COGNITIVE PSYCHOLOGY 9, 441–474 (1977)

Hierarchical Structure in Perceptual Representation

STEPHEN E. PALMER

University of California, Berkeley

A theoretical framework for perceptual representation is presented which proposes that information is coded in hierarchical networks of nonverbal propositions. The hierarchical structure of the representations implies selective organization: Some subsets of a figure will be encoded as integral, structural units of that figure, while others will not. A context-sensitive metric for the "goodness" of a part within a figure is developed, corresponding to the probability that the subset will be encoded as a structural unit. Converging evidence supporting this position is presented from four different tasks using simple, straight-line figures. The tasks studied are (a) dividing figures into "natural" parts, (b) rating the "goodness" of parts within figures, (c) timed verification of parts within figures, and (d) timed mental synthesis of spatially separated parts into unitary figures. The results are discussed in terms of the proposed theory of representation, the processes that operate on those representations, and the general implications of the data for perceptual theories.

What are the units of perception and how are they combined to form an integrated percept of a visual stimulus? The answers to these questions are important components in our understanding of perception and related cognitive processes. Various theoretical views have been proposed, attacked, and defended over the years. At one extreme lies the structuralist position that the perception of whole figures is nothing more than the concatenation of primitive perceptual elements. At the other extreme lies the Gestalt position that the perception of whole figures is an indivisable entity whose properties are not determinable from the properties of their components.

Historically, the structuralist and Gestalt schools of thought have been presented as a dichotomy; perception is either atomic or holistic with

Portions of this paper are based on a Ph.D. dissertation completed under the direction of Dr. D. A. Norman at the Department of Psychology, University of California, San Diego. The research was supported in part by a National Science Foundation Graduate Fellowship and Biomedical Support Grant NIH 3-S05-RR-0706-0851 to the author, and by Research Grants NS-07454-07 from the National Institute of Health and GB-32235 from the National Science Foundation to D. A. Norman. I wish to thank D. A. Norman, L. A. Cooper, D. E. Rumelhart, N. H. Anderson, and E. Rosch for their helpful comments and suggestions on various aspects of this research and on earlier drafts of this paper.

Requests for reprints should be sent to Stephen E. Palmer, Department of Psychology, University of California, Berkeley, CA 94720.

ISSN 0010-0285

no acceptable middle ground. Intuitively, however, there seem to be elements of truth in both positions. Whole figures do seem to have natural parts, yet there are properties of the whole which its parts do not share. For example, it is eminently reasonable to believe that the perceptual representation of a square includes the representation of lines as subparts. But it is also important to realize that the square has attributes of closedness and area that are not attributes of the component lines.

The purpose of this paper is twofold. First, a theoretical position is presented concerning the structural nature of perceptual representation. The theory synthesizes the holistic and atomic approaches, postulating numerous levels of representation in the form of hierarchical networks. At each level in a hierarchy, structural units are defined both holistically as a set of global properties and atomically as an organized set of parts. These parts are the structural units at the next-lower level in the hierarchy.

Second, a body of experimental evidence is reported that supports one important aspect of the theoretical foundation: namely, that perceptual representations are selectively organized data structures. The research strategy might well be described as "breadth first." Each of the four experiments explores a different technique for studying perceptual structure: parsing figures, rating the "goodness" of parts within figures, speeded search for parts within figures, and synthesizing figures from separate parts. Each experiment sheds some light on the structural nature of perceptual representation, but none answers all (or even most) of the questions that might be asked about performance of the particular task studied. Taken as a whole, however, the data provide converging evidence about the units of perception, how those units are structured, and how this structure is used in visual information processing tasks.

A Hierarchical Network Theory of Perceptual Representation

The foundation of the present model is that perceptual representations are highly organized data structures containing many embedded levels of detail (see also Baylor, 1971; Moran, Note 2; Winston, 1975). This view is in contrast to standard views that the perceptual representation of a complex figure, object, or scene can be adequately described in terms of a single holistic chunk, a concatenation of primitive perceptual units, or even a simple conjunction of these two levels. Rather, it is proposed that many levels of structure are required to capture all of the structural information. The object as a whole has certain global properties as well as a set of component parts with specific perceptual relationships between them. The parts have the same logical status as the whole. They too have certain global properties and a further set of component parts. This general orientation is similar to that of the Gestaltqualität school.

As an example, consider how one might characterize the perceptual structure of a standing person. As a whole, the person is a rather elongated, ellipse-shaped object (of some specific length-to-width ratio) that is oriented vertically and has some scalar size. This might be the most global level of representation for the person, one that might be constructed from just low spatial–frequency information. At a finer level of resolution the parts of the body are delineated. There are a head, torso, two arms, and two legs. Each of these parts, when considered as a whole, has global properties too. The head, for example, is a less-elongated ellipsoid (of some specific length-to-width ratio) that is oriented vertically with a scalar size dependent on the size of the body. But the analysis need not stop at this level either. The head contains further parts; it has eyes, ears, a mouth, and a nose. These parts too can be represented both globally and as a further set of parts. What emerges is a multileveled hierarchical structure of parts and wholes, each of which has a representation of holistic properties as well as component structure.

Since there is no formal difference between parts and wholes (except with respect to a given level of analysis), the term "structural units" will be used to denote these entities. Structural units are virtually the same as Miller's (1956) construct of "chunks." They are elements of mental representation that can be processed as a single entity, regardless of their internal complexity, at a global level of analysis.

Figure 1 shows a generalized format for perceptual networks in the proposed system. The structural units (SUs) are linked together by relationships into a hierarchy of propositions. The graphic conventions follow those in Norman and Rumelhart (1975), specifically those in the chapter by Palmer (1975a). Structural units (SUs), relations (Rs), global properties (Ps), and quantitative property values (Vs) are represented as nodes in the network. Relational nodes are shown in ellipses, and conceptual (or "perceptual," if you like) nodes are shown in angular brackets. The arcs connecting the nodes are labeled to identify the arguments of the represented relation.

At each level, structural units are defined by their global properties. Only one global property per node is shown in Fig. 1, although it is assumed that there would be many such properties. Global properties consist of quantitative values along perceptual dimensions, specified relative to some referent (e.g., the size of an eye might be represented relative to the size of the head). In pattern recognition, these quantitative values are used in assessing the degree of correspondence (goodness of fit) between the information encoded in the network and some to-be-identified sensory data (see Palmer, 1975a, for details). Each structural unit is also defined by its subordinate structural units (e.g., the head by the eyes, nose, mouth, etc.) and the relationships between them (e.g., the position of the eye relative to the head and/or to the other eye or the nose).

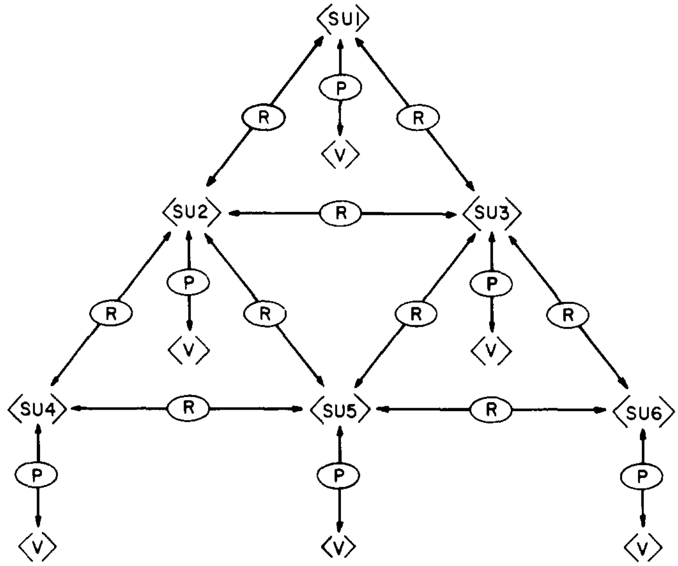

FIG. 1. A general format for representing perceptual information in hierarchical, relational networks. Structural units (SUs) at each level of the hierarchy are defined by their values (Vs) on global properties (Ps) and by their structural relationships (Rs) to other SUs.

The entire network dominated by a structural unit is called its "schema." The schema integrates all of the information known about the scene, object, or part into a systematic framework used during perceptual processing. I have proposed, for example, that the schema is used to generate contextual expectations that facilitate the identification process, to allocate attentional resources in analyzing finely resolved sensory data, and to direct eye movements to places where specific data are expected. [Similar ideas have been expressed by Hebb (1949), Hochberg (1968), and Neisser (1976).] These and other proposals are discussed more fully by Palmer (1975a), but are not directly relevant to the research presented here. The important point is the integration of structural units into hierarchical networks.

Stimulus Forms

The experiments reported below use simple, straight-line figures as stimuli. These figures are constructed by selecting 6 segments from the set of 16 possible segments shown in Fig. 2A. Note that these segments are quite similar to the configuration of 14 segments from which Rumelhart (Note 3; Rumelhart & Siple, 1974) constructed letters. Thus, the figures can be thought of as letter like. Some examples of the six-segment forms used in the experiments are shown in Figs. 2, 4, and 7. This kind of

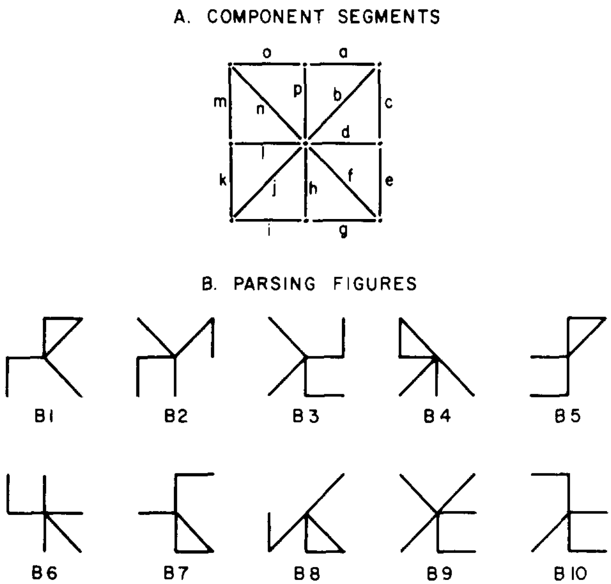

FIG. 2. Construction of stimulus figures. (A) The total set of 16 component segments from which figures are constructed by choosing a subset of 6 segments. (B) The 10 figures used in the parsing task (Experiment I).

stimulus has several desirable properties for the present research. They are novel, well defined, simple, not easily named, and yet complex enough to allow structural organization to be studied rigorously.

In terms of the network model of representation, these figures would produce structures something like the following. Consider form B1 in Fig. 2. The highest level SU (corresponding to SU_1 in Fig. 1) would represent the figure as a whole with global properties like symmetry, closedness, compactness, and so forth. At the next-lower level (corresponding to SU_2 and SU_3 in Fig. 1) there would be SUs for its high-level perceptual parts, say, the triangle (segments *a, b,* and *p*) and the inverted box-like portion (segments *f*, *k*, and *l*). Each of these would be associated with global properties like those suggested above. Relationships between these parts, their relative positions and so forth, would be encoded in the role of R_3 in Fig. 1. At the next-lower level, angles would be represented with global properties such as angular size, orientation, and position. At still lower levels would be SUs for the individual line segments with global properties such as length, orientation, and position. For present purposes there is no need to consider the lowest possible level in terms of individual points. This characterization is intentionally imprecise in certain ways. It is offered only to give the general lines along which representations might be constructed for these particular forms.

Selective Organization

One of the basic assumptions of the theory of perceptual representation outlined here is its hierarchical structure. Hierarchical structure implies selective organization. Organization is present because simpler elements are grouped into higher-order parts within the networks. This organization is selective because not all possible groupings are encoded as parts. The hypothesis that perceptual representations are selectively organized, then, amounts to proposing that some subsets of elements in a figure will be represented as structural units while others will not be. The experiments reported in this paper are concerned with some testable consequences of the selective organization hypothesis. Other aspects of the proposed theory of representation have yet to be explored.

If selective organization is to be studied, there must be some basis for predicting and analyzing figural organization. Rather than relying on intuitions, a model was developed for the "goodness" of parts within figures. The construction of this model is based on the following sorts of considerations. Suppose form B of Fig. 2 is viewed as being composed of the triangle (a, b, p) and the inverted box (f, k, l). When the figure is organized in this way, certain aspects of the figure are foregrounded (perceptually prominent) while others are backgrounded. For example, when the triangle is viewed as a part, the fact that p is connected to a is relatively more apparent than the fact that it is connected to l. If the same figure is structured such that (k, l, p) is a part, the situation is reversed. Now the fact that p is connected to l is relatively more apparent than the fact that it is connected to a. In general, when elements are perceived in larger structural groups, the relationships among the within-group elements are prominent while the relationships among the between-group elements are not. In terms of the hierarchical representations proposed above, this can be roughly reflected by proposing that relationships among SUs within the same superordinate SU are explicitly represented in the network, but those among SUs of different superordinates are not. In Fig. 1, for example, relations are encoded between SU_4 and SU_5 because they are both elements of SU_2. But relations are not encoded explicitly between SU_4 and SU_6 because they are not contained within the same immediately superordinate SU. Perhaps this is why the same figure "looks different" when organized in different ways; alternative groupings result in different information being perceptually encoded about the figure.

If this were the whole story, different organizations would result in different but equally plausible perceptions. Yet there seem to be some organizations of figures into parts that stand out as "natural" and "obvious." One way to conceptualize this is as follows. It seems reasonable to suppose that some information about a figure is more important than other information. The fact that two segments of a figure are directly connected to each other is presumably more important than the fact that two other segments are not. Similarly, the fact that two segments are close together is more important than that two other segments are far apart. This notion provides the key to defining the general notion of "goodness of parts within figures." A "good" part is one that foregrounds important information and backgrounds unimportant information. A "bad" part is one that foregrounds unimportant information and backgrounds important information. In the network representations, this means simply that a good organization is one that directly encodes important facts but not unimportant ones.

A measure of "goodness." To make this general notion of goodness of parts within figures more concrete, an algebraic model was developed. The "elements" to be grouped are the segments of the figures. The "relationships" between those elements are based on the Gestalt principles of grouping: proximity, closedness, connectedness, continuity, and so forth. The basic plan is to represent the importance of the information contained in these relationships as numbers and to formulate an algebraic function that integrates these numbers in some appropriate fashion.

We begin by assuming that a number can be assigned to a pair of segments that represents its value along each relational dimension. For example, the closer together two segments are to each other, the higher would be their value along the proximity dimension. For each dimension, higher values correspond to perceptually more important information within that dimension. Next, we assume that these values can be combined to obtain a "total value" for the relationships between any pair of segments across all dimensions. This integration is expressed as a weighted sum of the component scale values. That is, the total value, R, of the relationships between two segments, i and j, is

$$R(i,j) = \sum_{h=1}^{d} w_h s_h(i,j), \tag{1}$$

where $s_h(i,j)$ is the scale value of the (i,j) pair along grouping dimension h, w_h is the weight assigned to dimension h, and d is the number of relevant dimensions. For example, a given pair of segments will have a scale value for each grouping dimension: proximity, connectedness, continuity, etc. (The method for computing these values is described in the Appendix). Each of the dimensions will have some weighting parameter associated with it, whose magnitude depends on its salience for perceptual organization. The total value of the relationship between the pair of segments, then, is simply the sum of the products of weights and scale values over all grouping dimensions.

Note that each segment has some relational value, $R(i,j)$, to each other segment in the figure. To combine these values in a context-sensitive way, we must distinguish between two classes of relationships. First, there are those relationships that hold between pairs within the group of segments whose goodness is being computed. Second, there are those relationships that hold between pairs containing one group member and one nongroup member. The former are referred to as "within-group relationships" and the latter as "between-group relationships." The numbers representing these relations are integrated into the goodness measure by computing the average difference between the values for within-group relationships and between-group relationships, computed for all within-group segments. More precisely, consider any part, P, composed of a set of p segments, $P = (S_1, S_2, \ldots, S_p)$, whose goodness is to be computed within a figure, F, composed of f segments, $F = (S_1, S_2, \ldots, S_p, \ldots, S_f)$. That is, P is a proper subset of F whose elements are listed as the first p elements of F. The goodness of P within F, then, is

$$G(P|F) = \frac{\sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \neq i}}^{p} \sum_{k=p+1}^{f} (R(i,j) - R(i,k))}{p(p-1)(f-p)}, \tag{2}$$

where i and j are nonidentical elements of P, and k is an element of F but not P. Thus, the $R(i,j)$ values represent within-group relationships, and the $R(i,k)$ values represent between-group relationships.

It is easy to see that this formulation has the basic properties required. The more important the within-group relationships (i.e., the greater the $R(i,j)$ terms), the greater $G(P|F)$ will be. The more important the between-group relationships (i.e., the greater the $R(i,k)$ terms), the smaller the $G(P|F)$ will be. Note that contextual influences occur through the $R(i,k)$ terms such that the same part will have different goodness values in different figures.

The algebraic model can be related to the general network model along the lines suggested earlier. First, we assume that the propositions encoded in the network reflect information of varying importance. Second, we assume that the effect of grouping a set of elements into a larger structural unit is to encode directly the within-group relations but not the between-group relations. The result is some relative gain or loss of important information in the representation, depending on whether the encoded relations (modeled by the $R(i,j)$ terms) are more important than the nonencoded relations (modeled by the $R(i,k)$ terms). Thus, the algebraic model measures the relative informativeness of different part structures for a given figure. The goodness measure is by no means a necessary consequence of the network representation since other measures could be derived from this kind of representation. Moreover, it is not unique to the network approach since other types of representations are compatible with it. The two are, however, closely related.

Three comments should be made about the goodness measure before presenting the experiments. First, it is not a process model but a descriptive measure. It is used to predict the likelihood that various parts will be encoded as structural units in the perceptual representation of a figure. Given these likelihoods, process models can be developed for specific tasks, but such models are quite distinct in nature and purpose from the goodness measure described here. Second, the model was constructed specifically for the type of stimuli used in the present research. In its present form it is clearly inappropriate for continuous, curved figures such as the Shepard and Cermak (1973) forms. Third, the goodness measure was developed more as a tool for conducting research than as a model whose descriptive power is to be evaluated relative to other possible models. For this reason, no attempt was made to adapt alternative models (e.g., Garner, 1974; Hochberg & Brooks, 1960; Vitz & Todd, 1971) to the stimuli for comparison.

EXPERIMENT I

Parsing Figures

Perhaps the simplest way to collect evidence about the existence of part structure in perceptual organization is to ask subjects to divide figures into their "natural parts." It is reasonable to expect that, if perceptual representations are akin to the hierarchical networks discussed earlier, then people will find this task relatively easy to perform and will tend to parse the figures in the same way. If perceptual representations have no such structural organization (e.g., if they are simple lists of atomic components), then one would expect people to find the task artificial and/or to parse the figures arbitrarily.

This experiment examines the way people divide six-segment figures into three-segment parts. Given that there is consistency in the organizations chosen, preferences for groupings should conform in a relatively simple way to the goodness measure defined earlier. For the present experiment, the goodness of any parse is defined as the average goodness of the two three-segment parts comprising it. The model then predicts that the probability of choosing a given parse for a figure will be some monotonically decreasing function of the rank-order goodness of that parse within the set of all possible parses for that figure. The null hypothesis is that the chosen groupings will be randomly distributed among the set of all possible parsings.

FIG. 3. Proportion of parses chosen as a function of the rank-order goodness of the 10 possible parses for each figure.

Method

Stimuli. Ten connected, six-segment figures were selected from a set of 25 figures generated randomly. These figures are shown in Fig. 2. (The same figures are used in Experiments II and III.) The figures were drawn on 3 × 3 dot matrices with two unfilled matrices next to each figure.

Procedure. Subjects were given the following instructions: "You are to divide these figures into two mutually exclusive parts composed of three segments each. All six segments of the figure must appear in one and only one of the two parts. The parts you choose should be the 'best' or 'most natural' or 'most obvious' parts you see within the figure. Draw the two parts in the frames provided to the right of each figure."

Subjects. Sixteen subjects participated in the experiment. These subjects were undergraduate volunteers from the University of California at San Diego. For their participation, eight of the subjects received payment and eight received credit in an introductory psychology course.

Results and Discussion

The goodness of each possible parse for each figure was computed using the algebraic model described above. Since the best single predictor of goodness is proximity (see Experiments II and III) only this factor was used in the calculations. The 10 possible parses for each figure were then rank-ordered according to the average goodness of their component parts.

Subjects' responses were categorized according to the rank order of the chosen parse within the set of 10 possible parses for that figure. Figure 3 shows the probability of choosing a parse as a function of its predicted rank-order goodness. Clearly, the choice probabilities deviate strongly from the uniform distribution predicted by random choices [χ^2 (9) = 601.49].[1] For 5 of the 10 figures, all 16 subjects generated the same parse. For 3 figures two parses were used, and for the remaining 2 figures

[1] Unless otherwise specified, all statistical results presented are significant at $p < .01$.

three parses were used. Moreover, the regularities in subjects' choices conform to the predictions from the model, approximating an exponential function of the predicted rank orders. Of the five parses on which all subjects agreed, four were ranked first and one second by the model predictions. Over all figures, the most frequently chosen parse was ranked first in six cases, second in three cases, and third in one case. Although there is some room for improvement in the model, its success in predicting parsing performance is nontrivial, especially given that only one predictor variable was used and no parameters were estimated.

The present data support Gestalt claims about natural grouping phenomena in perception (e.g., Wertheimer, 1958). They go beyond traditional Gestalt demonstrations in that a well-defined model is able to account for the parsing data. There is no obvious way in which the Gestalt laws of grouping, by themselves, could predict the relative probability of particular parsings except intuitively. The goodness measure and its extension for parsing provide rules for integrating information from grouping factors in a context-sensitive way. It proves to be a good predictor of parsing performance, even when only the proximity factor is considered.

It may seem that a simpler description of the parsings is to postulate perceptual operators such as a "triangle finder," a "box finder," and so forth. If it is assumed that these perceptual operators are performed (or simply completed) in a particular sequence, say, triangles first, then boxes, etc., the parsings can be predicted from this order. Indeed, such an hypothesis is consistent with the data; triangles and boxes are frequently used parts. This approach has two major drawbacks, however. First, given that a box is present in a figure, it may not be encoded as a part, depending on the other segments in the figure. Virtually any part can be camouflaged by some contextual arrangement. Experiments II and III address this issue directly. Second, the perceptual operator hypothesis gives no clues about *why* triangles and boxes are found before other parts. Unless one were to appeal to frequency principles, a problematic claim for several reasons, the answer to this question would probably involve something akin to the Gestalt grouping parameters [e.g., amount of information (Attneave, 1954)]. That is, triangles and boxes are found quickly because their segments are highly compact, connected, and so forth. The "simplicity" of the perceptual operator hypothesis, then, is misleading because its power is derived from implicit principles rather than explicit ones as in the present model.

The obvious question at this point is how perceptual parsing is accomplished. Because it is a static descriptive measure, the algebraic model does not address this important issue. One wants to know, for example, whether parsing is a synthetic process of building up larger and larger groups by putting together smaller ones or an analytic process of dividing larger groups into smaller ones. Or perhaps it is some interactive combination of the two. Unfortunately, the present data provide no information about these problems. One must study the specific time course of the parsing process in order to evaluate such proposals about how the parsed structure evolves. This is a topic for future investigation.

EXPERIMENT II

Subjective Goodness Ratings of Parts Within Figures

The second task used to study the importance of structural organization in perceptual representation is that of having subjects rate the goodness of parts within figures. The conceptual model for goodness of parts makes two qualitative predictions about such rating data. First, different sets of segments within the same figure will be rated differently (see Fig. 4A). Some sets of segments will be consistently rated as good parts, while other sets of segments will be consistently rated as bad parts. Second, the same set of segments will be rated differently in the context of different figures (see Fig. 4B). A given set of segments will be consistently rated as a good part within one figure, while the same segments will be rated as a bad part within another figure. The latter prediction is dictated by the context-sensitive nature of the model. Both types of

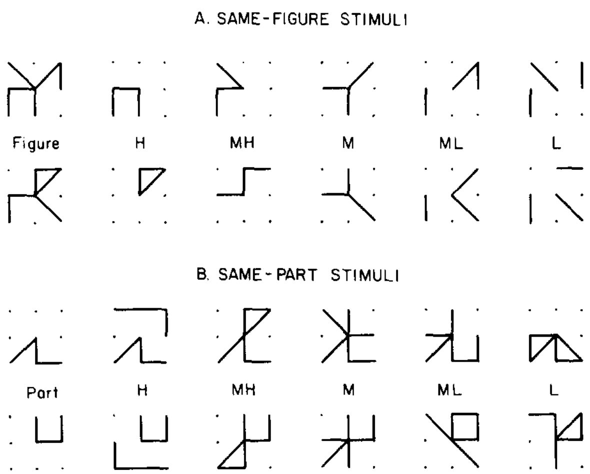

FIG. 4. Construction of stimuli for goodness rating task (Experiment II) and part verification task (Experiment III). (A) Two examples from the same-figure set in which different three-segment parts are paired with the same six-segment figure. (B) Two examples from the same-part set in which different six-segment figures are paired with the same three-segment part. For both sets, the goodness of the part within the figure is categorized as high (H), medium high (MH), medium (M), and medium low (ML), or low (L).

effects are crucial to the general approach taken in modeling perceptual organization, regardless of the specific formulation of the goodness metric.

Quantitatively, the algebraic model can be used to fit the observed goodness ratings for each part in each figure. Parts with high goodness values in the model should receive high goodness ratings from the subjects, and parts with low goodness values should receive low goodness ratings.

Method

Stimuli. Two sets of stimuli were constructed. One set consisted of 10 six-segment figures, each of which was paired with five three-segment parts (see Fig. 4A). This is the "same-figure" stimulus set. (The figures in this set are the same as those in Experiment I.) The other set consisted of 10 three-segment parts, each of which was paired with five six-segment figures containing the parts (see Fig. 4B). This is the "same-part" stimulus set.

The same-figure stimulus set was generated as follows: For each figure, the goodness measure was calculated for each of the 20 possible three-segment parts. (The derivation of weights and scale values used in selecting the stimulus set will not be elaborated here. Suffice it to say that they were based primarily on proximity and connectedness, and were chosen to give intuitively plausible values for individual pairs of segments.) The 20 parts were then rank-ordered according to their goodness within the figure. The five parts chosen for each figure were those with ranks 1, 5, 10, 15, and 20, corresponding to high (H), medium-high (MH), medium (M), medium-low (ML), and low (L) goodness values. Two examples from this stimulus set are shown in Fig. 4A.

The same-part stimulus set was generated as follows: 10 open, connected three-segment parts were chosen. The goodness of each part was computed within various figures until five figures were found which differed substantially in terms of the goodness of the part within them. The figure with the highest goodness value was always a figure in which the part was not connected to any other segments in the figure. Two examples from this stimulus set are shown in Fig. 4B.

Given the two stimulus sets, there were 100 figure-part pairs to be rated. The pairs were drawn on five pages such that each repeated figure (in the same-figure set) and repeated part (in the same-part set) appeared only once per page. The pairs from the two stimulus sets were thoroughly intermixed on the pages.

Procedure. Each subject rated the goodness of the figure-part pairs on two separate occasions at least 1 day apart. The order of the pages was balanced both between and within subjects.

Subjects were given written instructions to "judge how 'good' or 'natural' or 'obvious' the part is within the figure with which it is paired." Ratings were made on an 11-point scale from 0 (very bad part) to 10 (very good part). Subjects were given the 5-page booklet and were told to look over the stimuli to get a feel for the range of goodness levels present. They were instructed to use the entire range of the response scale.

Design. There is a three-factor, orthogonal design for each of the two stimulus sets described above with two observations per cell. The factors are Figures, Goodness Levels, and Subjects. For each of the 10 figures there were five levels of goodness, and each subject rated every figure/part pair twice.

Subjects. The subjects were the same as those in Experiment I.

Results and Discussion

The mean goodness ratings are plotted in Fig. 5. For both sets of stimuli, subjects' ratings increase monotonically with increasing goodness levels.

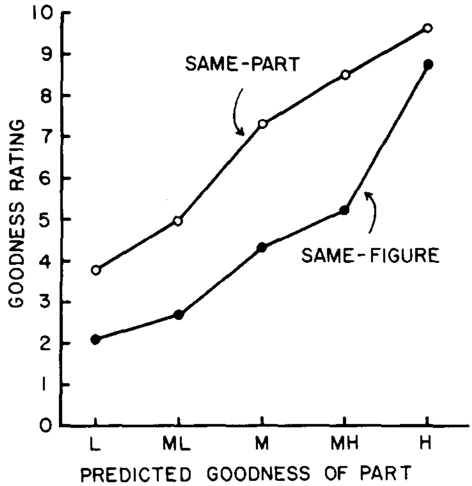

FIG. 5. Mean goodness ratings as a function of the predicted goodness of parts within figures for same-figure and same-part stimuli. See text and Fig. 4 for a description of the stimulus sets and goodness levels.

Separate analysis indicate that the predicted level of goodness is a highly significant factor, both for the same-figure stimuli [$F(4,28) = 86.16$] and the same-part stimuli [$F(4,28) = 94.90$]. This is true for each individual subject as well as for the group data. Also present for same-figure and same-part stimuli are a large effect due to Subjects [$F(7,400) = 29.41$, 16.84 for the same-figure and same-part stimuli, respectively] and smaller effects due to Figures [$F(9,63) = 7.02, 8.10$], the Figure–Goodness interaction [$F(36,252) = 7.46$, 5.79], the Subject–Goodness interaction [$F(28,400) = 8.68$, 10.02], the Subject–Figure interaction [$F(63,400) = 1.57$, 1.92], and the Subject–Figure–Goodness interaction [$F(252,400) = 1.79$, 1.19 ($p > .10$)].

The results clearly indicate that different subsets of segments within the same figure are not judged to be equally good parts. This could be due simply to the different properties of the parts: their dispersion, connectedness, and so forth. But it is apparent from the same-part stimuli that even the same set of segments is not judged to be equally good as part of different figures. This result demonstrates conclusively that the goodness of a part depends not only on the properties of the part itself, but also on the relationship between the segments of the part and the other segments of the figure.

The difference between the two stimulus sets is presumably due to the relatively good intrinsic structure of the parts in the same-part set. These parts are all connected and relatively compact (see Fig. 4B). Thus, it is unlikely that they would receive ratings as low as those for the disconnected and widely dispersed segments in the L and ML levels of the same-figure set. Even so, the range of ratings made for the same part in different figures is substantial.

Mean ratings for individual figure-part pairs in the H, M, and L conditions of both stimulus sets are shown in Table 1 along with the construction of the stimuli themselves. (The letters specifying the figures and parts refer to the labeling of segments provided in Fig. 2A.) The strength of the goodness effect is evident in the fact that there is virtually no overlap between the goodness levels for different figures in both sets. The mean ratings for all 100 figure-part pairs were used to fit the algebraic model for goodness. Best-fitting values for the dimensional weighting

TABLE 1

STIMULI AND RESULTS FROM EXPERIMENTS II AND III

Same-figure stimuli

	High goodness			Medium goodness			Low goodness		
Figure[a]	Part[a]	Rating[b]	RT[c]	Part	Rating	RT	Part	Rating	RT
(abfklp)	(abp)	9.06	976	(flp)	3.37	1342	(afk)	1.69	1924
(bchkln)	(hkl)	8.69	994	(chl)	3.13	1834	(ckn)	1.31	1763
(cdghjn)	(dgh)	5.94	838	(dhn)	3.81	1621	(cgj)	1.19	1804
(fhjlmn)	(fhj)	9.06	839	(jmn)	4.19	1902	(fjm)	2.87	1107
(abhilp)	(abp)	9.31	795	(ahp)	4.31	1367	(bil)	2.81	1492
(dfhlmp)	(dfh)	8.94	904	(fhp)	4.69	1114	(fmp)	2.31	1525
(afghlp)	(fgh)	9.13	1072	(fhp)	4.13	1540	(afl)	3.19	1359
(bfghjk)	(fgh)	9.25	885	(ghi)	4.50	1574	(bgk)	1.31	1757
(bdghjn)	(dgh)	9.00	810	(bhj)	3.62	1362	(bgj)	3.37	1253
(dghjpo)	(dgh)	8.81	713	(jpo)	7.75	1343	(gjo)	1.62	1907

Same-part stimuli

Part[a]	Figure[a]	Rating[b]	RT[c]	Figure	Rating	RT	Figure	Rating	RT
(cdp)	(cdgikp)	9.94	755	(cdhjlp)	8.75	917	(bcdhop)	3.50	1202
(alp)	(aegilp)	9.81	751	(acfhlp)	5.25	1887	(ablmnp)	3.06	1493
(fgl)	(acfglo)	9.81	868	(abfglp)	8.56	930	(fghjln)	2.87	1835
(flm)	(acfilm)	9.81	901	(dfjlmp)	5.56	1495	(dfhlmn)	3.19	1440
(bdf)	(bdfkmo)	9.87	812	(bdfklm)	8.75	1136	(bdfgip)	5.00	1208
(cdn)	(cdgikn)	9.81	842	(cdhijn)	8.62	945	(bcdfln)	4.12	1445
(ghj)	(acghjo)	9.87	790	(dghjnp)	6.37	1386	(fghjkl)	2.62	1252
(fln)	(acfiln)	9.69	996	(abflnp)	6.94	1073	(fhjlmn)	3.44	1310
(jkn)	(aegjkn)	9.81	1109	(bdejkn)	7.31	1340	(fjklmn)	4.81	1552
(hip)	(cehimp)	8.19	1092	(bdfhip)	7.75	1284	(fhilnp)	5.50	1093

[a] Figures and parts are listed as segments according to the labels in Fig. 2A.
[b] Mean ratings on a scale from 1 to 10.
[c] Mean reaction times (in milliseconds) based on raw latencies.

parameters were estimated statistically by a stepwise multiple regression technique. The dimensional values for each part in each figure were determined by specific physical measures of the stimuli (see Appendix). The dimensional values used as predictor values were proximity, connectedness, continuity, similarity in orientation, and similarity in length for pairs of segments. The regression program fit the model to the data by estimating optimal values of the weighting parameters for each dimension that accounted for a significant portion of the variance in the rating data. The dimensions incorporated into the final regression equation were proximity and connectedness. These two variables yielded substantial agreement between the model predictions and data [$r = .86$, $t(98) = 16.99$]. Thus, 74% of the variance from 100 data points was accounted for by just three estimated parameters, two dimensional weights and an additive constant.

EXPERIMENT III

Part Verification Reaction Time

The results of Experiments I and II are subject to the objection that such conscious judgments may have little in common with unconscious perceptual processing. To investigate the importance of part structure in real-time perceptual processing, a choice reaction-time paradigm was developed. The paradigm is a "part-probe" task in which subjects are required to decide whether or not the segments of a part probe are contained within a presented figure. [A similar paradigm was independently developed by Reed (1974).] The part-probe task is reminiscent of the Gestalt "hidden figures" task (Gottschald, 1926), although the response measures are different. In the present experiment the main dependent variable is reaction time (RT), whereas in the Gestalt studies the dependent variable is the probability that subjects manage to find the part.

The critical independent variable in this experiment is the goodness of the three-segment probes within their respective six-segment figures. The logic underlying the paradigm is as follows. The part probes with high goodness values will tend to be encoded as single SUs at a high level in the hierarchical network representing the figure. Part probes with low goodness values will tend to be encoded as multiple SUs at lower levels in the representation of the figure, since the segments will be embedded in more than one higher-order SU. If SUs are processed as integral units, then a "good" probe will be verified more quickly and/or accurately than a "bad" probe because the good probe will require less SUs to be verified than the bad probe. This prediction about the relative efficiency of finding good and bad parts within figures should hold for both same-figure and same-part stimuli, as described in Experiment II. The same-part stimuli provide a natural control condition for the intrinsic properties of the part in isolation, since the same part is used as the probe for all goodness levels.

Method

Stimuli. The stimuli were a subset of those generated for Experiment II; only the H, M, and L goodness levels were used. "Positive" pairs were those in which the part was actually contained within the figure. They were formed for the same-figure stimuli by pairing each figure with its own parts and for the same-part stimuli by pairing each part with its own figures (as shown in Fig. 4 and Table 1). "Negative" pairs were those in which the part was not contained within the figure. They were formed for the same-figure stimuli by pairing each figure with each of the parts for another figure, e.g., by replacing the upper figure with the lower figure (and vice versa) in Fig. 4. The goodness level of negative pairs was then defined by the goodness of the part with its own figure in the positive trials. Negative pairs were formed for the same-part stimuli in an analogous fashion; each part was paired with the figures for another part. Again, the goodness levels of negative pairs was defined as the goodness of the figure when paired with its own part. With this method of stimulus construction, each figure and each part appeared in an equal number of positive and negative trials. Moreover, there were equally many positive and negative trials for each level of goodness.

The figure-part pairs were presented on a Tektronix oscilloscope with a fast-decay phosphor (decays to 90% in .63 msec). Each figure measured .75 by .75 in., subtending an angle of approximately 1.5° of visual angle when viewed from the subjects' position. The figure and part appeared side by side, with the figure in the left position. The two stimuli were .5 in. apart horizontally, so the entire display subtended about 4° of visual angle.

Procedure. Subjects were given instructions regarding the nature of the task. They were instructed to respond as quickly as possible while keeping their overall error rate at or below 5%. Before the first session began, subjects were given 40 practice trials. If any subject had remaining questions, they were answered at this point.

Each trial consisted of the following sequence. A ready signal was presented for 500 msec. This signal was a pair of three-by-three dot matrices (the endpoints of the 16 possible line segments) in the position where the figures would appear. The dots remained on while the stimulus pair was presented. The display was terminated when the subject made a response. Because subjects usually knew when they had made a mistake, no specific feedback was given after the response. After a 2-sec delay the next trial began. The first 60 trials were followed by a 1-min mandatory rest period.

All subjects were tested individually on 2 days. Assignment of responses ("yes"/"no") to hands (left/right) was changed on the second day for all subjects and was balanced for order across subjects.

Design. There are five substantive factors for each stimulus set: Days (first and second), Correct Response (yes and no), Goodness Level (H, M, and L), Figures (the 10 figures or parts, depending on the stimulus set), and Subjects. All five factors were combined orthogonally with one observation per cell.

Subjects. The eight subjects were those in Experiments I and II who received payment for their participation.

Results and Discussion

Reaction time (RT) data in this (and the following) study were treated as follows. Using correct responses only, cell means were computed for all substantive factors other than Figures. "Wild" observations were defined as reaction times more than 5 standard deviations above or 3 standard deviations below the corresponding cell mean. (Asymmetrical cutoffs were used because the RT distributions were positively skewed.) Less than 0.5% of the data typically fell into the wild category. Cell means were then recomputed, discarding both incorrect and wild responses. These missing data were then estimated with the mean of the appropriate cell in the design.

Analyses of variance were performed on the RT data after applying a logarithmic transformation. The transformation normalized within-cell distributions and homogenized variances across conditions. All graphs of RT data are plotted in logarithmic units and thus conform to the pattern of results shown by the statistical tests.

Mean reaction times are plotted in Fig. 6. It is readily apparent that there are large effects due to goodness of the part probes in the figures. Positive (yes) responses to H pairs in the same-figure set are 500 msec faster than responses to the M pairs [$F(1,7) = 275.10$], and are 600 msec faster than responses to the L pairs [$F(1,7) = 128.84$]. For the same-part set, positive responses to the H pairs are 300 msec faster than responses to the M pairs [$F(1,7) = 43.43$] and 450 msec faster than responses to the L pairs [$F(1,7) = 65.13$]. These RT differences are highly significant for each individual subject as well as for the group data. Thus, the major prediction of the experiment is confirmed: Subjects can verify part probes much more quickly when they are good parts within their figures. This is true even for the same-part stimuli in which the segments of the part probe are held constant.

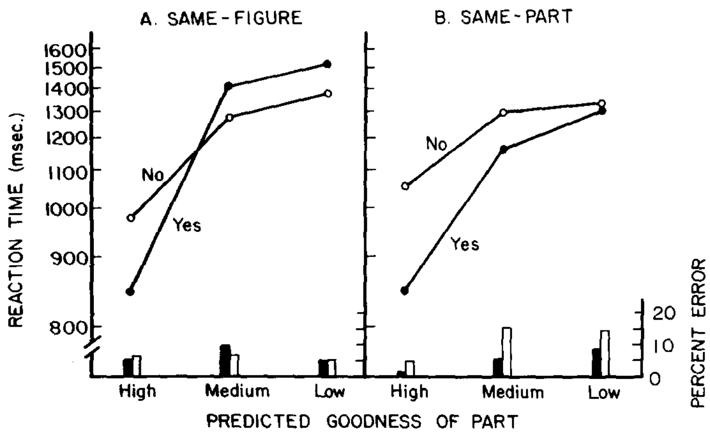

FIG. 6. Mean reaction time (logarithmic scale) for verification of parts within figures when presented simultaneously as a function of the goodness of parts within figures. (A) Results for the same-figure stimuli; (B) results for the same-part stimuli. Histograms at the bottom show error rates for "yes" (black) and "no" (white) responses at each goodness level.

The RT difference between the positive M and L pairs is not significant for the same-figure stimuli [$F(1,7) = 1.71, p > .10$] and is only marginally significant for the same-part stimuli [$F(1,7) = 9.83, p < .05$]. This is in contrast to the large difference between M and L goodness levels in the rating data (see Fig. 5). The lack of corresponding effects in the verification task suggests that only the "best" (H) parts are likely to be encoded as structural units. That is, while the M parts are judged "better" parts than the L parts, neither are "good enough" to be perceived as units under the present circumstances.

Comparisons of positive and negative responses within goodness levels also indicate a qualitative difference between good (H) and bad (M and L) parts. Positive responses are significantly faster than negative responses only for the H pairs in both the same-figure [$F(1,7) = 29.83$] and the same-part conditions [$F(1,7) = 23.13$]. In the same-figure condition, negative responses are marginally faster than positive ones for both the M and L pairs [$F(1,7) = 6.11$, 7.18, respectively, $p < .05$]. In the same-part condition, however, there is no difference between positive and negative responses for either the M or L pairs ($p > .10$).

The results for negative trials also show clear differences between good and bad probes. For the same-figure stimuli, RTs are significantly faster for H pairs than for either M pairs [$F(1,7) = 48.98$] or L pairs [$F(1,7) = 81.20$]. The corresponding comparisons for the same-part stimuli are also significant [$F(1,7) = 26.22$, 92.99, respectively]. In the same-figure conditions these differences are probably due to the intrinsic properties of the part probes. The segments of the H probes tend to be more connected and closer together than those of the M and L probes (see Fig. 4). In the same-part stimuli the differences may be due to unintentional covariation between goodness levels and the number of segments common to probes and figures. A more complete discussion of these effects appears in Palmer (1974).

Error rates for the same-figure stimuli are quite constant. Neither goodness levels nor response types produced any significant differences. Differences did occur, however, with the same-part stimuli. More errors were made on M and L pairs than on H pairs for both positive trials [$F(1,7) = 9.33$, 10.72, respectively, $p < .05$] and negative trials [$F(1,7) = 28.00$, 11.07, respectively, $p < .05$]. No differences between positive and negative error rates were evident for any of the goodness levels.

Mean (untransformed) RTs for individual figure-part pairs are shown in Table 1 along with the rating data from Experiment II. As in the rating data, there is virtually no overlap between the H pairs and the M or L pairs, but unlike the rating data there is substantial overlap between the M and L pairs. Note that there is a very general consistency between the ratings and RTs, due primarily to the fact that very high ratings correspond to very short RTs. The correlation between these two experimental measures of goodness of parts within figures is clearly significant [$r = -.82$, $t(58) = 11.09$], but is not as strong as might have been expected.

The algebraic model was fit to the RTs for positive trials by the stepwise multiple regression technique described in Experiment II. Proximity again proved to be the most important single predictor variable, accounting for 54% of the variance. Continuity was also included in the final regression equation, but accounted for only an additional 2% of the variance. The best-fitting regression equation for these two variables yielded a correlation of $r = .75$ [$t(58) = 8.52$] between the predicted and observed results.

A model of part verification. There are two major results from the present study that must be accounted for by any model of part verification. First, positive responses to H figure-part pairs are consistently faster than those to M or L pairs. This is a large effect in both stimulus sets and is consistent with Reed's (1974) data for sequential presentation. Second, positive responses are faster than negative responses only for the H pairs.

One conclusion that can be reached from these effects is that the process of part verification, whatever it is, cannot be a point-by-point (template) comparison or even a segment-by-segment comparison. Such models cannot account for either of the effects without appealing to some sequential ordering construct whereby the points or segments belonging to the good parts are matched before those of bad parts. In order to explain why such an order is used rather than others, some assumptions about selective organization must be invoked in one form or another.

One relatively simple model for part verification can be constructed through a matching procedure that operates on the type of hierarchical networks described earlier. The basic notions are that (a) H parts are likely to be encoded as single, higher-order units in such networks while M and L parts are not, and (b) the matching of a part network to a figure network proceeds in a "top-down" fashion, from larger units to smaller units. A brief sketch of such a process is given below.

The part verification process begins at the highest structural unit (SU) in the network representing the part probe. An attempt is made to find an identical SU within the network representing the figure. This process consists of a global similarity match between the highest SU in the part probe and a corresponding SU at the same level of resolution in the figure. (In this particular task, corresponding nodes are determined by the global location parameters associated with them.) The precise nature of the similarity value is presently unspecified, but might be based on global properties associated with the SUs: general shape, orientation, closedness, number of parts, etc. Given this similarity value, a high- and low-threshold decision model is used for the comparison. If the similarity of the SUs is greater than some upper criterion, then verification is achieved with a positive result. If the similarity is less than some lower criterion, verification is achieved with a negative result. If the similarity value lies between these two criteria, the result is indeterminant, and more highly resolved information is required.

Higher resolution is achieved by sequential verification of the subparts of the indeterminate SU. The verification of these lower-level SUs proceeds exactly as described above. Corresponding SUs are compared by a sequence of rapid, global similarity matches. Each outcome is either positive, negative, or indeterminate. For each indeterminate result, still higher resolution is obtained by sequential verification of even lower level SUs in the same way just described.[2] Naturally, at the lowest functional level (in this case, the individual segments) no higher resolution is required since there are no differences below this level.

If sequential verification of subparts of an SU is required, its result is determined by the logical conjunction of positive results on all subparts. That is, confirmation requires that all subparts be present. Thus, the sequential verification of subparts at any given level is proposed to be a serial, self-terminating process. If any subpart is indeterminate, the possible discrepancy must subsequently be pursued by lower level comparisons, and the status of the superordinate SU remains indeterminate until confirmation or disconfirmation is achieved. One of the critical assumptions for part verification is that the lower criterion is placed at a very low value. This assumption is necessary because obtaining a low similarity score from a global level does not imply that the segments are not present in the figure. If the part structure of the figure is inappropriate for the probe, the global similarity score between the probe and any part of the figure will be low, but cannot (or should not) disconfirm the probe since the probe segments may be embedded in several different parts at lower levels. Thus, while confirmation may be obtained through either high-level or low-level comparisons, disconfirmation should be obtained at the level of individual segments. By using a very low value for the lower criterion, disconfirmations almost always take place at the functional terminal nodes.

The two major results of the experiment are accounted for as follows.

(*i*) *Positive responses to H pairs are faster than positive responses to*

[2] The decision aspects of the model are similar in important aspects to Atkinson and Juola's (1974) two-stage decision theory. In both models, the initial stage consists of a fast, high-level similarity judgment with high- and low-threshold criteria for positive and negative responses. The second stage consists of lower-level, sequential judgments if the results of the first stage are insufficient. Because of the current proposal that each individual comparison in the second stage is the same as a first-stage comparison, except at a lower level, the present theory is applicable to any depth of hierarchy. Atkinson and Juola's formulation is, therefore, a special case in which there is only one subordinate level in the hierarchy.

M or L pairs. It is assumed that the H part probes are almost always encoded as structural units within the representation of the figures. Since verification proceeds downward from the highest level, these probes will usually be confirmed by the initial global match operation, resulting in a short-latency positive response. The M and L part probes, however, are seldom encoded as single, structural units in the figure representation. It follows that they will seldom be verified at the highest level and will almost always require sequential verification of subparts at a lower level in the hierarchy. The additional matches required for M and L probes predict longer latencies than for the H probes.

(*ii*) *Positive responses are faster than negative responses only for the H probes*. Because of the very low value of the lower criterion, disconfirmation of negative pairs is virtually never achieved in high-level matches. Thus, negative responses to H probes require additional matching operations at lower levels. Positive responses to H probes, however, are usually achieved at the highest level. Clearly, the model predicts that positive H responses will be faster than negative H responses. For M and L probes, the relative speed of positive and negative responses is difficult to predict. It depends on the probability of confirming the positive probes at high levels, the probability of disconfirming the negative probes at high levels, and the number of common segments in the negative probes. These factors may vary considerably with different stimuli and presentation conditions. The results show that positive responses to M and L probes are sometimes faster than, sometimes the same as, and sometimes slower than negative responses, depending on the stimulus set.

The process model of part verification just described is not specific on a number of points. First, there is the problem of the global similarity values. Given two corresponding SUs, how is the similarity between them to be assessed? I have deliberately sidestepped this issue because it is probably the thorniest problem of perception and cognition. The model simply postulates that some appropriate measure of similarity is computed. Second, there is the problem of the sequence of comparisons for lower-level SUs when a higher-level SU fails to be confirmed or disconfirmed on the basis of global similarity. Presumably the selection process is not random, but it still might be based on many different factors. Third, there is the problem of whether the network search is to proceed depth first or breadth first. For example, if a comparison for an SU at level n is indeterminant, and if the first of the resulting lower-level comparisons at level n-1 is also indeterminant, does the process then perform the rest of the comparisons at level n-1 (breadth first) or does it continue downward to perform comparisons at level n-2 (depth first) before proceeding with the rest of the level n-1 comparisons? All of these issues must be resolved to construct a well-defined process model for this task. At the present time, however, attempting to do so would be premature. The data do not provide enough information to make intelligent choices, and further experiments will have to be designed specifically to get at these issues.

EXPERIMENT IV

Mental Synthesis of Figures from Parts

The final experiment investigates the importance of structural organization in a "mental synthesis" task. In mental synthesis, subjects are presented with two mutually exclusive, three-segment parts in spatially separated displays (see Fig. 7). They are instructed to construct a six-segment figure by imagining the right display superimposed on the left display. When they have completed the construction of the imaginary figure, they make a response whose latency is measured. After a brief delay, a test figure is presented on the left display. The subjects must then decide whether or not this presented figure is the same as the figure synthesized from the parts. The latency of this discrimination response is also measured.

The critical manipulation concerns the goodness of the component parts to be synthesized. If the initial parts correspond to the part structure of the whole figure, all that is required is to move the one part to the other and to encode the relationships between them (i.e., to "hook up" the two parts). If the initial parts do not correspond to the part structure of the whole figure, simply moving one part and linking it to the other part is not sufficient. The segments must then be regrouped into

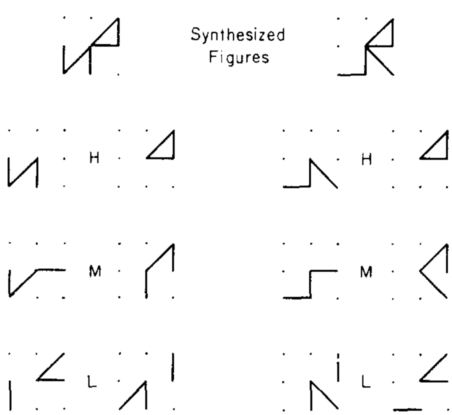

FIG. 7. Construction of stimuli for mental synthesis task (Experiment IV). At the top are two figures that are synthesized by combining the pairs of parts shown below them. The part pairs are high (H), medium (M), or low (L) goodness parts of the synthesized figures.

the appropriate part structure in order to "see" the synthesized figure as it would be perceived when presented as a whole. If this latter process is not completed correctly, the execution of the test discrimination will require reorganization of the test stimulus (thus increasing the test RT) or will suffer in accuracy.

The part synthesis task was developed, in part, to contrast with other mental operations which do not require structural manipulations. Mental rotation is the most important example of such nonstructural transformations. The results of these experiments (Cooper, 1975; Cooper & Shepard, 1973; Shepard & Metzler, 1971) suggest the operation of processes quite like actual rotation of an object in physical space. Too often it is assumed that this reflects a basic property of images: that they are representations which can be operated on just as their referents would be in the real world. If this were true, one might expect that subjects should be able to "slide" one image onto the other and to "see" the resulting figure as soon as they were superimposed. This analysis of the synthesis task predicts that no differences will be obtained due to the goodness of the initial configuration of parts.

Method

Stimuli. Sixteen connected stimulus figures were constructed, each of which has an obvious organization into two three-segment parts. This pair of parts was designated the high goodness (H) organization. A medium goodness (M) organization was constructed by finding two internally connected, three-segment parts from the same figure whose grouping differed from the H organization. A low goodness (L) organization was constructed by finding two internally disconnected three-segment parts whose grouping also differed from the H organization. Several examples from the stimulus set are illustrated in Fig. 7.

The 16 figures were grouped into 8 pairs of similar figures such as the two figures shown in Fig. 7. The "different" trials were formed by using the parts of one figure for synthesis, and then testing with the other figure in the similar pair. Thus, each figure and each pair of parts were presented equally often in "same" and "different" trials.

The parts were presented side by side on a CRT display, separated by .75 in. In all other respects the initial stimulus display was identical to the display used in Experiment III. Both spatial orderings of the parts were used. The test figure was always presented on the left position of the display.

Procedure. The subjects were instructed regarding the nature of the task. The instructions stressed that they should take as long as they needed to synthesize the parts into a figure, so that they could make a rapid discrimination when the test figure appeared. This strategy was emphasized to discourage the likelihood of subjects trying to remember the separate parts and then to verify them when the test figure was presented.

Each trial consisted of the following sequence. First, a ready signal composed of two 3×3 dot matrices was presented in the positions where the parts would appear. After 500 msec the parts were presented within the dot matrices and remained there until the subjects made their initial response indicating that they had completed synthesis. This "synthesis" response was made with the hand designated for "same" responses to the test figure. The test figure was presented in the left dot matrix 500 msec after the synthesis response. The figure remained on the screen until the subject made the discrimination response. After a 2-sec delay, the next trial began.

At the start of each session, subjects were given 24 practice trials with stimuli different from the experimental figures, but similarly constructed. The 192 experimental trials were then presented with a 1-min rest period after each block of 64 trials. Each subject participated in two sessions, one per day for 2 days. As in the previous experiment, response-hand assignments were changed on the second day and were balanced for order across subjects.

Design. There are five substantive factors in the design: Days (first and second), Correct Response ("same" and "different"), Goodness (H, M, and L levels), Figures, and Subjects. All factors were combined orthogonally with two observations per cell. (The replications were actually the two spatial orderings of the part pairs.)

Subjects. The eight subjects were the same as those in Experiment III.

Results and Discussion

Synthesis latencies. For the synthesis latencies, the Correct Response factor is presumably irrelevant. Subjects had no knowledge of the test stimulus until after making the synthesis response. Therefore, this factor is not included in analyses of synthesis latencies, yielding a four-way factorial design with four observations per cell.

Mean synthesis latencies are shown in Fig. 8. It is immediately apparent that the parts are synthesized more quickly and accurately when they are appropriate to the part structure of the figure (the H pairs) than when they are not (the M and L pairs). These latency differences of almost 3 sec are highly significant in the group data [$F(1,7)$ = 123.40, 377.87 for H vs M and H vs L comparisons, respectively]. While the magnitude of the

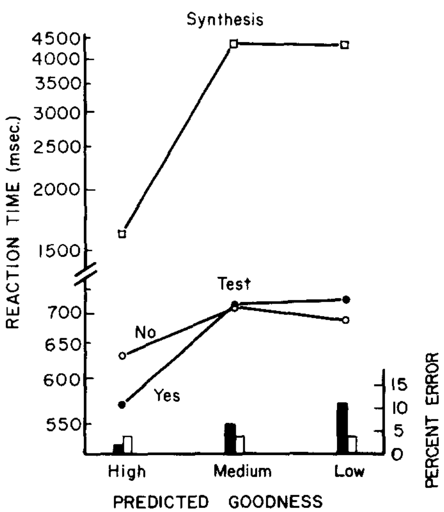

FIG. 8. Mean reaction time (logarithmic scale) to synthesize parts into figures and to make same/different test discriminations as a function of the goodness of parts within figures. Histograms at bottom show error rates for "yes" (black) and "no" (white) responses in the test discrimination at each goodness level.

effects varies substantially for different subjects, it is highly reliable for each individual, F ratios ranging from 134.23 to 746.56, df = 1,192. There is no significant difference between the M and L pairs in the group data ($F < 1$), although one subject was significantly faster on M pairs than L pairs, and one subject showed the reverse effect.

The sheer magnitude of the synthesis latencies for M and L pairs is noteworthy. On the first day, subjects took about 5 sec (on the average) to synthesize the inappropriate parts. Latencies of more than 10 sec were not uncommon for some of the more difficult figures. Clearly, synthesis is not an easy task given inappropriate parts, as the reader may confirm using the examples shown in Fig. 7.

Introspectively, the difficulty lies not so much in moving the segments as in regrouping them into the appropriate structure. There seems to be a curious period in which the observer knows which segments are present in the composite figure, but does not know what the figure "looks like." To determine the perceptual structure of the synthesized figure, the segments must be consciously organized. Subjects report actively looking for obvious parts—triangles, boxes, composite lines, and so forth—in an attempt to group the segments appropriately. This process can accurately be described as problem solving. There are exploratory hypotheses, false leads, dead ends, backtracking, and fresh starts. More than anything else, it seems to be this grouping process that accounts for the difference between appropriate and inappropriate initial parts.

Test latencies. If the parts were always correctly synthesized, and if the resulting figure had the same properties as a real figure, test latencies would be constant across goodness levels and same responses would be faster than different responses. The RT functions should be flat across goodness levels because the same figure/figure discrimination is being made, regardless of the goodness of the initial parts. Same responses should be faster than different responses, in agreement with the results of an earlier experiment for sequential figure/figure matching (Palmer, 1974). In contrast to these expectations, the group results indicate that both same and different test responses are made more quickly and accurately following synthesis of the H pairs than following synthesis of the M and L pairs [same, $F(1,7)$ = 15.93, 37.56; different, $F(1,7)$ = 24.27, 14.44, for H vs M and H vs L comparisons, respectively].

The group data, however, do not give an accurate picture of individual performance. Perhaps the only generalization that can be made from the data is that subjects differ remarkably in performing the discrimination response. Figure 9 shows the data from three subjects. Subject MB is the only subject whose data conform to a priori expectations based on complete synthesis—flat RT functions across goodness levels with same responses being faster than different re-

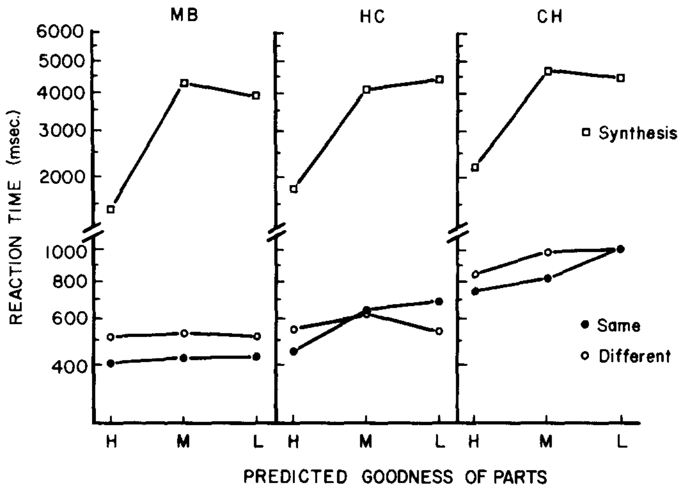

FIG. 9. Mean reaction time (logarithmic scale) for individual subjects to make synthesis and test discrimination responses as a function of the goodness of parts within figures.

sponses. (The author's data also showed this pattern of RTs, but are not included in the present results.) Subjects HC and CH illustrate two other RT functions for comparison. In fact, no two subjects showed the same RT functions. The individual differences are especially puzzling because there are no obvious differences in their synthesis latencies that would explain the resulting differences in test RTs.

One aspect of the present procedure may have contributed to the anomolous test RT results. Subjects did not know the figures before the experiment began. Thus, it was easy for them to believe they had synthesized the parts correctly (i.e., knew what the resulting figure looked like), when they had actually arrived at an inappropriate organization. In other words, the subjects lacked a good criterion for knowing when they were prepared for the test figure. If they knew the figures beforehand, their task would be to recognize the figure by synthesizing it. As soon as they had decided which figure was presented, they would be ready for the test discrimination. This procedural modification should produce more stable test RTs along the lines of MB's data. Recognizing the figure should require appropriate organization and, therefore, would require much the same processing as in the present task.

The test RT data give some indication that synthesis was not completed successfully for all initial parts. Recently, Klatzky and Thompson (Note 1) have independently begun to study mental synthesis with an emphasis on just this problem—the conditions under which synthesis can be performed successfully. They reasoned that if subjects manage to synthesize parts of a figure completely, test RTs in a same/different task should be independent of the number of parts presented for synthesis. Klatzky and Thompson used either one, two, or three initial parts, where the one-part condition corresponds to presenting the whole figure, thus requiring no synthesis. With simple geometric forms (e.g., triangles, parallelograms, etc.) as figures and lines or angles as parts, test RTs were invariant over the number of parts presented. However, with schematic faces as figures and eyes, noses, and mouths as parts, test RTs increased linearly with the number of parts presented. Thus, complete synthesis was obtained for the geometric figures but not for the faces. Since the present forms are all relatively simple geometric figures, it is reasonable to expect that synthesis is complete when good parts are used. The present data show, however, that even with the same figures and the same number of initial parts, synthesis tends to be less successful if bad parts must be synthesized.

A model of synthesis. An intuitive analysis of the synthesis task suggests three component processes: moving the segments in the right display to the left display, linking these segments to those already present in the left display (i.e., determining local connectedness and relations), and grouping the linked segments into perceptual parts. No claim is made that these processes are independent. Moving is prerequisite to linking and linking is prerequisite to grouping. In fact, linking may be considered part of the grouping process, but there are undoubtedly other grouping operations being performed.

Moving the segments seems to be the easiest of the three components. With the H pairs, movement of the right part is usually performed all at once. With the L pairs it is frequently performed separately for each of the disconnected components. Thus, ease of movement may depend on the proximity and connectedness of the segments in the part. The difficulty of linking should depend on the number of linkages required. There are fewest connections to be made in the H pairs, an intermediate number in the M pairs, and most in the L pairs. The grouping operation is clearly easiest for the H pairs, since they are already grouped appropriately. It is arguable that grouping is more difficult for the M pairs than for the L pairs. Because the segments of the L parts are internally disconnected, there is a clear strategy for grouping them. The two connected segments of one part are grouped with the single disconnected segment of the other part (see Fig. 7). This strategy, in fact, gives the same grouping as in the H pairs. For M pairs, however, the strategy is not readily applicable. It is still true that two segments of one part should be grouped with one segment of the other part. But because both parts are internally connected, there is no clear basis for deciding which pair of segments in a part should be grouped together.

This analysis leads to the following account of the results. The H pairs are synthesized most quickly because they are easy to move, require few linkages, and need virtually no regrouping operations. The M and L pairs are synthesized with approximately equal speed because of trade-offs between different processes. While L pairs are possibly harder to move and link than M pairs, they are easier to regroup.

Synthesis versus rotation. The present results demonstrate the importance of perceptual (or imaginal) operations other than "analog mental movement" in performing the synthesis task. In this respect, synthesis differs markedly from mental rotation, in which the results are largely compatible with analogical operations (Cooper, 1975; Cooper & Shepard, 1973; Shepard & Metzler, 1971). The crucial difference between the tasks is that, in mental rotation, perceptual structure is constant, while in synthesis it is not. Rotation requires operations on the location and orientation parameters of the representation, whatever that representation is, without affecting its basic structural organization. Synthesis requires not only operations on location parameters, but also the modification of the structural organization present in the representation. Thus, the different results from the two tasks reflect a contrast between two different types of operations, parametric and structural, rather than a contradiction between similar types of operations (see Palmer, 1975b).

GENERAL DISCUSSION

The experimental results reported here provide substantial evidence for selective organization in perception and imagery. To them, one can add a number of findings in the Gestalt literature on laws of grouping (Wertheimer, 1958), perceptual similarity (Goldmeier, 1972), and hidden figures (Gottschaldt, 1926). More recently there have been studies demonstrating organizational effects in perceptual memory (Bower & Glass, 1976) and imagery (Baylor, 1971; Moran, Note 2; Klatzky & Thompson, Note 1). Taken together, these findings leave little doubt about the importance of selective organization in perceptual representation and processing.

The present results have several important implications for perceptual theory. First, they point to the need for representing information about the perceptual organization of elements into structural units. Given the stimuli and tasks used in the experiments, at least three levels of structural units are required: the whole figure, the multisegment parts, and the individual line segments. With more complex figures and/or more demanding tasks, many more levels may be required. The requirement of multiple levels of structural units rules out standard template or iconic forms of representation, since only one level of functional unit is defined for a pattern. It also rules out any feature-list representation that has only one level of structural features. Rumelhart's (Note 3) particular representational scheme for letters is of this type because the only structural features are component line segments. Such representations fail to account for the data because there is no mechanism to predict that any particular subset of the figure will be processed any differently than any other subset of the figure. The present results are inexplicable without such a mechanism.

Another implication of the results is that the encoding of structural units must be context sensitive. Whether or not a given group of elements will be encoded as a unit must depend strongly on the total set of elements within the figure, not just on the presence of elements of that group. A process that blindly encodes all instances of a particular subset (e.g., a "three-sided box" in the present figures) would not meet the criterion of context sensitivity. It would have to encode such units selectively, only under conditions where that subset is a better group than other possible groupings within the figure. Without some context-sensitive mechanism, the findings for same-part stimuli in Experiments II and III cannot be accounted for.

Another perceptual issue for which the data have implications concerns the way in which subparts are processed. The results of Experiment III show that "good" parts can be verified more quickly and accurately than "bad" parts. Moreover, this is true even when the same segments are good and bad parts within different figures. These findings require that the processing of good parts be more efficient and accurate than that of bad parts, regardless of the intrinsic properties of the part. A further constraint on the nature of part processing is that positive responses are faster than negative responses only for the good parts of the figure. This suggests that good parts may be processed in a qualitatively different way from bad parts. Perhaps the simplest general type of process capable of accounting for these results is one in which good parts can be matched holistically and in parallel, while bad parts must be matched componentially and serially.

The last important implication of this research concerns the relationship between perceptual and imaginal representation. In Experiment IV it was shown that mental synthesis of bad parts took more time and was less successful than synthesis of good parts. This imagery effect is obviously related to the perceptual effects of the previous three experiments. It adds to the current evidence showing that imagery is similar in many important respects to perception (e.g., Segal, 1971, Cooper, 1975; Kosslyn, 1975). Imaginal representations seem to have the same kind of organizational structure as perceptual memory representations, and this organization affects the way in which images are formed and processed. This claim must not be confused with any claim that the image is like the stimulus itself (cf. Pylyshyn, 1977) since this would entail a statement that the structure involved in the perception of a figure is the same as the structure of the stimulus itself. Indeed, the distal stimulus has no explicit structure. If it can be said to have any structure at all, it is only implicit structure. That is, the stimulus object has only the *potential* for different organizations, and it has all of these organizations equally. It is only when the stimulus is perceived or imagined that one organization is emphasized over others, and that emphasis is in the mind, not in the world. Certainly, one does not want to claim that the distal stimulus itself changes when a person organizes it differently in a perceptual or imaginal act.

The parsing, goodness rating, part verification, and mental synthesis tasks provide a diverse set of converging operations for studying structural organization in perception, perceptual memory, and imagery. The results argue strongly that a structural component is necessary in theories of perceptual derived representations. The relational network theory of perceptual representation discussed at the outset is one relatively simple and obvious way to build in such a structural component. Whether it will prove to be the best way is not clear since neither the theory nor the nature of the phenomena is yet sufficiently well specified. It is clear, however, that cognitive theories of perception and imagery must deal with organizational structure and its role in processing more precisely than they have in the past.

APPENDIX

Computation of Dimensional Associations for Goodness-of-Part Analysis

For each component segment, x, there is a centerpoint, C_x, and two endpoints, $E1_x$ and $E2_x$. From the locations of these points, the location, length, and orientation of the line segment are computed as follows.

Location. The location of x is specified as the location of C_x.

Length. The length of x is the Euclidean distance between $E1_x$ and $E2_x$.

Orientation. The orientation of x is the direction between $E1_x$ and $E2_x$ as specified by two values: a "frame" (either horizontal–vertical or diagonal) and a "marker" denoting the value within the frame.

Associations of a pair of segments, i and j, are computed along the dimensions of connectedness, dispersion, continuity, similarity in orientation, and similarity in length. These associational values are $s(i,j)$ and $s(i,k)$ terms in Eq. (1) and are computed as follows.

Connectedness. If i and j share an endpoint, the connectedness of i and j evaluates to 1; otherwise it evaluates to 0.

Dispersion (or proximity). The dispersion of i and j is the Euclidean distance between the location of i and the location of j. (Proximity is defined as the complement of dispersion.)

Continuity. If i and j form a compound line (i.e., they are connected segments with the same orientation), the continuity of i and j is 1; otherwise it is 0.

Similarity in orientation. If i and j have the same orientation, their similarity in orientation is 2. If they have the same frame but different markers, it is 1. If they have different frames, it is 0.

Similarity in length. If i and j have the same length, their similarity in length is 1; otherwise it is 0.

The computational definitions for these dimensions are somewhat arbitrary. The dispersion of two segments might be defined as the distance between the closest endpoints rather than the distance between the centerpoints. Similarity in orientation might be defined simply as a function of the difference between the orientations of the two segments. It can also be argued that the psychological values of these associations should be some monotone transformation (e.g., exponential or power) of the computed physical values. Such alternatives have not been explored at present.

Fitting the model by multiple regression. The model for goodness of parts within figures [see Eq. (2)] was fit to the rating and reaction time data by a stepwise multiple regression technique. The scale values for dimensions were determined a priori from the physical measures defined above. In order to understand how multiple regression was used to estimate the model weighting parameters, it is useful to derive a new form for the goodness equation. If Eqs. (1) and (2) are combined, and if the weights are separated from their scale values, the expression for the goodness of a part P within a figure F becomes

$$G(P|F) = \sum_{h=1}^{d} w_h \left[\frac{\sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \neq i}}^{p} \sum_{k=p+1}^{f} (s_h(i,j) - s_h(i,k))}{p(p-1)(f-p)} \right]$$

This form of the model equation shows that for each relevant dimension, h, the sum of the differences between scale values may be calculated for an entire figure. Each sum is then multiplied by the weighting parameter for the appropriate dimension. Thus, a figure's context-sensitive value along each dimension can be computed independently of the weighting parameter. These values are then used as predictor variables in the multiple regression analysis. The regression program then fits the model to the data of interest by statistically estimating optimal values for the weighting parameters.

REFERENCES

Atkinson, R. C., & Juola, J. J. Search and decision processes in recognition memory. *In* D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology.* San Francisco: W. H. Freeman, 1974. Vol. 1.

Attneave, F. Some informational aspects of visual perception. *Psychological Review*, 1954, **61**, 183–193.

Baylor, G. W. *A treatise on the mind's eye: An empirical investigation of visual mental imagery*. Unpublished doctoral dissertation, Carnegie–Mellon University, 1971.

Bower, G., & Glass, A. Structural units and the redintegrative power of picture fragments. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, **2**, 456–466.

Cooper, L. A. Mental transformation of random two-dimensional shapes. *Cognitive Psychology*, 1975, **7**, 20–43.

Cooper, L. A., & Shepard, R. N. Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press, 1973.

Garner, W. R. *The processing of information and structure*. Potomac, MD: Erlbaum, 1974.

Goldmeier, E. Similarity in visually perceived forms. *Psychological Issues*, 1972, **8** (1, whole No. 29).

Gottschaldt, K. Über den Einfluss der Erfahrung auf die Wahrnehmung von Figuren: 1. *Psychologische Forschung*, 1926, **8**, 261–317.

Hebb, D. O. *Organization of behavior*. New York: Wiley, 1949.

Hochberg, J. In the mind's eye. In R. N. Haber (Ed.), *Contemporary theory and research in visual perception*. New York: Holt, Rinehart & Winston, 1968.

Hochberg, J., & Brooks, V. The psychophysics of form: Reversible-perspective drawings of spatial objects. *American Journal of Psychology*, 1960, **73**, 337–354.

Kosslyn, S. M. Information representation in visual images. *Cognitive Psychology*, 1975, **7**, 341–370.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, **63**, 81–97.

Neisser, U. *Cognition and reality*. San Francisco: W. H. Freeman, 1976.

Norman, D. A., & Rumelhart, D. E. (Eds.). *Explorations in cognition*. San Francisco: W. H. Freeman, 1975.

Palmer, S. E. *Structural aspects of perceptual organization*. Unpublished doctoral dissertation, University of California, San Diego, 1974.

Palmer, S. E. Visual perception and world knowledge. In D. A. Norman, D. E. Rumelhart, & LNR Research Group (Eds.), *Explorations in cognition*. San Francisco: W. H. Freeman, 1975. (a)

Palmer, S. E. The nature of perceptual representation: An examination of the analog/propositional debate. In R. C. Schank & B. L. Nash-Webber (Eds.), *Theoretical issues in natural language processing*. Arlington, VA: Tinlap, 1975. (b)

Pylyshyn, Z. What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, 1973, **80**, 1–24.

Pylyshyn, Z. Imagery and artificial Intelligence. In C. Wade Savage (Ed.), *Perception and cognition: Issues in the foundations of psychology*, Vol. IX of Minnesota Studies in the Philosophy of Science. Minneapolis, MN: University of Minnesota Press, in press, 1977.

Reed, S. K. Structural descriptions and the limitation of visual images. *Memory and Cognition*, 1974, **2**, 329–336.

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

Segal, S. J. Processing of the stimulus in imagery and perception. In S. J. Segal (Ed.), *Imagery: Current cognitive approaches*. New York: Academic Press, 1971.

Shepard, R. N., & Cermak, G. W. Perceptual–cognitive explorations of a toroidal set of free-form stimuli. *Cognitive Psychology*, 1973, **4**, 351–377.

Shepard, R. N., & Metzler, J. Mental rotation of three-dimensional objects. *Science*, 1971, **171**, 701–703.

Vitz, P. C., & Todd, T. C. A model of the perception of simple geometric figures. *Psychological Review*, 1971, **78**, 207–228.

Winston, P. H. Learning structural descriptions from examples. In P. H. Winston (Ed.), *The psychology of computer vision*. New York: McGraw-Hill, 1975.

Wertheimer, M. Principles of perceptual organization. In D. C. Beardslee & M. Wertheimer (Eds.), *Readings in perception*. Princeton, NJ: Van Nostrand, 1958.

REFERENCE NOTES

1. Klatzky, R. L., & Thompson, A. *Studies of visual synthesis—Mental construction of wholes from parts*. Paper presented at 16th Annual Meeting of the Psychonomic Society, Denver, 1975.
2. Moran, T. P. *The symbolic nature of visual imagery*. Paper presented at the Third International Conference on Artificial Intelligence, Stanford University, August 1973.
3. Rumelhart, D. E. *A multi-component theory of confusion among briefly exposed alphabetic characters*. (Tech. Rep. No. 22). La Jolla, CA: Center for Human Information Processing, 1971.

(Accepted May 24, 1977)