

Human Factors and Behavioral Science:

Textons, The Fundamental Elements in Preattentive Vision and Perception of Textures

By B. JULESZ* and J. R. BERGEN*

(Manuscript received September 23, 1981)

Recent research in texture discrimination has revealed the existence of a separate "preattentive visual system" that cannot process complex forms, yet can, almost instantaneously, without effort or scrutiny, detect differences in a few local conspicuous features, regardless of where they occur. These features, called "textons", are elongated blobs (e.g., rectangles, ellipses, or line segments) with specific properties, including color, angular orientation, width, length, binocular and movement disparity, and flicker rate. The ends-of-lines (terminators) and crossings of line segments are also textons. Only differences in the textons or in their density (or number) can be preattentively detected while the positional relationship between neighboring textons passes unnoticed. This kind of positional information is the essence of form perception, and can be extracted only by a time-consuming and spatially restricted process that we call "focal attention". The aperture of focal attention can be very narrow, even restricted to a minute portion of the fovea, and shifting its locus requires about 50 ms. Thus preattentive vision serves as an "early warning system" by pointing out those loci of texton differences that should be attended to. According to this theory, at any given instant the visual information intake is relatively modest.

* Bell Laboratories.

©Copyright 1983, American Telephone & Telegraph Company. Copying in printed form for private use is permitted without payment of royalty provided that each reproduction is done without alteration and that the Journal reference and copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free by computer-based and other information-service systems without further permission. Permission to reproduce or republish any other portion of this paper must be obtained from the Editor.

I. INTRODUCTION

In this article we give an overview of some insights into the workings of the human visual system gained during two decades of research at Bell Laboratories, and culminating in the discovery of a few local conspicuous features that we call textons. Textons appear to be the basic units of preattentive texture perception,¹ when textures are viewed in a quick glance with no further effort or analysis. Although this article goes beyond texture perception into preattentive vision in general, studies of texture discrimination led to the basic insights presented here and provide excellent demonstrations of the main findings. Based on our findings we propose a novel theory of vision in which the preattentive visual system inspects a large portion of the visual field in parallel and detects only density differences in textons. It then directs focal attention to these loci of texton differences for detailed scrutiny.

Now, after 20 years of research, when we know what textons are and their role in vision is clarified, we can save the reader from following the rather difficult steps that led to their discovery. [The reader interested in the history of these developments, and in the sophisticated mathematics necessary to generate textures with certain stochastic constraints, should turn to the original articles referred to in a recent review by one of us¹ and to the Appendix.] Here we follow an axiomatic treatment. The main findings are presented in Section II as heuristics (similar to axioms, but not necessarily totally independent), immediately followed by many demonstrations. The reader can test the power of these newly acquired heuristics by being able to predict and then verify which texture pairs will be perceived to be different, and which will appear as a single texture. The reader can thus understand the new theory of vision without mathematical knowledge.

Section III emphasizes the essentially local nature of texture perception. Section IV relates the psychologically identified textons to some neurophysiological results concerning local feature analyzers in primate cortex. Section V extends the texton theory from texture perception to the discrimination of briefly presented patterns. In Section VI a model of human vision is proposed that postulates two different modes of visual system function. Section VII discusses some implications of this model.

II. HEURISTICS: DEFINITION OF TEXTONS AND THEIR INTERACTIONS IN PREATTENTIVE VISION

Visual textures are defined as aggregates of many small elements. The elements can be either dots of certain colors (e.g., black, white, grey, red) or simple patterns. For purposes of this article, we consider

only elements that do not overlap, and are placed at either regular or random positions, in identical or in random angular orientations.

Usually in our demonstrations two textures (composed of two different elements) are placed side-by-side, or one is embedded in the other, as shown in Fig. 1. When the reader cursorily inspects Fig. 1, an area made up of +’s will appear to stand out from the surrounding texture composed of L’s. Indeed, without scrutiny, that is without detailed element-by-element inspection, the reader might not notice that a third area composed of T-shaped elements is also embedded in the texture of L’s. We call this effortless perceptual segregation of the texture composed of +’s from the surrounding texture of L’s *preattentive texture perception*. On the other hand, if texture discrimination requires element-by-element scrutiny, as is the case of finding the T’s in the L’s, we call this way of looking with scrutiny *focal attention*. We will show many other preattentively indiscriminable texture pairs (e.g., Figs. 3c and 6b), which, because they do not segregate, often are not even perceived as containing different elements until this is pointed out.

Although in all texture perception the preattentive system is dominant, the role of focal attention can be even further reduced by brief presentation. The reader who is not convinced by the qualitative difference between preattentive and attentive texture discrimination

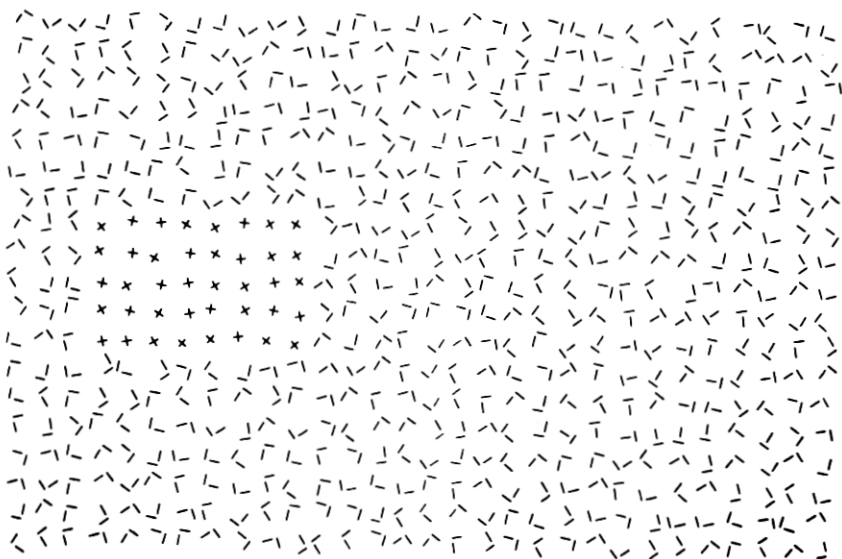


Fig. 1—"Preattentive texture discrimination" is shown between areas composed of +’s and L’s, while element-by-element scrutiny, called "focal attention" is required to find the T’s embedded in the L’s.

might inspect Fig. 1 through a camera shutter set at $\frac{1}{10}$ second exposure time.

Heuristic 1: Human vision operates in two distinct modes

1. Preattentive vision—parallel, instantaneous, without scrutiny, independent of the number of patterns, covering a large visual field, as in texture discrimination.
2. Attentive vision—serial search by focal attention in 50-ms steps limited to a small aperture, as in form recognition.

Heuristic 2: Textons

1. Elongated blobs—e.g., rectangles, ellipses, line segments with specific colors, angular orientations, widths, and lengths.
2. Terminators—ends-of-line segments
3. Crossings of line segments

Heuristic 3: Preattentive vision directs attentive vision to the locations where differences in the density (number) of textons occur, but ignores the positional relationships between textons.

Before we discuss the implications of these heuristics, let us apply them to a few pairs of elements and predict whether the texture pairs formed from these elements will yield preattentive texture discrimination or not. This application of the rules also helps to clarify them. For instance, elongated blobs of different widths or lengths are different textons, as Fig. 2a demonstrates. The larger sized R's containing longer and wider line segments form a texture that segregates (i.e., is preattentively discriminable) from its surround, which is composed of smaller R's with shorter and narrower line segments.

Similarly, elongated blobs of different orientations are different



Fig. 2—Preattentive texture discrimination based on texton differences between line segments of (a) length and width and (b) angular orientation. (*Nature*, March 12, 1981¹)

textons. Indeed, in Fig. 2b the texture pair composed of the same sized R's having two different orientations in the two textures, yields preattentive discrimination. Obviously, the same elongated blob shape with the same orientation yields different textons if the colors (e.g., black, gray, white, red, green, etc.) are different.

Now, let us predict what would happen if we took an R and a mirror-image R, as shown in Fig. 3a, and formed a texture pair by throwing them in random orientations. Obviously, without randomizing the orientations, the two textures would yield texture discrimination, since even though their widths and lengths agree, some of the line segment textons have different orientations in the R and in its mirror image, though the widths and lengths agree, as shown in Fig. 3b. However, if the two elements are thrown at random orientations, then the two textures formed have the same average density of textons (i.e., in some area of integration the number of line segments with the same color, width, length, and orientation is identical). Therefore, the preattentive visual system should not be able to direct focal attention to loci of texton differences that form the boundary between the two regions. Indeed, an inspection of Fig. 3c yields a single, uniform texture. It requires laborious element-by-element inspection for several seconds

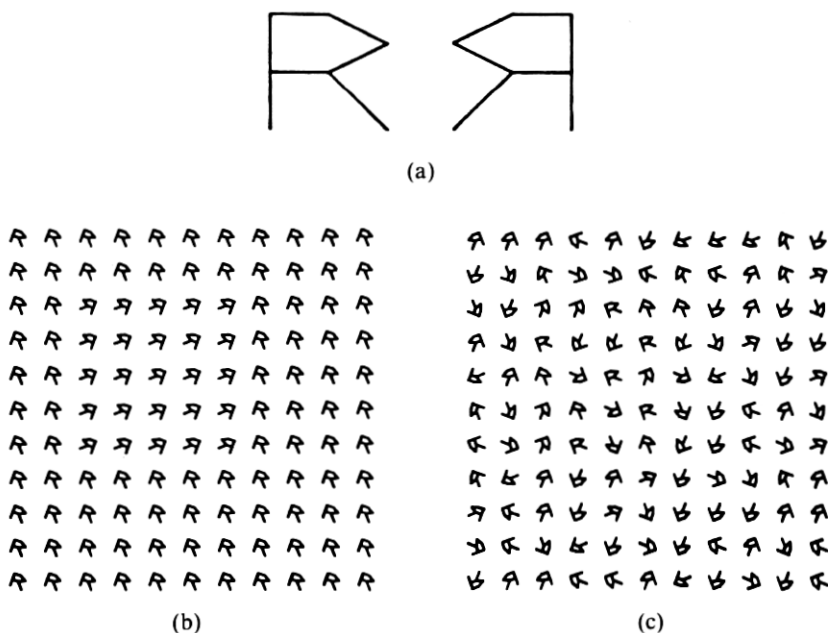


Fig. 3—Demonstration of how the heuristics given in text predict why (a) R and its mirror image in aggregates yield texture discrimination (b), or are indistinguishable (c). (*Perception*, 1973⁵)

to find the boundary between the array of R's and mirror-image R's. Obviously, in a 100-ms presentation discrimination of these textures is impossible.

Let us note that if one were to select a pair of elements without knowing the rules given above, most probably the resulting texture pair would be discriminable. Only through the joint effort of our colleagues (D. Slepian, M. Rosenblatt, E. Gilbert, L. Shepp, H. Frisch, T. Caelli, and J. Victor) from 1962 to 1978 were some elegant methods found that yielded indistinguishable textures, even though their elements appeared very different.

In the next examples we stress the importance of terminator textons. For instance, in Fig. 4a the two elements are composed of three identical line segments (i.e., same orientation, width, and length). The only difference is in the number of their ends-of-lines (terminators). The triangle-shaped element has no open ends, while the "dual" element has three ends-of-lines. One should expect texture segregation, given such a large difference in terminator number, and as Fig. 4b demonstrates, this is the case.

As a matter of fact, discrimination is so strong that a single element can be preattentively detected among 35 dual elements, as shown in Fig. 4c. This arrangement is now routinely used by us in studying

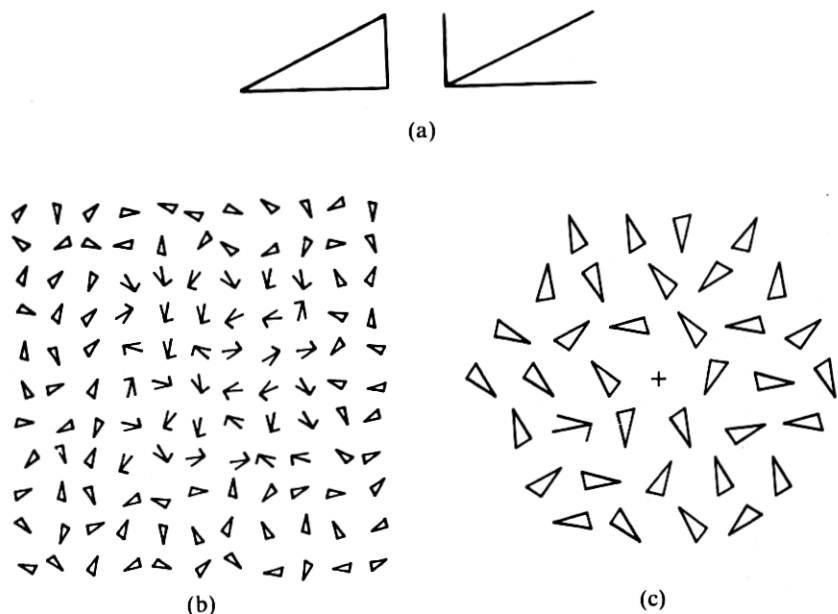


Fig. 4—Demonstration of how the heuristics given in text predict preattentive texture discrimination (b) and even discrimination of a single element among many (c), based on terminator number difference (zero versus three) between elements (a). (*Nature*, March 12, 1981¹)

pattern discrimination in preattentive vision, as discussed in Section V. Here we note only that when there is a texton difference (as in Fig. 4c) detecting one element in the midst of 35 other elements is almost as easy as detecting the difference between two elements (shown in Fig. 4a) for presentation times as brief as 100 ms.

In the next example, both members of the element pair of Fig. 5a are again composed of the same five line segments (each corresponding line segment in the two elements has identical width, length, and orientation, respectively) but one element contains only two ends-of-lines, whereas the other contains five. This large difference in terminator numbers should yield texture segregation, and inspection of Fig. 5b demonstrates that it does. Figure 5c consists of the same texture pair as Fig. 5b, except that the texture containing the five terminators is now the surround. Although, as predicted, the large difference in terminator numbers again yields texture segregation, the appearance of the boundary between the two regions is different for Fig. 5b and 5c.

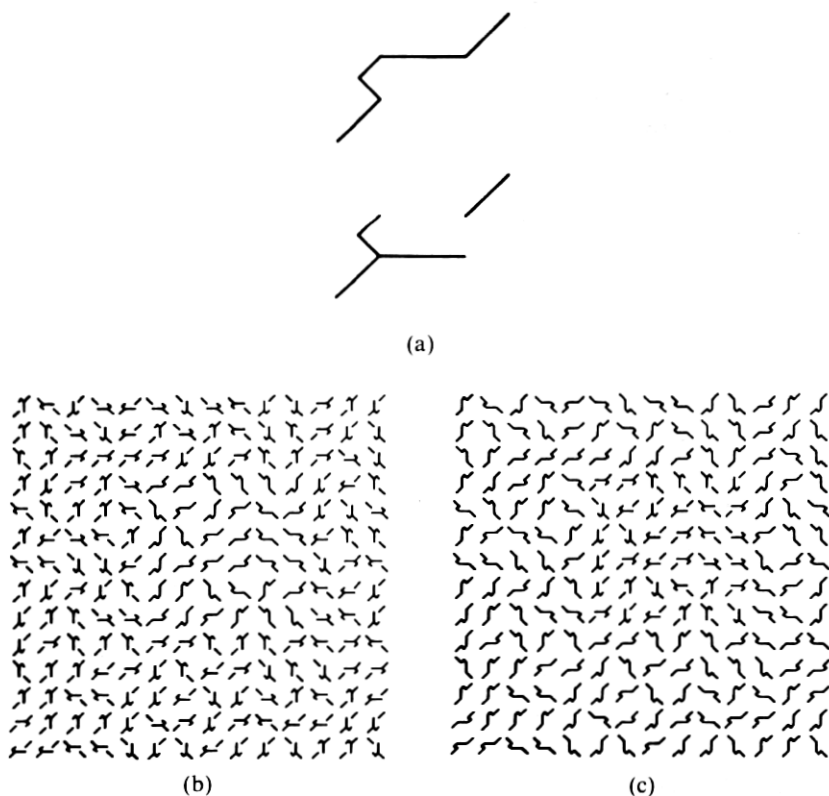


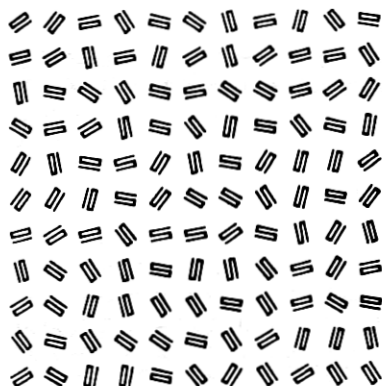
Fig. 5—Similar to Fig. 4 except the terminator number between elements is two versus five.

The next example, shown in Fig. 6a, consists of the "S"- and "10"-shaped elements, that in isolation appear quite different. However, the two contain the same number of line segment textons (three identical horizontal and two identical vertical line segments) and both contain two ends-of-lines. The fact that the positional relationship between these textons is different (as it is in Fig. 3b) can be perceived only by the attentive visual system (yielding the percept of an S versus a 10). However, according to Heuristic 3 the preattentive system can count only the density (number) of textons and ignores their relative positions. So, according to our rules, a texture pair composed of these elements contains the same average density (number) of textons, and thus should be indistinguishable. Surprising as it may seem, the texture pair is indeed preattentively indistinguishable as demonstrated by Fig. 6b. [Readers who find this demonstration of the distinction between preattentive and focal vision not adequately convincing without brief presentation should note the contrast between the attentively different percepts of Fig. 6a, and the texture pair in Fig. 6b, which remains difficult to distinguish even with element-by-element scrutiny.]

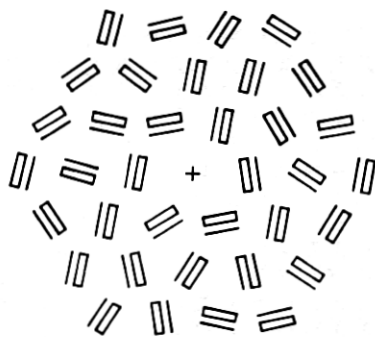
Finally, let us demonstrate the third texton, the crossing of elongated blobs (line segments). Figure 7a shows the conspicuous difference between a texture pair that segregates based on the presence or absence of elements having crossing versus not-crossing line segments.



(a)



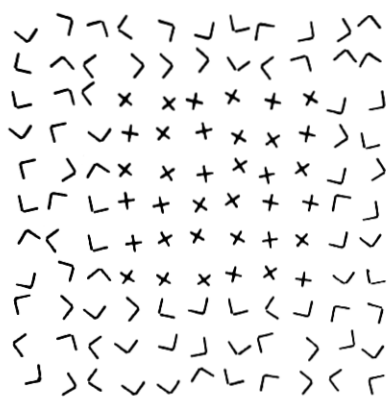
(b)



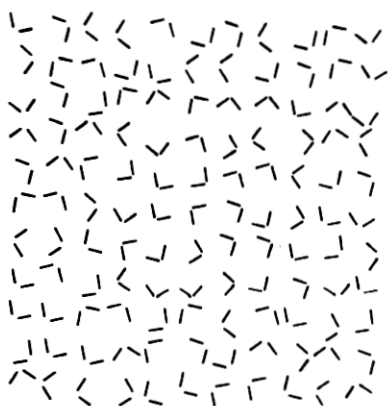
(c)

Fig. 6—Demonstration of how the heuristics given in text predict why (a) the differently appearing S- and 10- shaped elements in aggregates (b) and one S in 10's (c) are indistinguishable. (*Nature*, March 12, 1981¹)

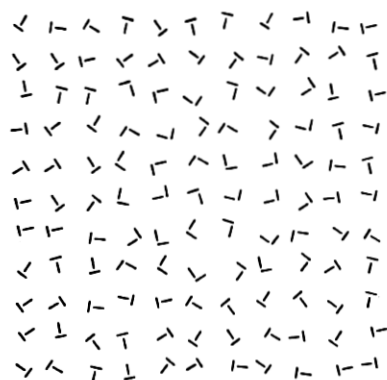
If the elements have identical textons, including crossing (or not-crossing line segments) the texture pairs become preattentively indistinguishable. The positional relationship between the line segment textons is unnoticed by the preattentive system. The difference in gap size between the L-shaped elements in Fig. 7b yields a preattentively indistinguishable texture pair. Particularly interesting is the demonstration in Fig. 7c where T- versus L-shaped elements yield an indistinguishable texture pair. Although we have kept a small gap between the perpendicular line segments that make up the L's and T's, preattentive discrimination of texture pairs composed of these elements is impossible even when the gaps are not resolvable. Apparently, the difference of a single end-of-line terminator is not adequate to yield texture segregation. Finally, Fig. 7d depicts a preattentively indistinguishable texture pair, where, with scrutiny, it is obvious that the



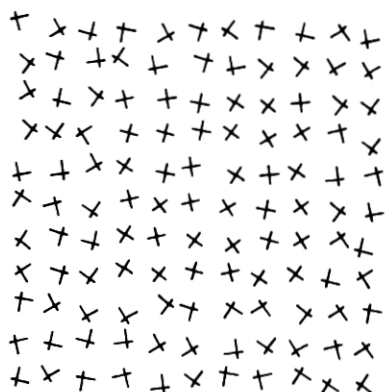
(a)



(b)



(c)



(d)

Fig. 7—Demonstration that crossing of line segments is a texton.

elements contain line segments that either cross at midpoint or cross far from the midpoint.

The last two examples are given in Figs. 8a and b and Figs. 9a, b, and c. From the element pairs containing the same textons, the reader can predict that although their elements in isolation appear very different, the resulting texture pairs will be indistinguishable.

In all these demonstrations the texture elements consisted of line segments. For line segments the definition of terminators (ends-of-lines) and their crossings are straightforward. For elongated bars with substantial width these definitions are less direct. Particularly difficult is the notion of terminators, because instead of terminators some combination of white elongated bars in a black surround with black elongated bars in white surround might suffice. So, we are not certain whether terminators are independent textons. Nevertheless, as a first approximation these three heuristics work remarkably well.

III. PREATTENTIVE TEXTURE PERCEPTION IS ESSENTIALLY A LOCAL PROCESS

The essence of all the findings reported in the previous section can be summed up as follows: In texture perception the preattentive visual system utilizes only local conspicuous features, textons, and these textons are not coupled to each other (i.e., a vertical and horizontal line segment do not cohere to form an L or T). The preattentive system utilizes globally only the textons in the simplest possible way by counting their numbers (densities). This might surprise many of our readers who assume that texture perception utilizes complex global statistical interactions between textural elements.

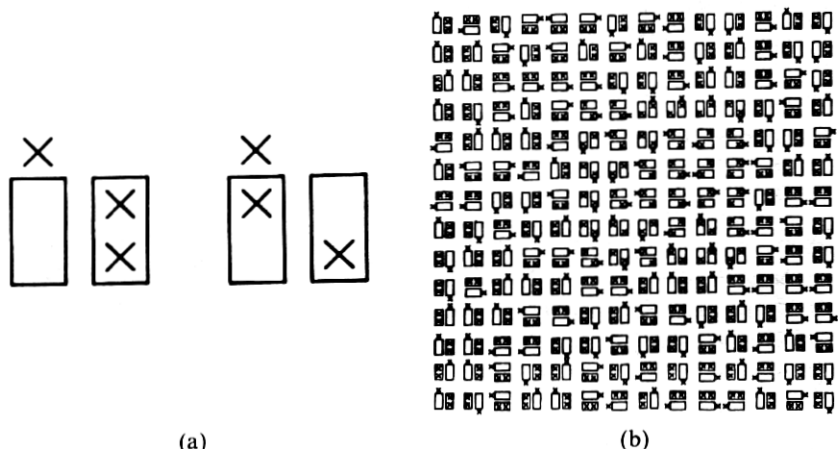
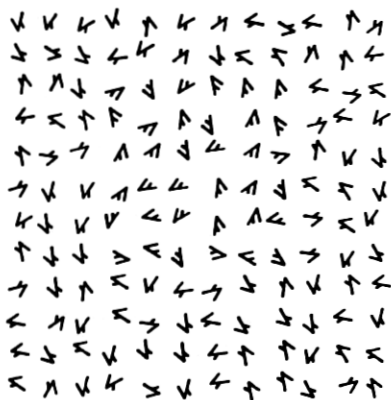


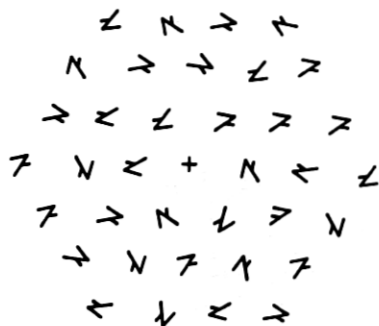
Fig. 8—Since the element pair in (a) is composed of the same textons, the texture pair (b) composed of these elements is preattentively indistinguishable. (*Philosophical Transactions*, 1980²⁶)



(a)



(b)

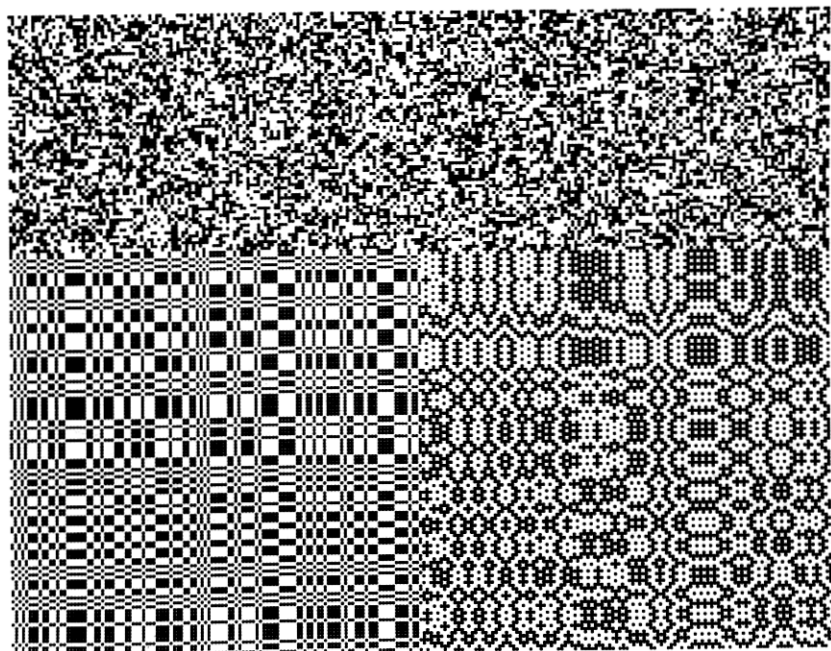


(c)

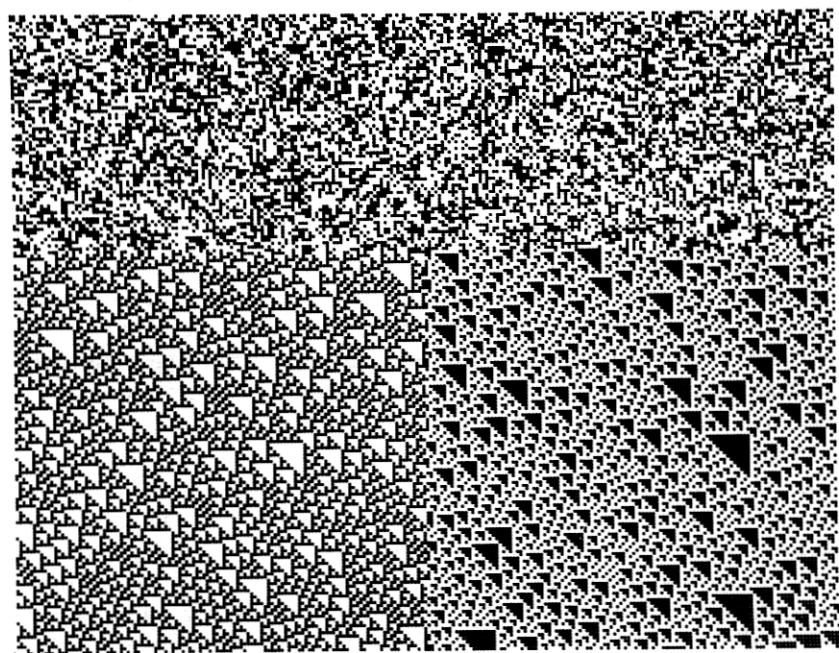
Fig. 9—Similar to Fig. 8, showing that aggregates of elements composed of the same textons cannot be preattentively discriminated. (*Philosophical Transactions*, 1980²⁵)

One of the simplest *global* computations routinely performed on images by engineers, and recently by psychologists in vision research, is to determine the images' Fourier power spectra. This process involves the decomposition of the images into one-dimensional sinusoidal luminance gratings whose specific amplitudes, spatial frequencies, phases, and angular orientations depend on the spatial characteristics of luminance distributions across the entire image. The amplitude of the spectral components ignoring phase determine the power spectra. When Fourier power spectra of textures are taken, it is a common misconception that differences in these will reveal differences in texture granularity. That the preattentive visual system does not perform Fourier analysis is demonstrated next.

Figure 10a consists of three areas that have identical Fourier power spectra (invented by Julesz, Gilbert and Victor²) and yet appear as very distinct textures. [The mathematically sophisticated reader might appreciate that the three areas have identical third-order statistics, and differ only in their fourth-order statistics. Those interested in the definition of *n*th-order statistics should consult Refs. 3 and 4 and the Appendix.] Figure 10b also consists of three areas with identical Fourier power spectra, and again these areas appear conspicuously different. Conversely, in Fig. 11a the lower left quadrant of the bottom



(a)



(b)

Fig. 10—Discriminable texture pairs with identical Fourier power spectra (*a* has even identical third-order statistics) based on local granularity (texton) differences. (*Biological Cybernetics*, 1981²)

right array has a very different power spectrum from the remainder of the array, yet no preattentive texture discrimination results.⁵ The derivation of this texture pair is presented in three steps. The top left array, in Fig. 11a, consists of 4x4 dot elements (8 black and 8 white) with 6-dot periodicity. The bottom left array contains this periodic array in one quadrant, but the 2-dot-wide gaps are filled by a checkerboard screen, while the rest is covered with uniformly random black and white dots. The bottom right array is similar to the bottom left array, except the 2-dot-wide gaps between the periodic patterns are now randomly speckled with dots. Obviously, the periodic patterns in the lower quadrant of the bottom right array in Fig. 11a yield a very different Fourier power spectrum from the rest, which has a flat (white noise) spectrum. The reason that this texture pair is indistinguishable can be easily understood in the light of the texton theory. The periodic patterns are not different from the surrounding random-dot array in the density of elongated blob textons, and therefore are indistinguishable. Indeed, if the 4x4 dot micropattern consists of vertical stripes, which contain textons different from the surrounding random-dot array, as shown in Fig. 11b, the periodic quadrant embedded in randomness is easily perceived.

In all these densely packed dot textures, discrimination is based on local granularity differences that correspond to differences in the density (number) of elongated blobs of certain sizes and orientations. Global statistical descriptors of textures, including the Fourier power spectrum, apparently are ignored in preattentive vision.

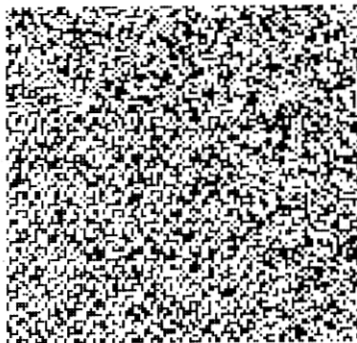
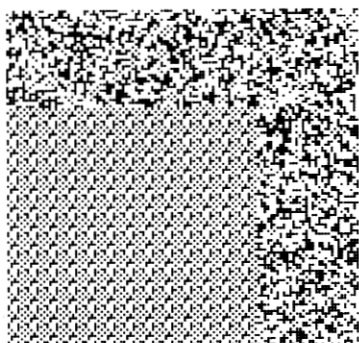
IV. TEXTURES AND NEUROPHYSIOLOGICAL FEATURE ANALYZERS

We have seen how elongated blob textons are crucial in preattentive texture perception. These human psychological findings have a parallel in primate neurophysiology. Neural units have been found by Hubel and Wiesel⁶ in the visual cortex of monkeys that fire optimally for elongated blobs of specific width, length, and orientation. These neural units in the cortex have *retinal receptive fields* consisting of elongated, blob-shaped, excitatory regions, which are surrounded by inhibitory regions. Some of these elongated blob detecting units—which fire optimally for black elongated blobs surrounded by white flanking areas—are called simple “off” detectors. Other neural units are excited optimally by white elongated blobs surrounded by black. These are called simple “on” detectors. The exact shape of the receptive fields of these simple neural units varies a great deal, and is of secondary importance. The important property of these cortical units is that the weighting of the excitatory and inhibitory areas of their receptive fields is about equal, so that for homogeneous stimuli they do not fire.

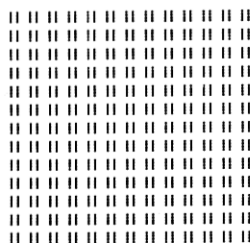
It should be stressed that the textons reported here were found by



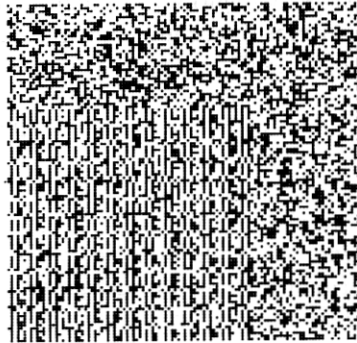
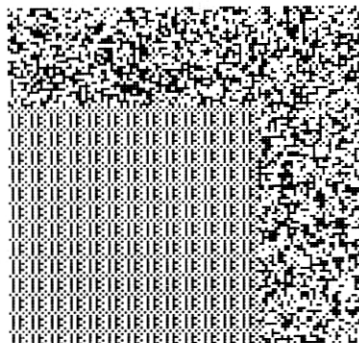
PERIODIC TEXTURE
4X4 WITH 6 PERIOD
SEED=.124 FR=.5



(a)



PERIODIC TEXTURE
4X4 WITH 6 PERIOD
BAR TEXTONS, FR=.5



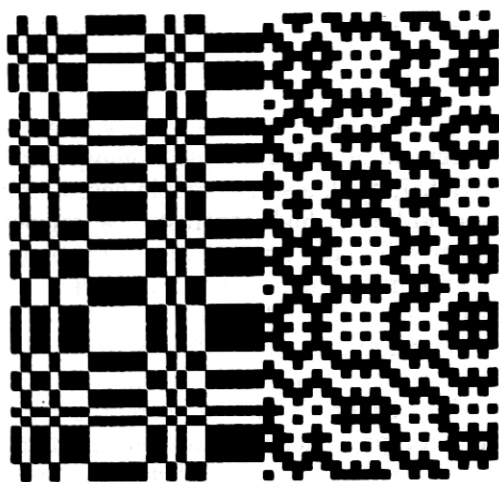
(b)

Fig. 11—Demonstration that the preattentive system cannot perform even such a simple global computation as Fourier power spectra, as described in text. (*Biological Cybernetics*, 1978⁶)

psychological methods, and imply that simple neural units found as early as the striate cortex of the monkey might be used in texture perception. However, the relationship between a texton—for example, a perceived line segment—and a Hubel and Wiesel type of neural feature analyzer with a receptive field whose excitatory center matches the shape of the line segment is not a simple isomorphism. As we pointed out years ago (Ref. 7, p. 3), a single simple neural unit might respond equally for a broad line of high contrast or a narrow line of low contrast, while perceptually one can preattentively perceive both the width and contrast of a line segment. Thus, obviously a perceived line segment is encoded by many neural units of similar orientations but tuned to different widths, and having different firing thresholds. It is some combination of these units that would correspond to a perceived line segment. Until more is known about the relationship between perception and neurophysiology, the textons must be defined as perceptual entities, that is conspicuous local features as we actually perceive them. Nevertheless, even though textons and neural units are not simply related, one can easily conceptualize how a “perceptual analyzer” could be built from known neural analyzers that could extract, say, a line segment texton. The question of whether terminators and crossings of line segments—which have been regarded as textons—could be related to the complex and hypercomplex neural analyzers found by the neurophysiologists remains to be seen.⁶

David Marr, in his primal-sketch model of machine vision, also incorporated such elongated blob detectors, by assuming that the neurophysiological findings had direct relevance to vision.⁸ The work reported here followed an opposite trend. It took almost two decades to find evidence for the utilization of simple cortical units in texture perception. Caelli and Julesz found the first elongated blob textons that could account for texture discrimination locally, when all global statistical properties of the texture pairs were kept identical.⁹ Later demonstrations such as Figs. 10a and b illustrate even more strikingly the importance of local blob textons.

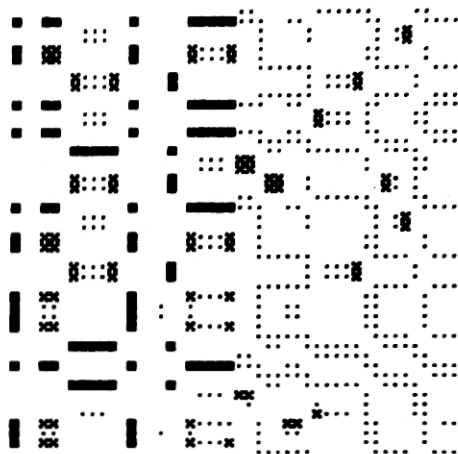
To demonstrate the possible role of the Hubel and Wiesel type of neural units in preattentive texture perception, we developed a computer program called TEXTONS that filters any image with a pool of elongated bar-shaped receptive fields. Each pool of filters consists of “on” and “off” types having the same width, length, orientation, and firing threshold and placed at each point of the array. Figure 12 (bottom) shows the three largest response levels of a pool of 3x3 dot square-shaped receptive fields as this pool processes the texture pair of Fig. 10a, shown also in Fig. 12 (top). These filters have 2x2 dot excitatory centers flanked by one-dot-wide inhibitory margins, as shown in Fig. 12 (right). Of course, there are several pools consisting



(a)



(b)



(c)

Fig. 12—Automatic texture segregation, shown in (c), by applying a texton filter (b) to texture pair (a) (also shown in Fig. 10a).

of filters having receptive fields with some other dimensions and orientations that would be even more effective in segregating the two textures of Fig. 12 (top). Here we stress again that the combination of several filters would be required to yield the best texture segregation, corresponding to human texture discrimination. This combination of filters would correspond to a texton detector.

What our psychological findings show, however, could not have been

guessed by physiologists and theoreticians of artificial intelligence. In preattentive texture perception the various textons are not coupled, that is their relative positions are ignored. T- and L-shaped pairs of line segments cannot be discriminated preattentively in textures. Marr thought that elongated blobs and terminators would form some higher molar unit, which he called place tokens.⁸ However, in preattentive vision no such higher interactions are found; the textons appear to be independent of each other.

V. EXTENSION OF THE TEXTON THEORY TO RAPID PATTERN DISCRIMINATION

The success of the texton theory in predicting phenomena of texture perception is the result of the spatial complexity of the patterns. This complexity over a large area exceeds the capacity of focal attention and thus allows the preattentive system to dominate. This same deemphasis of focal attention can be achieved in simpler patterns by very brief presentation. We will show that under these conditions the same texton theory can be applied.¹⁰

Because brief temporal presentation is required, the stimuli used in these experiments can be produced only in the laboratory. Consequently, they cannot be demonstrated as the texture discrimination results have been. Thus, in this section we present the main findings as curves describing observers' performance.

The stimuli used in these experiments are shown in Fig. 13. In Fig. 13a there are 35 T's and one L arranged on a hexagonal grid with slight random positional jitter added. In Fig. 13b the T's have been replaced by + elements, and in Fig. 13c only two of the 36 possible positions actually contain an element. In all cases, a disk surrounding the central fixation marker is kept empty. Stimuli of this type are presented for 40 ms, followed by a blank interval of variable duration and a 40-ms erasing field. This erasing field consists of elements, which are the union of the two being discriminated, arranged in the same way as the test field. Use of this erasing technique allows restriction of the inspection interval to times shorter than the duration of the retinal afterimage. The times used are all too short to allow eye movements to be initiated during the presentation. In half of the presentations the test field consists of all identical elements, while in the other half one element is different, as in the examples shown. The task of the observer is to discriminate between these two conditions.

Results obtained using the three stimuli of Fig. 13 are shown in Fig. 14. On the abscissa is the time in milliseconds between the onsets of the test and erasing fields, or the stimulus onset asynchrony (SOA). On the ordinate is the percentage of correct discrimination. The results

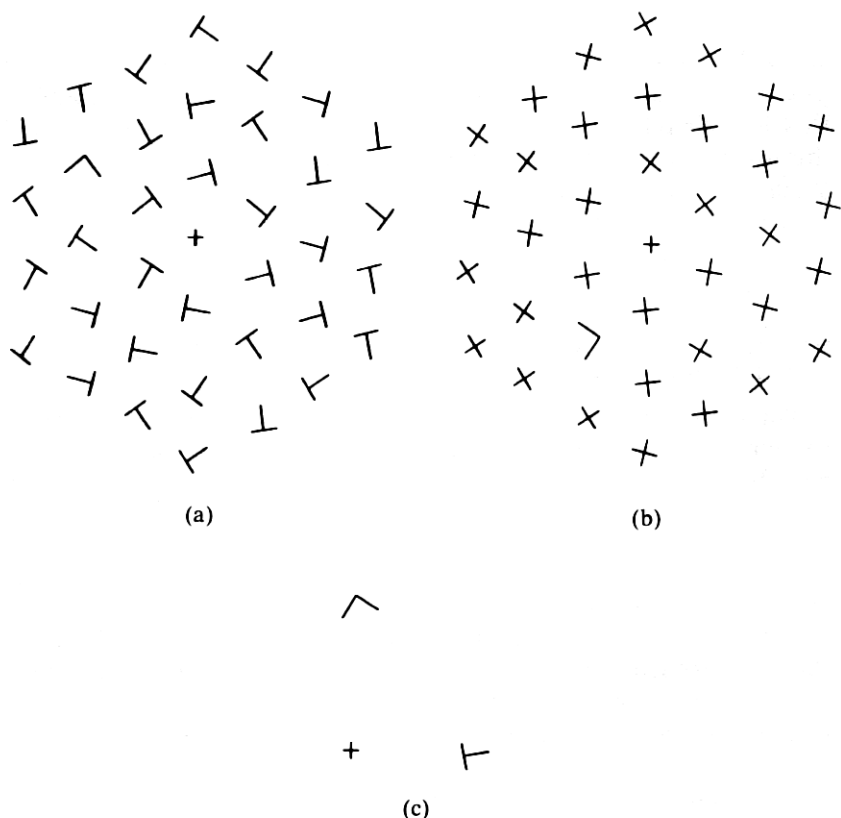


Fig. 13—The three types of stimuli used to obtain the results of Fig. 14. In all cases the task is to discriminate these from a control stimulus in which all elements are identical.

are very different for the case in which the elements share the same textons (T vs L, solid circles) from that in which they contain different textons (+ vs L, open circles). Note that in the T vs L case, not only does performance never exceed 65 percent correct, but it takes over 300 ms to reach this asymptote, while in the + vs L case the asymptote is reached within 200 ms. In fact, by the time the asymptote in the T vs L case is reached, the afterimage resulting from the test flash has largely disappeared. In the case in which only one T and one L are presented (filled squares), the results closely resemble those for 35 +'s and one L. Perceptually, the difference between the same-textons and different-textons cases is simply that the L in the field of T's stands out almost as if presented alone on a blank field, while the same L in a field of T's must be sought out. Attention is rapidly shifted to the L in the former case, while in the latter the search process is apparently still going on after 300 ms have passed. When there are only two elements to choose from, this search time is very brief.

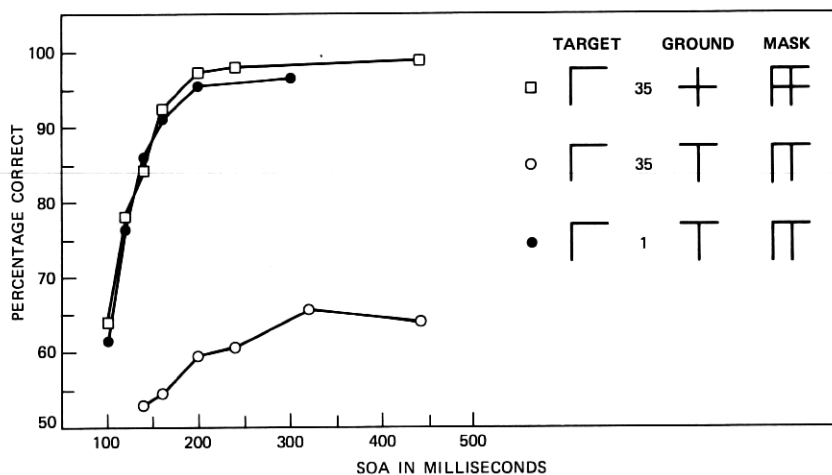


Fig. 14—Results of discrimination experiments using Fig. 13. The open circles, filled circles, and squares correspond to stimuli of Fig. 13a,c,b, respectively. SOA is the abscissa, and percent correct discrimination is the ordinate.

It is interesting to note that the observed asymptotic level of about 65 percent correct is what would be expected if seven or eight of the possible positions could be searched in the time available. Combining this number with the afterimage persistence time of 300–400 ms^{11,12} gives a figure of about 50 ms per position inspected.¹⁰

This process of sequential inspection seems to be essentially independent of the overall angular subtense of the stimulus. Figure 15 shows results from an experiment in which the observer is required to distinguish a stimulus consisting of six T's and one L, or vice versa, from one in which all elements are identical. The stimulus was uniformly contracted so as to fall entirely within the fovea (<3 degrees across), or dilated to extend almost 14 degrees across, with no systematic variation in performance.

Another way of describing this is to say that the measurements are independent of the distance from which the stimulus is viewed, assuming that all of the elements remain resolvable: This independence suggests two important points. First, the fovea is not better than the near periphery in the extraction of this type of visual information. Second, the aperture of attention changes its spatial scale according to the size of the feature being sought. Thus, the same number of sequential fixations of attention are needed when the stimulus is reduced in size uniformly, because the sizes of the features upon which the discrimination is based are proportionally reduced. This extension of the scope of the texton theory from texture perception to rapid pattern discrimination suggests a model of vision in general as described in the following section.

VI. A MODEL OF THE "TWO VISUAL SYSTEMS"

When a visual scene changes suddenly in time or space, and our attention encompasses the entire visual scene, only those areas in which density differences in textons occur are conspicuous. These textons are elongated blobs with specific colors, widths, lengths, orientations, terminators, and crossings between them. Furthermore, because binocular disparity, movement disparity, and flicker are locally conspicuous features that can be detected in a brief presentation,^{7,11,13} they, like color, are also properties of elongated blob textons.

Focal attention is directed to areas of spatial or temporal texton changes. The preattentive process appears to work in parallel and extends over a wide area of the visual field, while scrutiny by local or foveal attention is a serial process, which at any given time is restricted to a small patch. Focal attention can be shifted in 50-ms steps, four times faster than the fastest scanning eye movements. Furthermore, the aperture of focal attention can vary in size and can be a minute portion of the fovea, that is, extending to only a few minutes of arc (as shown in Fig. 15). Therefore, if the visual environment is rich in detail even when slowly changing in time, or is rather lacking in spatial detail but changes rapidly, we perform the major portion of our spatio-temporal processing in the preattentive state.

The focus of visual attention seems to be characterized by a texton class as well as a spatial locus. In particular, just as it apparently is impossible to attend simultaneously to two different places, it also seems impossible simultaneously to attend to very different sizes of features. This fact has been noted previously by other psychologists.¹⁴ Stimuli widely separated in space produce cortical responses which are far apart. Similarly, stimuli of differing sizes often generate re-

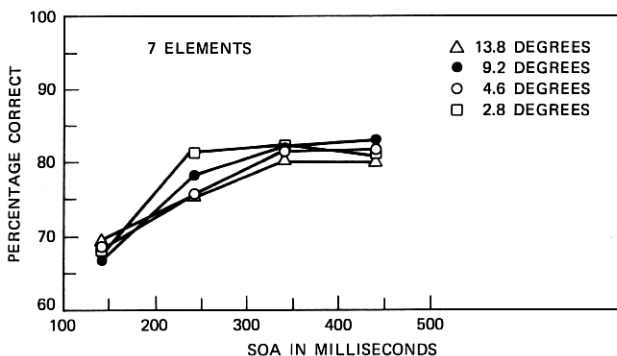


Fig. 15—Size invariance while the angular subtense of seven elements is varied from 2.8 degree of arc diameter to 13.8 degree of arc. Findings imply that the aperture of focal attention can be as small as a few minutes of arc.

sponses in different cortical areas.^{6,15} These results seem to imply that the focus of attention is restricted to a very small region of visual cortex, and that stimuli producing responses far apart in the cortex cannot be attended simultaneously.

The essence of our findings is illustrated in Fig. 16. The left array contains a texture composed of "L"-shaped elements (formed by two perpendicular line segments with a gap), except for one "+" shaped and one "T" shaped element (formed by two perpendicular line segments which cross, or have a gap, respectively). The "+" shaped element (target) differs from the many surrounding L's in one texton, namely the "crossing", and perceptually stands out immediately. On the other hand, the T-shaped target can be detected only after some search, by directing the aperture of attention to the target itself.

The right array of Fig. 16 is identical to the left but illustrates our model of vision. The parallel preattentive system instantly detects the location of texton differences and directs the aperture of focal attention to this location, as indicated by the dotted disk around the "+". Since the T contains the same textons as its surround, its detection requires the aperture of attention (symbolized as a "cone" of a search-light) to scrutinize the texture elements in sequence. Therefore, this

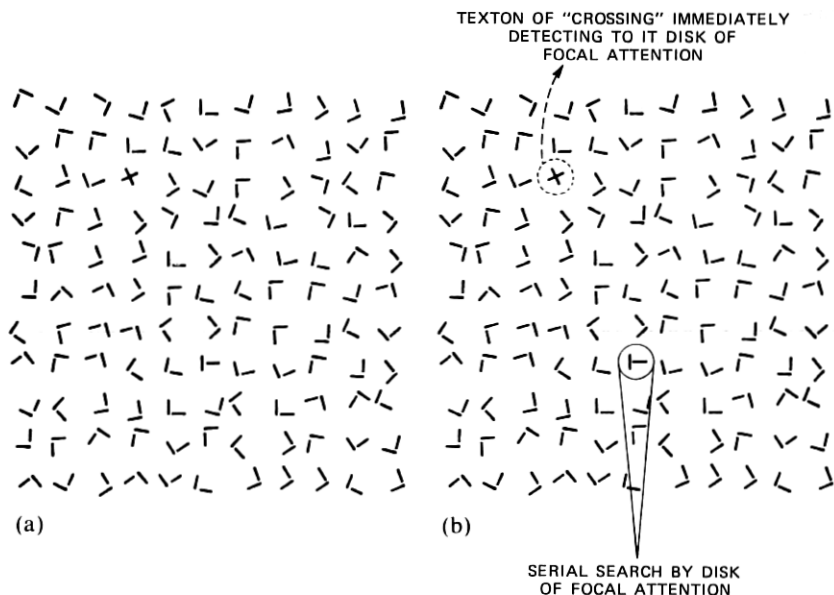


Fig. 16—Model of the two visual systems (b), showing how the preattentive system directs the aperture of focal attention to the loci of texton differences [the + in the L's in (a)], while without such texton differences [the T in the L's in (a)] focal attention requires time-consuming search.

serial search for the T-shaped target depends on the number of texture elements and may take considerable effort and time. However, after the T has been found, and the aperture of focal attention surrounds it, both the + and the T targets are seen with the same clarity. Obviously, form recognition, restricted to the aperture of focal attention, does not depend on the way attention has been directed to the targets. Whether a local difference in textons quickly directed focal attention to the target, or in the absence of texton differences it required time-consuming search to find the target, is immaterial for processing of the target by the attentive visual system.

This mode of behavior of the preattentive and attentive visual systems can also be observed in texture perception, when the reader inspects Fig. 1. The preattentive system immediately detects the texton differences at the boundary of the + and L aggregates, and a quick inspection by focal attention of a few elements on the two sides of the boundary lets the observer conclude that the two areas must contain +'s and L's. Only detailed scrutiny will reveal that the area believed to contain L's only has a region of T's as well.

In summary, the reason that texture discrimination is such a revealing process for showing the workings of the two visual systems is that textures usually cover wide areas of the visual field, while the texture elements are a small portion of the textural area. When the observer is inspecting an extended field, there is an "uncertainty region" in which the relative spatial position of local features is ignored. This is very different from a resolution limit due to visual acuity. In all of the indistinguishable texture pairs, the line segments which make up the texture elements are clearly resolved; nevertheless, if these textons fall within this uncertainty region, it is impossible to tell a T from an L. Many physiologists and psychologists have proposed two visual systems, one ambient and the other focal.¹¹⁻²⁰ Yet, without the notion of textons, whose spatial and temporal changes are detected by the preattentive system, which in turn directs focal attention to these loci, the model of the two visual systems is not complete. We hope that the model outlined here gives some useful insights into human vision.

VII. IMPLICATIONS AND CONCLUSIONS

Some conspicuous local features called textons have been identified by psychological means. These textons, particularly the elongated blobs, are quite similar to features found to stimulate the simple neural units in the striate cortex of the monkey, which are selectively tuned to elongated blobs of certain colors, orientations, width, and length.

Our findings, that in preattentive vision objects are distinguished only through their texton decompositions, might be of considerable

importance. Since in preattentive vision these textons are not coupled, and furthermore the resolution of texton properties—i.e., the perceptual threshold for color, width, length, and orientation differences—is rather limited, the number of distinguishable textons is within practically useful bounds. (For example, the width of periodic bars can be judged with an error of 4 to 6 percent,²¹ while accuracy of bar orientation is measured to be only 6 degrees of arc.²²) This limitation makes practical the devices that simulate preattentive vision. This contrasts with attentive vision for which virtually an infinite number of recognizable patterns exist whose biological, social, or intellectual interest to the observer is unknown. Whether additional textons will be discovered remains to be seen. But as long as they remain independent of the previously isolated textons, the model outlined here will not be importantly affected.

The main implication of our findings is as follows: A considerable amount of vision is carried out by the preattentive system whose workings appear to be much simpler than that of the attentive system. This is important in judging the information requirements of the human visual system realistically. Furthermore, it is important to realize that even in the attentive mental state, with all its prodigious processing powers, complex feats of form recognition are restricted to a small spatial aperture, often as small as a few minutes of arc. Also, changing the position or extent of the aperture of focal attention requires considerable time. The shortest time is about 50 ms when eye movements are prevented, and as long as about 200 ms if saccadic eye movements are necessary.

This dichotomy between preattentive and attentive mental states, the first limited in its power of information processing, the latter limited in its spatial extent, gives a model of human vision that could be exploited in visual communication. Here we do not want to invent specific methods, but only indicate some obvious possibilities. With the advent of fast, perhaps parallel computers, the textons that direct the human observer's attention could be simultaneously extracted by hardware. Detailed images need only be presented in such areas.

Also, one could program computers to extract local features other than textons. For instance, a parallel computer might rapidly detect the difference between an L and a T, rather than between a + and an L. If an observer's attention were directed by such a machine, whose capabilities are very different from human preattentive vision, perhaps a new way of inspecting the visual environment could be made available and possibly learned.

The textons reported here help to discriminate textures, mainly surfaces of objects, without the need of complex familiarity cues. Such an early separation of the visual environment into figure and ground,

or objects and their backgrounds, is a fundamental operation of visual perception. Lack of understanding of this process is, as of now, the greatest bottleneck in machine vision, which in turn is necessary in extending the capabilities of robots.

Regardless of the feasibility of such ambitious schemes, the finding that texton differences can be almost instantaneously perceived over large areas of the visual field can be practically exploited in traffic signs and in directing attention to select areas of visual displays. Traditionally, flickering or static colored lights have been used as traffic signs, or in instrument panels. Now we can add other texton classes—for instance, gaps to increase the terminator number—to enhance visibility. For example, in Fig. 17 we show how a single gap introduced in the conventional alphabet draws attention to the word STOP, which otherwise would require a long time to be segmented and detected. Such slight modification of the alphanumeric characters (amounting to a new “font”) might be beneficial in improving legibility. For instance, dyslexic children—children who cannot distinguish well between similar characters with different symmetric transformations such as b, d, or p—might greatly benefit if a gap or stroke were added to one of the characters, so that all characters would differ in at least one texton.

It should be stressed that the textons of preattentive vision only draw attention to certain areas, and we do not claim that these same textons are also the building blocks of form vision. If they were, our findings would prove preattentive vision to be the basis of attentive vision. Even if textons are restricted to vision in the preattentive state, we feel that to know those conspicuous features that grab our attention, wherever they appear, is of interest to everyone who wants to communicate through visual means.

VIII. ACKNOWLEDGMENTS

We thank our many colleagues who contributed to this research effort throughout the years, and who are mentioned in Section II and in the references. We also thank Max V. Mathews for his helpful comments while reading the manuscript. We are indebted to Walter



Fig. 17—Demonstration that the introduction of textons into the alphabet (here through increasing the terminator numbers by adding a gap) can help to segment and detect certain areas in a dense letter array.

Kropfl who developed the display hardware, and to Peter Burt who wrote the GENTEX program permitting the rapid generation, display, and manipulation of texture arrays. We thank our summer student, Franklin Schmidt, for developing the TEXTONS program.

REFERENCES

1. B. Julesz, "Textons, the Elements of Texture Perception, and Their Interactions," *Nature*, 290 (March 12, 1981), pp. 91-7.
2. B. Julesz, E. N. Gilbert, and J. D. Victor, "Visual Discrimination of Textures with Identical Third-Order Statistics," *Biol. Cybernetics*, 31 (1978), pp. 137-40.
3. B. Julesz et al., "Inability of Humans to Discriminate Between Visual Textures that Agree in Second-Order Statistics—Revisited," *Perception*, 2 (1973), pp. 391-405.
4. B. Julesz, "Experiments in the Visual Perception of Texture," *Sci. Am.*, 232 (April 1975), pp. 34-43.
5. B. Julesz, "A Theory of Preattentive Texture Discrimination Based on First-Order Statistics of Textons," *Biol. Cybernetics*, 41 (1981), pp. 131-8.
6. D. H. Hubel and T. N. Wiesel, "Receptive Fields and Functional Architecture of Monkey Striate Cortex," *J. Physiol.*, 195 (1968), pp. 215-43.
7. B. Julesz, *Foundations of Cyclopean Perception*, Chicago: University of Chicago Press, 1971.
8. D. Marr, "Early Processing of Visual Information," *Philos. Trans. R. Soc. London Ser. B*, 275 (1976) pp. 483, 524.
9. T. Caelli and B. Julesz, "On Perceptual Analyzers Underlying Visual Texture Discrimination: Part I," *Biol. Cybernetics*, 28 (1978), pp. 167-75.
10. J. R. Bergen and B. Julesz, "Discrimination with Brief Inspection Times," *J. Opt. Soc. Am.*, 71, No. 12 (December 1981), p. 1570. Also see "Parallel Versus Serial Processing in Rapid Pattern Discrimination," *Nature*, 303 (June 23-29, 1983), pp. 696-8.
11. B. Julesz, "Binocular Depth Perception of Computer-Generated Patterns," *B.S.T.J.*, 39 No. 5 (September 1960), pp. 1125-62.
12. E. Averbach and G. Sperling, "Short-term Storage of Information in Vision," in C. Cherry (ed.) *Information Theory, Fourth London Symposium*, London: Butterworth, 1961, pp. 196-211.
13. B. Julesz and J. J. Chang, "Interaction Between Pools of Binocular Disparity Detectors Tuned to Different Disparities," *Biol. Cybernetics*, 22 (1976), pp. 107-19.
14. G. Sperling and M. J. Melchner, "Visual Search, Visual Attention and the Attention Operating Characteristics," in J. Requin (ed.) *Attention and Performance VII*, Hillsdale, NJ: Erlbaum, 1978, pp. 675-86.
15. S. M. Zeki, "The Functional Organization of Projections From Striate to Prestriate Cortex in the Rhesus Monkey," *Cold Spring Harbor Symposia on Quantitative Biology*, 15 (1976), pp. 591-600.
16. R. Held et al., "Locating and Identifying: Two Modes of Visual Processing," *Psychol. Forsch.*, pp. 44-62; 299-348 (1967-1968).
17. C. B. Trevarthen, "Two Mechanisms of Vision in Primates," *Psychol. Forsch.* 31 (1968), pp. 299-337.
18. J. E. Hoffmann, "Hierarchical Stages in the Processing of Visual Information," *Perception and Psychophysics*, 18 (1975), pp. 348-54.
19. A. Treisman and G. Gelade, "A Feature-Integration Theory of Attention," *Cognitive Psychol.*, 12 (1980), pp. 97-136.
20. B. Julesz, "Visual Pattern Discrimination," *IRE Trans. Inform. Theory*, IT-8 (February 1962), pp. 84-92.
21. F. W. Campbell, J. Nachmias, and J. Jukes, "Spatial-Frequency Discrimination in Human Vision," *J. Opt. Soc. Am.*, 60 No. 4 (April 1970), pp. 555-9.
22. J. P. Thomas and J. Gille, "Bandwidths of Orientation Channels in Human Vision," *J. Opt. Soc. Am.*, 69, No. 5 (May 1973), pp. 652-60.
23. M. Rosenblatt and D. Slepian, "Nth Order Markov Chains With Any Set of N Variable Independent," *J. Soc. Indust. Appl. Math.*, 10 (1962), pp. 537-49.
24. T. Caelli, B. Julesz, and E. N. Gilbert, "On Perceptual Analyzers Underlying Visual Texture Discrimination: Part II," *Biol. Cybernetics*, 29 (1978), pp. 201-14.
25. B. Julesz, "Spatial Nonlinearities in the Instantaneous Perception of Textures with Identical Power Spectra," *Phil. Trans. R. Soc. Lond. B*, 290 (1980), pp. 83-94.
26. B. Julesz, "Perceptual Limits of Texture Discrimination and Their Implications to

- Figure-Ground Separation," in E.L.J. Leeuwenberg and H.F.J.M. Buffart (eds.), *Formal Theories of Visual Perception*, New York: Wiley, 1978, pp. 205-16.
27. H. L. Frisch and F. H. Stillinger, "Contribution to the Statistical Geometric Basis of Radiation Scattering," *J. Chem. Phys.*, 38 (1963), pp. 2200-7.
28. N. Wiener, "Extrapolation, Interpolation and Smoothing of Stationary Time Series, With Engineering Applications," New York: Cambridge University Press.

APPENDIX

It required two decades of research efforts to discover that preattentive texture perception depends on local features alone and that global higher-order statistical parameters can be ignored. In 1962, Julesz asked mathematicians to generate stochastic texture pairs that would be identical in their first ($n-1$)th order statistics, but different in the n th- and higher than n th-order statistics.²⁰ The n th-order statistics are similar to the well-known n th-order joint probability distribution of n samples. The n samples are n points of a texture selected at random. However, in random geometry the shape of the n samples is of importance.

These n points can be regarded as the vertices of an n -gon. The n -gon (or n th-order) statistics are obtained when these n points (having the same n -gon shape) are selected at random, and statistics indicate that these n points have certain color values. For instance, the second-order statistics can be obtained if a 2-gon (dipole, or needle) is randomly thrown at the texture and the probability is determined that the two end-points of the dipole—of given lengths and orientations—fall on certain color combinations: e.g., black and black; or black and white; or black and gray, etc.

In the intervening years many such stochastic textures were discovered, particularly with identical first- and second-order statistics, but different third- and higher-order statistics.^{3,23-25} As a matter of fact, the texture pairs in Figs. 3-6, and 8-10 have this property. The finding that many of these *iso-second-order* texture pairs differing only in third- and higher-orders are indistinguishable suggests that the preattentive visual system cannot compute statistical difference beyond the second order. The recent finding by Julesz, demonstrated in Fig. 11a, suggests that the preattentive visual system cannot even process second-order statistical parameters.⁴ From the second-order statistics the autocorrelation function can be uniquely determined—as a matter of fact, for two-tone textures composed of black and white dots, the second-order (dipole) statistic is the autocorrelation function^{26,27}—and the Fourier transform of the autocorrelation is the Fourier power spectrum.²⁸ Therefore, all the texture pairs with identical second-order statistics also have identical power spectra. The finding that texture segregation can be obtained in *iso-second-order* textures, after it was established that the preattentive system cannot process third-order statistics (and, as Fig. 11a demonstrates, not even second-order statis-

tics), implies that this segregation must be based on *local* density differences. Finally, it was proposed that the density changes of certain local conspicuous features, the textons, explain preattentive texture discrimination.^{1,25}

AUTHORS

James R. Bergen, A.B. (Mathematics and Psychology), 1975, University of California, Berkeley; Ph.D. (biophysics and theoretical biology), 1981, University of Chicago; Postdoctoral Fellow, Bell Laboratories, 1981-82; RCA, 1983—. Mr. Bergen's work concerns the quantitative analysis of information processing in the human visual system. At the University of Chicago he was involved in the development of a model of the spatial and temporal processing which occurs in the early stages of the system. At Bell Laboratories, his work has concentrated on the effect of visual system structure on the extraction of information from a visual image. Mr. Bergen recently joined RCA Laboratories in Princeton, NJ.

Bela Julesz, Diploma, 1950 (Electrical Engineering), Technical University, Budapest; Ph.D., 1956, Hungarian Academy of Sciences; Bell Laboratories, 1956—. Mr. Julesz taught and did research in communications systems for several years prior to 1956. Since joining Bell Laboratories, he has devoted himself to visual research, particularly depth perception and pattern recognition. He is the originator of the random-dot stereomage technique and of the method of studying texture discrimination by constraining second-order statistics. He has written extensively in the area of visual and auditory perception and is the author of *Foundations of Cyclopean Perception*. Mr. Julesz was Head of the Sensory and Perceptual Processes Department from 1964 to 1982, and in 1983 was made Head of the Visual Perception Research Department. He has been visiting professor of experimental psychology at M.I.T. and other universities. In February 1983 he received the MacArthur Prize Fellow Award in Experimental Psychology and Artificial Intelligence. He was a Fairchild Distinguished Scholar at the California Institute of Technology from 1977 to 1979. Fellow, AAAS, OSA, and the American Academy of Arts and Sciences; Corresponding Member of the Goettingen Academy of Sciences.

