The gaze selects informative details within pictures [1]

NORMAN H. MACKWORTH [2] AND ANTHONY J. MORANDI [3]

HARVARD UNIVERSITY

The visual fixations of 20 Ss viewing each of two pictures were measured. Each picture was later divided into 64 squares, and 20 other Ss judged their recognizability on a 10-point scale. Both measures gave high readings for unusual details and for unpredictable contours. Although they were judged to be highly recognizable, all the redundant (or predictable) contours received few fixations. Areas of mere texture scored low on both measures. The relations between fixation densities and estimated recognizability suggest that a scene may be divided into informative features and redundant regions. Not only do the eyes have to be aimed, they are usually aimed intelligently, even during the casual inspection of pictures.

The main experimental question is: Do adults look at dominant details in the stimulus with foveal vision? If, as an alternative, adults learned to perceive whole patterns, it would presumably not matter where they looked on the picture. There would then not be much correlation between the particular areas selected by different individuals.

Perception is no passive sampling from external events. It is so closely interlocked with memory that Hake (1957) has claimed that there is no perception without recognition. Bartlett's work (1932) has also indicated that remembering is based on reconstructing from mental schemata rather than from playing back unmodified recordings. Such a selective process implies that the organism actively controls the stimuli to which it is sensitive by using the mismatch between input and plan (Miller, Galanter, & Pribram, 1960). These control functions depend upon the establishing of a program by prior learning, because without such a plan there can be no basis for selectivity toward stimuli (Bruner, 1965).

The empirical question is, therefore, whether these cortical effects alter the eye tracks across pictures. Ten years ago, Hake (1957) stated that "...Vision does systematically select parts of stimulation in perception" and that the parts selected are those which lead to the greatest coherence. Those aspects are chosen which produce the greatest consistency of stimulation—those which produce the greatest resemblance between present and past stimulation and which lead to the most efficient prediction about future stimulations." More recently, this approach has been taken further by considering visual perception as a selective process distinguishing signal from noise in the visual environment (Hake et al, 1966).

If adults have learned to pick out the important details for analysis, they might be expected to agree to some extent on which were the informative areas. Group records of fixation density show that the same specific areas are usually selected by the line of sight (Mackworth & Bruner, 1966). We might expect regions that are rated highly informative to receive more fixations than others, in line with Senders' (1966) finding that rapidly changing instrument dials are fixated more frequently. Perhaps comparisons between different pictures would also reveal the general characteristics of these important regions by a content analysis of those stimuli that dominate the others in the scene.

Objectives

The main purpose of our paper is to determine whether key regions exist within pictorial displays. Preliminary attempts are also made to explain why some stimuli are more important than others within the displays.

Procedure

The approach was to make a direct comparison between (1) the visual choices, shown by the position of the gaze while the Ss were looking at an entire picture, and (2) the verbal estimates of the relative importance of the different regions within the same picture. The studies included the two pictures given in Fig. 1. The first, called the "eyes" picture, shows a pair of eyes behind a crimson mask, with a bright background of yellow and orange. The second, called the "map" picture, is an astronaut's view of Baja California seen against a dark blue sea. (The pictures shown to the Ss did not contain the superimposed squares or Ds).

The visual scanning choices were determined by recording the density of visual fixations or pauses made by the line of sight while the picture was being viewed by 20 Harvard or Radcliffe students. The eye track of each S was photographed by a stand eye-camera on a copy of the stimulus display (Mackworth, 1967). By superposing the individual eye tracks from all Ss, we could calculate what percentage of the total number of fixations made by the group of Ss fell within each separate square inch of the 8 x 8 in. picture that subtended a visual angle of 16° at the selected viewing distance. Every S was shown each of the two color photographs separately for 10 sec. The task then assigned to them while
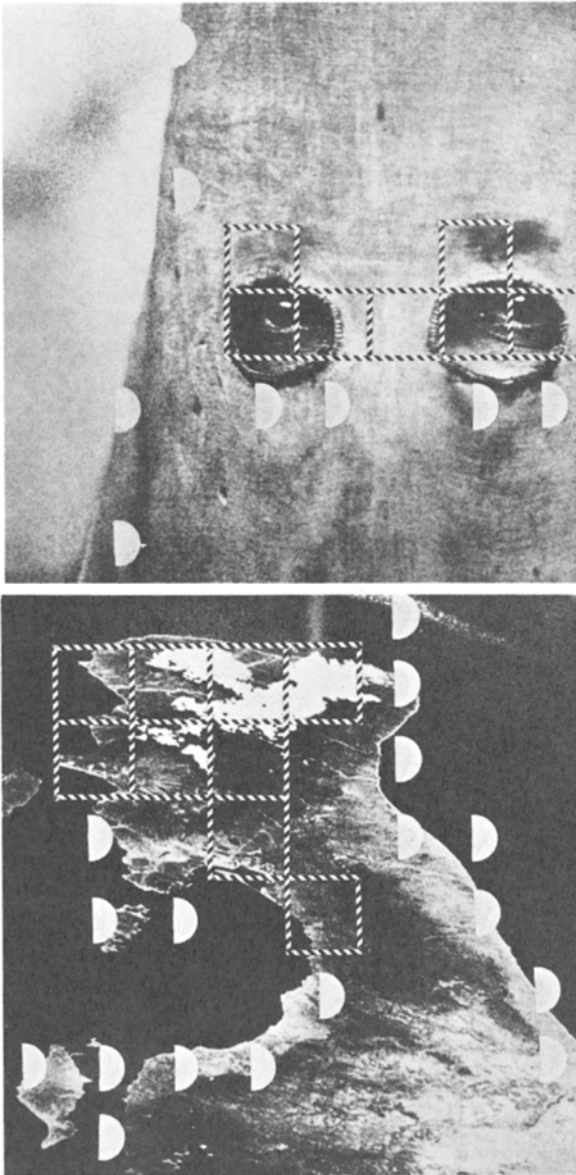
Fig. 1. Stimulus displays with group responses superimposed for all areas verbally rated above average. Note that all high densities of fixations (enclosed by dotted lines) were on unique edges or areas. Conversely, all low densities of fixations (marked by solid D's) were on redundant contours. Disregarded by direct vision, these redundant contours were being processed by peripheral vision.

in the eye-camera apparatus was to decide which picture they preferred.

The verbal estimates of the relative importance of the different regions within each picture were expressed as the average ratings or subjective scale values for each of the 64 squares comprising the same picture. These ratings were obtained from another group of 20 Harvard or Radcliffe students; none of these Ss had taken part in the previous eye-camera research. Each picture was dissected into the 64 1 in. squares used as a matrix for determin-

ing the density of eye fixations. These separate squares were given to each S, who was asked to rate the squares individually for "informativeness." Each square was to be placed upon one of the 10 rungs of a scale which was drawn on an answer sheet and ranged from "completely informative" to "completely uninformative"; these estimates were numbered 0-9 on the scale. Informativeness was defined in terms of "recognizability"—in the sense of how easy a square would be to recognize on another occasion. The Ss did not see the picture as a whole beforehand.

Results

The main result was that a few regions in each picture dominated the data. These outstanding areas, which have been emphasized by dotted lines in Fig. 1, gave high readings on both measures. Each dotted square received more then 2 percent of the visual fixation density, and it also fell above the midpoint on the scale of readings for estimated informativeness. (Completely random scanning would have allocated 1.6 percent of all fixations to each of the 64 squares in the 8 by 8 matrix.)

Figure 2 summarizes the general relationship between the mean estimated informativeness for a given square and the percentage of the total visual fixations that fell in that square, as calculated from the group scatterplots. Squares that were rated high in informativeness averaged many times the number of fixations per square as those rated lower in informativeness.

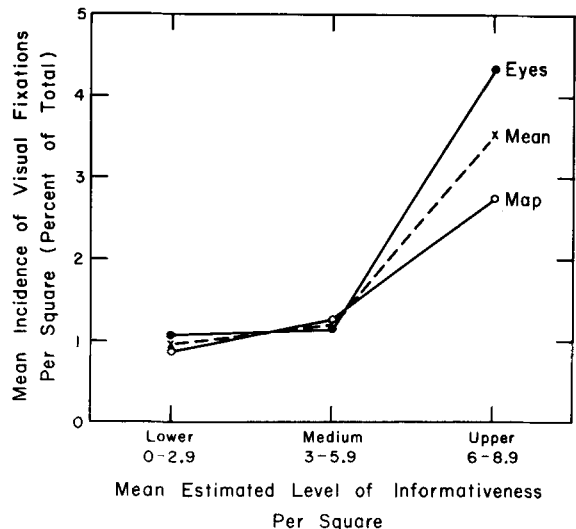An extreme instance of the tendency for informative

Fig. 2. Relation between the informativeness ratings given squares and their average incidence of visual fixations. The squares from each picture were grouped according to their rated informativeness. Then the incidence of visual fixations for the squares in each group were averaged to obtain the mean fixation incidence for that group of squares.

On Textures		On Contours	
Redundant	Non-redundant	Redundant	Non-redundant
Sea 0.5	Land 1.4	Predictable Coast 0.9	Unpredictable Coast 2.8
			Clouds 4.7
Mask 0.6	Nil	Mask Edge 1.0	Eyes 6.9

On Textures		On Contours	
Redundant	Non-redundant	Redundant	Non-redundant
Sea 0.8	Land 2.3	Predictable Coast 6.0	Unpredictable Coast 5.9
			Clouds 7.2
Mask 1.3	Nil	Mask Edge 5.0	Eyes 8.4

areas to receive clusters of visual fixations was noted in a content analysis of the eyes picture in Fig. 1. The squares actually presenting the eyes were the only areas in the picture rated above 8 on the informativeness scale. The same four squares each had fixation readings between 5 and 8 percent. Mackworth and Bruner (1966) have noted that individual visual fixations show higher percentages than are obtained from group scatterplot data, which are usually easier to obtain. But highly informative areas can cluster the fixations on particular regions within a display so markedly that even the group visual concentrations may be as high as five times the 1.6 percent that represents a completely random distribution of fixations. As many as 60 percent of all fixations fell on the 10 squares across the eyes or just above them; that is, two-thirds of all fixations were crowded into one-tenth of the total area. This clustering had the other effect, that three-quarters of all the squares in the picture were so seldom fixated that their fixation densities fell at or below 2 percent per square inch. They are left unmarked in Fig. 1. (A similar analysis follows for the map display.)

Each fixation was scored for informativeness by being given a number according to the estimated importance of the square in which it fell (Mackworth & Bruner, 1966). The sum of these readings was determined for all the fixations within each successive 2 sec period. It was then found that the value for this sum did not increase in successive periods. Therefore, it must be assumed that the Ss placed their fixations just as effectively during the first 2 sec as they did during the last 2 sec of the 10 sec exposure. There was rapid rejection of unwanted areas at the very beginning of the period of inspection. Attention started with inhibition, in the sense of the processing and rejection of unwanted stimuli (Berlyne, 1965; Spinelli & Pribram, 1966).

This processing and rejection must have been mediated by peripheral vision, because three-quarters of the area of the eyes picture and nearly two-thirds of the map picture were scarcely fixated at all, and received only 2 percent or fewer fixations per square inch (see unmarked regions of Fig. 1). This early sifting of the input within the first 2 sec allowed no time for each different region within the display to be examined by the fovea, which is only 2 degrees wide. The display arrangements were

such that this would have entailed machinelike scanning of each of the 64 squares in turn. This improbable process would have called for at least 20 sec rather than the 2 sec actually required; the reason the hypothetical 64 fixations would need so much time is that pictures are usually inspected at a speed of three fixations per second (Mackworth & Bruner, 1966).

Peripheral vision edited out the redundant stimuli in the pictures. The perceptual process is constantly trying to find simplifying regularities and consistencies to detect and discard unwanted redundant stimuli which may overload input channels. Such lawfulness, or constraint, between adjacent details allows people to predict a portion of the field from a knowledge of some other region (Attneave, 1954; Bevan, 1958). Redundancy can arise from overemphasis by an extra cue along a further dimension as well as from unnecessary details (Staniland, 1966).

A content analysis of the details within each picture was therefore essential. The square inches of each picture were grouped into those that contain textures and those that contain contours, and Table 1 makes this distinction. The mean incidence of fixations per square was noticeably lower for most textures than for contours. Table 2 shows that textures were also judged less informative than contours. Between various textures, Table 1 shows how the rougher, less predictable textures of the land mass attracted slightly more fixations than did the smoother textures of the sea. The smooth textures in the eye picture were also seldom fixated.

The most marked effect of predictability was seen when various contours were compared, especially in the map picture. As with the eyes picture, a few squares were outstanding on both fixation density and judged recognizability. Figure 1 shows that these squares contained unpredictable contours of coastline. Their high fixation densities averaged 2.8 percent, which was four times the incidence for the smooth coastline at the right. There was no overlap in the distributions between the average fixation readings for each of the outstanding squares showing unpredictable coastline patterns and those for other squares displaying simpler coastline shapes. On the rating estimates, the unpredictable coastlines were all above the midpoint of the scale, ranging between 5.3 and 7.4.

Predictable coastline contours were rated higher

than their fixation density readings suggested they deserved. Subjective ratings, therefore, give too high a score for the predictable lines in the pictures, at least in terms of fixation density. All contours, whether predictable or not, were rated equally high on the scale. On the map, for example, the right-hand coastline, which was smooth and predictable, averaged 6.3, whereas the less predictable lower left-hand coastline averaged 5.8, and the quite predictable upper left-hand coastline averaged 5.9. All squares which the ratings scored too high have been marked D (for disregarded) in Fig. 1. These were squares which scored above the midpoint of 5.1 on the recognizability scale but received only 2 percent or less of the fixations.

This three-way comparison between the subjective rating, the fixation density, and the pictorial content has shown that the discrepancies between fixation density and estimated values were a blessing in disguise. Comparisons between fixation readings and scale values for each square inch have provided a new experimental method of identifying the areas in stimulus displays that do not attract the gaze, even though they may contain features that are easily recognized. The solid D markings in Fig. 1, for example, have virtually reoutlined the predictable parts of the coastline. This approach may therefore prove to be a way of estimating and locating redundancy in contours. For example, the redundant regions of the predictable contours under the eyes and along the straight border of the mask also show some D marking.

Discussion

Attneave (1954, 1959) was correct when he asserted that most of the informational content of a drawing is found where the contours change their direction. But his exposition of the informational aspects of visual perception gave too much emphasis to the role of straight lines or redundant contours joining such points, mainly because subjective measures tend to give greater importance to such cues than is shown by visual fixation on realistic pictures in color. It may be that Attneave overemphasized contours because (1) he extrapolated to textured material from line drawings, and (2) in his exposition of shape perception he also compared television and human scanning procedures. (Although he disclaimed that human perception depended upon any process of this kind, Attneave did use machine scanning as an illustration.) Comparisons between television and human scanning can be dangerous, because the television flying-spot would represent for human Ss only a marked case of tunnel vision. Simple, predictable contours providing boundaries between colors or textures do not play a large part in direct human foveal scanning. These predictable outlines and regular textures are processed by the

peripheral retina, and they therefore restrict the general area of attention that the fovea explores in its search for the unusual. Similarly, artists often use blank or recurring items to keep the line of sight on salient features.

Redundant and predictable contours in textured pictures are processed and discarded by the peripheral retina in less than 2 sec. Research on the visual processes of cats has demonstrated a possible physiological mechanism for this rapid process, because certain cortical cells are directly activated by the peripheral retina. These cells in the cortex respond electrically only to straight lines or negligibly curved boundaries with a particular orientation, just as other cells respond only to lines that show changes in direction, again with the appropriate orientation (Hubel & Wiesel, 1965). At first one might believe that the reason for discarding smooth, predictable, and redundant contours was merely that such outlines would not only be redundant but would also be easier to perceive by peripheral vision. Against this possibility one has to place the rejection of redundant textures; these textures are also viewed by peripheral vision, and yet the visual acuity demands here may be very high. But with textures, it can easily be demonstrated that one fixation is sufficient to identify a whole area comprising very fine but regular details. Quite different visual fixation results are obtained when the task is changed from a judgment about the relative coarseness of two adjacent textures to an inspection of each texture for flaws. Many more fixations then occur per unit area (Mackworth & Hiebert, 1967).

The distinction between unpredictable and simple outlines resembles findings by Berlyne (1966) that Ss spend longer times looking at the more complex than at the less complex member of a pair of patterns. Zusne and Michels (1964) and Michels and Zusne (1965) have found that angular complexities in polygon shapes attracted more fixations than did straight lines. A check test with the map display reversed from left to right ruled out possible spatial effects in the map picture. The unique coastline patterns still attracted more fixations. This concentration of fixations was therefore not due to the left side of the picture dominating the right. Note also that in the eyes picture there was the opposite trend of the right side dominating the left.

In a comparison of Attneave's views on perception with those of the Gibsons, Bevan (1958) noted that both approaches imply that the faithfulness with which the perception matches the object depends upon the efficiency with which the perceptual mechanism scans and encodes the input. Therefore, both theories stress the correspondence between the external environment and our perception of it. The individual becomes more sensitive to the variables

of the stimulus array as he learns differential responses to relevant cues.

Our visual fixation data clearly support the Gibsons' (1953, 1955, 1963a, b) approach to perception. Our records have demonstrated that details in the pictures determine to a considerable extent the adjustments made by the adult visual sensory mechanism, as Hake (1957) also predicted. But this confirmation of the Gibsonian views on the importance of perceptual differentiation does not in any way weaken Bruner's case for more accurate classification and better internal models of reality as features of perceptual learning (Bruner, 1957; Bruner, Goodnow, & Austin, 1961; Bruner, Olver, & Greenfield, 1966). The use of more precise mental models implies no retreat from the visual world (Zinchenko, Chzhi-Tsin, & Tarakanov, 1963). On the contrary, the presence of more accurate schemata suggests an even more detailed cross-check between such internal models and external stimulus patterns. In adults, this matching process inevitably guides the fovea toward the unusual, specific features in the display. Simultaneously, redundant and more predictable features are relegated to the periphery of the retina; the peripheral retina therefore quickly screens off the predictable features and leaves the fovea free to process the unpredictable and unusual stimuli.

Conclusions

A few outstanding areas within pictures received high concentrations of the gaze and these regions of the pictures were also judged to be very recognizable. These dominant regions always contained unpredictable contours or unusual details.

Simpler, predictable contours were estimated to be equally recognizable, but these outlines were seldom fixated.

Half to two-thirds of each picture received very few fixations; in particular, areas of texture seldom reached the fovea, especially when the textures were smooth and therefore predictable.

Peripheral vision edited out predictable contours, as well as regions of smooth texture, so rapidly that the unusual and informative areas were fixated in 2 sec or less from the start of the presentation.

Adults fixated details in pictures even during a relatively casual inspection of the scenes. This result supports the Gibsonian view that perceptual learning improves ability to discern distinctive features in the stimulus array.

A general approach has been devised which pinpoints the salient features of pictures most likely to be fixated. The combination of (1) objective measurement by visual fixation patterns with (2) subjective estimates of recognizability or informativeness can also indicate the predictable areas of texture and find redundant contours. Both these

latter types of stimuli are less likely to be fixated because they are so uninformative.

References

Attneave, F. Applications of information theory to psychology. New York: Henry Holt and Co., 1959.

Attneave, F. Some informational aspects of visual perception. Psychol. Rev., 1954, 61, 183-193.

Bartlett, F. C. Remembering. London: Cambridge, University Press, 1932.

Berlyne, D. E. Curiosity and exploration. Science, 1966, 153, 25-33.

Bevan, W. Perception: Evolution of a concept. Psychol. Rev., 1958, 65, 34-55.

Bruner, J. S. On perceptual readiness. Psychol. Rev., 1957, 64, 123-152.

Bruner, J. S. The cognitive consequences of early sensory deprivation. In P. Solomon, et al. (Eds.), Sensory deprivation. Cambridge, Mass : Harvard University Press, 1965.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. A study of thinking. New York: John Wiley & Sons, Inc., 1961.

Bruner, J. S., Olver, R. R., & Greenfield, P. M. Studies in cognitive growth. New York: John Wiley & Sons, Inc., 1966.

Gibson, E. J. Improvement in perceptual judgment as a function of controlled practice or training. Psychol. Bull., 1953, 50, 401-431.

Gibson, E. J. Perceptual learning. Annu. Rev. Psychol., 1963a, 14, 29-56.

Gibson, J. J. The useful dimensions of sensitivity. Amer. Psychol., 1963b, 18, 1-15.

Gibson, J. J., & Gibson, E. J. Perceptual learning: Differentiation or enrichment? Psychol. Rev., 1955, 62, 32-41.

Hake, H. W. Contribution of psychology to the study of pattern vision. USAW WADC Tech. Rept. 1957, 57-621.

Hake, H. W., Rodwan, A., & Weintraub, D. Noise reduction in perception. In K. R. Hammond (Ed.), The psychology of Egon Brunswik. New York: Holt, Rinehart and Winston, Inc., 1966.

Hubel, D. G., & Wiesel, T. N. Receptive fields and functional architecture in two non-striate areas (18 and 19) of the cat. J. Neurophysiol., 1965, 29, 229-289, especially pp. 285-286.

Mackworth, N. H. A stand camera for line-of-sight recording. Percept. & Psychophys., 1967, 2, 119-127.

Mackworth, N. H., & Bruner, J. S. Selecting visual information during recognition by adults and children. Mimeographed monograph, Harvard University Center for Cognitive Studies, 1966.

Mackworth, N. H., & Hiebert, Joyce. Measuring the useful field of view during visual search. (In preparation) 1967.

Michels, K. M., & Zusne, L. Metrics of visual form. Psychol. Bull., 1965, 63, 74-86.

Miller, G. A., Galanter, E. H., & Pribram, K. H. Plans and the structure of behavior. New York: Henry Holt & Co., 1960.

Senders, J. W. A re-analysis of the pilot eye-movement data. I.E.E.E. Trans. on Human Factors in Electronics, Vol. HFE-7, No. 2, June, 1966, 103-106.

Spinelli, D. H., & Pribram, K. H. Changes in visual recovery functions and unit activity produced by frontal and temporal cortex stimulation. EEG clin. Neurophysiol., 1966, 20, 44-49.

Staniland, A. C. Patterns of redundancy: A psychological study. New York: Cambridge University Press, 1966.

Zinchenko, V. P., Chzhi-Tsin, V., & Tarakanov, V. V. The formation and development of perceptual activity. Soviet Psychol. & Psychiat., 1963, 2, 3-12.

Zusne, L., & Michels, K. M. Non-representational shapes and eye movements. Percept. mot. Skills, 1964, 18, 11-20.