Integrating the second term by parts yields,

$$T = \left[\int_0^{\infty} t(F_1 - F_4)(F_2 - F_3) - \int_0^{\infty} (F_1 - F_4)(F_2 - F_3)dt. \right.$$

Again, if finite variances are assumed, the first term is equal to zero. Since $F_1 < F_4$ and $F_2 < F_3$, the second integral is positive and hence $T < 0$. It should be noted that the expressions derived allow for calculations of T and $E(M_{11}) - E(M_{12})$ if the cumulative distribution functions for the components are known or can be estimated.

REFERENCES

Atkinson, R. C., Holmgren, J. E., & Juola, J. F. Processing time as influenced by the number of elements in a visual display. *Perception and Psychophysics*, 1969, **6**, 321–326.

Biederman, I. Perceiving real-world scenes. *Science*, 1972, **177**, 77–80.

Dondieri, D. C., & Zelnicker, D. Parallel processing in visual same–different decisions. *Perception and Psychophysics*, 1969, **5**, 197–200.

Egeth, H. E., Jonides, J., & Wall, S. Parallel processing of multielement displays. *Cognitive Psychology*, 1972, **3**, 674–698.

Eichelman, W. H. Familiarity effects in the simultaneous matching task. *Journal of Experimental Psychology*, 1970, **86**, 275–282.

Estes, W. K. Interactions of signal and background variables in visual processing. *Perception and Psychophysics*, 1972, **12**, 278–285.

Estes, W. K., & Wessel, D. L. Reaction time in relation to display size and correctness of response in forced-choice visual signal detection. *Perception and Psychophysics*, 1966, **1**, 369–373.

Holmgren, J. E., Juola, J. F., & Atkinson, R. C. Response latency in visual search with redundancy in the visual display. *Perception and Psychophysics*, 1974, **16**, 123–128.

Kinchla, R. A. Detecting target elements in multielement arrays: A confusability model. *Perception and Psychophysics*, 1974, **15**, 149–158.

Krueger, L. E. Search time in a redundant visual display. *Journal of Experimental Psychology*, 1970, **83**, 391–399.

Pollatsek, A., Well, A. D., & Schindler, R. M. Familiarity facilitates the visual processing of letter strings. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, **1**, 328–338.

Reicher, G. M. Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 1969, **81**, 275–280.

Sternberg, S. The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and performance II: Acta Psychologica*, 1969, **30**, 276–315.

Sternberg, S., & Knoll, R. L. The perception of temporal order: Fundamental issues and a general model. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic Press, 1973.

Treisman, A. M. Strategies and models of selective attention. *Psychological Review*, 1969, **76**, 282–299.

Wolford, G. S. Perturbation model for letter identification. *Psychological Review*, 1975, **82**, 184–199.

Wolford, G. S., Wessel, D. L., & Estes, W. K. Further evidence concerning scanning and sampling assumptions of visual detection models. *Perception and Psychophysics*, 1968, **3**, 439–444.

REFERENCE NOTE

1. Sternberg, S. *Scanning a persisting visual image versus a memorized list*. Paper presented at the Eastern Psychological Association, 1967.

COGNITIVE PSYCHOLOGY **9**, 353–383 (1977)

Forest Before Trees: The Precedence of Global Features in Visual Perception

DAVID NAVON

University of Haifa, Haifa, Israel

The idea that global structuring of a visual scene precedes analysis of local features is suggested, discussed, and tested. In the first two experiments subjects were asked to respond to an auditorily presented name of a letter while looking at a visual stimulus that consisted of a large character (the global level) made out of small characters (the local level). The subjects' auditory discrimination responses were subject to interference only by the global level and not by the local one. In Experiment 3 subjects were presented with large characters made out of small ones, and they had to recognize either just the large characters or just the small ones. Whereas the identity of the small characters had no effect on recognition of the large ones, global cues which conflicted with the local ones did inhibit the responses to the local level. In Experiment 4 subjects were asked to judge whether pairs of simple patterns of geometrical forms which were presented for a brief duration were the same or different. The patterns within a pair could differ either at the global or at the local level. It was found that global differences were detected more often than local differences.

The Principle of Global Precedence

Do we perceive a visual scene feature-by-feature? Or is the process instantaneous and simultaneous as some Gestalt psychologists believed? Or is it somewhere in between? The Gestaltists' view of the perceptual system as a perfectly elastic device that can swallow and digest all visual information at once, no matter how rich it is, is probably too naive. There is ample evidence that people extract from a picture more and more as they keep looking at it (e.g., Helson & Fehrer, 1932; Bridgen, 1933; Yarbus, 1967). But does this mean that interpreting the picture is done by integrating information collected in a piecemeal fashion? Is the perceptual whole literally constructed out of the percepts of its elements?

This paper is based on parts of a doctoral dissertation submitted to the Department of Psychology, University of California, San Diego. The research was supported by Grant No. NS07454 from the National Institutes of Health. I am greatly indebted to Lynn Cooper, David Rumelhart, and many members of LNR research group at UCSD for lots of useful comments. Special thanks to Donald Norman for his guidance, really valuable suggestions, comments, and careful review of many drafts. I also thank Donald Broadbent, Ralph Haber, and Yaakov Kareev for useful comments on a draft. Requests for reprints should be addressed to: David Navon, Department of Psychology, University of Haifa, Haifa 31999, Israel.

ISSN 0010-0285

My approach to the problem is, in a sense, in the tradition of the early studies of *Aktualgenese* (see review in Flavell & Draguns, 1957). The idea put forward in this paper is that perceptual processes are temporally organized so that they proceed from global structuring towards more and more fine-grained analysis. In other words, a scene is decomposed rather than built up. Thus the perceptual system treats *every* scene *as if* it were in a process of being focused or zoomed in on, where at first it is relatively indistinct and then it gets clearer and sharper.

Some Definitional Framework

The interpreted contents of a scene can be viewed as a hierarchy of subscenes interrelated by spatial relationships (cf. Winston, 1973; Palmer, 1975). The decomposition of a scene into parts, each of which corresponds to exactly one node of the hierarchical network, is conceivably done in accordance with some laws of Gestalt such as proximity, connectedness, good continuation, and so forth (see Palmer, Note 1). This statement does not explain the process of decomposition, but it gives an idea of its product. The globality of a visual feature corresponds to the place it occupies in the hierarchy: The nodes and arcs at the top of the hierarchy are more *global* then the nodes and arcs at the bottom. The latter are said to be more *local*. We cannot claim, however, that one visual feature is more global than another one, unless we know that both correspond to actual nodes in the network. The operational test is to try to construct networks in which feature "x" will dominate feature "y" (or vice versa). If this can be done in just one direction, it can be argued that *if* "x" and "y" constitute perceptual units, *then* "x" must be more global than "y" (or vice versa). For an experimental test of the principle of global precedence one should use as stimuli figures in which the spatial hierarchy is intuitively transparent.

It is claimed that processing of a scene proceeds from the top of the hierarchy to the bottom; that is to say, it is *global-to-local*. It follows that the global features of a visual object that is within an observer's effective visual span (i.e., none of its parts is either viewed peripherally or below the threshold of visual acuity) will be apprehended before its local features.

As an example, consider how a picture like the one in Fig. 1 may be processed. The structuring of the picture that comes first in the process of perception is something like L (blob-1, frame), where blob-1 is the region where the figure is, and L is the spatial relationship that holds between blob-1 and the overall frame. At this point more processing effort is directed to the analysis of blob-1, so that the structure of the scene is refined into something like L (R[blob-2, blob-3], frame), where blob-2 is the region of the crescent, blob-3 is the region of the star, and R is the spatial relationship of those two. During the next stage more effort

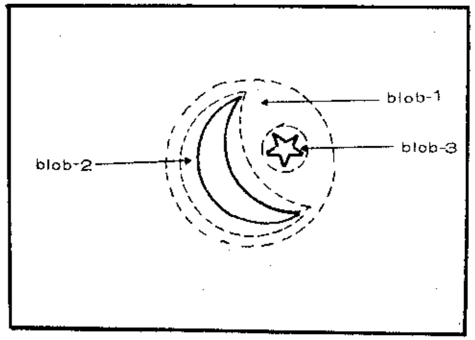

FIG. 1. An example of a simple picture and its parts.

goes into differentiating blob-2 and blob-3 until they are recognized as a crescent and a star, respectively, so that the final structure is: L (R[crescent, star], frame).

The view presented here does not amount to a distinction between stages of attention (cf. Neisser, 1967). It is, rather, a claim about perceptual analysis of whatever is *attended to*. Note that perceptual processing is viewed as a unified process, in that both the "where" and the "what" questions are answered while the scene is structured. Spatial organization is treated as a sort of crude figural analysis which sometimes may even be sufficient for recognition.

Functional Importance of Global-to-Local Processing

In most real situations the task of the human perceptual processor is not just to account for given input but also to select which part of the surrounding stimulation is worth receiving, attending to, and processing. The constraints imposed by the optical limits of our eyes and by the nature of the surroundings have a twofold implication for the processing structure in the visual domain. One, the resolution of most of the stimuli in the picture plane (or the largest part of their visible surface) is low by default. The crude information extracted from the low-resolution parts of the visual field should be used for determining the course of further processing. Two, in an ecology where uncertainty is the rule, there is little to be gained from being set for a particular type of input. The system should be flexible enough to allow for gross initial cues to suggest the special way for processing a given set of incoming data. These two observations suggest that a multipass system, in which fine-grained processing is guided by prior cursory processing, may be superior to a system that tries to find a coherent structure for all pieces of data simultaneously.

One important function of the first pass is that of locating the stimuli, an obvious prerequisite for any figural analysis. Note, however, that finding the location of stimuli provides the system with some very global figural information.

Since perception is basically dynamic, there is often only time for partial analysis because of the constant change in input. In that case a rough idea about general structure is more valuable than few isolated details. Furthermore, often we do not place the same importance on every portion of the input. It was found (Yarbus, 1967; Mackworth & Bruner, 1970) that people tend to spend more time on the more informative sectors of the picture. In this case we ought to have the results of some initial gross analysis in order to determine which part of the field is likely to bear more on our behavior or thinking. But even when we focus on an important sector of the field, we may not need to build a very elaborate structure for it. Details are detected only to the degree that they are essential for determining contents.

As pointed out by Palmer (1975) and Norman and Bobrow (1976) and supported by much empirical evidence, perceptual processing must be both input-driven and concept-driven. That is, the activity in the system is triggered by the sensory input but is guided by expectancies formed by context and early indications from sensory data. Thus, perception is regarded as a two-way process: The hypotheses about what a stimulus may be interact with what the stimulus actually is in determining what the stimulus is finallly perceived to be. Now, since local features often serve as constituents in more global structures, the identification of the global features is a very useful device of narrowing down the range of candidates for accounting for a certain local region. Moreover, as pointed out by Palmer (1975, Figure 11.6), sometimes the identification of a part of the picture merely on the basis of its own features is almost impossible, yet it can be easily recognized within the appropriate context. In general, the more definite the output of global analysis, the more concept-driven local analysis is, so the less effort has to be expended in overcoming deficiencies in data quality. This is a substantive advantage in view of the limited acuity of the visual sensory mechanisms.

The point that spatial organization precedes interpretation of details is essential for resolving ambiguities stemming from rotation, projection, and interposition. By the time a particular stimulus in the field is interpreted, the more global processes have generated a hypothesis about the angle from which that stimulus is viewed. Thus, the expected object is not *any* instance of a category, but rather an actual object as seen from a certain point of view and probably partly concealed either by other objects or by parts of itself.

Some Empirical Evidence

It is well supported that perceiving the whole facilitates the perception of its parts.

The word–letter phenomenon (Reicher, 1969; Wheeler, 1970) is an excellent demonstration of how the mere presence of a higher-level perceptual unit improves later forced-choice recognition of its individual constituents over the case when they are presented alone (or just being focused upon, as shown by Johnston and McClelland, 1974). Whether it is the discriminability or the codability of the constituents which is enhanced is yet an open question.

Selfridge (in Neisser, 1967, p. 47) illustrated how the same pattern can be interpreted as two different letters depending on the context. In many cases it appears that the perceptual system ignores details that are inconsistent with the interpretation indicated by the context or even completes features that are missing in the actual scene. Pillsbury (1897) demonstrated how readers may not be disturbed at all by omission or substitution of letters in texts they read. Warren (1970) reported about a similar effect in speech perception. Huey (1908) and recently Johnson (1975) provided some more examples in this vein. Palmer (1975a) showed that interpretation of ambiguous elements of a picture tends to conform to the semantic structure of the whole scene, even when it involves some distortion or deletion of few details.

In those examples, the perception of the global unit or the overall theme is more veridical than the perception of the elements. However, note that in all of these examples (excluding, perhaps, the word–letter phenomenon) the whole is more predictable than the elements, especially when the target element is incongruent to some extent with the rest. Therefore, it is not clear whether the veridical identification of the whole is due to very potent extraction of global features or to highly redundant inference made on the basis of a sample of local features. (See Johnson, 1975; Rumelhart & Siple, 1974).

What is the evidence that global features or relationships are perceived *first*? First, there are indications that people can take advantage of peripheral information. The subjects in an experiment by Williams (1966) were able to utilize peripherally viewed size or color to direct their search for targets. Rayner (1975) showed that readers seem to perceive the gross shape of peripherally viewed words to the right of the word being fixated at the time. Since peripheral information must be of low resolution then to the extent that recognition is aided by peripheral cues, the precedence of the gross features falls out.

Second, even within the angular span that can be perceived with high acuity in just one fixation, there seems to be progression with exposure time from very gross global perception to very fine-grained recognition. In many early studies of the development of percepts (comprehensively reviewed in Flavell & Draguns, 1957) subjects were presented with visual stimuli for very short durations. The general finding is that as the duration of exposure got longer, subjects progressed from perceiving just the location of the object, through differentiating figure and ground, then to some inaccurate apprehension of the global form, and finally to good figural sensation. In another experiment (Navon, Note 2) subjects were presented for a brief duration with a picture of a clock with Greek letters as hour markers and arms in sleeves as hands. Whereas all of them identified the clock correctly, recognition of the details was below chance level. Thus, not only is the perception of global structure earlier than detailed figural analysis, but it is often sufficient for identifying an object or a scene with a fair amount of confidence.

Another relevant finding comes from motion perception. Since motion perception has to keep up with the continual change in the visual field, it must be affected mostly by those properties of the visual stimuli that are processed first. It was found (Navon, 1976) that in situations of ambiguous apparent motion, figural identity of the elements did not have any effect on determining the type of motion experienced, whereas more global features did.

On the developmental level, Meili–Dworetzki (1956) had children of different ages respond to several ambiguous figures in which whole and parts suggest different interpretations (e.g., a man made out of fruit). She found that children perceived wholes at an earlier age than parts. On the other hand, Elkind, Koegler, and Go (1964) devised a set of figures that produced the opposite effect. The source of conflict in those findings resides, obviously, in the stimuli. Ambiguous figures may vary in the relative plausibility of their alternative interpretations. If young children tend to make just one decision about a stimulus, then they are likely to overlook the duality of ambiguous figures and to detect just the more salient aspect, whichever the case may be.

It seems, thus, that the general problem with the experimental treatment discussed so far is lack of proper control over the stimulus material. Global and local structures may differ in *complexity, salience, familiarity, recognizability,* or *relative diagnosticity* for determining the identity of the whole, and they do differ in some of these properties in all the studies mentioned so far. Hence, the two major principles of the experimental attack I used were: (a) control of all these properties of global and local features; and (b) independence of global and local features, so that the whole cannot be predicted from the elements and vice versa.

GLOBAL PRECEDENCE IN A DIFFUSE-ATTENTION SITUATION: EXPERIMENTS 1 AND 2

The best way to equate the properties of global and local features is to use stimuli in which the set of possible global features is identical with the set of possible local ones. For this purpose I constructed large letters that were made out of small letters (see Fig. 5A). When one looks at

these stimuli in normal viewing conditions, one cannot miss either the identity of the whole stimuli or the fact that they are made out of letters whose identities are also definitely recognizable.

I constructed a task in which visual perception is slightly restricted both by visibility conditions and by limited attention (or using the terminology of Norman & Bobrow, 1975, the quality of the data and the availability of processing resources). My prediction was that in such a situation subjects' performance will be insensitive to the figural identity of the local features. Perceptual awareness was measured by means of an indirect method: an intermodality Stroop task. Stroop tasks are named after Stroop (1935) who found that when subjects have to name the color of an ink in which a word is written, their responses are inhibited when the word is a name of color different from the ink color.

Subjects were asked to respond to a name of a letter while looking at a visual stimulus of the type shown in Fig. 5A. The rationale was that recognition of a visual stimulus may interfere with discrimination of an equivalent auditory stimulus or with performing the appropriate responses for it. Experiment 1 was done in order to test the validity of this rationale and to determine the optimal parameters for such auditory–visual interference. One of the problems with Stroop tasks, which is especially pronounced in an intermodality task, is ensuring that the subject is actually exposed to the secondary channel or aspect of the stimulus. I have taken several measures in order to do that, and they are described below.

Experiment 1

Method

Apparatus. The equipment consisted of a display Tektronix oscilloscope with a fast decay phosphor (decays to 90% in .63 msec), a typewriter keyboard, a Krohn–Hite 3550R filter, a Shure microphone mixer, two Hewlett–Packard 350D attenuator sets, a chin rest, and a pair of headphones. Auditory and visual stimuli were generated and controlled by a PDP-9 computer equipped with digital-to-analog converters. The subject sat alone in an acoustically isolated booth in front of a table wearing the headphones; his chin was on a chin rest and his hands were on the keyboard. The display oscilloscope was positioned in front of the subject at eye level. Viewing distance was 50 cm. The intensity of the picture on the oscilloscope was adjusted so that when plotting a test square with side of 13 mm containing 51 × 51 dots, the luminance of the square was 1.37 cd/m². The room illumination was such that the luminance of the periphery of the screen was .65 cd/m².

Design and procedure. The major characteristics of the experimental task are schematized in Fig. 2. The temporal structure of the stimuli is presented in Fig. 3. The subject listened through the headphones to a sequence of utterances of equal duration evenly spaced in time.

The utterances could be either of the names of the letters H and S (namely "ach" and "es") sequenced at random with equal probabilities. While listening to the sequence of auditory stimuli, the subject was monitoring the oscilloscope. A visual stimulus of the set shown in Fig. 2A could either be flashed on the oscilloscope or not (randomly

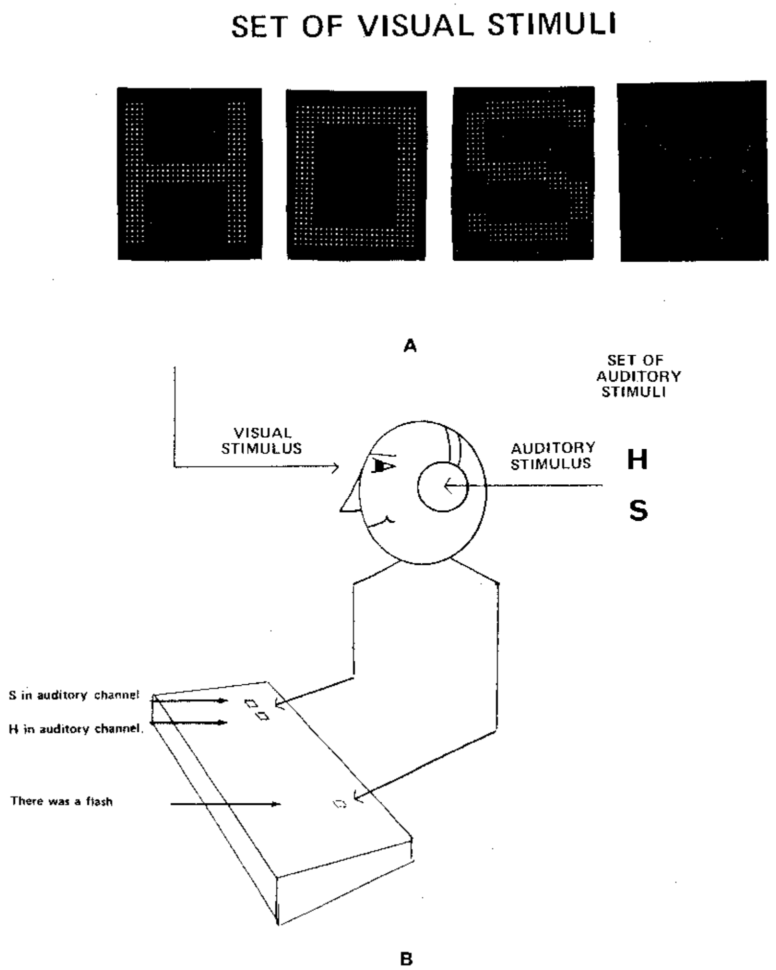

FIG. 2. A schematization of the experimental task in Experiment 1.

with probability .5) in very close temporal proximity to the auditory stimulus. The primary task of the subject was to indicate after each utterance which of the two letters he had heard by depressing either of two keys with the second or third finger of his right hand. The secondary task was to respond by depressing a key with his left hand to the appearance of any visual stimulus regardless of what it was. The subject was instructed to perform his response to the visual display only after he had made the auditory discrimination response. It was also emphasized that although he only had to respond to the *presence* of the visual stimulus, he would be questioned about the identity of the visual stimulus later. He was not told before the experiment what the visual stimuli would look like.

Each trial was preceded by a short warning beep. Simultaneously with the onset of the beep a square frame with a side slightly larger than the longer side of the visual stimuli appeared on the screen and persisted until the scheduled offset time of a visual stimulus, regardless of whether or not such a stimulus was actually presented. The warning signals

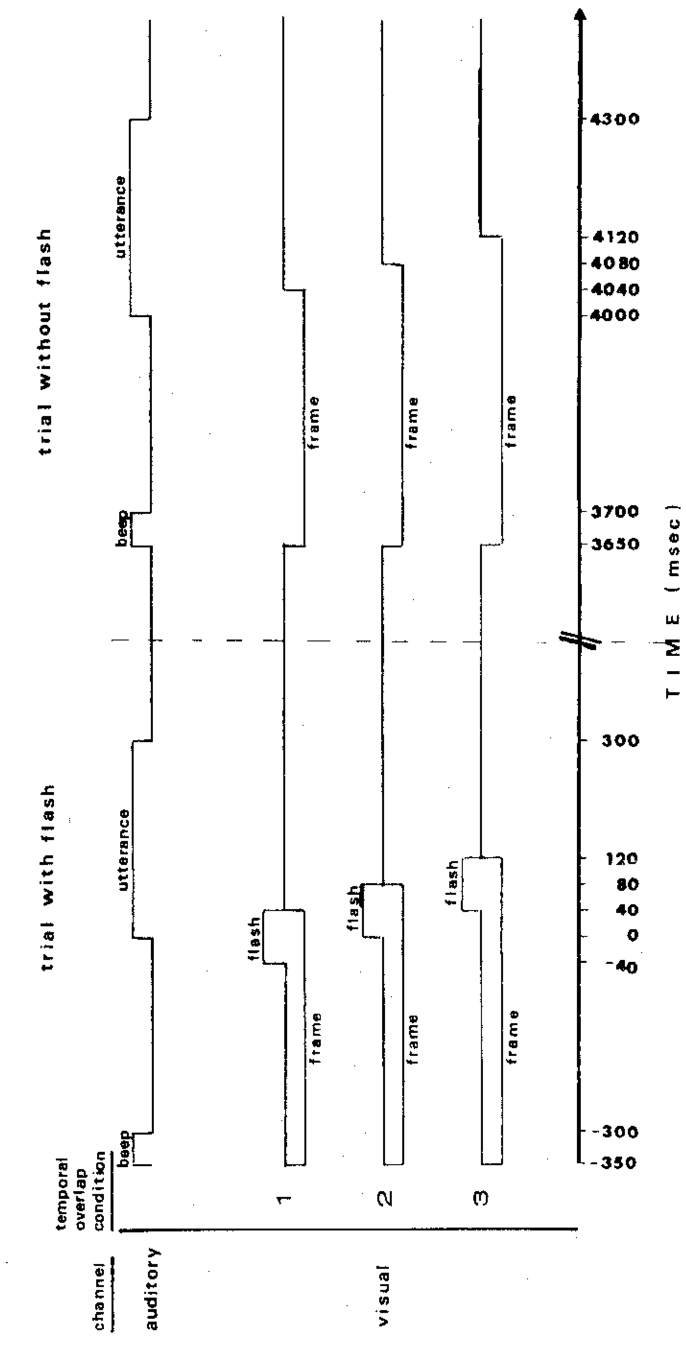

FIG. 3. A diagram of the temporal structure of the stimuli in Experiment 1.

served to minimize both the temporal and spatial uncertainty of the subject with regard to the stimuli. Accuracy and latency for both the auditory discrimination response and the visual detection response were recorded. Latency was measured from the start of the auditory stimulus.

Each subject was run individually in one session of 504 trials. After every block of 72 consecutive trials the subject was given a rest period of about 40 sec. The necessary randomizations were done for each block independently. The auditory stimuli and the presence or absence of visual stimuli were factorially crossed in each block, as were the visual stimuli with the auditory ones in those trials where visual stimuli appeared. The visual stimulus was said to be *consistent* with the auditory stimulus if they both consisted of the same letter; it was said to be *conflicting* with the auditory stimulus if they were different letters (namely an H and an S); and it was said to be neutral with regard to the auditory stimulus if it consisted of a rectangle.[1] It falls out that these levels of consistency were randomized and balanced in each block.

There were three conditions of temporal overlap between the auditory and the visual stimuli. As seen in Fig. 3, the exposure duration of the visual stimuli was always the same, but the delay of their onset with respect to the start of the auditory stimuli was either −40, 0, or 40 msec. The overlap conditions were administered in different blocks. The order of the administration of the conditions for half of the subjects was: 1, 2, 3, 1, 2, 3, in blocks 2 through 7, respectively. For the other half, the reverse order was used. The first block was considered to be practice and the temporal overlap used during it was identical to the one in condition 2.

Stimuli. The auditory stimuli were generated in the following way: Each stimulus was uttered by a male native English speaker, and its wave form was digitized by the PDP-9 computer at a sampling rate of 10 khz. It was stored in digitized form and converted back into its analog form and played to the subject whenever needed. The quality of the sound and the signal-to-noise ratio were sufficient to preclude any possibility of acoustic confusion.

The longest vertical diameter of each of the visual stimuli was 28 mm, thus subtending about 3° 12′ visual angle with viewing distance of 50 cm. The side of the square frame was 33 mm, thus subtending about 3° 47′ visual angle.

Subjects. Eight subjects were used, all undergraduates at the University of California, San Diego, who participated in the experiment as part of their course requirement. The subjects were also paid a monetary bonus that depended heavily on accuracy for both auditory discrimination and visual detection responses and slightly on the speed of the first one. The subjects were asked to try to be as fast as they could without making errors at all. All had normal vision or fully corrected vision.

Results and Discussion

The error percentages for the individual subjects were mostly between 1 and 2% and never exceeded 4%. Latencies for incorrect responses did not depart much from latencies for correct responses. There was no indication of a speed–accuracy tradeoff. Thus, it was decided to use in the analysis all the latency data for both correct and incorrect responses.

Letters were considered as a random factor in the analysis and preliminary tests were performed to determine for each systematic source whether or not its interaction with the letters factor and its triple interaction with the letters and subjects factors, should be included in the error term. (See Winer, 1971, pp. 378–384). In all the tests described here and later in the paper, those interaction terms were found nonsignificant, thus the error term for each within-subject source was its interaction with subjects.

[1] All the subjects in this experiment and the following one referred to the rectangle in their later verbal description as the letter O.

The mean latencies to auditory discrimination responses for each of the three temporal overlap conditions are plotted in Fig. 4 as a function of the consistency between the auditory and the visual input. The effect of consistency on latency is highly significant, $F(2,12) = 39.09$; $p < .001$. The agreement among individual subjects with respect to the order of mean latencies for the different levels of consistency is very high (Kendall coefficient of concordance of .89). The different conditions of temporal overlap not only vary with regard to their effect on mean latency, $F(2, 12) = 26.30$; $p < .001$, but also interact with the variable of consistency,

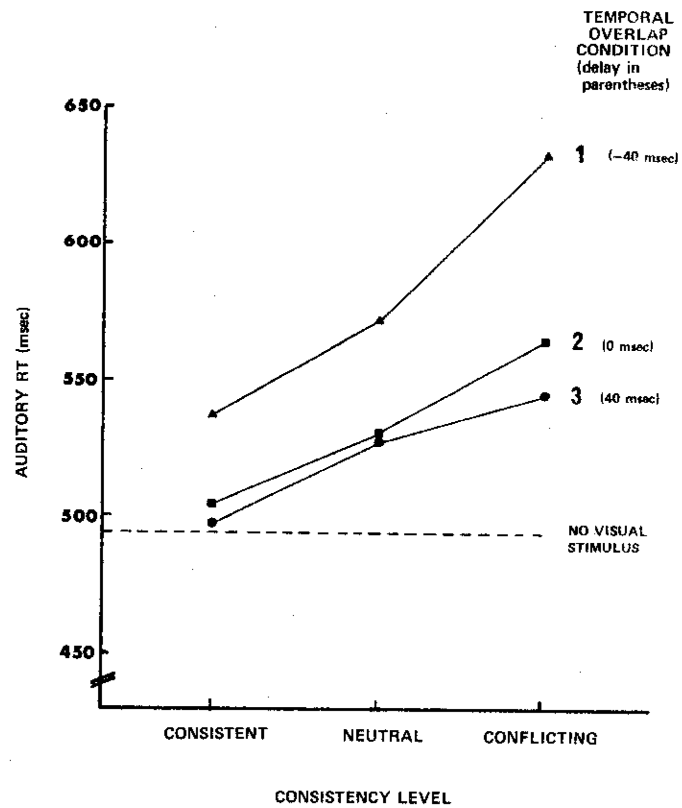

FIG. 4. Mean auditory latencies in Experiment I as a function of consistency level and temporal overlap condition. The delay is of the visual stimulus with respect to the auditory stimulus.

TABLE 1

MEAN LATENCIES (MILLISECONDS) IN EXPERIMENT 1 TABULATED BY CONSISTENCY LEVELS AND TEMPORAL OVERLAP CONDITIONS

Temporal overlap condition (delay in parentheses)	Consistency: Consistent		Neutral		Conflicting	Total
1 (−40)	538	≪	572	≪	633	581
	≫		≫		≫	≫
2 (0)	505	<	531	≪	565	534
	=		=		=	=
3 (40)	498	≪	528	=	547	524
Total	514	≪	544	≪	581	

Note. Results of posthoc pairwise comparisons are represented by equality (=) and inequality (<) signs. An equal sign denotes nonsignificant comparison. One inequality sign denotes comparison significant to the .05 level. Two inequality signs denote comparison significant to the .01 level.

$F(4,24) = 3.45$; $p < .025$. A closer inspection of the data by means of Newman–Keuls procedure for post hoc comparisons (See Table 1) suggests that the differences between all pairs of levels of consistency are significant and that there is a significant difference between the negative delay of the flash and the two other delays. The difference between the last two is not significant. It also appears that the effect of consistency is stronger when the auditory stimulus starts after the onset of the flash.

Experiment 2

Method

This experiment has the main characteristics of Experiment 1 with regard to procedure, apparatus, and setting. The major difference is in the set of visual stimuli. Two sessions were administered, a test session and then a control session. The visual stimuli used in the test session were a large H, S, or rectangle. These *global* characters were made out of *local* characters: small Hs, Ss, or rectangles. The shapes of the local and global characters were identical. The arrangement of the centers of the small characters making up a large one was the same as the arrangement of the dots making up the small characters proper. The set of actual stimuli is shown in Fig. 5A.

The global level was factorially crossed with the local level, and each visual stimulus was presented twice with each auditory stimulus, within each block. The test session consisted of one practice block and three test blocks. The control session included two parts of two blocks each. In the first part the stimuli used were the ones used in Experiment 1. The stimuli used in the second part (See in Fig. 5B) were just single-element characters of the stimuli from the test session presented at the center of the display field. The square frame used for delimiting the field of the visual stimulus was smaller during the second part of the control session. The ratio between its size and the size of the stimuli was the same as in the test session and in the first part of the control session.

The temporal overlap between the auditory and the visual stimuli was identical to that in condition 1 of Experiment 1: The flash was 80 msec long and was turned on 40 msec before the start of the auditory stimulus.

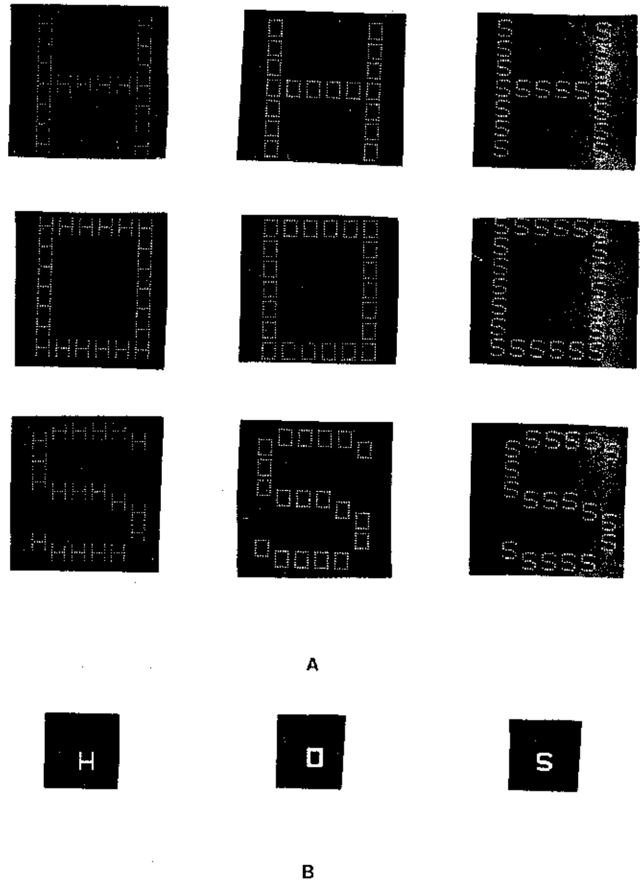

FIG. 5. The set of stimuli used in the test session of Experiment 2 is presented in A. The set of stimuli used in the second part of the control session is presented in B.

Eighteen subjects were run, none of whom had served in Experiment 1. The *size* of the visual display was varied between subjects. For nine of them the size of the whole stimulus was the same as in Experiment 1; the size of the element characters (or of the whole stimulus in the second part of the control session) was 1/8 of that size. The other nine subects were presented with stimuli that were 1.5 as large as the respective stimuli the first nine subjects were presented with.

Results

The same type of analysis as in Experiment 1 was applied. The data of two subjects, one from each size condition, were eliminated because their error rates were too high (greater than 4%).

The mean latencies to auditory discrimination responses during the test session are plotted in Fig. 6 as a function of the consistency between the auditory input and both the global level and the local level of the visual

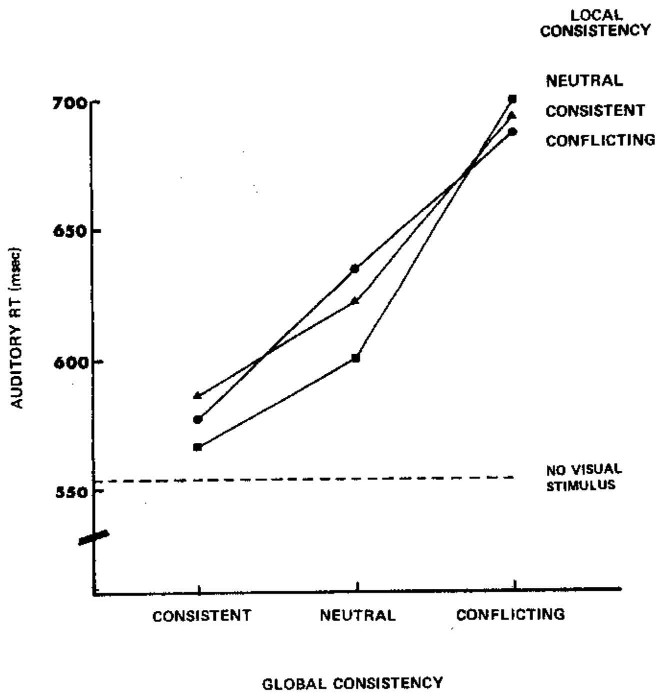

FIG. 6. Mean auditory latencies in the test session of Experiment 2 as a function of global consistency and local consistency.

input. The effect of consistency of the global level with the auditory stimuli is highly significant, $F(2,28) = 94.90; p < .001$, whereas the consistency of the local level does not have any significant effect, $F(2,28) = 1.64; p > .20$.

The agreement among individual subjects with respect to the order of mean latencies is very high for the global level but very low for the local ones (Kendall coefficients of concordance were .94 and .09, respectively). The post hoc pairwise comparisons between the levels of global consistency done by means of Newman–Keuls procedure were all significant at the .01 level.

No other factor in the design, including size, was found to be significant, except for the factor of subjects and for two quadruple interactions.

The mean latencies to auditory discrimination responses during each of the parts of the control session are plotted in Fig. 7 as a function of the consistency between the auditory and the visual input. The effect of consistency is highly significant, $F(2,28) = 103.74; p < .001$, and it does not interact with the type of stimuli, $F(2,28) = 0.17$. The effect of the type of stimuli proper fell short of significance, $F(1,14) = 2.89; p > .10$. The agreement among individual subjects with respect to the order of mean

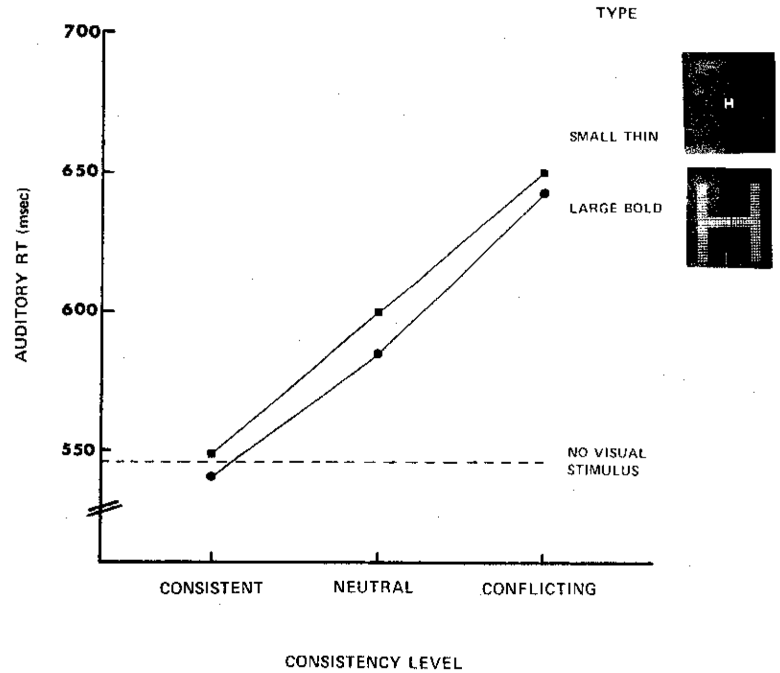

FIG. 7. Mean auditory latencies in the control session of Experiment 2 as a function of consistency level and type of stimuli.

latencies was high for either of the two types (Kendall coefficients were .79 for the larger bold type and .74 for the small thin type). The post hoc pairwise comparisons were all significant at the .01 level. The factors of subjects and letters were found significant as well as five interactions involving at least one of these factors.

The difference between overall mean latency in the test session and overall mean latency in the control session was nonsignificant, $F(1,15) = 1.21$; $p > .25$, for trials without visual input but significant, $F(1,15) = 10.44$; $p < .01$, for trials with visual input. The significant drop in average latency from test to control may be due to the fact that the subjects may have been more practiced during the control session since it was administered after the test session. Another possibility is that as the visual stimuli in the control session were simpler, they required less processing, thus enabling earlier completion of auditory processing and response selection.

Discussion

The interference effect applied just to the global visual pattern and not at all to the elements of which it was made. The same effect holds for stimuli as small as the elements when they stand alone in the visual field. Hence it is not the smaller size of the elements per se that makes them relatively or absolutely unnoticed.

The results of this experiment suggest that there are situations in which visual processing is carried just to a limited depth. The global pattern is apprehended but not its components. All but three subjects did not even notice that the stimuli were made of small letters. When asked after the experiment was over, they said that the stimuli may have been made of dots or of blocks.

There seems to be no sensory limit responsible for the superficial account for the visual data; the second part of the control session of this experiment, as well as the results of a later experiment, Experiment 3, indicate that one can voluntarily attend to the local features as well. The visual system seems rather to have made a decision to neglect the processing of the elements in view of the structure of the task, although it might have had enough capacity for performing a more thorough analysis. I believe that such economy in processing effort is very characteristic of human vision. Since usually the informative contribution of local features over the information gained by processing global structures is small or negligible, whereas the reverse is not true, the system will quit after a scene is interpreted on the global level.

THE INEVITABILITY OF GLOBAL PROCESSING: EXPERIMENT 3

We have already seen that in certain conditions when no specific demands are made to the viewer with regard to what should be recognized, his visual system is tuned to pick just the identity of the global pattern. What happens, however, when the viewer is told what to focus at and what to ignore? Can the viewer control his own perceptual processes? And if he does, can he ignore any aspect at will?

The basic experimental idea is to use visual stimuli of the type used in Experiment 2, this time without any auditory stimuli, and to have subjects respond either just to the global level or just to the local level. The assumption is that if they are aware of the other level too, it should interfere with their performance. However, if subjects were allowed to look at the stimuli as long as they want, they probably would not be able to concentrate well on either of the levels uniquely. Therefore, I used brief presentation with postexposure masking. In addition, because when subjects know exactly where the stimulus will appear they might focus on a spot that contains just one element, I introduced spatial uncertainty into the task.

Method

Apparatus. The apparatus was the same as in Experiments 1 and 2, without the acoustical equipment.

Stimuli. This experiment used the set of visual stimuli from the test session of Experiment 2 (see Fig. 5A) in the small size condition.

Design and procedure. In each trial of the experiment the subject saw a stimulus of the type presented in Fig. 5A. Each subject was run through two attention conditions: In the *global-directed* condition the subject was supposed to indicate whether the global character was H or S. (In this condition stimuli 2, 5, and 8 from Fig. 5A were not used.) In the *local-directed* condition the subject was supposed to indicate whether the element character was H or S. (In this condition stimuli 4, 5, and 6 from Fig. 5A were not used.) Each trial was preceded by a 50-msec warning beep. The beep started simultaneously with the onset of a fixation point at the center of the oscilloscope that remained in the field for 500 msec. The stimulus followed the offset of the fixation point immediately, and it appeared randomly at either of the four quadrants of the oscilloscope. The distance between the center of any of the stimuli and the fixation point was 24 mm. The stimulus was presented for 40 msec and was immediately masked by a square of 33 × 33 dots.[2] The mask remained on the oscilloscope until the subject depressed either of two keys with his second or third finger of his right hand to indicate his response. The subject's response initiated the generation of the next trial which started about 3 sec later. Both accuracy and latency measured from the onset of the stimulus were recorded for each trial.

Each subject was run individually in one session. There were 288 trials in a session. After every block of 36 trials the subject was given a rest period of about 20 sec. The necessary randomizations were done for each block independently. Each of the six possible stimuli appeared six times in a block in a random fashion, and each spatial position at which the stimulus could appear was used nine times in a block in a random fashion. One attention condition was administered in the first six blocks, and the other one in the last six blocks. Half of the subjects received the global-directed condition first, and the other half received it last. The first two blocks for each condition were considered as practice blocks.

Subjects. Fourteen subjects were run. All of them were undergraduates at the University of California, San Diego, who participated in the experiment as part of their course requirement. The subjects were also paid a monetary bonus that depended heavily on accuracy and slightly on speed. The subjects were asked to be as fast as they could while making as few errors as possible. None of the subjects had participated in Experiments 1 or 2, and all had normal vision or fully corrected vision.

Results

The overall error percentage was 3.3%. However, unlike the earlier experiments, the pattern observed in latency data was apparent in error percentages as well. Therefore, only latencies for correct responses were analyzed. Mean latencies to correct responses are plotted in Fig. 8 as a function of the attention condition and the consistency of the stimuli in the attended and supposedly unattended levels. The factors of attention condition, consistency, and their interaction were found to have significant effects. The F values for these factors were $F(1,12) = 855.85$ ($p < .001$), $F(2,24) = 72.48$ ($p < .001$), and $F(2,24) = 16.59$ ($p < .001$), respectively.

Post hoc pairwise comparisons between the different levels of consistency were performed by means of Newman–Keuls procedure for each

[2] In fact, the mask was not strictly immediate because the dots were sequentially plotted at a rate of 10 μsec per dot.

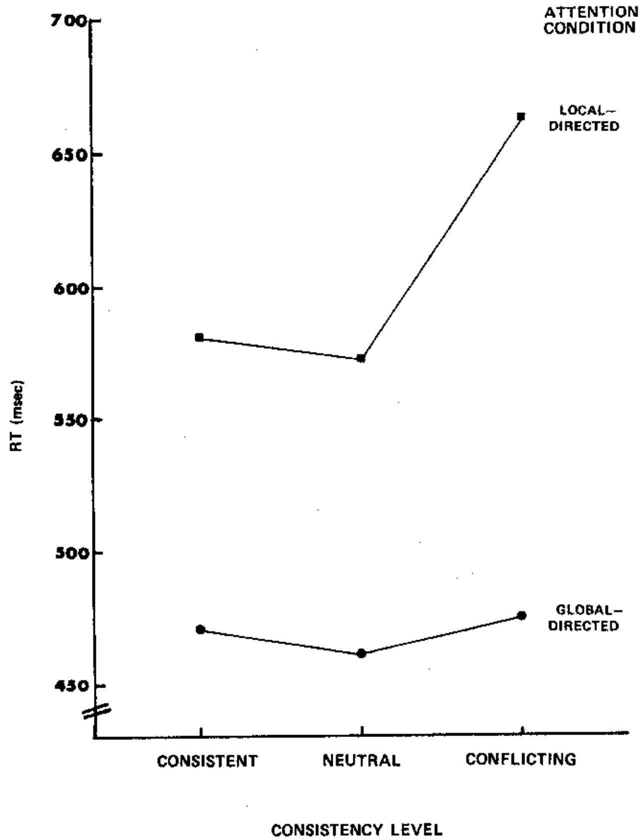

FIG. 8. Mean latencies in Experiment 3 as a function of consistency level and attention condition.

of the attention conditions. The mean latencies in the neutral, consistent, and conflicting consistency conditions were 573, 581, and 664 msec, respectively, for the local-directed attention condition and 462, 471, and 477 msec, respectively, for the global-directed one. All the differences are nonsignificant at the .05 level, except for the ones between the conflicting consistency condition to each of the other ones when attention is directed at the elements, $p < .01$. The agreement among individual subjects with respect to the order of the mean latencies for the different levels of consistency was high for the local-directed condition (Kendall coefficient of .78) but very low for the global-directed condition (Kendall coefficient of .12). In the local-directed condition the mean latency in "conflicting" trials was higher than the mean latency in "consistent" trials for each of the 14 subjects. (The Binomial probability of this occurring by pure chance is .0001). In the global-directed condition the same relationship holds for just eight of the subjects (Binomial probability of .395).

Since the difference in accuracy rate between the two attention conditions was very small (1.8%) and nonsignificant, $F(1,12) = 2.86$; $p > .1$, the results cannot be attributed to poor resolution of the elements. Although this experiment lacks a size control of its own, an indirect control is supplied by the results of the control session of Experiment 2 (see Fig. 7): If size per se had had an effect, the smaller visual stimuli should probably have delayed more the responses to the auditory stimuli and/or interfered less with them, but they do not.

The order of the attention conditions interacts significantly with both attention condition, $F(1,12) = 47.59$; $p < .001$, and consistency, $F(2,24) = 7.70$; $p < .005$, but not with the interaction of the latter two. Neither interaction is easy to interpret.

Discussion

The results of this experiment indicate that the global pattern is responded to faster than the elements. Moreover, whereas people can voluntarily attend to the global pattern without being affected by the local features, they are not able to process the local features without being aware of the whole. The subjects' ability to restrict their attention to the whole suggests either that they can "turn off" recognition processes after the categorization of the global pattern has been completed, or that perceptual processes proceed but have no effect on the response based on the results of global analysis, or that the subjects can voluntarily degrade the quality of the local sensory data (for example, by changing the focus of the eye lense). Either way, the finding that attention cannot be efficiently *diverted* from the whole may be interpreted as a support to the notion that global processing is a necessary stage of perception prior to more fine-grained analysis.

All the experiments reported thus far demonstrate the potence and vividness of the global percept. Nevertheless, the local features could still be processed if a deliberate attempt were made to do so. The next experiment purports to examine whether or not there are conditions under which local features are less likely to be processed (or alternatively, processed less thoroughly) than global features, although both levels are equally critical for performance.

GLOBAL PRECEDENCE IN NEAR-THRESHOLD CONDITIONS: EXPERIMENT 4

If processing of global features precedes processing of the local ones, then at a certain moment after the onset of a visual stimulus, a viewer must have an idea of the gross structure yet no idea of the finer details. If we managed to capture his percept at that moment, we would be able to test the global precedence hypothesis. Certainly, there are difficulties stemming from the fact that, although we may be reasonably successful in cutting the lines of input by replacing the stimulus with a masking field, there is no guarantee that we can eliminate all unprocessed or half-processed "traces." Since the perceptual system may be parallel to some degree, the processes of local analysis may already have begun before the global analysis is completed. However, even if this is the case, the order of processing should be reflected in higher likelihood of error in local analysis than in global analysis.

The question calls for a task that requires the subject to place equal importance on both levels. It also requires critical duration of exposure that is not long enough to enable good perception of both levels, but not short enough to disrupt any perception at all. The stimuli should not be too complex, because the effects must be attributed to perceptual processing rather than to memory. The shapes of the global pattern and of the local elements must be equally complex and all the elements in a given stimulus must be the same. It is highly important to make sure that any effect is not due to factors such as size or spatial uncertainty.

One experimental task that meets all those requirements is the making of the "Same–Different" judgments on pairs of simple patterns of geometrical forms. The possible global patterns were chosen to be very similar in all respects to the possible local elements.[3] The patterns of a pair could differ *either* on the local level *or* on the global level.

Method

Design and procedure. The experimental task was to make "Same–Different" judgments on pairs of patterns of the type presented in Fig. 9. Each pattern consisted of nine squares grouped in three clusters of three squares each. The squares were arranged within each cluster so that their centers of gravity fell on the vertices of an imaginary isosceles triangle pointing upwards, downwards, to the right, or to the left. The same is true for the spatial arrangement of the three clusters with respect to each other: The imaginary squares circumscribing the clusters could relate to each other in precisely the same ways as the squares within a cluster, only three times magnified.

In a given pattern the particular arrangement of the squares within the clusters (namely the orientation of the imaginary isosceles triangles) was identical across the three clusters. This will be called the *local* configuration. The local configuration was varied orthogonally to the particular spatial arrangement among the three clusters. The configuration formed by the three clusters will be called the *global* configuration.

There were 16 patterns in total. The two patterns presented on a trial were the same 50% of the time. When they were different from each other, the difference was either in the global configuration (25% of the time) or in the local configuration (25% of the time), but not in both levels. Examples for the three possible combinations are shown in Fig. 9.

The subjects were carefully instructed about the shape of the patterns and they were told that they should respond "same" only in case of perfect congruence between the two

[3] This was true for the stimuli of Experiment 1 and 2, too. However, there the distance between two adjacent elements was somewhat arbitrary. As will be seen, in the stimuli of this experiment even the spatial relationships of the components were identical for the global pattern and the elements.

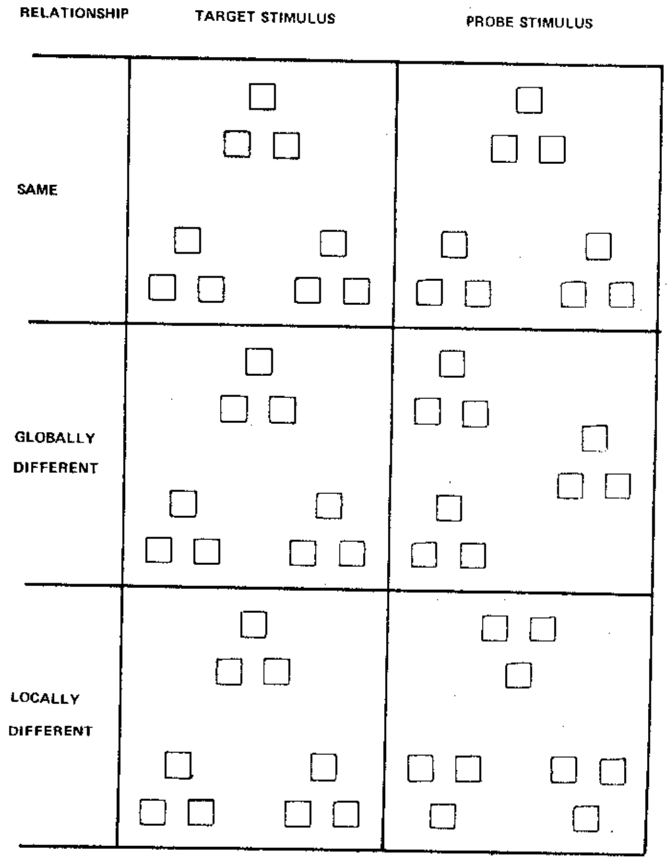

FIG. 9. Examples of pairs of patterns shown in each of the three types of trials in Experiment 4.

patterns. Each subject was run through a practice session, a test session, and a control session. Each session consisted of 192 trials. The practice session and the test session were completely identical with respect to procedure, only the practice session served to determine for each subject a presentation time for the test session that would be likely to produce an overall rate of 70–80% correct responses. For this purpose six different presentation durations were used during the practice session in decreasing order and the percentage correct for each duration was computed. The data from the practice session were not otherwise analyzed.

During the control session the procedure was the same, only the figures to be compared were clusters of just three squares in either of the four arrangements they could have within the nine-squares patterns. Two different conditions were administered in the control session: *Spatial Certainty*, where the clusters appeared always at the center of the field where the patterns appeared during the test session; and *Spatial Uncertainty*, where the clusters appeared randomly at either of the eight positions in which they could fall during the test session.

Several variables were varied between subjects. There were two modes of presentation: *sequential mode*, in which a target stimulus was presented for a fixed brief duration followed by a probe stimulus terminated by the subject's response; and *simultaneous mode*, in which both stimuli were presented at the same time for a fixed brief duration.

There were two conditions of readiness for the onset of the display: *temporal certainty*, in which the subjects were given both an auditory warning and a fixation point before each trial; and *temporal uncertainty*, in which subjects did not get any warning and the inter-trial intervals were Poisson-distributed.

In the sequential mode the size of the patterns was varied: the *large size* was three times as large as the *small size*, so a single square of the large size had the same side as the imaginary square circumscribing a cluster of the small size. In the simultaneous mode only small size patterns were used because the display oscilloscope could not accommodate two large patterns, but the orientation of the two patterns to each other was varied. In the *horizontal comparison* condition the patterns were presented side by side. In the *vertical comparison* condition they were presented one above the other. In both conditions the space between the two figures was the size of a figure. Examples of the displays in each of the conditions and sessions are presented in Fig. 10.

The temporal structure of a trial in each of the conditions is schematized in Fig. 11. Both accuracy and reaction time were recorded. In the sequential mode, reaction time was measured from the onset of the probe stimulus. In the simultaneous mode, reaction time was measured from the onset of the postexposure mask.

Each subject was run individually. The subject was dark adapted for 5 min, then the practice session began followed by the test session and then the control session. The subject was given a rest period of about 1 min after every block of 32 consecutive trials. In every block, each pattern appeared twice as a target stimulus, once in a Same-trial, and once in a Different-trial. (In the simultaneous mode the right pattern or the upper pattern was considered as the target.) The order of presentation within a block was randomized. A given

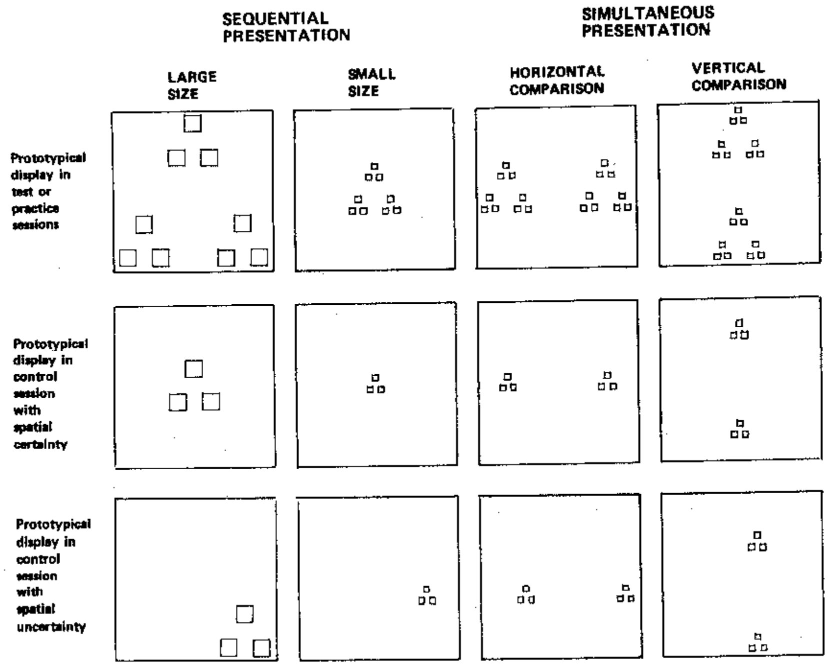

FIG. 10. Examples of the displays in each of the conditions and sessions of Experiment 4.

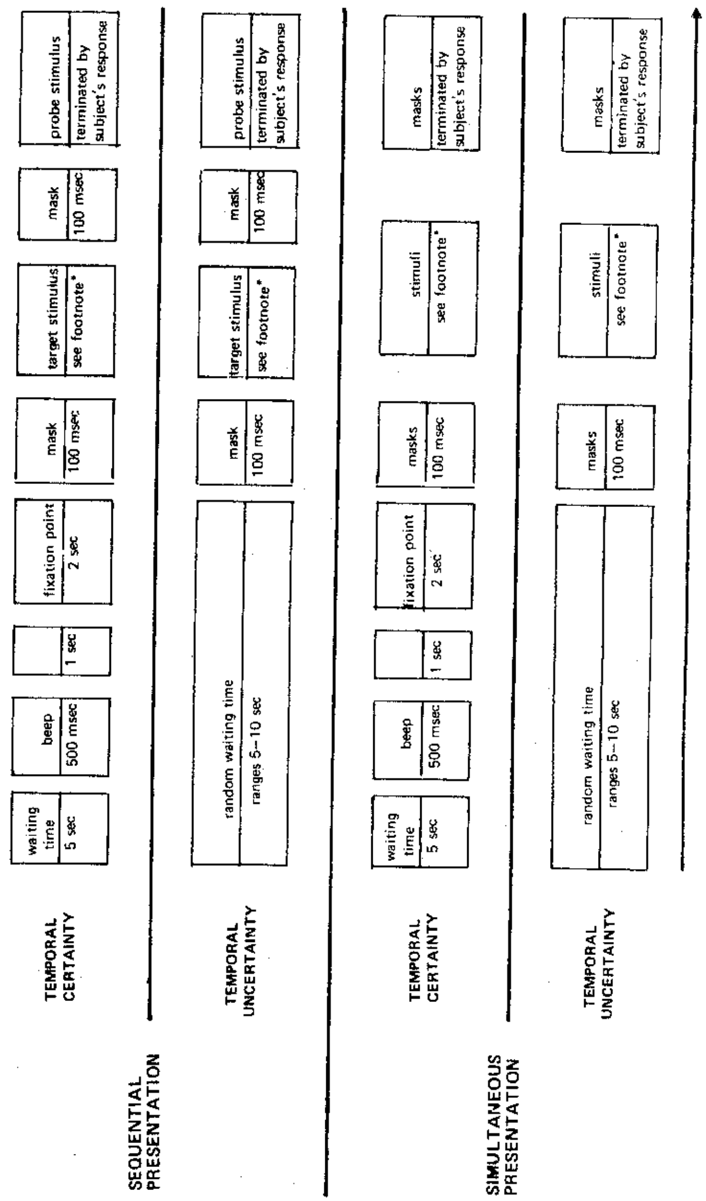

FIG. 11. A schematized diagram of the temporal structure of a trial in both modes of presentation and both conditions of temporal certainty in Experiment 4. The exposure time was adjusted for each subject individually to achieve an overall level of about 70–80% correct. The durations actually used during the test and control sessions ranged from 48 to 100 msec in the sequential mode and from 100 to 300 msec in the simultaneous mode. During the practice session 6 different exposure durations were used (140, 120, 100, 80, 60, and 40 msec in the sequential mode, and 400, 330, 260, 190, 120, and 50 msec in the simultaneous mode).

pattern was paired in different blocks with different patterns as probe stimuli, so that across the six blocks of a session each pattern was shown twice with each of the patterns from which it differed on just one aspect, either globally or locally: once as a target and once as a probe. In the control session a block consisted of 24 consecutive trials, and each cluster appeared as a target once in a Same-trial and once in a Different-trial in each sequence of eight consecutive trials. In every block each cluster was shown twice with each of the other clusters, once as a target and once as a probe. The condition of spatial certainty was administered in odd-numbered blocks, and the condition of spatial uncertainty was administered in even-numbered blocks.

Apparatus. The equipment was the same as in Experiment 3. The subject sat in front of a keyboard. Two different keys were assigned for "Same" and for "Different" responses. Half of the subjects used their dominant hand (two were left-handed) to respond "Same" and the others used it to respond "Different." The display oscilloscope was positioned in front of the subject at about eye level. Viewing distance was about 100 cm. The side of a square of the small size was about 2.4 mm, thus subtending about 0° 8′ visual angle, and it was plotted as 10 equally spaced dots. The side of the imaginary square circumscribing a cluster was three times as long, and the side of the imaginary square circumscribing a whole pattern was nine times as long. The masking stimulus contained about 3000 dots and the side of a masking square was 21 mm.[4] The large size patterns were three times as long. The illumination of the booth was dim to avoid reflection of the subject's image in the oscilloscope's screen. The intensity of the picture on the oscilloscope was adjusted so that, when plotting a test square with side of 13 mm containing 51 × 51 dots, the luminance of the square was 1.46 cd/m² and the luminance of the periphery of the screen was .53 cd/m².

Subjects. Sixteen subjects were run, eight in each mode of presentation. They were assigned evenly to all the combinations of the between-subjects treatments. All the subjects were undergraduates at the University of California, San Diego, who participated in the experiment as part of their course requirement. The subjects were also paid a monetary bonus that depended on accuracy and slightly on speed. The subjects were asked to be as accurate as they could and, in case they were not confident, to make their best guess based on what they perceived. All of them reported normal vision with or without the aid of correcting lenses.

Results

Data from the two modes of presentation, sequential and simultaneous, were analyzed separately.

Accuracy data. The data analysis was done by an analysis of variance performed on arcsine square root transforms of the proportions of correct responses in each cell (See Winer, 1971). Same-trials were excluded from the major analysis because they were irrelevant for the purpose of the experiment.

The proportions of correct "Different" responses in the sequential mode were .88 for global difference and .72 for local difference, $F(1,4) = 23.16$; $p < .01$. The corresponding proportions in the simultaneous mode were .80 and .54, $F(1,4) = 12.75$; $p < .025$. The agreement among individual subjects is very high: The proportion of detecting global differences is higher than the proportion of detecting local differences for all the eight subjects in the sequential mode (the Binomial probability of this occurring by chance is .004) and for seven of the eight subjects in the simultaneous mode (Binomial probability of .035).

[4] It should be noted that the mask was not strictly immediate because the dots were sequentially plotted at a rate of 10 μsec per dot.

None of the between-subject variables (readiness for onset of trial, size of stimulus, or orientation of comparison) had any main effect. In the sequential mode the effect of type of difference did not interact with any other factor of the design. In the simultaneous mode one triple interaction that was hard to interpret was found to be marginally significant, $F(1,4) = 8.24$; $p < .05$.

The effect of the amount of difference between the two patterns was also inspected. The amount of difference within levels of globality was defined in terms of the angular disparity between the units being different in a given trial (either the clusters or the global configuration). The proportions of correct detections of the difference for the three levels of angular disparity and both types of difference are presented in Table 2. For the sequential mode neither the factor of amount of difference proper nor its interactions with any other factor are significant.

For the simultaneous mode, however, the main effect is significant, $F(2,8) = 6.70$; $p < .025$, and it also interacts significantly, $F(2,8) = 6.44$; $p < .025$, with the orientation of comparison, so that it exists in the side-by-side display but not in the vertical one. The results suggest that at least in the simultaneous mode a rotation of 180° produces a more discernible difference than a rotation of 90° in either direction ($p < .01$ in the post hoc pairwise comparisons by the Newman–Keuls procedure). It may well be, although it is not reliably indicated by the data, that this is true for rotation within clusters, whereas global disparity is likely to be detected regardless of its "degree."

Results in the control session. To show that the greater accuracy for global features is not due just to effects of size, the performance in the

TABLE 2

THE PROPORTIONS OF CORRECT "DIFFERENT" RESPONSES IN EXPERIMENT 4 TABULATED ACCORDING TO MODE OF PRESENTATION, TYPE OF DIFFERENCE, AND ANGULAR DISPARITY BETWEEN THE PROBE STIMULUS AND THE TARGET STIMULUS

Angular disparity (measured in degrees clockwise from target to probe)	Mode of presentation			
	Sequential Type of difference		Simultaneous Type of difference	
	Global	Local	Global	Local
90	.86	.68	.78	.52
180	.89	.78	.81	.61
270	.89	.69	.80	.49

test session should be compared to the performance in the control session. The idea is to compare detection of differences within clusters when a cluster is embedded within a larger configuration to detection of differences when the cluster stands alone. However, in different patterns, the clusters may appear in different locations, so one might argue that, while the subject always knows where to look for the global features, he is somewhat uncertain about the exact location of the local features, so he has to spend some time to locate them before processing. Therefore, the control session included both conditions of spatial certainty and spatial uncertainty. The spatial uncertainty in the control session was actually even larger than the spatial uncertainty in the test session (3 bits vs 1.37 bits).

Another problem is what index to use for comparing the test with the control. The percentages of correct detections of differences, global and local in the test and overall in the control, could have been used under the assumption that the subject had used the same decision criterion in both sessions. If he had not, he could gain, for example, more correct detections of differences in the control session just by making more false detections as well. However, this assumption seemed too risky. Therefore, it was decided to compare the proportion of total correct responses in either condition of the control session to the proportion of the sum of correct "Same-trials" and correct *global* difference detections in the test session, on the one hand, and to the proportion of the sum of correct "Same-trials" and correct *local* difference detections in the test session on the other hand. This way the decision criterion does not matter. Those proportions for both modes of presentation are presented in Table 3.

TABLE 3

COMPARISON OF ACCURACY MEASURES IN THE TEST AND CONTROL SESSIONS OF EXPERIMENT 4 FOR BOTH MODES OF PRESENTATION

Proportion of correct responses in:	Mode of presentation	
	Sequential	Simultaneous
Test session		
Same-trials plus global different-trials	.76	.73
Same-trials plus local different-trials	.71	.66
Control session		
Spatial certainty all trials	.82[a]	.88[b]
Spatial uncertainty all trials	.75[c]	.80[d]

[a] Significantly higher than .76 [$\chi^2(16) = 35.24$; $p < .01$].
[b] Significantly higher than .73 [$\chi^2(16) = 59.84$; $p < .001$].
[c] Not significantly different from .76 [$\chi^2(16) = 15.09$; $p > .50$].
[d] Significantly higher than .73 [$\chi^2(16) = 42.72$; $p < .001$].

The statistical comparisons were done by means of z tests for proportions within subjects and then by a χ^2 for combining the z values across subjects (See Winer, 1971, pp. 49–50). The results clearly indicate that the poor detection of local differences cannot be accounted for by the smaller size of the local features, since accuracy in the control session is greater than accuracy with regard to the global structure in the test session.

Latency data. In the sequential mode the mean latency for local differences was 1126 msec and that for global differences was 935 msec, $F(1,4) = 4.95$; $p < .10$. If the between-subject variables are excluded from analysis, $F(1,7) = 5.72$; $p < .05$. That relationship holds for seven out of eight subjects. (The binomial probability of this occurring by chance is .035.) However, in the simultaneous mode there was no clear pattern. The mean latency for local differences was 1345 and that for global differences was 1315, $F(1,4) = 0.93$, and only five out of eight subjects show the same relationship. In sum, the latency data, especially from the simultaneous mode, are not very suggestive. It should be remembered that those latencies are measured when the target stimuli are no longer on, so they correspond to decision processes more than to perceptual processing. The data do not seem to teach us very much about either of those processes.

Discussion

The data indicate clearly that global differences are more frequently detected than local differences. Since both levels are critical for the subjects' judgments and since the criterion of discrimination between the possible stimuli is very clear, it is not plausible that the effect is due to decision processes. If the effect resides in the perceptual system, the conclusion must be that the global configuration is more likely to be perceived on brief exposures than the local pattern. Since on longer exposure one can see both, it follows that local analysis is done relatively late in the process.

GENERAL DISCUSSION

There is now much evidence about the existence of receptors or channels sensitive to particular spatial frequencies, that is, stimulation with particular spatial separations between its highly stimulated regions (See Campbell, 1974). Low frequency channels have some different properties from high frequency ones (see e.g., Broadbent, 1975). One might try to interpret the findings of the experiments reported in this paper as reflecting the relative speed of operation of the lower versus higher spatial frequency channels in the range of spatial frequencies (viz., sizes) used in these experiments. However, if this is true, the performance of subjects should be a function of *absolute* size. However, small elements were not processed less frequently, less rapidly, or less accurately than the larger ones, unless they were also more local. Thus, *across* patterns size had no effect. However, it is still true that *within* a given structure the global features were larger than the local ones. That confoundedness raises the question whether or not the findings can be interpreted in terms of *relative* size within a given pattern. In other words, one may conclude from the results that in a given scene, the *larger* features have priority in attracting attention.

To the extent that globality and relative size are ecologically correlated, it would just be a matter of personal choice, which term to use. However, one can conceive of patterns in which some local features are larger than some more global structures that do not contain them. For example, consider a stimulus made up of two patterns of the type in Fig. 9, one beside the other, in which the side of a square in the first one is four times as long as the side of a square in the second one. The local level (within-cluster arrangement) of the first one is larger than the global level (between-cluster arrangement) of the second one. Such stimuli seem to offer a proper testing field for weighing the plausibility of the two interpretations.

Another question is whether or not we are justified in generalizing from the size ranges used in the reported experiments. If the perceiver is close enough to the forest, he will probably see a tree rather than a forest. In this case, however, the tree is seen foveally whereas everything else is seen peripherally. Global precedence will presumably hold when both global and local levels have the same visibility, be it high or low. This is the more general case, because usually the objects we attend to subtend fairly small visual angles. For example, when you are reading these words, your retinas encompass a portion of a much larger scene, such as a desk, an opposite wall, and so forth, but this is the stationary ground; the bulk of visual processing effort is probably expended on figuring out what the words are. Thus, usually what we direct our visual attention on is small enough to be subject to global precedence.

No attempt was made here to formulate an operational definition of globabity of visual features which enables precise predictions about the course of perception of real-world stimuli. What is suggested in this paper is that whatever the perceptual units are, the spatial relationship among them is more global than the structure within them (and so forth if the hierarchy is deeper). Thus, I am afraid that clear-cut operational measures for globality will have to paiently await the time that we have a better idea of how a scene is decomposed into perceptual units. However, the difficulty in providing objective criteria for determining what in a real-world situation corresponds to a certain psychological term that does have an objective definition within the laboratory should not deter us from using such a term. (Think, for example, of the terms "stimulus," "reinforcement," "familiarity," "utility," and many others). For the time being, we can use some Gestalt laws and a pinch of common sense.

SUMMARY

In this paper, I propose that perception proceeds from global analysis to more and more fine-grained analysis. The global precedence has a number of possible advantages such as utilization of low-resolution information, economy of processing resources, and disambiguation of indistinct details. Although evidence from the psychological literature supports the notion that global features are extracted earlier and/or better than local ones, in most previous research little attention has been given to the control over the complexity of global and local features.

In the first two experiments reported in this paper, subjects were asked to respond to an auditorily presented name of a letter while looking at a visual stimulus that consisted of a large character (the global level) made out of small characters (the local level). The subjects' auditory discrimination responses were subject to interference only by the global level and not by the local level. These results demonstrate that when visual stimulation is attended to, but is not the unique source of information that bears on behavior, visual processing may stop when the gross features have been accounted for.

If global precedence is not an inherent property of visual perception but just a consequence of an ad hoc policy of allocating resources, then one should be able to voluntarily force himself to skip global analysis and process just the elements. To answer this question I performed Experiment 3, in which subjects were presented with large characters made out of small characters. They had to respond either just to the global level (the large characters) or just to the local level (the small characters). Whereas the identity of the local level had no effect on global recognition, global cues which conflicted with the local ones inhibited the response to the local level. This result shows that in some circumstances, while people can voluntarily attend just to the global level of a scene, they cannot skip global processing, thus the latter is a necessary stage of perception.

In Experiments 2 and 3 the local level was ignored or could be ignored; nevertheless, it could be processed if a deliberate attempt was made to do so. In Experiment 4 the exposure was short enough to make local processing very difficult even though the subject was motivated to obtain its output. The subjects were asked to judge whether pairs of simple patterns of geometrical forms which were presented for a brief duration were the same or different. The patterns within a pair could differ either on the global or on the local level. It was found that global differences were detected more often than local differences. This may be interpreted as a

support to the idea that global processing is done before more local analysis is completed.

In sum, this is a diverse set of findings, and along with previous findings it constitutes a body of evidence supporting the notion of global precedence.

REFERENCES

Bridgen, R. F. A tachistoscopic study of the differentiation of perception. *Psychological Monographs*, 1933, **44**, (1), 153–166.

Broadbent, D. E. Waves in the eye and ear. *Journal of Sound and Vibration*, 1975, **41**, 113–125.

Campbell, F. W. The transmission of spatial information through the visual system. In F. O. Schmitt & F. G. Worden (Eds.), *The neurosciences third study program*. Cambridge, MA: M.I.T. Press, 1974.

Elkind, D., Koegler, R. R., & Go, E. Studies in perceptual development II: Part–whole perception. *Child Development*, 1964, **35**, 81–90.

Flavell, J. H., & Draguns, J. A microgenetic approach to perception and thought. *Psychological Bulletin*, 1957, **54**, 197–217.

Helson, H., & Fehrer, E. V. The role of form in perception. *American Journal of Psychology*, 1932, **44**, 79–102.

Huey, E. B. *The psychology and pedagogy of reading*. The Macmillan Co., 1908 (New edition: Cambridge, MA: M.I.T. Press, 1968).

Johnson, N. F. On the function of letters in word identification: Some data and a preliminary model. *Journal of Verbal Learning and Verbal Behavior, 1975*, **14**, 17–29.

Johnston, J. C., & McClelland, J. L. Perception of letters in words: Seek not and ye shall find. *Science*, 1974, **184**, 1192–1194.

Mackworth, N. H., & Bruner, J. S. How adults and children search and recognize pictures. *Human Development*, 1970, **13**, 149–177.

Meili-Dworetzki, G. The developments of perception in the Rorschach. In B. Klopfer (Ed.), *Developments in the Rorschach technique*. New York: Harcourt, Brace, & World, 1956. Vol. II.

Navon, D. Irrelevance of figural identity for resolving ambiguities in apparent motion. *Journal of Experimental Psychology: Perception*, 1976, **2**, 130–138.

Neisser, U. *Cognitive psychology*. New York: Appleton–Century–Crofts, 1967.

Norman, D. A., & Bobrow, D. G. On data-limited and resource-limited processes. *Cognitive Psychology*, 1975, **7**, 44–64.

Norman, D. A., & Bobrow, D. G. On the role of active memory processes in perception and cognition. In C. N. Cofer (Ed.), *The structure of human memory*. San Francisco: Freeman, 1976.

Palmer, S. E. Visual perception and world knowledge: Notes on a model of sensory–cognitive interaction. In D. A. Norman, D. E. Rumelhart, & the LNR Research Group, *Explorations in Cognition*. San Francisco: Freeman, 1975.

Palmer, S. E. The effects of contextual scenes on the identification of objects. *Memory and Cognition*, 1975a, **3**, 519–526.

Pillsbury, W. B. A study in apperception. *American Journal of Psychology*, 1897, **8**, 315–393.

Rayner, K. The perceptual span and peripheral cues in reading. *Cognitive Psychology*, 1975, **7**, 65–81.

Reicher, G. M. Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 1969, **81**, 274–280.

Rumelhart, D. E., & Siple, P. Process of recognizing tachistoscopically presented words. *Psychological Review*, 1974, **81**, 99–118.

Stroop, J. R. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 1935, **18**, 643–662.

Warren, R. M. Peceptual restoration of missing speech sounds. *Science*, 1970, **167**, 392–393.

Wheeler, D. D. Processes in word recognition. *Cognitive Psychology*, 1970, **1**, 59–85.

Williams, L. G. The effect of target specification on objects fixed during visual search. *Perception and Psychophysics*, 1966, **1**, 315–318.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1971. 2nd ed.

Winston, P. H. Learning to identify toy block structures, In R. L. Solso (Ed.), *Contemporary issues in cognitive psychology*: The Loyola Symposium. Washington, D.C.: Winston, 1973.

Yarbus, A. L. *Eye movements and vision*. New York: Plenum, 1967.

REFERENCE NOTES

1. Palmer, S. E., *Structural aspects of perceptual organization*. Unpublished doctoral dissertation, University of California, San Diego, 1974.
2. Navon, D., *Global precedence in visual recognition*. Unpublished doctoral dissertation, University of California, San Diego, 1975.

(Accepted January 10, 1977)