properties, a number of fixed or very slowly changing attributes. Some of these attributes represent the peculiarity of a particular object; the rest of these attributes, which may be called "features," determine the class to which the object belongs. The assignment of a class to an object is essentially the task of pattern recognition. The features, therefore, are invariant properties when we pass from one particular object to another object within a class, and also when we change the mode of presentation of the same object to our observation.

The traditional approach to pattern recognition is to start with various methods of taking "neutral" data and then later consider classification. For those readers who are not familiar with the practice of pattern recognition we prepared Appendix A-1, Data Taking. Appendix A-2 explains Fourier transformation and Fourier optics as an example of extracting "invariant" features from the data.

It would be a commonsensible view to consider the "raw" data about the properties as provided by our sensory organs or instruments of observation or both and by the class-assignment as made by our cognitive judgment on the basis of the raw data. This picture of two-stage process of cognition is often deliberately reproduced in the mechanical design of pattern recognition, but we should ask ourselves whether such a picture is applicable to the natural perception in animals and man. This is the first question posed in this chapter.

The second question pertains to the idea of a discrete number of self-identical objects. What we see with our eyes are patches of colors that blend continuously into one another, or more precisely, one two-dimensional extension with continuously changing hues and shades. What falls on our ears is equally a continuously extending one-dimensional flow. The segmentation of nature into discrete objects probably has its ontogenetic and phylogenetic psychological origin closely related to the primitive animistic picture of the world; however, we can explain to some extent this object-segmenting tendency by the physiological effect of the Mach band. We shall further see this problem in connection with the convergence of sensory informations in the association area of the cortex.

The third question we should keep in mind while reading this chapter is whether the traditional separation between dispassionate cognition and affective evaluation, between knowledge and value, and between epistemology and axiology, is so clear-cut as is usually assumed. The same theme will reappear also in Chapter 4 and in Epilogue E-3.

The fourth question, which is related to the first question but bridges us over to Chapter 3 on the controversy about the universal, is whether the concept we use in classification is imposed by experience of the external world or more or less pre-determined by the innate nature of our minds, or perhaps by something that transcends our actual world of experience.

Keeping these questions in mind, let us study a few interesting facts in the realm of sensory science and animal perception, which is the primordial type of pattern recognition. The mechanical methods of pattern recognition will profit from the study of the natural perception, and the sensory science and epistemology will profit from the effort to make a mechanical facsimile of the natural perception. Of course, we have to expect that one approach which is successful in natural pattern recognition may not always be the best approach in the mechanical counterpart.

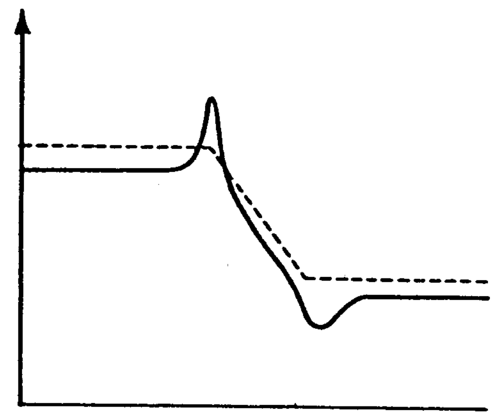

Figure 2.2.1. Mach's curve. The subjective intensity of Mach's curve shown here is the physical intensity plus a term proportional to the upward convexity of the physical intensity.

2.2 CONTOUR AND FIGURE

This section relates to the fact that when we ask someone, adult or child, sophisticated or naive, to draw a picture of something, he is likely to draw the contour lines of the object, whereas there are no such lines in physical signals from the object. This phenomenon has been a long-standing enigma, and there have been many would-be answers. But, the fact is that, apart from any explanations belonging to the conceptual level, a physiological effect underlying this phenomena has now become clarified, thanks to a series of investigations of which the very first dates back more than one century to Ernst Mach's 1865 paper [M-2, M-3]. The source of the following description is Mach's papers, Ratliff's book [R-2] and a personal conversation with Georg von Békésy (see also [V-3, V-4, K-5]).

What is known as the Mach band is the following phenomenon. Suppose the light intensity is distributed like the solid line of Figure 2.2.1. The left portion has a constant high light intensity, the central part represents a gradual decrease in intensity, and the right portion shows again a constant but lower intensity. In other words, the inclination is zero in the left portion and in the right portion, but it has a negative constant value in the central portion. If a human eye sees such a distribution of light, it sees an additional change of intensity as shown by a dotted curve on top of the solid curve. Where the inclination suddenly decreases, we see an additional increase of intensity, and where the inclination suddenly increases, we see an additional dip in intensity. Note that this statement need not be changed if we invert the positive x-direction of Figure 2.2.1.

If we want to see this effect by an experiment, the simplest way is to paint concentric zones on a round disk following Mach, so that the physical light intensity distribution along the radial direction becomes like the solid line of Figure 2.2.1 when we rotate the disk by an electric motor. We shall see, in addition to the three concentric zones, one bright band and one dark band between the zones. Mach expressed this effect by the formula

$$K = I - c\,\frac{\partial^2 I}{\partial x^2}, \qquad (2.2.1)$$

where I is the physical intensity of light, K is the intensity that the human eye sees, x is the x-axis of Figure 2.2.1 or the radial direction on Mach's disk. The second derivative becomes positive where the curve is concave upward, and negative where the curve is convex upward, hence the formula (2.2.1) represents the phenomenon explained above.

Mach's papers in physics and philosophy were well remembered by many readers, but his papers in physiology had quickly fallen into oblivion. But, the young physicist von Békésy remembered them, when a photographic record of a spectral line of a star started a controversy. The line in the photographic plate looked to the naked eye like two lines, suggesting that the two lines were caused by a double star due to the Doppler effect, whereas the photometric observation showed only one line. von Békésy gave immediately a solution to the controversy, by explaining the two lines as the Mach bands on both sides of a single line.

Our explanation used one dimension, but a similar psychological effect can naturally happen in a two-dimensional light distribution, and Mach already wrote a formula which is essentially

$$K = I - c\left(\frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}\right), \qquad (2.2.2)$$

where the operation in the parentheses is what is called the Laplacian. It is very interesting that the early pattern recognition engineers who struggled with picture processing introduced a digital equivalent of the Laplacian without knowing the works of Mach or von Békésy, in order to obtain the contour of an object.

An interesting anecdote is that Mach, who was known to be a pure positivist, was afraid that he would be criticized for introducing the "subjective" intensity (K) into his theory and tried to forestall such objections by declaring "even illusion is a fact that requires explanation." As a matter of fact, he even wrote that this effect must be caused by an interaction of neighboring cells on the retina. This theory at that time was only a conjecture, because neurophysiology in the modern sense then was still to be born.

This conjecture was proven to be true 90 years after Mach's theory by the discovery of "lateral inhibition" by Hartline and Ratliff [H-4, H-11] in 1950. When light falls on a spot on the retina, a positive signal is emitted through the optical nerve fiber originating from that spot on the retina. The lateral inhibition consists of suppressing the signals originating from the neighboring optical cells around the first irradiated point. This action can be expressed mathematically by the response function which is positive on the point and negative on the neighboring points (the von Békésy function). Hearing the news of this discovery, von Békésy immediately knew that this must be the mechanism of the Mach band.

Relegating a mathematical treatment to Appendix 2, we can qualitatively understand why it is so. The first derivative is essentially the difference of the values of a curve between two neighboring points. Repeating this process of difference taking twice, we can see that the second derivative (with a sign change) amounts to taking twice the value of the curve at a point and subtracting the value of the left-neighbor point and the value of the right-neighbor point. Therefore, it is easy to

Figure 2.2.2 *(a)* **is the original picture.**

understand that the lateral inhibition corresponds to the negative of the second derivative.

If we idealize the sensory response by mathematical formulae like (2.2.1) and (2.2.2), and make the transient zone (what we called middle portion in Figure 2.2.1) tend to zero, the additional increase and additional decrease would formally cancel each other. But insofar as there exists a finite transient zone, narrow as it may be, the contrast between the bright zone and the dark zone

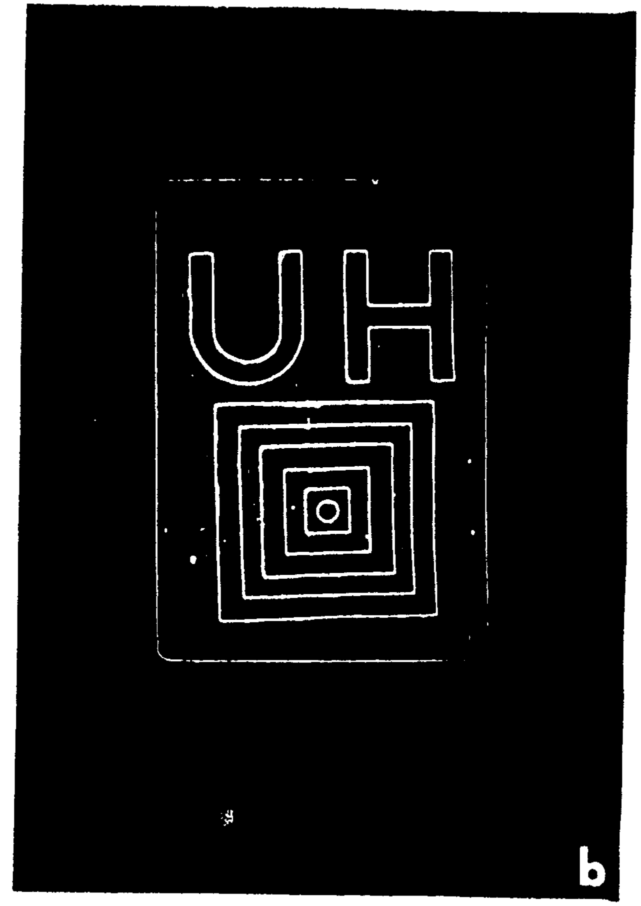

Figure 2.2.2 (b) is the contour of *(a)* obtained by the Fourier optics method simulating the Laplacian operator.

will make the contour very real. Even if the width of the transient zone becomes zero, the distance between the neighboring points previously mentioned is finite, and this finite distance will determine the distance between the bright zone and the dark zone. This distance is therefore physiologically determined. In any event, the contour vision is a real phenomenon from the point of view of sensory science, and not a simple subjective construction.

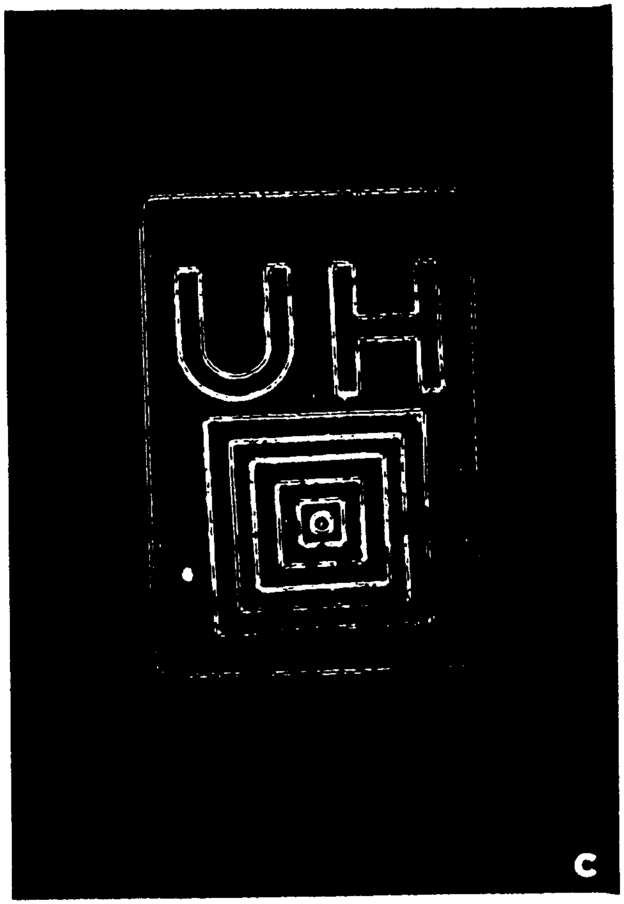

Figure 2.2.2 (c) **is the same as** ***(b)*** **but the horizontal distance between the maximum and the minimum of Mach's curve is assumed to be finite;** ***(b)*** **is the limiting case where the distance tends to zero.**

In computer simulation, we can do three things to see the contour: use the extended version of Laplacian keeping the distance between neighboring points finite; use the absolute value of the Laplacian (or even of the gradient, i.e., the first derivative along the steepest descent); or use

Fourier optics and apply the Laplacian to the amplitude instead of the intensity of the picture. Figure 2.2.2b is the contour obtained by the third method from the original Figure 2.2.2a in our laboratory in Hawaii. Figure 2.2.2c corresponds to the case where the maximum and minimum in the sense of Figure 2.2.1 are separated by a finite distance. In these pictures the intensity instead of the amplitude is photographed, therefore we see two bright lines (see Appendix 2).

von Békésy assumed a similar effect in the auditory sensation even before the actual discovery of lateral inhibition in optical sensation. Before von Békésy's work, people used to believe in Helmholtz's theory which claimed that a stationary acoustic wave was produced in the cochlea and that the sensitive hairlike cells, whose own frequency corresponded to the impinging wave, sensed that pitch. von Békésy showed that the acoustic wave produced in the cochlea is not a stationary wave, but a damped traveling wave. The envelope of this traveling wave has a maximum whose position depends on the frequency. The hair cell which is located near the maximum point senses the frequency. The trouble with this theory was that the maximum was not sharp enough to explain the extreme sensitivity of aural sensation on pitches. To explain this sharpness, von Békésy imagined something like the Mach band which would enhance the maximum. Unfortunately, this conjecture was proven recently not to be exactly tenable because the auditory nerve fibers are so connected that only neighbor cells on one side exert inhibition [D-2, R-3, S-6, Z-1]. As a result, the first derivative instead of the second derivative is added to the straight sensation. Yet von Békésy's idea that the sharpening of auditory sensation is affected by the inhibitive action of the neighboring cells seems to be still true.

In later years, von Békésy tried and succeeded in extending his theory of sharpening of sensation by lateral inhibition to all kinds of sensation other than the eye and the ear. When the author visited von Békésy's hospital bed immediately after his death in 1971 in Honolulu, there was on the bedside table an experimental setup consisting of a rotating disk very much like Mach's disks.

We have seen in Figure 1.1.3 that the picture represents a goblet or two faces according as our "attention" is shifted from one aspect to another. The object we "see" at one instant is called the "figure," and the suppressed aspect is called "ground." Now, from the point of view of the usage in ordinary language of the term contour, we have to say that the contour is to be considered to belong to the figure, not to the ground. Therefore, the contour from one instant to another changes its affiliation. But, all the same, the contour vision makes the figure more conspicuous. It is interesting to note that the term figure often means the contour of the conspicuous object, i.e., of the figure in the Gestalttheoretic vocabulary. For instance, when we say that she has a good figure, it is implied that we make abstraction of the hue of the skin, or of the ugliness of the face, or of the physiognomy.

If a picture is given like Figure 1.1.3, there is no reason to decide which one of the two alternatives is the "correct" interpretation, although there may be a difference in the degree of Prägnanz. But, if the same picture is a part of a larger picture which represents a group of bottles and glasses, the interpretation of the picture as a goblet may be judged to be the right one. Therefore, the decision of figure-ground may be said to be context dependent. Again, if the observer is unfamiliar with this type of goblet, he will not see the goblet aspect of the picture at all. Thus the notion of context should be extended not only to all the coexisting objects but also to the background of the observer.

Coming back to the context-free case, i.e., assuming that both aspects are equally valid, we have to note an important fact, already mentioned, that when one aspect is seen, the other aspect completely disappears. In other words, two alternative figures are not simultaneously tenable. Yet, it is not justifiable to say that one aspect is correct and the other is incorrect. Both have an equal right to reality although they are conflicting with each other. Such a situation fits perfectly well with Niels Bohr's epistemological principle of complementarity, which he introduced to explain the wave-particle duality of an elementary particle. The reality can have two

conflicting aspects, either of which is correct and consistent, yet both not simultaneously tenable.

The lesson we draw from this section is that already at the stage of peripheral organs, our sensation already modifies the incoming physical signals in such a way to facilitate pattern recognition in terms of discernible objects.

2.3 PSEUDOCONCEPT

We saw that the sensory organs of animals detect not only the simple variables of physical stimuli but also the first and second spatial derivatives of them when these are useful for the purposes of life. But, the derivatives are still relatively simple variables closely related to the physical variables. In nature, some of the peripheral organs send signals to the nervous center which are quite complicated functions of incoming physical stimuli. In higher animals, complicated functions of physical signals are mostly formed on the cortex of the cerebrum to represent "concepts," but in the lower animals whose brains are not developed, quite complicated functions seem to be formed at the peripheral levels which in higher animals would correspond to concepts. If we agree that the pattern is recognized by the "mind's eye," the "mind" must be said to be extended to the peripheral sensory organs. Let us describe briefly the famous research made by Lettvin et al. [L-3, L-4] on the frog's eye. It is a classical example showing that the goal-matching information selection and something similar to the formation of "concepts" start already right after the retina, i.e., before the signals reach the brain. Even a simple geometrical or physical variable corresponds to a concept, but what we call concepts in this section are those which correspond to very complicated functions of physical stimuli.

In the frog's eye, light falls on about one million receptors (rods and cones), and the signals go out of the eye in a bundle of optic nerve fibers which are actually the axons* of the so-called "ganglia" cells, of which there are about one half million. In the back layer of the eye, the receptors and the ganglia cells are connected by bipolar cells. There are also a total of 2 1/2 to 3 1/2 million horizontal connecting cells. As a result, one rod or cone is connected to a large number of ganglia cells, and conversely one ganglion cell is connected to many thousands of receptors. The two bundles of optic nerve fibers coming from the left and right eyes are crossed and connected to the optic tectum which lies on the dorsal side of the brain.

Lettvin and collaborators detected electric signals in individual axons in the optic nerves and in cells in the optic tectum. In the frog's eye, there is no fovea. Some ganglion cells have a receptive field (area that they "see") of only a few degrees in diameter (1° - 4°), and some react to a very wide field. In the optic nerve fibers, the geometric relations of receptive fields of ganglion cells are all mixed up, but they seem to be restored when they arrive in the optic tectum. The cells in the optic tectum, however, are no longer connected to definite receptive fields but react to the more global feature of the entire field.

Lettvin et al. discovered five major groups of optical fibers: Group 1, boundary detectors--the receptive fields of these fibers are 2° to 4° in diameter, and they respond to any sharp boundary between two areas with different greyness. The sharpness of the boundary is important, not the degree of contrast between the two greynesses. Motion of the boundary enhances the electric discharge, but is not necessary to cause the discharge. Group 2, convexity detectors--these fibers have receptive fields of 3° to 5°. They respond only to sharp boundaries between two areas of different degrees of greyness, provided the boundary is curved and the "inside" area is darker than the "outside" area (i.e., the darker area is convex), and provided the boundary has moved or is moving. In other words, these nerve fibers respond strongly to a small dark object that has moved or is moving. When such an object moves into the receptive field and stops, it causes a long-lasting discharge which stops if the light is turned off, and does not restart if the light is turned on again, unless the object moves again. Groups 1 and 2 belong to what is called unmyelinated fibers (i.e., fibers that are not covered with soft, white, fatty sheaths), and make up about 97% of the optic nerves. The signals travel at velocities between 20 and 50 cm/sec. Groups 3 and 4, which occupy the remaining 3% of fibers are myelinated, and the signals travel much faster at various velocities between 1 to 10 m/sec. Group 3, moving boundary detectors--in contrast to Group 1, these fibers respond only when the boundary is moving or changing. In contrast to Group 2, the discharge stops when the object stops. The discharge response becomes more frequent when the boundary is sharp, and the speed of the motion is high. Group 4, dimming detectors--respond by a prolonged discharge to any dimming of illumination in the entire receptive field, which is as large as 15°. They are insensitive to the existence of boundaries or motions of objects.

The fibers belonging to Group 5 are very few and their function is not studied in detail, except that their firing rate seems to be an indicator of the average illumination. The darker the illumination, the higher the firing rate. It is significant that the authors observed nerve fibers with very high speed (20 m/sec), which they assumed carried messages from the brain to the retina, in contrast to the great majority of nerve cells that carry messages in the opposite direction. It is also interesting that they anatomically recognized four different types of dendritic connections of ganglion cells at the back of the retina. Some cells have connections at one particular depth, some others have connections at more than one particular depth. Some have horizontal widespread connection and some others

---

*A nerve cell consists of a cell body, dendrites (signal input) and an axon (signal transmission and output). Both dendrites and axon-ends divide into small branches establishing contact with other nerve cells. A potential peak (spike) travels along an axon with various velocities peculiar to each cell. The intensity of response is inversely related to the interval between consecutive spikes. The propagation of signals by electric spikes, unlike the electric current in metals, is determined by the motion of ion concentration and is, therefore, much slower than the light velocity.

have more concentrated connection. They could establish with sufficient plausibility a correspondence between the functional types and the anatomic types. For instance, the dimming detector respondent to the general darkening of the wide field is identified with the one-level widespread type.

At the brain end of the optic fibers, the four major types of axons end at four different depths in the outer layer of the optic tectum. Below the outer layer there are nerve cells which receive signals from the optic fibers. Among these cells, Lettvin and collaborators recognized two major types of neurons. The first type is named "newness" neurons. If an object moves across the neurons' receptive fields, which are as wide as 30°, the neurons respond with frequencies dependent upon the jerkiness, velocity, direction of the motions, and the size of the object. The peculiarity of these cells is that they do not respond for at least 5 to 10 sec when we repeat the same motion of the same size object in the same direction. Even after 20 sec the response remains weak. If the first movement and the second movement are at right angles, there is no interference; the second movement is new.

The second type of neurons is called "sameness" neurons by those authors. If a small object, 1° or 2° in diameter, is brought into the field (which is almost as wide as the entire visual field), the nerve cell starts to "notice" after a short time. After the starting time, the response continues as long as the object is in the field and becomes intense every time the object makes any kind of jerky move. The authors report a particularly interesting phenomenon in the case where they showed two objects at once. At least with a subclass of sameness neurons, it was observed that they respond only to one or the other of the two objects at one time and not to both of them simultaneously, although their attention may shift from one to the other when one of them stops.

The frog's survival depends upon factors as the sky, pond, small insects, snakes, etc. It is obvious that the four types of ganglion cells are developed to send signals selectively pertinent to these vital factors. (Group 2 is often called "bug detector.") It is amazing that such a selection of variables takes place already right behind the retina instead of in the brain. In the optic tectum, there is already a prototype of memory and also a special faculty of concentrating attention to one object of concern.

The bug detectors send signals corresponding to the concept "bug-ness," although we cannot say that the peripheral organ literally forms a concept. Similarly, the frog can see the snake-like objects and pond-like objects. With this type of concept-like sensation on one extreme and the sophisticated genuine concepts in human intelligence on the other extreme, we can further recognize a third kind of class, which may be called behavioral categories of stimuli. For instance, in the T-maze experiment, an untrained rat will go up the entrance branch and turn to the left and to the right with an approximately equal probability (relative frequency). But, if a food pellet is placed consistently on the left branch, it starts to turn to the left with an increasingly higher probability, until finally it ends up turning consistently to the left. We cannot say that there is a particular inborn behavioral distinction between left and right, nor can we say that the rat has created a concept of left and right. But, its response pattern looks as if it learned the "concept" of left and right, and causally associated left with food.

The fact that the rat has extracted the "pseudoconcept" of left and right among other possibilities can be shown by simultaneously illuminating one or the other of the two exit branches randomly. It will soon discover that there is no relation between light and food, and will learn to turn to the left to get the food. But, on the other hand, if we do not fix the location of the food pellet and place it on the illuminated side, the rat will probably learn to turn to the illuminated branch.

What is important to remember at this point is that the selective formation of a pseudoconcept or a response category is intimately related to what is valuable for survival--in this case, the food pellet. In the T-maze experiment, and in many other similar experiments of learning and conditioned reflexes*, the animals learn to notice selectively such a feature or pattern in the situation in which they are placed and their responses based on that feature would lead to acquisition of a desirable "reward." Pattern recognition in humans is not basically different, but somewhat in the same way as science can become independent of technology, we are capable of carrying out a pattern recognition task which is not directly related to an immediate profit. For a further discussion of the subtle interrelation between the value-system and the concept formation, see Chapter 4, and Epilogue 3.

We have learned two important things in this section pertinent to the work of pattern recognition: (1) The choice of observational variables in animals is directly related to their value situation, (2) The concept formation is not exclusive to human intelligence, but its prototype exists already in lower animals and is often carried out not necessarily at the cortex. At this level, we cannot distinguish between observed variables and concepts.

Thus, for a computerized pattern recognition, the most vital thing is to choose good observational variables which may be quite different from the simple variables that physicists would use to describe the physical aspects of the incoming stimuli. The physicists' variables are chosen so as to make the physical laws not only possible but also simple. The value for them here is the scientific truth described by the physical laws.

2.4 SYNTHESIS OF SENSORY STIMULI

Since the pioneering investigation of Lettvin et al. opened a new era of neurophysiological study of perception, more than 20 years have passed. During this time a fantastic number of experimental results in this field have been accumulated. It is therefore

*A Russian scholar, Bongard [B-9], recently emphasized that animals can learn many things that do not exist as unconditioned reflexes, showing that Pavlovian theory is too narrow to explain the whole gamut of learning processes. Compare this idea, however, with the findings described in the next section, where some concepts are shown to be innate.

impossible for a layman like the present author even to try to make a brief survey of the new progress. It may be of some interest, however, to pick up randomly a few topics from the newer findings that have relevance to pattern recognition. The animals used now for experiments are usually monkeys and cats which certainly have more similarity to humans than the frog.

One of the most valuable projects that requires a great deal of tedious and careful labor is the mapping of the interconnections of neurons. The cortex can be divided into three areas: (1) the receptor area where sensory stimuli arrive, (2) the motor area where a new decision of action is emitted to the effectors, and (3) the association area where presumably complex information processing takes place. The neurons in this area are eventually connected to the receptor area and the motor area, but the flow of information within this area is highly complicated. In general, three types of connection are observed in the association cortex: divergent type--starting from one neuron or a few neurons, the signals diverge to many different regions; parallel type--signals from a small region are converged to a small region; convergent type--signals from different regions are conveyed to a single neuron or to a small region (see Chavis and Pandya [C-6]). The convergence type of connection is interesting, because, as we have seen, pattern recognition aims at finding one image in a field consisting of many "parts." In the converging type of connection, a neuron can receive signals not necessarily of the same kind.

The perception of depth (i.e., the 3-D view) is made possible primarily due to the disparity of the left and right visual signals, and, in fact, Bishop and others have discovered neurons which are sensitive to the left and right disparity (see for instance [B-10]). But a perfect 3-D image requires more than the disparity of binocular view. Sakata reports that a visual fixation neuron, which fires when a certain point is gazed at, shows a positive response at a certain direction and a negative response at a different direction. A similar discrimination is shown with regard to the distance of the object gazed at, which demonstrates that on such a neuron, not only the visual signal but also the information about the muscular tensions (which are governed by a motor decision) converge and help the depth sensation [S-13]. It is, at least, now clear that the perfect 3-D effect is produced by a synthesis of many different stimuli.

Apart from the visual fixation neurons, there are visual tracking neurons which fire when following a moving point. Sakata discovered that the same tracking neuron fires either by a moving point or by a background moving in the opposite direction and fires more strongly by both motions together [S-14]. This provides, as Sakata rightly points out, a neurophysiological backing to the idea of "induced motion" discovered by Gestalt psychologist K. Dunker (see p. 282 of [K-13]).

Helmholtz, in spite of his fame as an early proponent of neurophysiological explanation, is known to have insisted on the necessity of innate mental (instead of physical) activity to make the 3-D vision possible. The newer researches seem to show that what was called "mental" was just a result of confluence of many sensory (and perhaps even motor) signals. But, note that this visual faculty is not acquired nor learned but innate. Also note that this neurophysiological explanation does not deny the coexistence of mental activities simultaneous with physical activities.

More close to pattern recognition is the discovery by Perrett et al. [P-10] of visual neurons in the monkey that respond exclusively to human and simian faces. The responses were found to be relatively constant for such transformations as inversion, change in color, and change in size, but the response diminishes drastically when the front view is changed to the profile. The fact that different neurons of this group respond to different parts (eye, mouth or hair) of the face shows that these cells did not represent the last stage of synthesis, but are very close to the formation of the idea of "face-ness." The facial recognition is known to be present in newborn animals showing that this recognition is definitively an innate capability necessary for survival.

On a more abstract level, Sato and collaborators [S-15] discovered a small number of neurons which respond exclusively to one or another of four geometrical figures (cross, circle, triangle,

square). The response was found not to depend on size, color, illumination and, above all, on training but depends strictly on the mentioned kinds of figures. This is a striking capability of form-discrimination. If this new discovery stands all possible criticisms, it will show that there is an innate idea of cross, circle, triangle, and square. Could there be, however, as many kinds of neurons as there are possible geometrical forms? Is this capability limited to a small number of nice, simple forms with a high degree of Prägnanz? Could this give support to the philosophy of Cubism in painting?

It is meaningful to ask if there is anything in the neurophysiological findings that is necessarily bound up with the Aristotelian view of nature that the world consists of a discrete number of self-identical objects with fixed attributes (see Chapter 3). The fact of contour vision and the existence of visual fixation neurons and visual tracking neurons certainly encourage the use of the idea of self-identical objects. But, these neurophysiological findings by no means confirm or necessitate the Aristotelian view; they only invite us to assume the short-term sameness of a discerned "object." On the other hand, some neurons may be said to have inborn capacities of perceiving certain "forms," certain "---nesses," such as "face-ness, triangle-ness, etc.," whereas some other alien properties such as size, color, illumination are abstracted away. We should remember that Plato did not believe in particular objects because they are mixtures of fleeting properties and insisted on the permanence of innate "ideas." It is, of course, a far-fetched comparison, but note that the content of the neuronic responses is unique, innate, abstract, and seems to pre-exist and supercedes the concrete particular objects with multifarious, ever-changing, often deceiving properties. We thank Dr. Sakata for imparting some of the recent discoveries in neurophysiology. He let us know by a recent private communication that he had discovered a visual neuron of the monkey that is exclusively sensitive to the clockwise motion of an object and a neuron that is sensitive to the counterclockwise motion of an object. He said that so far he could not find neurons exclusively sensitive to vertical and horizontal motion. See Epilogue 3 for more about the recent findings in neurophysiology and our comments.

2.5 PERCEPTION AS MENTAL CONSTRUCT

In the last two sections we relied on the commonsensible view that the neurophysiological phenomena and the mental process have intimate correspondence. This assumption is very difficult to prove and forms the center of the long-standing philosophical controversy over the mind-body problem. The author has his own theory about this problem [W-33, W-34] but does not want to go into the philosophical arguments here. Rather in this book he will proceed with the assumption (which is allowed according to our theory) that there is a rough (but not exact) parallelism between the neurophysiological-behavioral description and the mental description. See Epilogue 2. Furthermore, we take the view that consciousness or awareness is not limited to man and that the ratio of consciousness to unconsciousness gradually diminishes when we go down the scale of complexity of organic beings.

From this point of view, human beings provide a very useful case because we can get verbal reports of inner experience, which should not be so different from the animal experience. In this respect, we differ from those philosophers who are too strongly influenced by the linguistic theories and the ordinary language philosophy. They tend to believe that our intelligence and thought are based on, and conditioned by, our language. Their theory will be shattered by the recent experimental discoveries that show that human children can recognize patterns and classify objects before they start to talk or understand others talking. This is not surprising from our continuous viewpoint of the animal-man transition. In this section, we shall include more of the human introspective reports in addition to purely non-mental findings.

One of the important facts we emphasized in the last three sections was that sensation and perception are not a passive and unbiased transmission of physical stimuli, but they are an active selective formation of valuable information. It is interesting to note that such a statement would have aroused less sensation in the 19th century than in the first half of this century. Remember that Goethe's Farbenlehre was based on the theory that the eye is an active agent rather than a passive recipient. Kant could advance his theory, which for the sake of simplification we might call a theory of innate ideas, without arousing a resistance it would have caused in this century. This century with its deluge of technological successes which were essentially based on the 19th-century classical physics brought about a new Zeitgeist, championed by the Wiener circle which preached unification of sciences under the banner of logical empiricism, but, the trend had to change quickly. The middle of this century may be epitomized as a reaction to empiricism, calling for a return, in a guise or another, to what may be roughly called the Kantian tendency. The author still remembers having participated in a conversation one evening at the AAAS (American Association for Advancement of Science) meeting at Denver in 1961, in which Lettvin tried to explain his discoveries to Herbert Feigl (the last remaining great member of the Wiener circle, and who had an open mind) saying that his discoveries about the frog's eye can be understood only in terms of the "Kantian" view, but not in terms of the passive philosophy of logical empiricism. In the field of philosophy of science, the trend toward the "Kantianism" gained more and more momentum, and in the field of perception psychologists have become more and more convinced that a simple "passive" view had to be abandoned.

The weakest form of the "active" theory of perception may be an emphasis on "attention." Among many others who expressed similar ideas in the 19th century, Franz Brentano held the view that the distinctive feature of the mind was its "Dirigiertsein," i.e. attention being directed on one object. Applied to visual perception, this may mean that the eye can actually see only one thing out of many things in the visual field, or one part out of the whole, or an aspect out of more than one. According to Brentano, the phenomenon of "directedness" can be seen in presentation (Vorstellang), judgment, and emotion. His often quoted idea of "intentional inexistence" (we can think of something that does not exist) gave the starting point to Edmund Husserl's phenomenology. In the mid-20th century, many psychologists rediscovered the importance of attention in various aspects of human mental activity.

One of the most interesting phenomena showing the surprising faculty of attention of human perception is what is called the "cocktail party effect." The human ear can hear discriminately one particular person's voice in a noisy room where many people simultaneously talk. Psychologists in experiments in visual perception discovered what they call "fragmentation." If a subject is forced to see the same view for a long time, some part of the view disappears completely from perception leaving only the part on which attention is concentrated. Again, when many characters are shown simultaneously, the human recognizer immediately reads characters one after another, whereas the mechanical recognizer programmed for reading characters would be unable to concentrate attention on one character at a time, and then shift the attention.

The concentration of attention on one part of the visual field is the simplest case, where the mechanical recognizer can be expected to be programmable to cope with the task. But, the overlapping in the same dimensionality, such as double exposure photography or the case of a cocktail party effect may be more difficult, if not impossible, for the mechanical recognizer.

The case of Gestalt-theoretical flip-flop of figure and ground may be included as an example of attention concentration and shifting. As we remarked before, it is important to note that it is possible to shift from one object or one aspect to another object or another aspect and that at one instant we cannot share attention to two objects or to two aspects.

It may be a little questionable to include the following problem as a case of attention concentration and shifting, but it is interesting to view the discovery of invariance in pattern recognition as a shift of attention to the invariant properties of the object and a rejection from attention of the non-invariant properties of the object. It is significant that a young child has no hesitation in classifying the same letter rotated in different angles as one thing, whereas little older

children have more difficulty in classifying them as the same letter.

It should be remembered that a prototype of "attention" and its "switching" were observed in the frog's eye in connection with "sameness" and "newness" neurons, and it may be justifiable to think that attention and its switching constitute the very fundamental nature of the nervous system.

The attention implies a nonpassive selection but is not really an active modification of incoming signals. The case of the Mach band is a definitive modification, however, it is physiologically determined and not an alteration of signals due to the mental activity of the observing subject. Some of the psychological illusions are definitively subject-dependent. Hallucinations are mental phenomena of seeing nonexisting objects and their contents are independent from outside optical stimuli. Even without going to the extreme of hallucinations, we can mention cases of our active modification. For instance, we shall read Q as O, without noticing that we modified our perception if the letter is placed in a position in a word where usually O stands instead of Q. A similar thing can be said about m and m, or ち and ち. Here "familiarity" plays an important role, and we can characterize such phenomena as "context-dependent," or Gestalt-dependent, or global sensing perception.

An important subclass of Gestalt-dependence is what some psychologists call "physiognomical perception." A secretary can often tell whether her boss has a bad mood today, without being able to say why she judges so. A dog will jump into a car, when it perceives that the master is going to make a long trip, out of fear that it may be left behind alone. This response happens even if the master takes a very careful deceiving technique. The fortune teller usually has a well developed sense of physiognomical perception.

Another example of "global perception," where things are seen that exist in no way in the details of the view, is used in the Rorschach test. In Figure 1.1.4, we saw an example of the Rorschach picture. The doctor asks, "What do you see in this ink-blot?" The answer is supposed to reveal the emotional, intellectual, and characterological trend of the subject. The objects of value, fear, desire, suspicion, suppression, familiarity, etc., will emerge in this test. We can see the emotion-laden complex through this test. This "what" in this question corresponds to $something_2$ in Wittgenstein's "I see $something_1$ for $something_2$." In the language of pattern recognition, it is the class into which an object is placed. (That is why it is subject to inductive ambiguity.) It is important to know that the answer is not dictated by the object [$something_1$] and is very much dependent on the emotional and evaluative situation of the subject.

We have repeated again in this section that perception is not a simple "passive" act. If this active aspect is explained only in terms of our innate categories, it is a Kantian view. We want to draw the attention of the reader to the fact that the active aspect is, in a deeper analysis, an emotion-dependent and value-dependent mental construct, not a colorless Kantian category (see Epilogue 2 for more about the modern cognitive psychology).