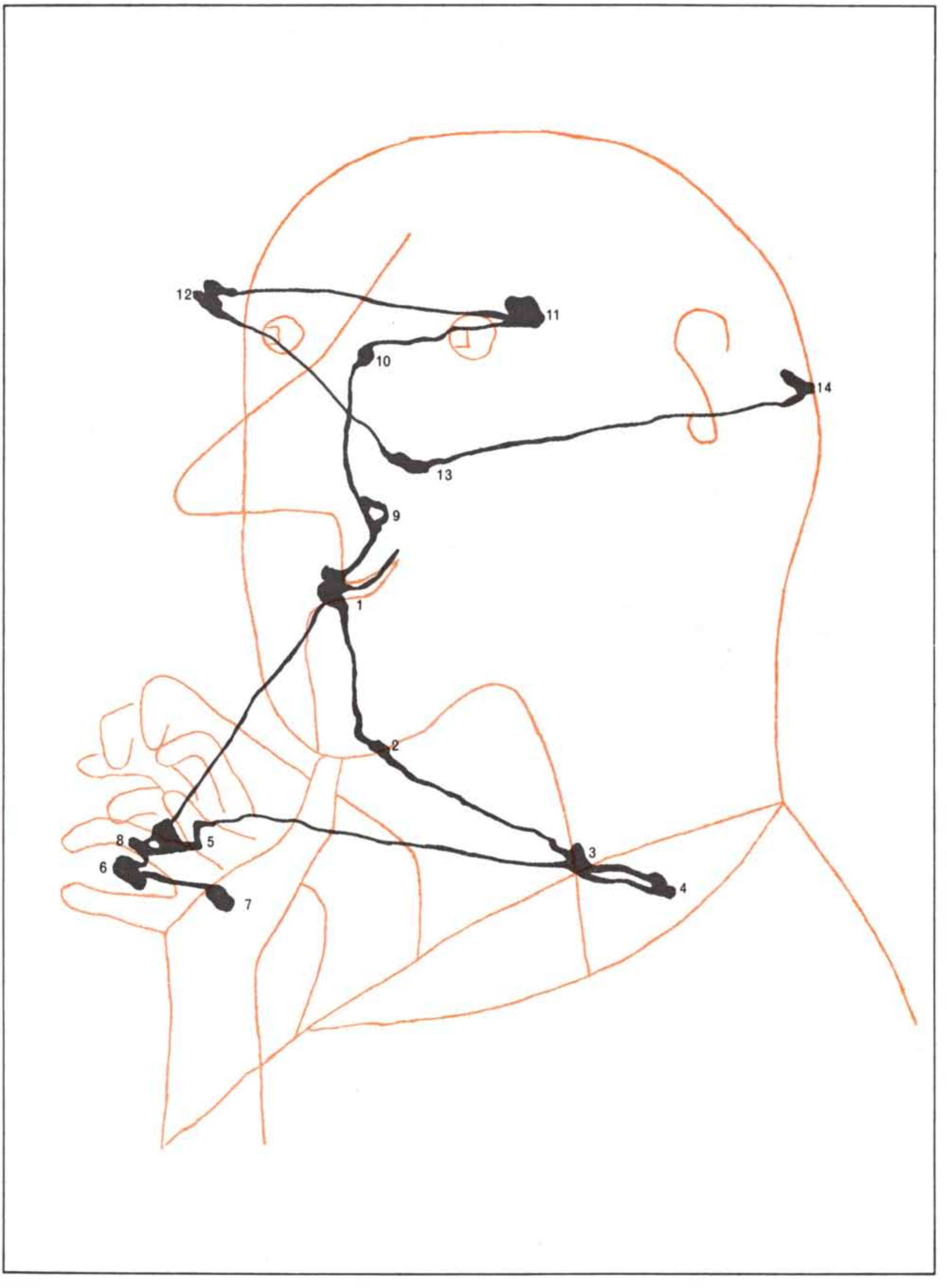
Eye Movements and Visual Perception

Author(s): David Noton and Lawrence Stark
Source: *Scientific American*, Vol. 224, No. 6 (June 1971), pp. 34-43
Published by: Scientific American, a division of Nature America, Inc.
Stable URL: https://www.jstor.org/stable/10.2307/24922750



EYE MOVEMENTS made by a subject viewing for the first time a drawing adapted from Paul Klee's "Old Man Figuring" appear in black. Numbers show the order of the subject's visual fixations on the picture during part of a 20-second viewing. Lines between them represent saccades, or rapid movements of eyes from one fixation to the next. Saccades occupy about 10 percent of viewing time.

Eye Movements and Visual Perception

Recordings of the points inspected in the scanning of a picture and of the path the eyes follow in the inspection provide clues to the process whereby the brain perceives and recognizes objects

by David Noton and Lawrence Stark

The eyes are the most active of all human sense organs. Other sensory receptors, such as the ears, accept rather passively whatever signals come their way, but the eyes are continually moving as they scan and inspect the details of the visual world. The movements of the eyes play an important role in visual perception, and analyzing them can reveal a great deal about the process of perception.

We have recently been recording the eye movements of human subjects as they first inspected unfamiliar objects and then later recognized them. In essence we found that every person has a characteristic way of looking at an object that is familiar to him. For each object he has a preferred path that his eyes tend to follow when he inspects or recognizes the object. Our results suggest a new hypothesis about visual learning and recognition. Before describing and explaining our experiments more fully we shall set the stage by outlining some earlier experiments that have aided the interpretation of our results.

Eye movements are necessary for a physiological reason: detailed visual information can be obtained only through the fovea, the small central area of the retina that has the highest concentration of photoreceptors. Therefore the eyes must move in order to provide information about objects that are to be inspected in any detail (except when the object is quite small in terms of the angle it subtends in the visual field). The eye-movement muscles, under the control of the brain, aim the eyes at points of interest [see "Control Mechanisms of the Eye," by Derek H. Fender, Scientific American, July, 1964, and "Movements of the Eye," by E. Llewellyn Thomas, Scientific American, August, 1968].

During normal viewing of stationary objects the eyes alternate between fixations, when they are aimed at a fixed point in the visual field, and rapid movements called saccades. Each saccade leads to a new fixation on a different point in the visual field. Typically there are two or three saccades per second. The movements are so fast that they occupy only about 10 percent of the viewing time.

Visual learning and recognition involve storing and retrieving memories. By way of the lens, the retina and the optic nerve, nerve cells in the visual cortex of the brain are activated and an image of the object being viewed is formed there. (The image is of course in the form of neural activity and is quite unlike the retinal image of the object.) The memory system of the brain must contain an internal representation of every object that is to be recognized. Learning or becoming familiar with an object is the process of constructing this representation. Recognition of an object when it is encountered again is the process of matching it with its internal representation in the memory system.

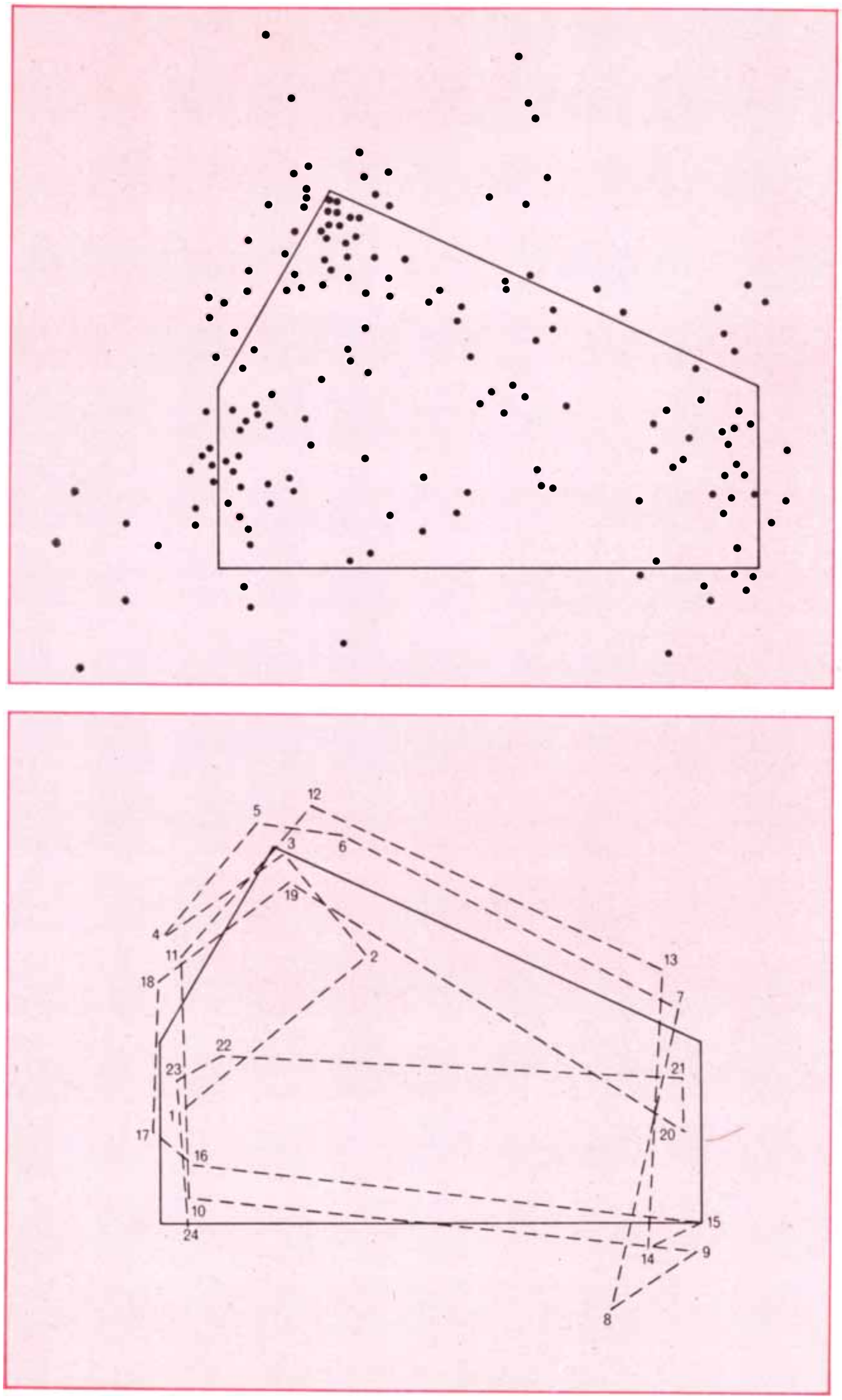
A certain amount of controversy surrounds the question of whether visual recognition is a parallel, one-step process or a serial, step-by-step one. Psychologists of the Gestalt school have maintained that objects are recognized as wholes, without any need for analysis into component parts. This argument implies that the internal representation of each object is a unitary whole that is matched with the object in a single operation. More recently other psychologists have proposed that the internal representation is a piecemeal affair—an assemblage of parts or features. During recognition the features are matched serially with the features of the object step by step. Successful matching of all the features completes recognition.

The serial-recognition hypothesis is supported mainly by the results of experiments that measure the time taken by a subject to recognize different objects. Typically the subject scans an array of objects (usually abstract figures) looking for a previously memorized "target" object. The time he spends considering each object (either recognizing it as a target object or rejecting it as being different) is measured. That time is normally quite short, but it can be measured in various ways with adequate accuracy. Each object is small enough to be recognized with a single fixation, so that eye movements do not contribute to the time spent on recognition.

Experiments of this kind yield two general results. First, it is found that on the average the subject takes longer to recognize a target object than he does to reject a nontarget object. That is the result to be expected if objects are recognized serially, feature by feature. When an object is compared mentally with the internal representation of the target object, a nontarget object will fail to match some feature of the internal representation and will be rejected without further checking of features, whereas target objects will be checked on all features. The result seems inconsistent with the Gestalt hypothesis of a holistic internal representation matched with the object in a single operation. Presumably in such an operation the subject would take no longer to recognize an object than he would to reject it.

A second result is obtained by varying the complexity of the memorized target object. It is found that the subject takes longer to recognize complex target objects than to recognize simple ones. This result too is consistent with the serial-recognition hypothesis, since more features must be checked in the more complex object. By the same token the result also appears to be inconsistent with the Gestalt hypothesis.

It would be incorrect to give the impression that the serial nature of object recognition is firmly established to the exclusion of the unitary concept advanced by Gestalt psychologists. They have shown convincingly that there is indeed some "primitive unity" to an object, so that the object can often be singled out as a separate entity even before true recognition begins. Moreover, some of the recognition-time experiments described above provide evidence, at least with very simple objects, that as an object becomes well known its internal representation becomes more holistic and the recognition process correspondingly becomes more parallel. Nonetheless, the weight of evidence seems to support the serial hypothesis, at least for objects that are not notably simple and familiar.



IMPORTANCE OF ANGLES as features that the brain employs in memorizing and recognizing an object was apparent in experiments by Leonard Zusne and Kenneth M. Michels at Purdue University. They recorded fixations while subjects looked at drawings of polygons for eight seconds. At top is one of the polygons; the dots indicate the fixations of seven subjects. Sequence of fixations by one subject in an eight-second viewing appears at bottom.

If the internal representation of an object in memory is an assemblage of features, two questions naturally suggest themselves. First, what are these features, that is, what components of an object does the brain select as the key items for identifying the object? Second, how are such features integrated and related to one another to form the complete internal representation of the object? The study of eye movements during visual perception yields considerable evidence on these two points.

In experiments relating to the first question the general approach is to present to a subject a picture or another object that is sufficiently large and close to the eyes so that it cannot all be registered on the foveas in one fixation. For example, a picture 35 centimeters wide and 100 centimeters from the eyes subtends a horizontal angle of 20 degrees at each eye—roughly the angle subtended by a page of this magazine held at arm's length. This is far wider than the one to two degrees of visual field that are brought to focus on the fovea.

Under these conditions the subject must move his eyes and look around the picture, fixating each part he wants to see clearly. The assumption is that he looks mainly at the parts of the picture he regards as being its features; they are the parts that hold for him the most information about the picture. Features are tentatively located by peripheral vision and then fixated directly for detailed inspection. (It is important to note that in these experiments and in the others we shall describe the subject is given only general instructions, such as "Just look at the pictures," or even no instructions at all. More specific instructions, requiring him to inspect and describe some specific aspect of the picture, usually result in appropriately directed fixations, as might be expected.)

When subjects freely view simple pictures, such as line drawings, under these conditions, it is found that their fixations tend to cluster around the angles of the picture. For example, Leonard Zusne and Kenneth M. Michels performed an experiment of this type at Purdue University, using as pictures line drawings of simple polygons [*see illustration on* *opposite page*]. From the fixations made by their subjects in viewing such figures it is clear that the angles of the drawings attracted the eyes most strongly.

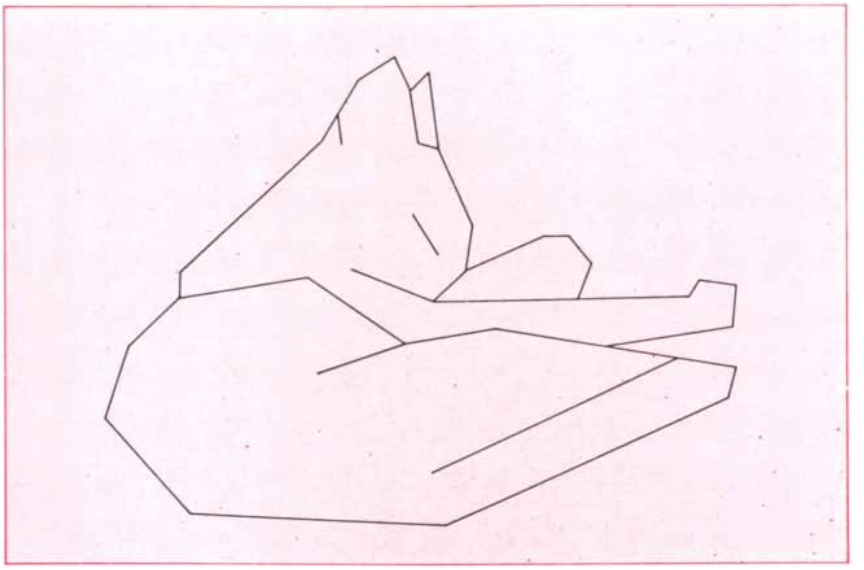
Our tentative conclusion is that, at least with such line drawings, the angles are the principal features the brain employs to store and recognize the drawing. Certainly angles would be an efficient choice for features. In 1954 Fred Attneave III of the University of Oregon pointed out that the most informative parts of a line drawing are the angles and sharp curves. To illustrate his argument he presented a picture that was obtained by selecting the 38 points of greatest curvature in a picture of a sleeping cat and joining the points with straight lines [*see top illustration at right*]. The result is clearly recognizable.

Additional evidence that angles and sharp curves are features has come from electrophysiologists who have investigated the activity of individual brain cells. For example, in the late 1950's Jerome Y. Lettvin, H. R. Maturana, W. S. McCulloch and W. H. Pitts of the Massachusetts Institute of Technology found angle-detecting neurons in the frog's retina. More recently David H. Hubel and Torsten N. Wiesel of the Harvard Medical School have extended this result to cats and monkeys (whose angle-detecting cells are in the visual cortex rather than the retina). And recordings obtained from the human visual cortex by Elwin Marg of the University of California at Berkeley give preliminary indications that these results can be extended to man.

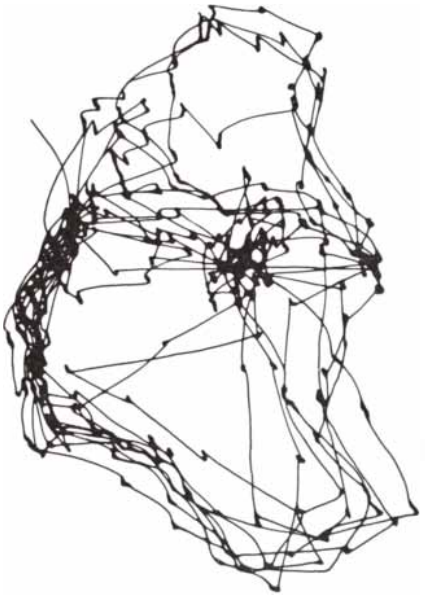
Somewhat analogous results have been obtained with pictures more complex than simple line drawings. It is not surprising that in such cases the features are also more complex. As a result no formal description of them has been achieved. Again, however, high information content seems to be the criterion. Norman H. Mackworth and A. J. Morandi made a series of recordings at Harvard University of fixations by subjects viewing two complex photographs. They concluded that the fixations were concentrated on unpredictable or unusual details, in particular on unpredictable contours. An unpredictable contour is one that changes direction rapidly and irregularly and therefore has a high information content.

We conclude, then, that angles and other informative details are the features selected by the brain for remembering and recognizing an object. The next question concerns how these features are integrated by the brain into a whole—the internal representation—so that one sees the object as a whole, as an object rather than an unconnected sequence of features. Once again useful evidence comes from recordings of eye movements. Just as study of the locations of fixations indicated the probable nature of the features, so analysis of the order of fixations suggests a format for the interconnection of features into the overall internal representation.

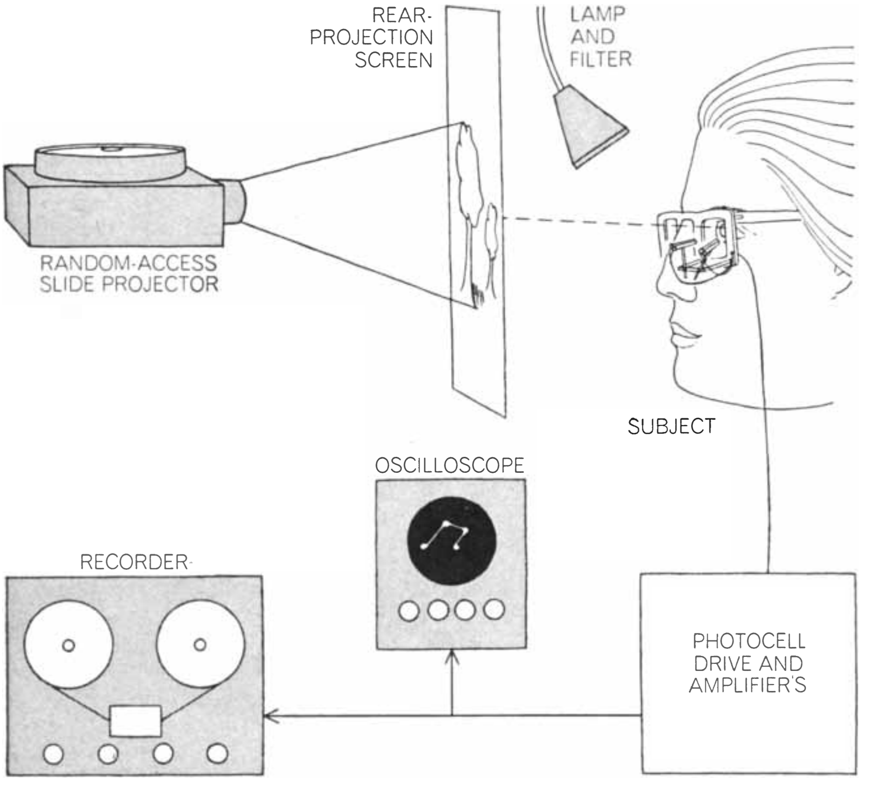
The illustration below shows the fixations made by a subject while viewing a photograph of a bust of the Egyptian queen Nefertiti. It is one of a series of recordings made by Alfred L. Yarbus of the Institute for Problems of Information Transmission of the Academy of Sciences of the U.S.S.R. The illustration



SHARP CURVES are also important as features for visual identification, as shown by Fred Attneave III of the University of Oregon in a picture made by selecting the 38 points of greatest curvature in a picture of a sleeping cat and joining them with straight lines, thus eliminating all other curves. The result is still easily recognizable, suggesting that points of sharp curvature provide highly useful information to the brain in visual perception.



REGULARITIES OF EYE MOVEMENT appear in a recording of a subject viewing a photograph of a bust of Queen Nefertiti. At left is a drawing of what the subject saw; at right are his eye movements as recorded by Alfred L. Yarbus of the Institute for Problems of Information Transmission in Moscow. The eyes seem to visit the features of the head cyclically, following fairly regular pathways, rather than crisscrossing the picture at random.



EXPERIMENTAL PROCEDURE employed by the authors is depicted schematically. The subject viewed pictures displayed on a rear-projection screen by a random-access slide projector. Diffuse infrared light was shined on his eyes; his eye movements were recorded by photocells, mounted on a spectacle frame, that detected reflections of the infrared light from one eyeball. Eye movements were displayed on oscilloscope and also recorded on tape.

shows clearly an important aspect of eye movement during visual perception, namely that the order of the fixations is by no means random. The lines representing the saccades form broad bands from point to point and do not crisscross the picture at random as would be expected if the eyes visited the different features repetitively in a random order. It appears that fixation on any one feature, such as Nefertiti's eye, is usually followed by fixation on the same next feature, such as her mouth. The overall record seems to indicate a series of cycles; in each cycle the eyes visit the main features of the picture, following rather regular pathways from feature to feature.

Recently at the University of California at Berkeley we have developed a hypothesis about visual perception that predicts and explains this apparent regularity of eye movement. Essentially we propose that in the internal representation or memory of the picture the features are linked together in sequence by the memory of the eye movement required to look from one feature to the next. Thus the eyes would tend to move from feature to feature in a fixed order, scanning the picture.

Most of Yarbus' recordings are summaries of many fixations and do not contain complete information on the ordering of the fixations. Thus the regularities of eye movements predicted by our hypothesis could not be definitely confirmed from his data. To eliminate this constraint and to subject our hypothesis to a more specific test we recently made a new series of recordings of eye movements during visual perception.

Our subjects viewed line drawings of simple objects and abstract symbols as we measured their eye movements (using photocells to determine the movements of the "white" of the eye) on magnetic tape and recorded them [*see illustration above*]. We thereby obtained a permanent record of the order of fixations made by the subjects and could play it back later at a lower speed, analyzing it at length for cycles and other regularities of movement. As in the earlier experiments, the drawings were fairly large and close to the subject's eyes, a typical drawing subtending about 20 degrees at the eye. In addition we drew the pictures with quite thin lines and displayed them with an underpowered slide projector, throwing a dim image on a screen that was fully exposed to the ordinary light in the laboratory. In this way we produced an image of low visibility and could be sure that the subject would have to look directly (foveally) at each feature that interested him, thus revealing to our recording equipment the locus of his attention.

Our initial results amply confirmed the previous impression of cycles of eye movements. We found that when a subject viewed a picture under these conditions, his eyes usually scanned it following—intermittently but repeatedly—a fixed path, which we have termed his "scan path" for that picture [*see illustration on opposite page*]. The occurrences of the scan path were separated by periods in which the fixations were ordered in a less regular manner.

Each scan path was characteristic of a given subject viewing a given picture. A subject had a different scan path for every picture he viewed, and for a given picture each subject had a different scan path. A typical scan path for our pictures consisted of about 10 fixations and lasted for from three to five seconds. Scan paths usually occupied from 25 to 35 percent of the subject's viewing time, the rest being devoted to less regular eye movements.

It must be added that scan paths were not always observed. Certain pictures (one of a telephone, for example) seemed often not to provoke a repetitive response, although no definite common characteristic could be discerned in such pictures. The commonest reaction, however, was to exhibit a scan path. It was interesting now for us to refer back to the earlier recordings by Zusne and Michels, where we observed scan paths that had previously passed unnoticed. For instance, in the illustration on page 36 fixations No. 4 through No. 11 and No. 11 through No. 18 appear to be two occurrences of a scan path. They are identical, even to the inclusion of the small reverse movement in the lower right-hand corner of the figure.

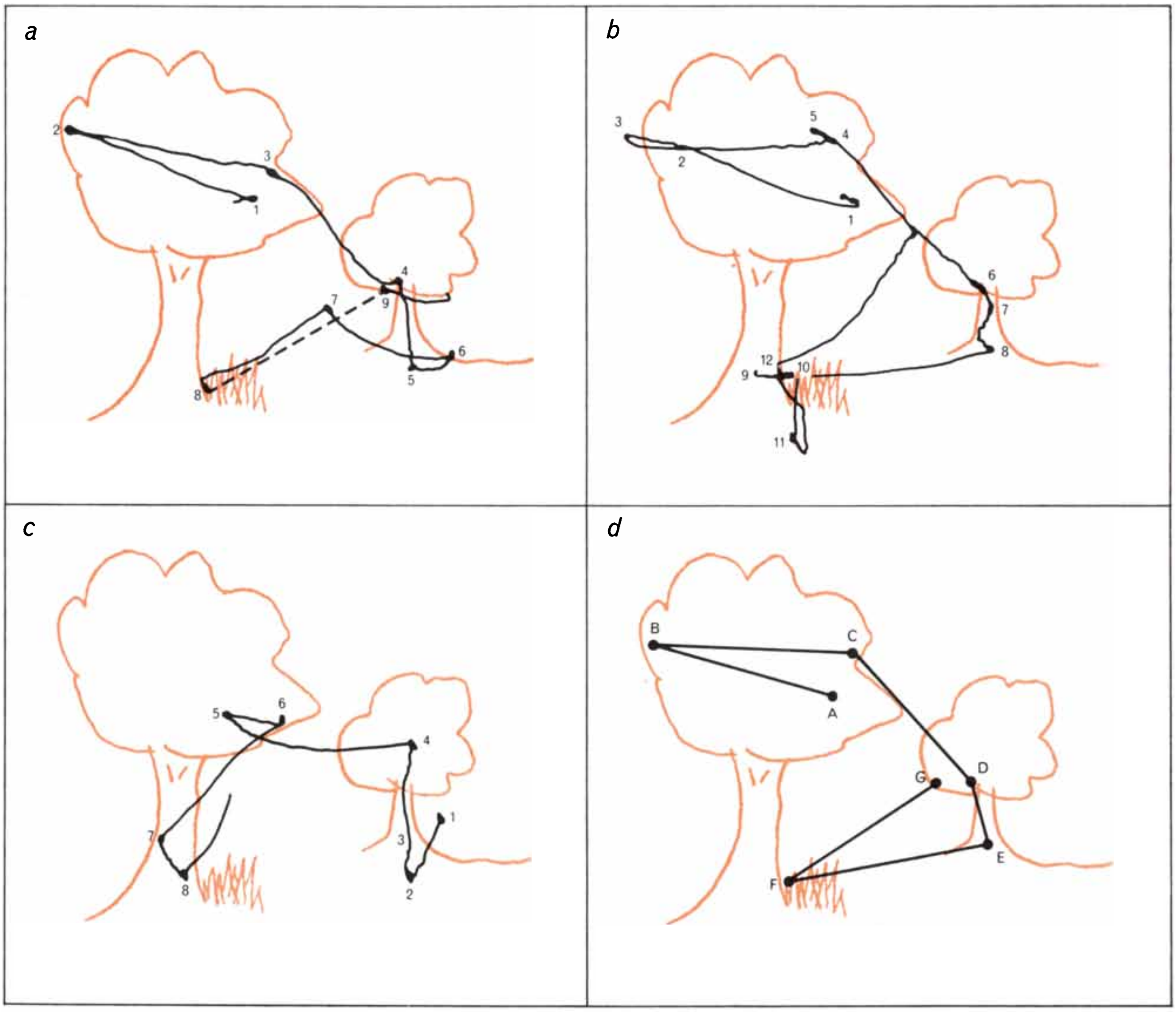
This demonstration of the existence of scan paths strengthened and clarified our ideas about visual perception. In accordance with the serial hypothesis, we assume that the internal representation of an object in the memory system is an assemblage of features. To this we add a crucial hypothesis: that the features are assembled in a format we have termed a "feature ring" [*see illustration on page 40*]. The ring is a sequence of sensory and motor memory traces, alternately recording a feature of the object and the eye movement required to reach the next feature. The feature ring establishes a fixed ordering of features and eye movements, corresponding to a scan path on the object.

Our hypothesis states that as a subject views an object for the first time and becomes familiar with it he scans it with his eyes and develops a scan path for it. During this time he lays down the memory traces of the feature ring, which records both the sensory activity and the motor activity. When he subsequently encounters the same object again, he recognizes it by matching it with the feature ring, which is its internal representation in his memory. Matching consists in verifying the successive features and carrying out the intervening eye movements, as directed by the feature ring.

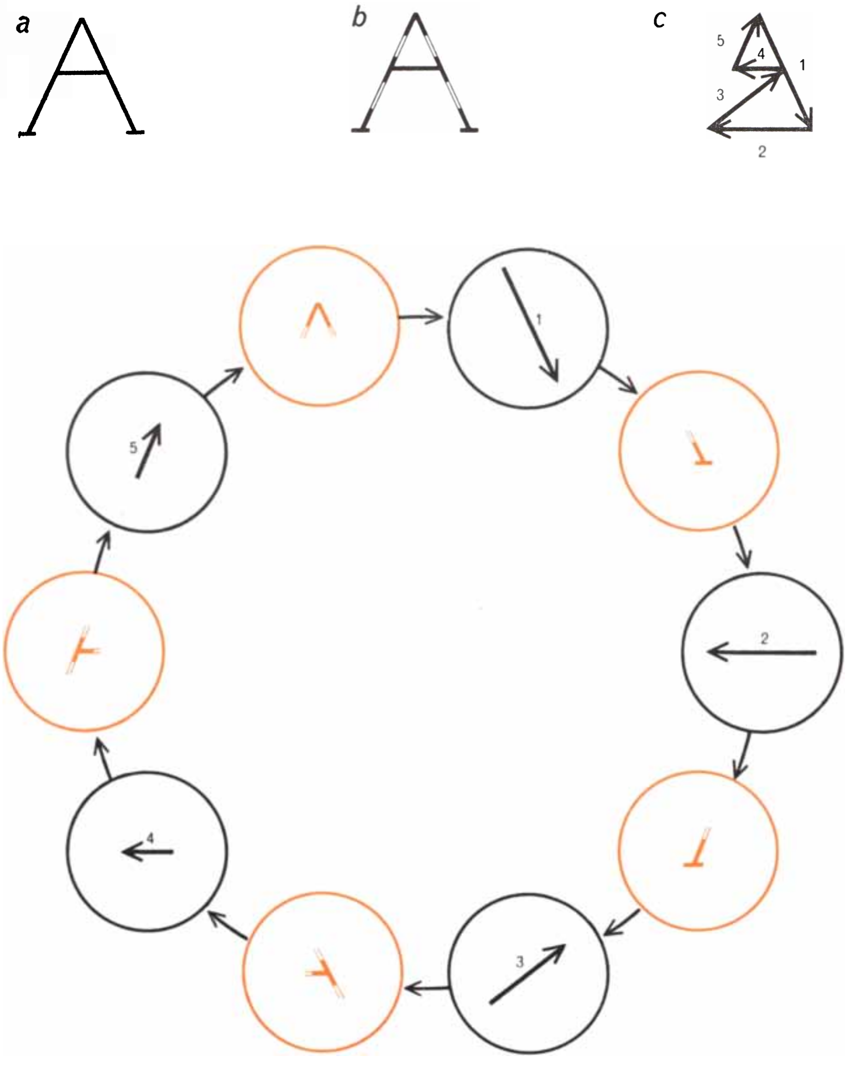
This hypothesis not only offers a plausible format for the internal representation of objects—a format consistent with the existence of scan paths—but also has certain other attractive features. For example, it enables us to draw an interesting analogy between perception and behavior, in which both are seen to involve the alternation of sensory and motor activity. In the case of behavior, such as the performance of a learned sequence of activities, the sensing of a situation alternates with motor activity designed to bring about an expected new situation. In the case of perception (or, more specifically, recognition) of an object the verification of features alternates with movement of the eyes to the expected new feature.

The feature-ring hypothesis also makes a verifiable prediction concerning eye movements during recognition: The successive eye movements and feature verifications, being directed by the feature ring, should trace out the same scan path that was established for the object during the initial viewing. Confirmation of the prediction would further strengthen the case for the hypothesis. Since the prediction is subject to experimental confirmation we designed an experiment to test it.

The experiment had two phases, which we called the learning phase and the recognition phase. (We did not, of



REGULAR PATTERN of eye movement by a given subject viewing a given picture was termed the subject's "scan path" for that picture. Two of five observed occurrences of one subject's scan path as he looked at a simple drawing of trees for 75 seconds are shown here (*a*, *b*). The dotted line between fixations 8 and 9 of *a* indicates that the recording of this saccade was interrupted by a blink. Less regular eye movements made between these appearances of the scan path are at *c*. Subject's scan path is idealized at *d*.



FEATURE RING is proposed by the authors as a format for the internal representation of an object. The object (*a*) is identified by its principal features (*b*) and is represented in the memory by them and by the recollection of the scan path (*c*) whereby they were viewed. The feature ring therefore consists of sensory memory traces (*color*) recording the features and motor memory traces (*black*) of the eye movements from one feature to the next.

course, use any such suggestive terms in briefing the subjects; as before, they were simply told to look at the pictures.) In the learning phase the subject viewed five pictures he had not seen before, each for 20 seconds. The pictures and viewing conditions were similar to those of the first experiment. For the recognition phase, which followed immediately, the five pictures were mixed with five others the subject had not seen. This was to make the recognition task less easy. The set of 10 pictures was then presented to the subject three times in random order; he had five seconds to look at each picture. Eye movements were recorded during both the learning phase and the recognition phase.

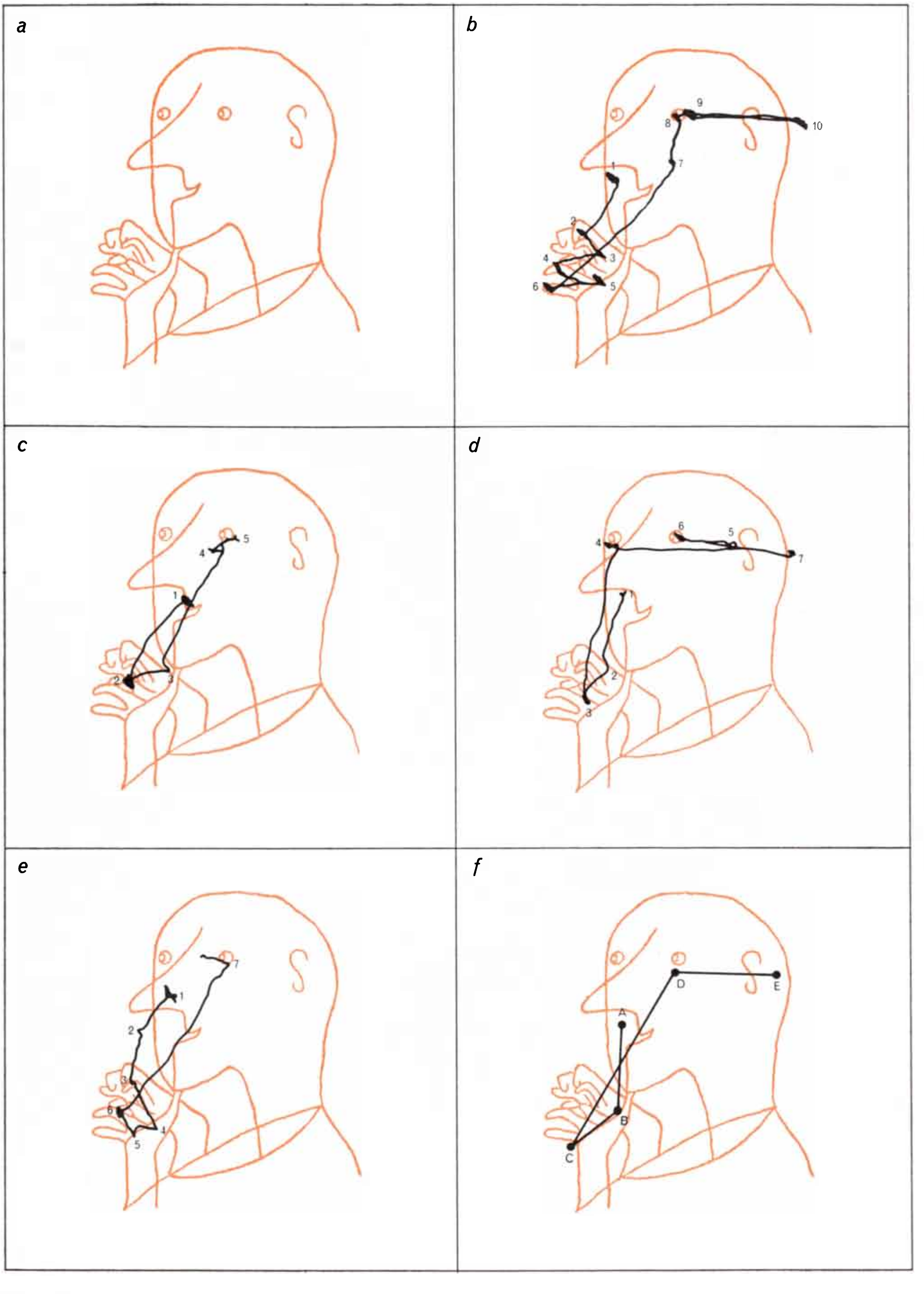
When we analyzed the recordings, we were pleased to find that to a large extent our predictions were confirmed. Scan paths appeared in the subject's eye movements during the learning phase, and during the recognition phase his first few eye movements on viewing a picture (presumably during the time he was recognizing it) usually followed the same scan path he had established for that picture during the learning phase [*see illustration on opposite page*]. In terms of our hypothesis the subject was forming a feature ring during the learning-phase occurrences of the scan path; in the recognition phase he was matching the feature ring with the picture, following the scan path dictated by the feature ring.

An additional result of this experiment was to demonstrate that different subjects have different scan paths for a given picture and, conversely, that a given subject has different scan paths for different pictures [*see illustration on page 42*]. These findings help to discount certain alternative explanations that might be advanced to account for the occurrence of scan paths. The fact that a subject has quite different scan paths for different pictures suggests that the scan paths are not the result of some fixed habit of eye movement, such as reading Chinese vertically, brought to each picture but rather that they come from a more specific source, such as learned feature rings. Similarly, the differences among subjects in scan paths used for a given picture suggest that the scan paths do not result from peripheral feature detectors that control eye movements independent of the recognition process, since these detectors might be expected to operate in much the same way in all subjects.

Although the results of the second experiment provided considerable support for our ideas on visual perception, certain things remain unexplained. For example, sometimes no scan path was observed during the learning phase. Even when we did find a scan path, it did not always reappear in the recognition phase. On the average the appropriate scan path appeared in about 65 percent of the recognition-phase viewings. This is a rather strong result in view of the many possible paths around each picture, but it leaves 35 percent of the viewings, when no scan path appeared, in need of explanation.

Probably the basic idea of the feature ring needs elaboration. If provision were made for memory traces recording other eye movements between features not adjacent in the ring, and if the original ring represented the preferred and habitual order of processing rather than the inevitable order, the occasional substitution of an abnormal order for the

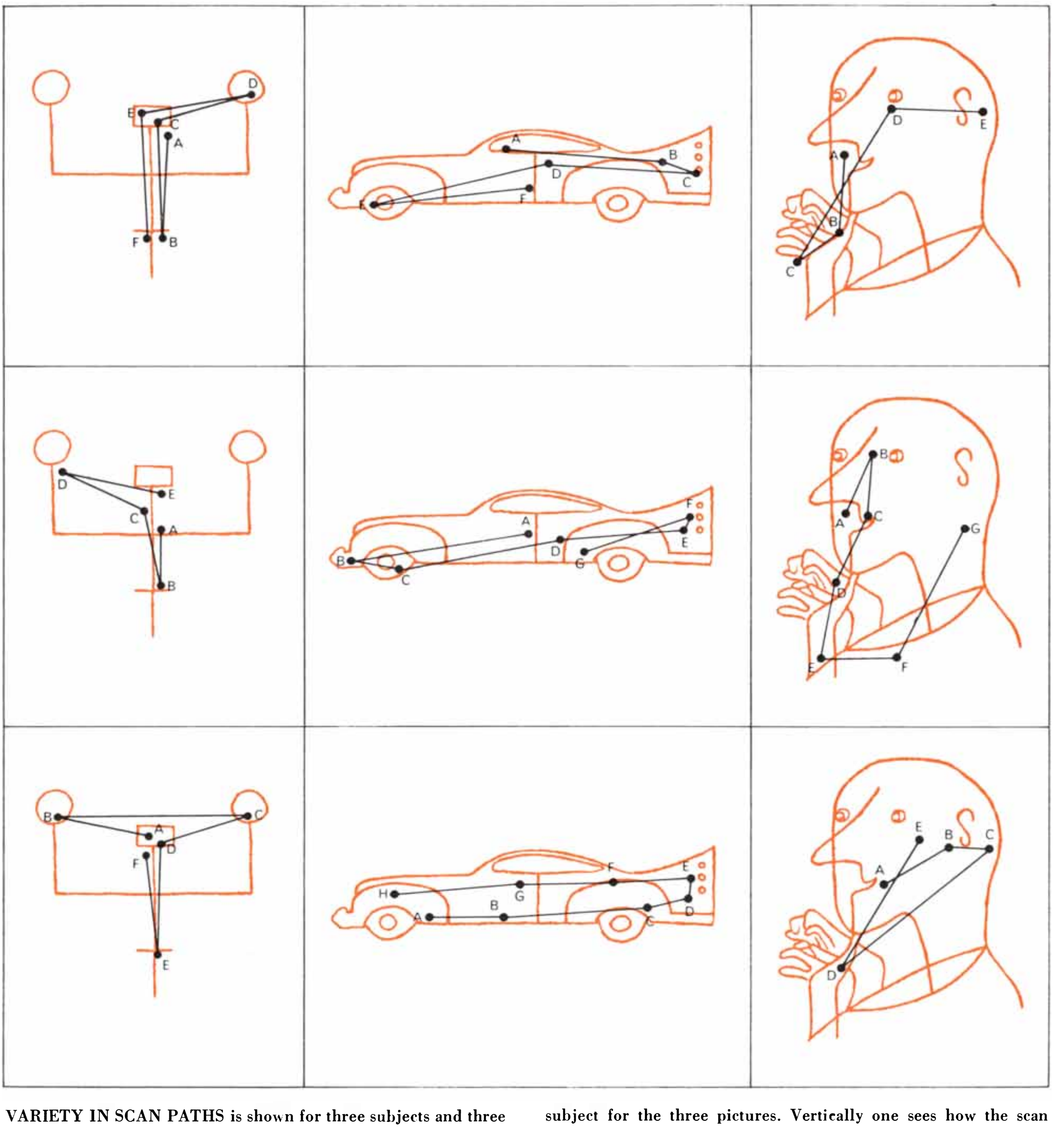
RECURRENCE OF SCAN PATH during recognition of an object is predicted by the feature-ring hypothesis. A subject viewed the adaptation of Klee's drawing (*a*). A scan path appeared while he was familiarizing himself with the picture (*b, c*). It also appeared (*d, e*) during the recognition phase each time he identified the picture as he viewed a sequence of familiar and unfamiliar scenes depicted in similar drawings. This particular experimental subject's scan path for this particular picture is presented in idealized form at *f*.


a
b
c
d
e
f
A
B
C
D
E

scan path would be explained [*see top illustration on opposite page*].

It must also be remembered that the eye-movement recordings in our experiments were made while the subjects viewed pictures that were rather large and close to their eyes, forcing them to look around in the picture to see its features clearly. In the more normal viewing situation, with a picture or an object small enough to be wholly visible with a single fixation, no eye movements are necessary for recognition. We assume that in such a case the steps in perception are parallel up to the point where an image of the object is formed in the visual cortex and that thereafter (as would seem evident from the experiments on recognition time) the matching of the image and the internal representation is carried out serially, feature by feature. Now, however, we must postulate instead of eye movements from feature to feature a sequence of internal shifts of attention, processing the features serially and following the scan path dictated by the feature ring. Thus each motor memory trace in the feature ring records a shift of attention that can be executed either externally, as an eye movement, or internally, depending on the extent of the shift required.

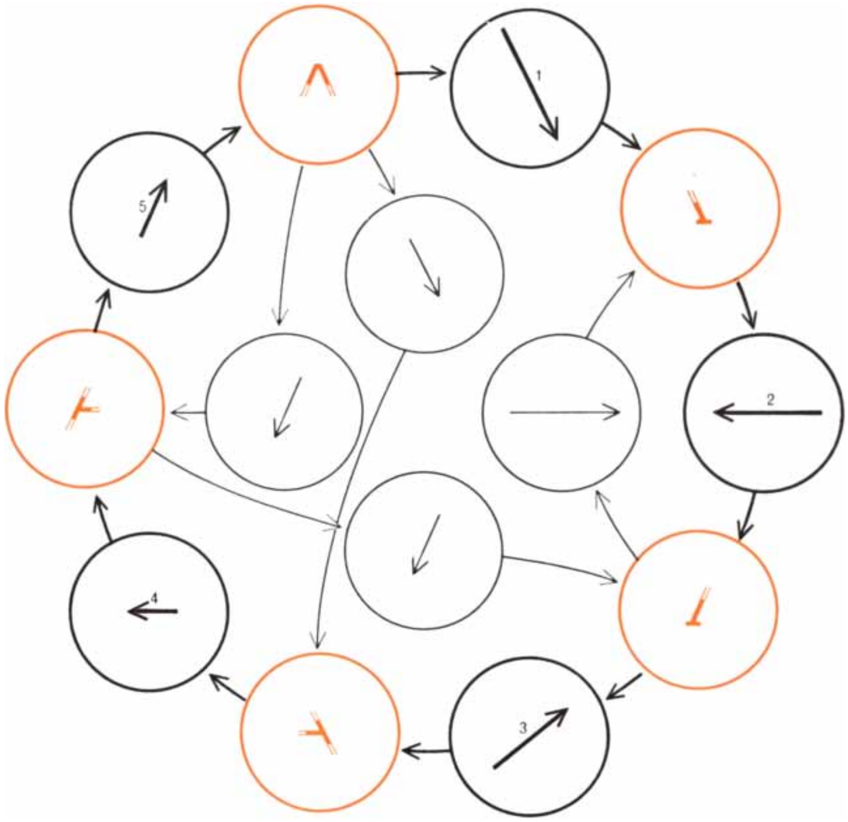
In this connection several recordings made by Lloyd Kaufman and Whitman Richards at M.I.T. are of interest. Their subjects viewed simple figures, such as a drawing of a cube, that could be taken in with a single fixation. At 10 randomly chosen moments the subject was asked



VARIETY IN SCAN PATHS is shown for three subjects and three pictures. Each horizontal row depicts the scan paths used by one subject for the three pictures. Vertically one sees how the scan paths of the three subjects for any one picture also varied widely.

to indicate where he thought he was looking. His answer presumably showed what part of the picture he was attending to visually. His actual fixation point was then recorded at another 10 randomly selected moments [*see bottom illustration at right*]. The results suggest that the subject's attention moved around the picture but his fixation remained fairly steady near the center of the picture. This finding is consistent with the view that smaller objects too are processed serially, by internal shifts of attention, even though little or no eye movement is involved.

It is important to note, however, that neither these results nor ours prove that recognition of objects and pictures is necessarily a serial process under normal conditions, when the object is not so large and close as to force serial processing by eye movements. The experiments on recognition time support the serial hypothesis, but it cannot yet be regarded as being conclusively established. In our experiments we provided a situation that forced the subject to view and recognize pictures serially with eye movements, thus revealing the order of feature processing, and we assumed that the results would be relevant to recognition under more normal conditions. Our results suggest a more detailed explanation of serial processing—the feature ring producing the scan path—but this explanation remains conditional on the serial hypothesis.

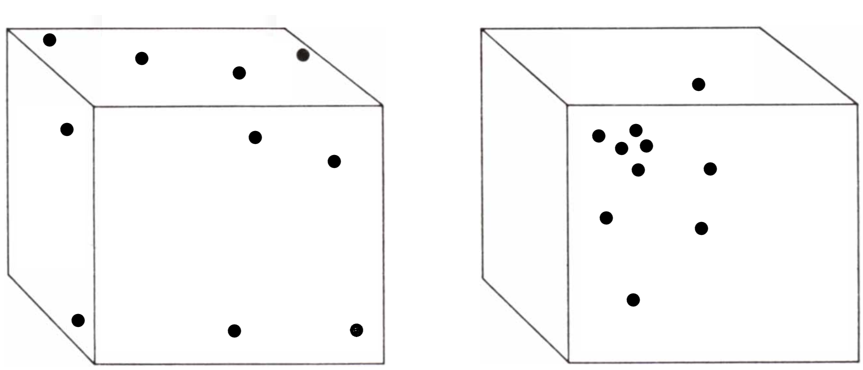


MODIFIED FEATURE RING takes into account less regular eye movements that do not conform to scan path. Several movements, which appeared in 35 percent of recognition viewings, are in center of this ring. Outside ring, consisting of sensory (*black*) and motor memory traces (*color*), represents scan path and remains preferred order of processing.

In sum, we believe the experimental results so far obtained support three main conclusions concerning the visual recognition of objects and pictures. First, the internal representation or memory of an object is a piecemeal affair: an assemblage of features or, more strictly, of memory traces of features; during recognition the internal representation is matched serially with the object, feature by feature. Second, the features of an object are the parts of it (such as the angles and curves of line drawings) that yield the most information. Third, the memory traces recording the features are assembled into the complete internal representation by being connected by other memory traces that record the shifts of attention required to pass from feature to feature, either with eye movements or with internal shifts of attention; the attention shifts connect the features in a preferred order, forming a feature ring and resulting in a scan path, which is usually followed when verifying the features during recognition.

Clearly these conclusions indicate a distinctly serial conception of visual learning and recognition. In the trend to look toward serial concepts to advance the understanding of visual perception one can note the influence of current work in computerized pattern recognition, where the serial approach has long been favored. Indeed, computer and information-processing concepts, usually serial in nature, are having an increasing influence on brain research in general.

Our own thoughts on visual recognition offer a case in point. We have developed them simultaneously with an analogous system for computerized pattern recognition. Although the system has not been implemented in working form, a somewhat similar scheme is being used in the visual-recognition system of a robot being developed by a group at the Stanford Research Institute. We believe this fruitful interaction between biology and engineering can be expected to continue, to the enrichment of both.



INTERNAL SHIFTS OF ATTENTION apparently replace eye movements in processing of objects small enough to be viewed with single fixation. A subject's attention, represented by statements of where he thought he was looking, moved around picture (*left*), whereas measured fixation point (*right*) remained relatively stationary. Illustration is based on work by Lloyd Kaufman and Whitman Richards at the Massachusetts Institute of Technology.