4 A Theory of Graph Comprehension

Steven Pinker
Massachusetts Institute of Technology

A striking fact about human cognition is that we like to process quantitative information in graphic form. One only has to look at the number of ways in which information is depicted in pictorial form—line, bar, and pie graphs, Venn diagrams, flow charts, tree structures, node networks, to name just a few—or to the great lengths that computer companies go to advertise the graphic capabilities of their products, to see that charts and graphs have enormous appeal to people. All of this is true despite the fact that in virtually every case, the same information can be communicated by nonpictorial means: tables of numbers, lists of propositions cross-referenced by global variables, labeled bracketings, and so on. Perhaps pictorial displays are simply pleasing to the eye, but both introspection and experimental evidence (Carter, 1947; Culbertson & Powers, 1959; Schutz, 1961a, 1961b; Washburne, 1927) suggest that, in fact, graphic formats present information in a way that is easier for people to perceive and reason about. However, it is hard to think of a theory or principle in contemporary cognitive science that explains why this should be so; why, for example, people should differ so strikingly from computers in regard to the optimal input format for quantitative information.

The goal of this chapter is to address this unexplained phenomenon in a systematic way. In particular, I propose a theory of what a person knows when he or she knows how to read a graph, and which cognitive operations a person executes in the actual process of reading the graph. This theory will be used to generate predictions about what makes a person better or worse at reading graphs, and what makes a graph better or worse at conveying a given type of information to a reader. In pursuing these goals, one must recognize a very pervasive constraint. Comprehending a graph (unlike, say, seeing in depth, uttering a sentence, or reaching for a target) is not something that anyone could argue is accomplished by a special-purpose mental faculty. Graphs are a recent invention and if they are an especially effective method of communication, it must be because they exploit general cognitive and perceptual mechanisms effectively. Any theory that hopes to explain the process of graph comprehension will have to identify the psychological mechanisms used in interpreting a graph, and which operating principles of each mechanism contribute to the overall ease or difficulty of the graph-reading process. Thus, any theory of graph comprehension will draw heavily on general cognitive and perceptual theory, and where our knowledge of cognitive and perceptual mechanisms is sketchy, we can expect corresponding gaps in our ability to explain the understanding of graphs.

I. WHAT IS A GRAPH?

There is a bewildering variety of graphs in current use, ranging from the line and bar graphs common in scientific journals, to drawings in popular magazines in which the thickness of two boxer's arms might represent the missile strength of the U. S. and Soviet Union, or in which the lengths of the rays of light emanating from a yellow disk might represent the price of gold in different months. Nonetheless, all graphs can be given a common characterization. Each graph tries to communicate to the reader a set of n-tuples of values on n mathematical scales, using objects whose visual dimensions (i.e., length, position, lightness, shape, etc.) correspond to the respective scales, and whose values on each dimension (i.e., an object's *particular* length, position, and so on) correlate with the values on the corresponding scales. The pairing is accomplished by virtue of the fact that any seen object can be described simultaneously by its values along a number of visual dimensions. For example, Fig. 4.1 represents a pairing of values on a nominal scale (countries) with values on a ratio scale (GNP) using objects (bars) whose horizontal position (a visual dimension) corresponds to a value on the first scale, and whose height (another visual dimension) corresponds to a value on the second scale.

Fig. 4.2 represents a pairing of values on an ordinal scale (months) with values on an interval scale (temperature) using objects (wedges) whose radial position represents the month, and whose darkness represents the temperature. This characterization, which can be applied to every graph I have seen, was first pointed out by Bertin (1967) in his seminal treatment of charts, graphs, and maps.

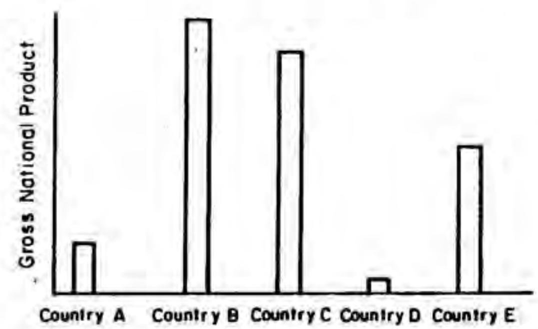


FIG. 4.1.

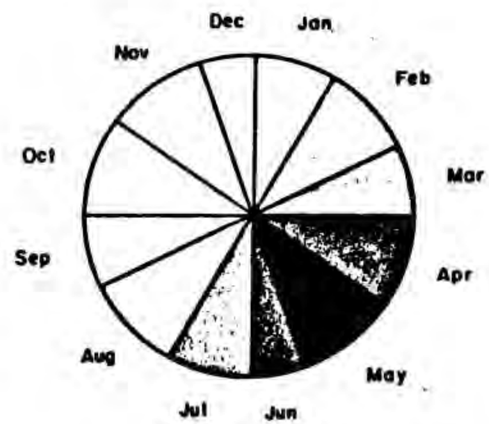


FIG. 4.2.

As Bertin points out, this characterization implies that a graph reader must do three things: (1) Identify, via alphanumeric labels, the conceptual or real-world referents that the graph is conveying information about (Bertin calls this "external identification"); (2) Identify the relevant dimensions of variation in the graph's pictorial content, and determine which visual dimensions corresponds to which conceptual variable or scale (Bertin's "internal identification"); and (3) Use the particular levels of each visual dimension to draw conclusions about the particular levels of each conceptual scale (Bertin's "perception of correspondence").

This simple observation implies that a graph reader must do two things. First, the reader must mentally represent the objects in the graph in only a certain way. In the case of Fig. 4.1, he or she must think of the bars in terms of their heights and their positions along the *x*-axis, but not necessarily in terms of the jagged contour formed by the tops of the bars, their distance from the edge of the page, and so on. Second, the graph reader must remember or deduce which aspects of the visual constituents of the graph stand for which of the mathematical scales that the graph is trying to communicate. In the theory to be described here, these two forms of knowledge are embodied in two types of mental representation: the *visual description*, which encodes the marks depicted on the page in terms of their physical dimensions, and the *graph schema*, which spells out how the physical dimensions will be mapped onto the appropriate mathematical scales. In using these structures to interpret a graph, a reader may obtain different sorts of information from it: the exact value of some scale paired with a given value on another scale, the rate of change of values on one scale within a range of values on another, a difference between the scale values of two entities, and so on. I will use the term *conceptual question* to refer to the particular sort of information that a reader wishes to extract from a graph, and *conceptual message* to refer to the information that the reader, in fact, takes away from it (cf. Bertin, 1967).

In the rest of the chapter, I characterize each of these representations explicitly, propose ways in which they are constructed and transformed in the course of reading a graph, and attempt to garner principles from perceptual and cognitive research dictating which aspects of these mental processes and representations affect the ease of extracting a message from a graph. These proposals will be justified by reference to concrete instances of graphs and other visual displays whose degree of intuitive difficulty is explained by the proposals, and to a number of experiments designed to test the proposals. Finally, a framework for further theoretical and applied research on graph comprehension will be outlined.

II. THE VISUAL ARRAY

The information in a graph arrives at the nervous system as a two-dimensional pattern of intensities on the retinas. I will use the term *visual array* to refer loosely to those early visual representations that depict the input in a relatively unprocessed, pictorial format (cf. the "primal sketch" and "2½ dimensional sketch" of Marr & Nishihara, 1978, and the "surface array" of Kosslyn, Pinker, Smith, & Shwartz, 1979). Information in this form is, of course, far too raw to serve as a basis for comprehending the meaning of the graph. For that, we need a representational format that can interface easily with the memory representations embodying knowledge of what the visual marks of the graph signify. Such memory representations cannot be stated in terms of specific distributions of light and

dark (or even lines and edges) as would be represented in the visual array, because vastly different intensity distributions (differing in size, orientation, color, shape, lightness, etc.) could all be equivalent exemplars of a given type of graph. Thus, the representation that makes contact with stored knowledge of graphs must be more abstract than a visual array.

III. THE VISUAL DESCRIPTION

A fundamental insight into visual cognition is that the output of the mechanisms of visual perception is a symbolic representation or "structural description" of the scene, specifying the identity of its parts and the relations among them (see Winston, 1975; Marr & Nishihara, 1977; Palmer, 1975, Pylyshyn, 1973). In this description, the various aspects of the scene, such as its constituent elements, and their size, shape, location, color, texture, and so on, together with the spatial relations among them, will be factored apart into separate symbols. As a result, each higher-level cognitive process need only refer to the symbols representing the aspect of the scene that is relevant to its own computations. I will use the term *visual description* to refer to the structural description representing a graph, and *visual encoding processes* to refer to the mechanism that creates a visual description from a visual array pattern.

Many such "languages" for visual descriptions have been proposed (Hinton, 1979; Marr & Nishihara, 1977; Miller & Johnson–Laird, 1976; Palmer, 1975; Winston, 1975). Most of them describe a scene using propositions, whose *variables* stand for perceived entities or objects, and in which *predicates* specify attributes of and relations among the entities. It is assumed that the visual encoding mechanisms can detect the presence of each of these predicates in the visual array (see Ullman, 1984, for explicit proposals concerning the sorts of mechanisms that are necessary to accomplish this). For example, one-place predicates specify a simple property of an object, such as *Circle* (*x*) (i.e., "*x* is a circle"), *Convex* (*x*), *Curve* (*x*), *Flat* (*x*), *Horizontal* (*x*), *Linear* (*x*), *Small* (*x*), and so on. Two-place predicates specify the relations between two objects, such as *Above* (*x*, *y*) (i.e., "*x* is above *y*"), *Adjacent* (*x*, *y*), *Below* (*x*, *y*), *Higher* (*x*, *y*), *Included-in* (*x*, *y*), *Points-toward* (*x*, *y*), *Parallel* (*x*, *y*), *Part* (*x*, *y*), *Near* (*x*, *y*), *Similar* (*x*, *y*) *Top* (*x*, *y*), and so on. Three and higher-place predicates indicate relations among groups of objects, such as *Between* (*x*, *y*, *z*) (i.e., "*x* is between *y* and *z*"), *In-line* (*x*, *y*, *z*), and so on. Parameterized predicates take a number of variables and a number of quantitative constants, such as *Area* (*x*, α) (i.e., "*x* has area α"), *Width* (*x*, α), *Location* (*x*, α, β), *Lightness* (*x*, α), *Orientation* (*x*, α), and so on. These predicates may also be appropriate for specifying continuous multidimensional attributes of

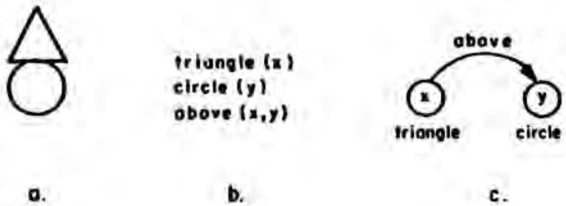


FIG. 4.3.

objects. For example, any member of a class of shapes ranging from a flattened horizontal ellipse through a circle to a flattened vertical ellipse can be specified by two parameters, representing the lengths of the major and minor axes of the ellipse, thus: *Ellipse* (x, α, β).

As is fitting for a discussion of graphs, I will use a graphic notation for visual descriptions. Each variable in a description will be represented by a small circle or *node* in which the variable name is inscribed (for simplicity's sake, I usually omit the variable name in these diagrams); each one-place predicate will simply be printed next to the nodes representing the variables that they are true of; and each two-place predicate will be printed alongside an arrow linking the two nodes representing the predicate's two arguments. Thus, a particular scene represented as the visual array in Fig. 4.3a will be represented as the visual description in Fig. 4.3b, or its graphic counterpart in Fig. 4.3c.

Constraining the Visual Description

If, as argued, a visual array representation is unsuitable for the computations involved in extracting information from a graph, an unconstrained visual description is not much better. Since any visual array can be described in an infinite number of ways, a theory that allowed any visual description to be built from a visual array would be unable to predict what would happen when a given individual faced a given graph. For example, the array in Fig. 4.4a can give rise not only to the visual description in Fig. 4.4c, but to the descriptions in Fig. 4.4b as well.

Clearly, if it is not to be vacuous, the theory must specify *which* visual description is likely to be constructed in a given situation, based on our knowledge of how the human visual system works. Of course, these constraints are simply the totality of our knowledge on perception. In the following section, I select four broad principles, each grounded in basic perceptual research, which constrain the form of visual descriptions in ways that are relevant to graph comprehension. These principles will bear a large explanatory burden in the theory to be outlined here, since later I

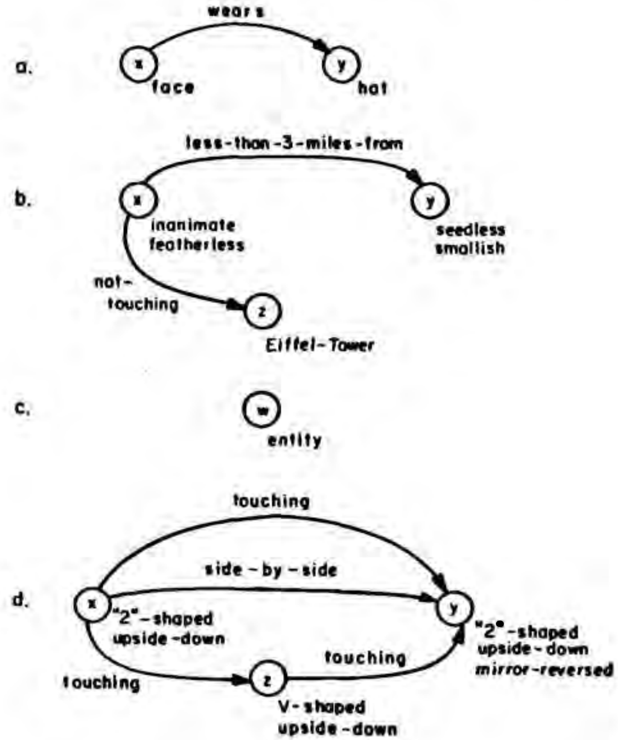


FIG. 4.4.

claim that a prime determinant of the difficulty of a graph will be whether the visual description specifies explicitly the visual dimensions and groupings that the graph maker recruited to symbolize the mathematical scales involved in the message of the graph.

A. The Indispensability of Space

It has long been known that an object's spatial location has a different perceptual status than its color, lightness, texture, or shape. Kubovy (1981) has addressed this issue systematically, and calls the two spatial dimensions of vision (plus the time dimension) *indispensable attributes*, analogous to the dimensions of pitch and time in audition. He defines the term "indispensable attribute" as an attribute with the following properties:

1. Perceptual Numerosity. The first constraint on a visual description must be on what is to count as a variable or node. Variables should stand for perceptual units of some sort, and not for any arbitrary subset of the light reflected from a scene (e.g., the set of all light patches whose dominant wavelength is divisible by 100). Kubovy points out that our perceptual systems pick out a "unit" or an "object" in a visual scene as any set of light patches that share the same spatial position, but *not* as a set of light patches that share some other attribute such as wavelength, intensity, or texture. Thus, Fig. 4.5a will give rise to the visual description in Fig. 4.5b, which partitions the array into three variables according to spatial location, rather than that in Fig. 4.5c, which partitions the array into two variables according to surface markings.

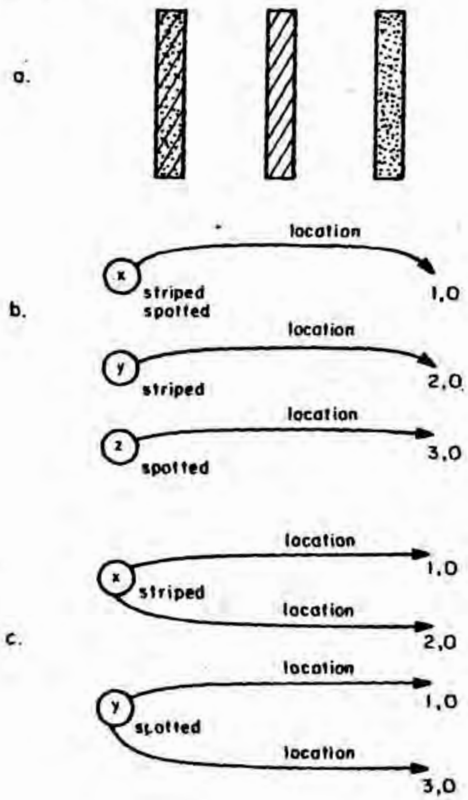


FIG. 4.5.

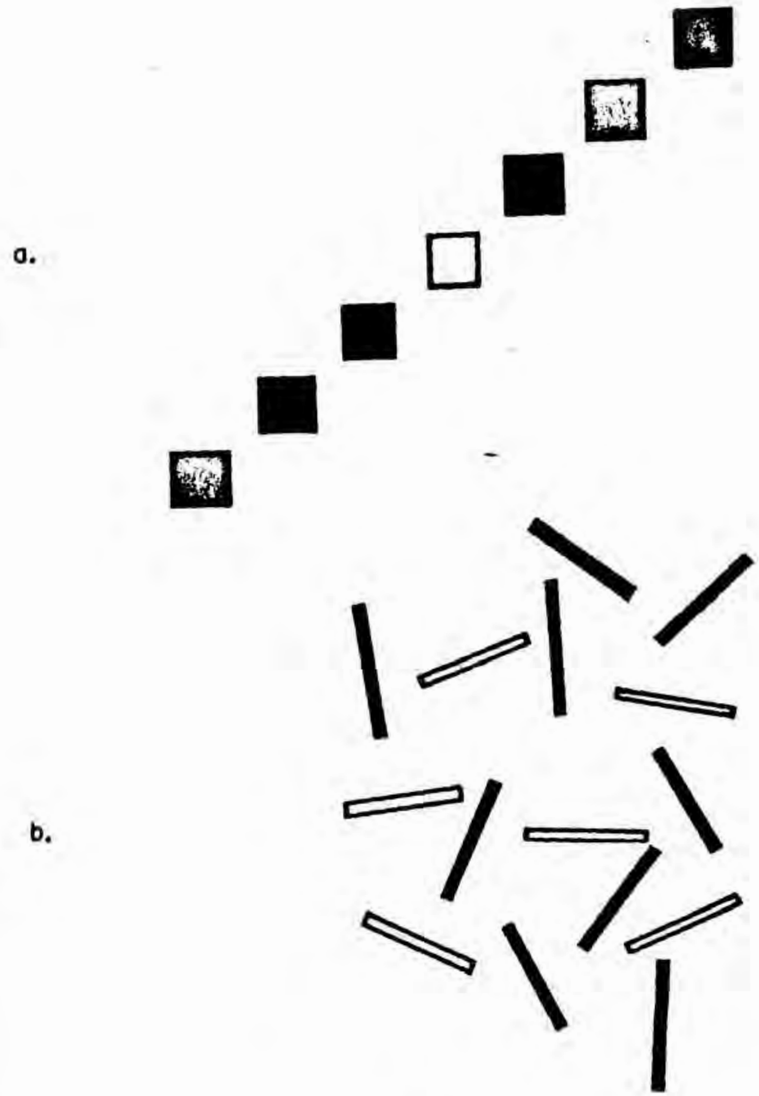


FIG. 4.6.

2. Configural Properties. The second constraint on a visual scene is the choice of predicates available in assembling a visual description. Naturally, there will be predicates corresponding to all perceptible physical dimensions (e.g., *bright* (x), *red* (x), *shiny* (x), *lightness* (x, α); in addition, there will be "configural" or "pattern" predicates corresponding to higher-order functions defined over the physical dimensions. Kubovy points out that most configural properties in a sensory modality are defined over the indispensable attributes, which in the case of static visual objects are the two spatial dimensions. As a consequence, there exist many predicates for spatial shapes (each of which can be defined by certain well-defined changes in relative horizontal and relative vertical positions in a pattern), but few for nonspatial "shapes" defined by analogous well-defined changes in other dimensions. For example, the array in Fig. 4.6a contains elements whose heights increase with their horizontal position (lightness varying randomly); the array in Fig. 4.6b contains elements whose lightnesses increase with their orientations (position varying randomly). However, the increase is immediately noticeable only in Fig. 4.6a, where the increase is of one spatial dimension with respect to another, not in Fig. 4.6b. Correspondingly, there exists a predicate *diagonal* (x) that can be used to describe the scene in Fig. 4.6a, but nothing analogous for describing the scene in Fig. 4.6b, whose elements would probably be specified individually. Note that as long as one member of a pair of related dimensions is spatial, there may be configural predicates available; when neither member is spatial, configural predicates are unlikely. Thus, the elements in Fig. 4.7 get darker with height, a change that, unlike that in Fig. 4.6b, is quickly noticeable, and may be captured by a single predicate (e.g., *lightness-gradient* (x)).

FIG. 4.7.

3. Discriminability and Linearity. It has been known for a century that physical variables are not in general perceived linearly, nor are small differences between values of a physical variable always noticed. In the visual description, this corresponds to numerical variables [e.g., *height* (x, *17*)] being distorted with respect to the real-world entities they represent, or to distinct numerical variables sharing the same value when the represented entities in fact differ [e.g., *lightness* (x, *17*); *lightness* (y, *17*) for two boxes differing slightly in lightness]. Kubovy remarks that indispensable attributes afford finer discriminations and more linear mappings then dispensable attributes, and indeed, the Weber fraction for spatial extent is 0.04, and the Stevens exponent is 1.0, both indicating greater accuracy for the representation of spatial extent than for the representation of other physical variables used in graphs.

4. Selective Attention. As a consequence of (1), each variable may have associated with it a unique pair of coordinates representing its location. This means that location could serve as an *index* or accessing system for visual information. This is a form of selective attention, and Kubovy summarizes evidence supporting the hypotheses that attention is more selective for indispensable attributes (two-dimensional location in vision) than for other visual attributes (see also Ullman, 1984). For example, one cannot easily attend to any visible object with a given shape, regardless of location (see Posner, Snyder, & Davidson, 1980). In the theory outlined in this chapter, selective attention according to location will consist of a mechanism that activates various encoding mechanisms to process a given spatial region of the visual array, in order to encode more predicates into the visual description or to verify whether a given predicate is true of the entity at that location. As we shall see, these mechanisms will play an important role in the "question-driven" or "top–down" processing of graphs.

B. Gestalt Laws of Grouping

The principles associated with the indispensability of space in vision place constraints on the parts of an array that variables may stand for, on how numerical variables represent physical continua, and on how predicates are encoded or verified with respect to the visual array. What is needed in addition is a set of principles governing how variables representing visual entities will be related to one another in visual descriptions, that is, how the atomic perceptual units will be integrated into a coherent percept. A notable set of such principles is the Gestalt Laws of Perceptual Organization (Wertheimer, 1938). These laws dictate that distinct static perceptual elements will be seen as belonging to a single configuration if they are near one another ("proximity"), similar in terms of one or more visual dimensions ("similarity"), smooth continuations of one another ("good continuation"), or parallel ("common fate") in the 2D plane. In terms of the

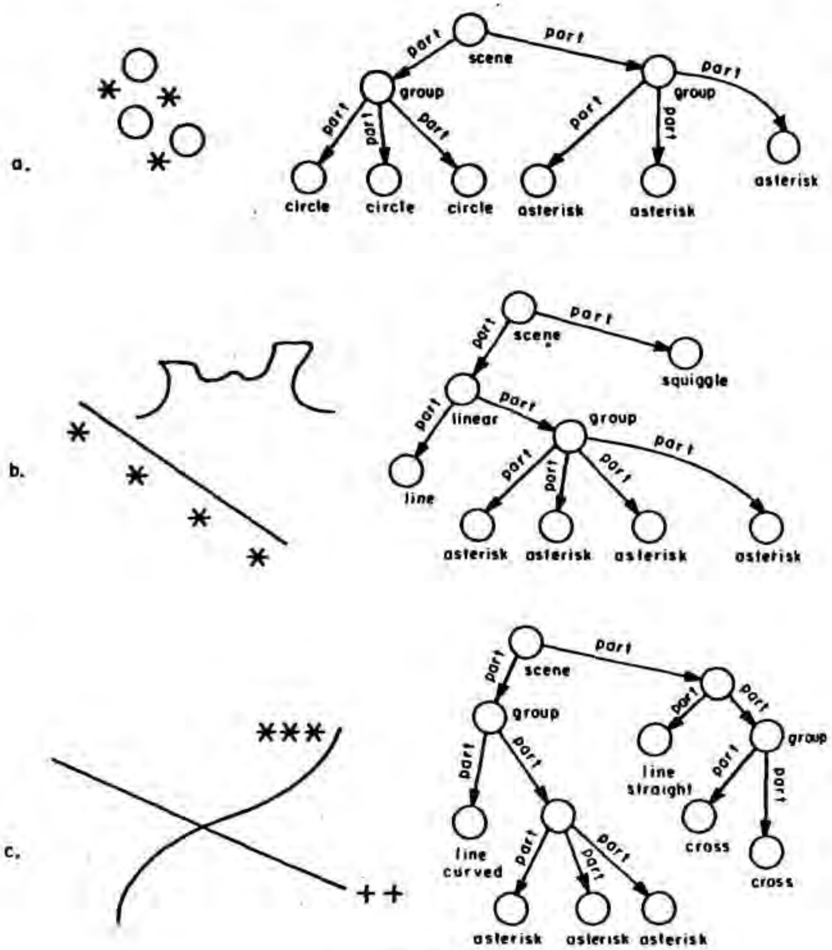


FIG. 4.8.

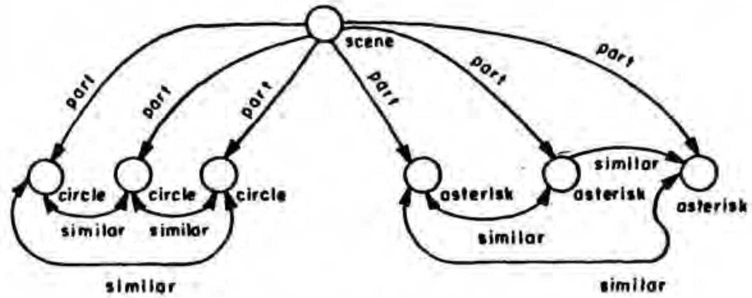


FIG. 4.9.

visual description, these principles will determine how variables are linked via the "part" relation in structures like those in Fig. 4.8a (where the law of similarity links asterisks to asterisks and circles to circles), Fig. 4.8b (where common fate links the asterisks to the line, and similarity links the asterisks to one another), and Fig. 4.8c (where good continuation keeps the straight and curved lines distinct, proximity links the asterisks and crosses to their respective lines, and similarity links asterisks to asterisks and crosses to crosses). Fig. 4.8 also shows how each collection of objects would be represented in a visual description.

There is another way of indicating the effects of grouping within visual descriptions. That is to link each member of a group to every other member using either the relation that gave rise to the grouping, or simply the relation "associated with." Thus, the visual array in Fig. 4.8a could also be represented as in Fig. 4.9: This notation can be used to indicate that the variables are grouped together perceptually, but not so strongly as to be a distinct perceptual unit. In the rest of this chapter, I use both notations for grouping, though no theoretical distinction need be implied by the choice.

C. Representation of Magnitude

Implicit in the earlier discussion of the psychophysics of visual dimensions was the assumption that these dimensions are represented by continuous interval scales in visual descriptions. Though the fine discriminations and smooth magnitude estimation functions found in psychophysical experiments strongly warrant this assumption, there is reason to believe that quantity can be mentally represented in other ways as well. First, there is evidence from experiments on the absolute identification of values on perceptual continua that people cannot remember verbal labels for more than about seven distinct levels of a perceptual continuum (Miller, 1956), and that in making rapid comparisons between remembered objects, subjects' reaction times are insensitive to the precise values of objects belonging to distinct, well-learned categories (Kosslyn, Murphy, Bemesderfer & Feinstein, 1977). Findings such as these suggest that quantity can also be represented (indeed, in memory *must* be represented, in certain circumstances) by one of a set of seven or so discrete symbols each specifying a portion of the range of quantities. These symbols could be signified by the Roman numerals I through VII.

Second, it is useful to distinguish between ratio values, where quantity is represented continuously but the units are arbitrary, and *absolute* values, where the units are well defined. The perception of pitch is a notorious example where a precise mental representation of a dimension is possible, but where for a majority of people, no absolute units can be assigned to the stimuli. Length, on the other hand, is an example of a continuum which people can judge either in ratio terms (e.g., one object being 1.7 times as long as another), or in terms of the well-known inches–feet–yards scale (e.g., Gibson, Bergman & Purdy, 1955). Indeed, whether subjects in magnitude estimation experiments are asked to use a well-learned versus their own arbitrarily selected modulus for estimated magnitude apparently affects their judgments (Stevens, 1961). Thus, interval descriptions must discriminate between these two forms of magnitude, which I will refer to an "interval-value" and "absolute-value," though ordinarily, visual descriptions will only contain "interval-value" propositions.

Finally, as every commercial sign maker can attest, values on a continuum that are extreme in comparison with values of that continuum for other objects in a scene are very likely to be perceptually encoded (as opposed to less extreme values, which are apt to be encoded only if attended to). To account for this salience principle, relatively extreme values will be represented redundantly in visual descriptions: in ordinary propositions such as *height* (x, α), as before, and also by special one-place predicates indicating the extremeness of the value along the particular dimension, such as *tall* (x), *bright* (x), *short* (x), and so on. When capacity limitations of visual descriptions are discussed later, it will be assumed that these special predicates have a very high probability of being encoded in the visual description.

D. Coordinate Systems

To express a unidimensional quality like lightness, one need specify in advance only the origin and the units of the scale to be used. However, for objects that vary along a number of continua, such as the position of an object on a two-dimensional piece of paper, or rectangles in a set varying in height and width, one has to specify how the variation will be partitioned

into dimensions and how each dimension will be represented. This is the issue of which *coordinate system* is appropriate to represent an object in a set varying along several dimensions. This involves questions about whether a polar or a rectangular coordinate system is used, whether there is a single or multiple origins, and so on. In their influential paper on shape recognition, Marr and Nishihara (1978) proposed that memory representations of shape are specified with respect to object-centered cylindrical coordinate systems. Furthermore, the coordinate systems are *distributed*: Instead of there being a global coordinate system with a single origin and set of axes, there is a cylindrical coordinate system centered on the principal axis of the object (e.g., in the case of an animal, its torso), in which are specified the origins and axes of secondary coordinate systems each centered on a part of the object attached to the principal axis (e.g., the animal's head and limbs). These secondary coordinate systems are, in turn, used to specify the origins and axes of smaller coordinate systems centered on the constituent or attached parts of the secondary part (e.g., the thigh, shin, and foot of the leg), and so on. I will adopt here the following aspects of Marr and Nishihara's theory: (1) Shapes and positions are mentally represented principally in polar or rectangular coordinates (the former is just a slice of a cylindrical coordinate system orthogonal to its axis; the latter is just a slice of a cylindrical coordinate system including its axis). (2) The locations of the different elements of a scene are represented in separate, local coordinate systems centered upon other parts of the scene, not in a single, global coordinate system. This means that in the visual description, the specification of locations (and also of directions and of parameterized shapes) of objects will be broken down into two propositions, one specifying the object upon which the coordinate system will be centered, the other specifying the extent or value of the object within the coordinate system, as in Fig. 4.10.

In fact, it is generally more perspicuous to indicate the extent along each dimension, and the location of the axis of the coordinate system corresponding to that dimension, separately, as in Fig. 4.11.

The important question of which objects may serve as the coordinate system for which other objects is only beginning to be answered in the vision literature, but the following condition seems to be a plausible first approximation: A spatial property of object *a* will be mentally specified

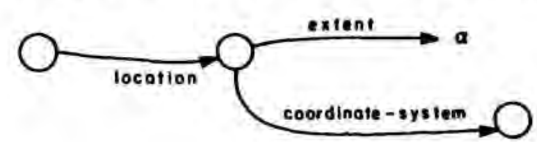


FIG. 4.10.

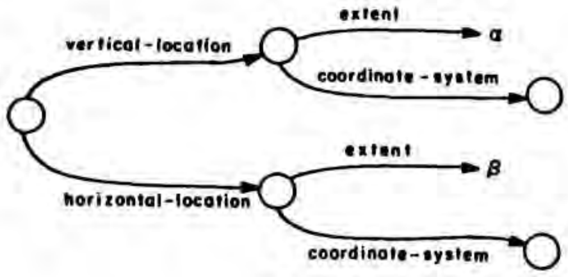


FIG. 4.11.

in a coordinate system centered on object *b* when: (1) *b* is larger than *a*, and (2) *a* and *b* are perceptually grouped according to one or more of the Gestalt laws.

IV. PROCESSING CONSTRAINTS ON VISUAL DESCRIPTIONS

Since, with deliberate effort, people can probably encode an unlimited number of properties (e.g., the angle formed by imaginary lines connecting a standing person's right thumbnail, navel, and right kneecap), visual descriptions can, in principle, be arbitrarily large. In practice, however, two factors will limit the size of visual descriptions:

1. Processing Capacity. Most models of cognitive processing have restrictions on the processing capacity used to maintain the activation of nodes in a short-term visual description (Anderson & Bower, 1973; Newell & Simon, 1973). Specifically, it is claimed that between four and nine nodes may be kept active at one time, fewer if processing resources are being devoted to some concurrent task. This limitation reflects the well-known finiteness on human immediate memory and processing capacity (Miller, 1956).

2. Default Encoding Likelihood and Automaticity. As mentioned, any predicate in a persons's visual repertoire can be added to a visual description in response to higher-level processes testing for the presence of a particular predicate applied to a particular variable (e.g., "is *x* a square?"). However, before these top–down processes come into play, a number of predicates will be assembled into a visual description, because they are "just noticed." Different predicates have different probabilities of being encoded under these "default" circumstances. Presumably, some predicates innately have a high default encoding likelihood [e.g., *enormous* (*x*), *dazzling* (*x*)], whereas the default encoding likelihood of others is determined by familiarity and learned importance. Shiffrin and Schneider (1977) and Schneider and Shiffrin (1977) propose that when a person frequently assigns a visual pattern into a single category, he or she will come to make that classification "automatically," that is, without the conscious application of attentional or processing capacity. Translated into the present vocabulary, this means that frequently encoded predicates will have a high *default encoding likelihood*. A number of experiments applying Shiffrin and Schneider's proposals to the learning of visual patterns confirm that the recognition of patterns becomes rapid, error-free, and relatively insensitive to other attentional demands as the patterns become increasingly well practiced.

Therefore, it is important to distinguish among several sizes of visual descriptions. A description that is assembled automatically by purely data-driven (as opposed to top–down or conceptually driven) encoding processes will be called the "default visual description." Its composition will be determined by the relative "default encoding likelihoods" of the various predicates satisfied by the visual array. In contrast, a description that is shaped by conceptual processes testing for the presence of visual predicates at particular locations in the array will be called an "elaborated visual description." Visual descriptions can also be classified in terms of whether short-term memory limitations are assumed to be in effect. A small visual description such as can be activated at a given instant will be called the "reduced visual description"; a visual description that includes all the predicates whose default encoding likelihood are above a certain minimum, plus all the predicates that are successfully tested for by top–down processes, will be called the "complete visual description." The complete visual description will correspond to the description encoded by a hypothetical graph reader with unlimited short-term memory, or to the description integrating the successive reduced descriptions encoded by a normal graph reader over a long viewing period. One way to think quantitatively of the size of the default visual description that a person will encode is to suppose that the probability of a given true predicate's entering into a visual description is a function of its default encoding likelihood multiplied by a constant between zero and one corresponding to the amount of capacity available (i.e., not devoted to other concurrent tasks). When the constant is one, the resulting description will be a "complete" visual description; as the constant decreases with decreasing available processing capacity, the size of the description will be reduced accordingly. I adopt the final assumption that the level of activation of a node begins to decrease steadily as soon as it is activated, but that the reader can repeatedly re-encode the description by reattending to the graph (this simply corresponds to the process of decay and rehearsal in short-term memory, see Crowder, 1976). Since encoding is probabilistic, the description will differ in composition somewhat from one encoding to the next.

V. AN EXAMPLE

Now that we have some constraints on the size and composition of visual descriptions, we can examine how the visual appearance of a particular graph might be described mentally. This will be the first step in working through an example of how a graph is understood according to the current theory. The example, shown in Fig. 4.12, is a bar graph plotting the price per ounce of a precious metal called "graphium" over a 6-month period. A complete "default" visual description is shown in Fig. 4.13. (Dotted lines represent propositions, omitted for the sake of clarity, that may be deduced from nearby propositions for similar parts).

Most aspects of this visual description are motivated by the constraints outlined in the previous section. The scene is parsed into subscenes, each occupying a distinct location in the visual array (though for readability's sake, the locations for the subscene nodes will not always be printed in subsequent examples). This parse is done according to the Gestalt principles, yielding separate nodes for the "L"-shaped framework and for the group of bars. By those same principles, the framework is connected by the "part" predicate to nodes representing its vertical and horizontal segments, and each of these is linked by "near" predicates to nodes representing the conceptual meaning of that text. Of course, the meaning of

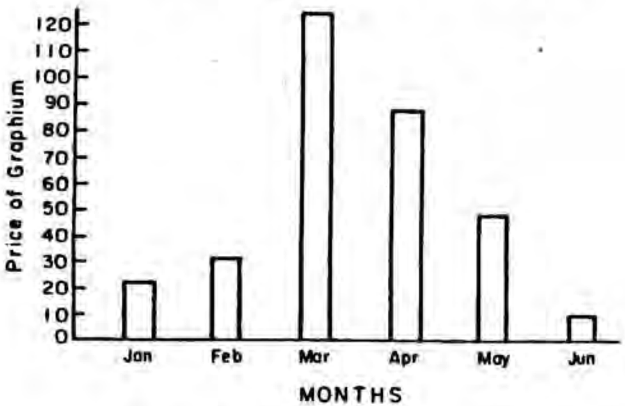


FIG. 4.12.

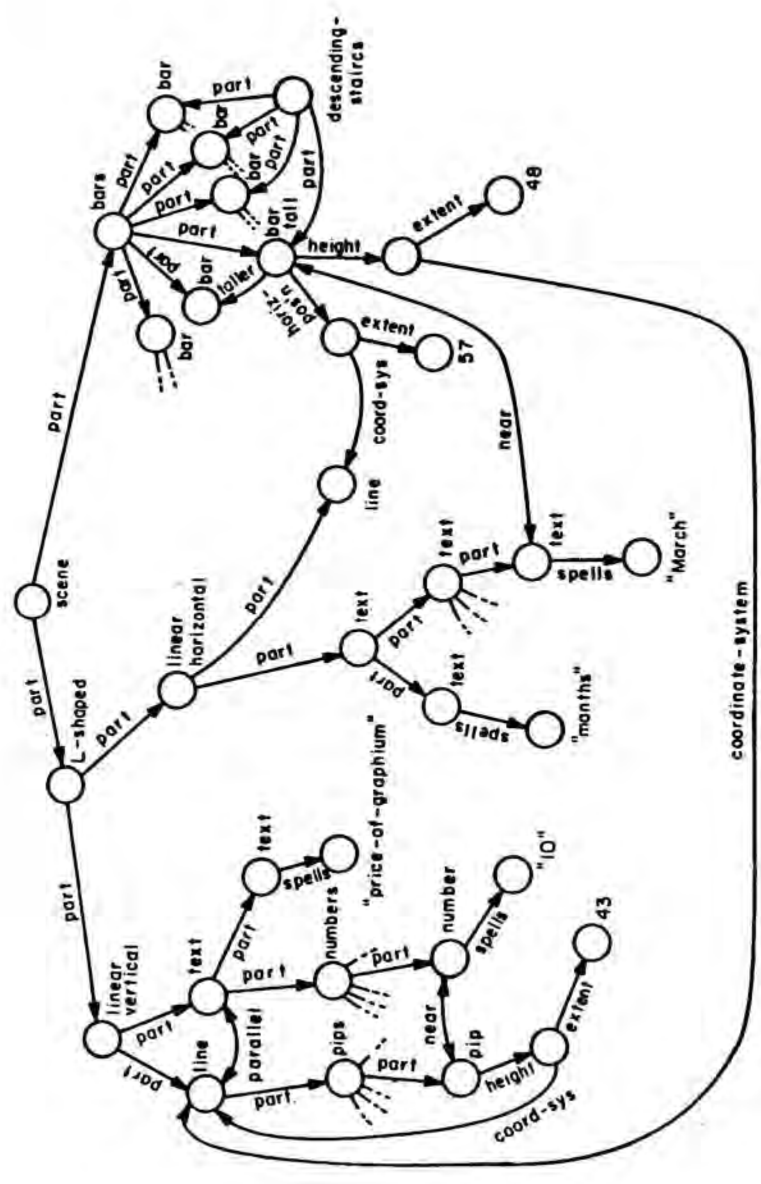


FIG. 4.13.

expressions such as "price of graphium" is, in all likelihood, mentally represented by an assembly of nodes linked in complex ways to the nodes representing the visual appearance of the text, but since the process of reading text is not of concern here, this simplified notation will suffice (the predicate associated with these "meaning" nodes will be replaced within quotation marks to indicate that they are not in fact unitary predicates). Predicates for the "bar" shape are attached to each bar node; the "tall" predicate is attached to the salient tallest bar; a pair of particularly discrepant bars is connected by the predicate "taller-than"; and the set of four progressively shorter bars is grouped together under its own node with its own shape predicate "descending-staircase." Finally, the height and horizontal position of each bar is specified with respect to a coordinate system centered on the appropriate framework segment, due to the framework's being larger than the bars and associated with them by proximity and common fate.

VI. CONCEPTUAL MESSAGES, CONCEPTUAL QUESTIONS

We now have an example of the immediate input to the graph comprehension process. Before specifying the process, it would be helpful to know what its output is as well. One can get a good idea of what that output must be simply by looking at a graph and observing what one remembers from it in the first few moments of seeing it or after it has just been removed from view. In the case of the graph in Fig. 4.12, one might notice things like the following: (1) the price of graphium was very high in March; (2) the price was higher in March than in the preceding month; (3) the price steadily declined from March to June; (4) the price was $20/ounce in January; (5) the price in June was *x* (where *x* is a mental quantity about half of that for January, about a fifth of that for May, etc.). Basically, we have a set of paired observations here, where the first member can be a particular value of the independent variable (e.g., "March"), a pair of values (e.g., "March vs. February"), or a range of values (e.g., "the last 4 months"). The second number of each pair can be a ratio value (e.g., a value *x* along some mental ratio scale), an absolute value (e.g., "$20/ounce"), a difference (e.g., "larger"), a trend (e.g., "decreasing"), or a level (e.g., "high"). (See Bertin, 1967, for a taxonomy of such questions). This information can be expressed in a representation consisting of a list of numbered entries, each specifying a pair (or, for more complex graphs, an *n*-tuple) of variables, the extent or type of each independent variable (e.g., ratio-value, pair, range), and the value (or difference or trend) of the corresponding dependent variable. Thus, the conceptual message representing the infor-

mation which we are assuming has been extracted from the graph in Fig. 4.12 will look like this (the intuitive meaning of each entry can be made clearer by assuming the entry is a sentence beginning with the word *when*):

1: V_1 absolute-value = March,	V_2 level = high
2: V_1 pair = March & February,	V_2 difference = higher
3: V_1 range = March–June,	V_2 trend = decreasing
4: V_1 absolute-value = January,	V_2 absolute-value = $20/oz.
5: V_1 absolute-value = June,	V_2 ratio-value = x.

In general, conceptual messages will be of the following form:

i:	V_a	ratio-value or absolute-value or pair or range	$= \alpha$,	V_b	ratio-value or absolute-value or pair or range

$= \beta, \ldots$

i designates the *i*th of an arbitrary number of entries (in principle), V_a designates the *a*th of an arbitrary number of variables, and α designates a specific value in a form appropriate to the entry (e.g., a "higher" or "lower" primitive symbol if the entry specifies a difference between values of the second variable corresponding to a pair of values of the first).[f] Note that the variables are differentiated by subscripts instead of being named by their real-world referents (e.g., month); this was done in recognition of people's ability to extract a great deal of quantitative and qualitative information (indeed, virtually the same information) when a graph has *no* labels at all, leaving the referents of the variables unknown. When the referents are known, the conceptual message can indicate this with entries such as the following:

6: V_1 = months, V_2 = price-of-graphium.

Presumably, when the reader has integrated all the information he or she wishes to extract from the graph, he or she can make the message representation more economical by replacing each V_i by its associated referent symbol.

[f] It is possible to have several equations in an entry refer to the same variable, e.g.: 17: V_1 absolute-value = 14, V_1 ratio-value = 132, V_1 level = high, V_2 level = low.

From here, it is a simple matter to devise a notation for conceptual questions. (Recall that a conceptual question is a piece of information that the reader desires to extract from a graph). One can simply replace the α or β in the generalized entry by the "?" symbol, indicating that that is the unknown but desired information. Thus, if a person wishes to learn the price of graphium during the month of April, we posit that he or she has activated the representation

7: V_1 absolute-value = April, V_2 absolute value = ?.

If the reader wishes to learn the trend of graphium prices during the first 2 months, he or she sets up the representation

8: V_1 range = January-February, V_2 trend = ?.

If the reader wishes to learn the month in which graphium prices were low, he or she activates.

9: V_1 absolute-value = ?, V_2 level = low,

and so on.

VII. THE GRAPH SCHEMA*

So far, the theory has implicated the information flow diagram in Fig. 4.14. Now, one must specify the unknown component labeled with a "?." From the flow chart, we can see what this component must do: (1) It must specify how to translate the information found in the visual description into the conceptual message, and (2) It must specify how to translate the request found in a conceptual question into a process that accesses the relevant parts of the visual description (culminating as before in one or more entries in the conceptual message). Furthermore, since (1) and (2) will involve different sorts of translations for different types of graphs (e.g., for line graphs versus bar graphs), the unknown component will also have to: (3) Recognize which type of graph is currently being viewed. The structure that accomplishes these three tasks will be called a *graph schema*, and it, together with the processes that work over it, will be discussed in this section.

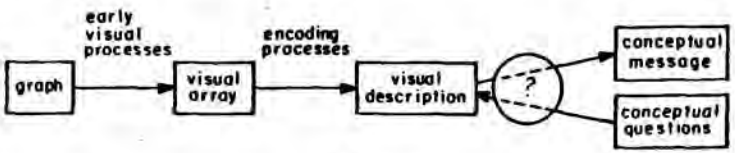


FIG. 4.14.

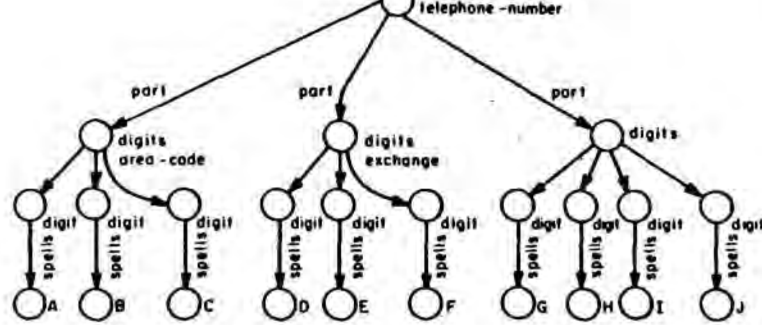


FIG. 4.15.

A. Schemas

I take a "schema" to be a memory representation embodying knowledge in some domain, consisting of a description which contains "slots" or parameters for as yet unknown information. Thus, a schema can specify both the information that must be true of some represented object of a given class, and the sorts of information that will vary from one exemplar of the class to another (see Minsky, 1975; Winston, 1975; Norman & Rumelhart, 1975; Bregman, 1977; Schank & Abelson, 1977). To take a simple example unrelated to graphs, Fig. 4.15 could be a schema for telephone numbers, specifying the number and grouping of the digits for any number but not the identity of the digits for any particular number, these being represented by the parameters *A-J*.[2]

This schema can be *instantiated* for a given person, becoming a representation of his or her telephone number, by replacing the parameters labeling the lowermost nodes by actual numerical predicates. In doing so, one is using the schema to *recognize* a candidate character string as a telephone number, by matching the schema against a visual description of the candidate string. The visual description of an as yet unrecognized number will be identical to the schema, except that it lacks the conceptual nodes such as "area code" and "exchange," and that it contains constants in place of parameters. Once the schema is instantiated by the visual description, one can use it to *retrieve* desired information about the telephone number using a node-by-node net searching procedure (i.e., one can quickly

---

[2] These uppercase parameters, which stand for unknown predicates, should not be confused with lowercase variables, which stand for perceptual entities and correspond to nodes in the visual description (although usually, the variable itself is omitted and only the node is depicted).

find "the first digit of the exchange" without searching the entire string, by starting at the top node and following the appropriate arrows down until the bottom node labeled by the desired number is reached). The double labeling of nodes is what allows schemas to be used both for recognition and for searching: a visual description of a to-be-recognized pattern will contain labels such as "digit," but not "area code," so the "digit" labels in the schema are necessary for recognizing the object. However, the search procedures will be accessing conceptual labels such as "area code," so these are necessary, too.

B. Graph Schemas: A Fragment

It seems, then, that a schema of this sort for graphs might fulfill two of our three requirements for graph knowledge structures: recognizing specific types of graphs, and directing the search for desired pieces of information in a graph. What we now need is some device to *translate* visual information into the quantitative information of the type found in the conceptual message. These devices, which I will call *message flags*, consist of conceptual message equations, usually containing a schema parameter, which are appended to predicates (nodes or arrows) in the graph schema. When such a node or arrow is instantiated by a particular visual description for a

a.

b.

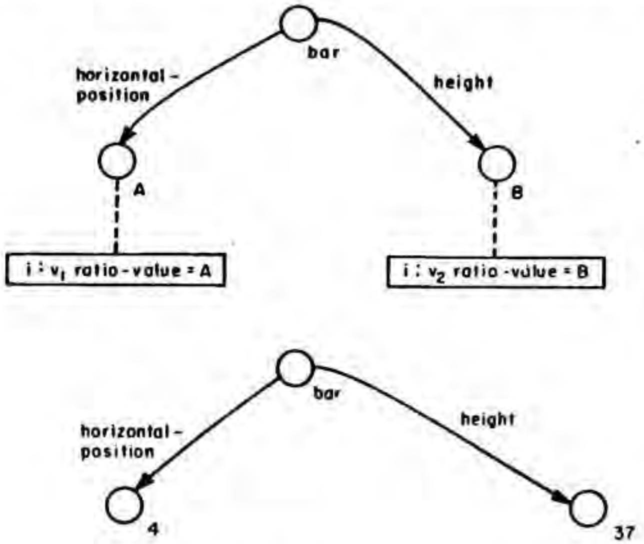


FIG. 4.16.

graph, the parameters in the message flag are replaced by the corresponding value in the instantiated schema, and the equation is added to the conceptual message. Fig. 4.16a illustrates equation flags for a fragment of a bar graph schema (the flags are enclosed in rectangles and are attached to the nodes they flag by dotted lines).

When a reader encounters the graph represented by the fragment of a visual description in Fig. 4.16b (the numbers representing values along a mental ratio scale with arbitrary units), he or she can instantiate the schema (i.e., replace the parameters A and B by the values 4 and 37), and add an entry to the conceptual message. All equations sharing a given *i* prefix are merged into a single entry, and each *i* is replaced by a unique integer when the entry is added to the conceptual message. Thus, the following entry is created:

$$1: V_1 \text{ ratio-value} = 4,\ V_2 \text{ ratio-value} = 37$$

This informal sketch should give the reader a general idea of how the graph schema is used in conjunction with the visual description to produce a conceptual message. In the sections following, I present a comprehensive bar graph schema and define more explicitly the processes that use it.

C. A Bar Graph Schema

Fig. 4.17 presents a substantial chunk of a schema for interpreting bar graphs. It is, intentionally, quite similar to the visual description for a bar graph in Fig. 4.13. The graph is divided into its L-shaped framework and its pictorial content, in this case, the bars. The framework is divided into the abscissa and the ordinate, and each of these is subdivided into the actual line and the text printed alongside it. In addition, the pips crosshatching the ordinate, together with the numbers associated with them, are listed explicitly. The height and horizontal position of each bar are specified with respect to coordinate systems centered on the respective axes of the framework, and each bar is linked to a node representing its nearby text. An asterisk followed by a letter inside a node indicates that the node, together with its connection to other nodes, can be duplicated any number of times in the visual description. The letter itself indicates that each duplication of the node is to be assigned a distinct number, which will appear within the message flags attached to that instance of the node.

The message flags specify the conceptual information that is to be "read off" the instantiated graph schema. They specify that each bar will contribute an entry to the conceptual message. Each entry will equate the ratio value of the first variable (referred to in the description as "IV," for Independent Variable) with the horizontal position of the bar with respect to the abscissa, and will equate the ratio value of the second variable (the

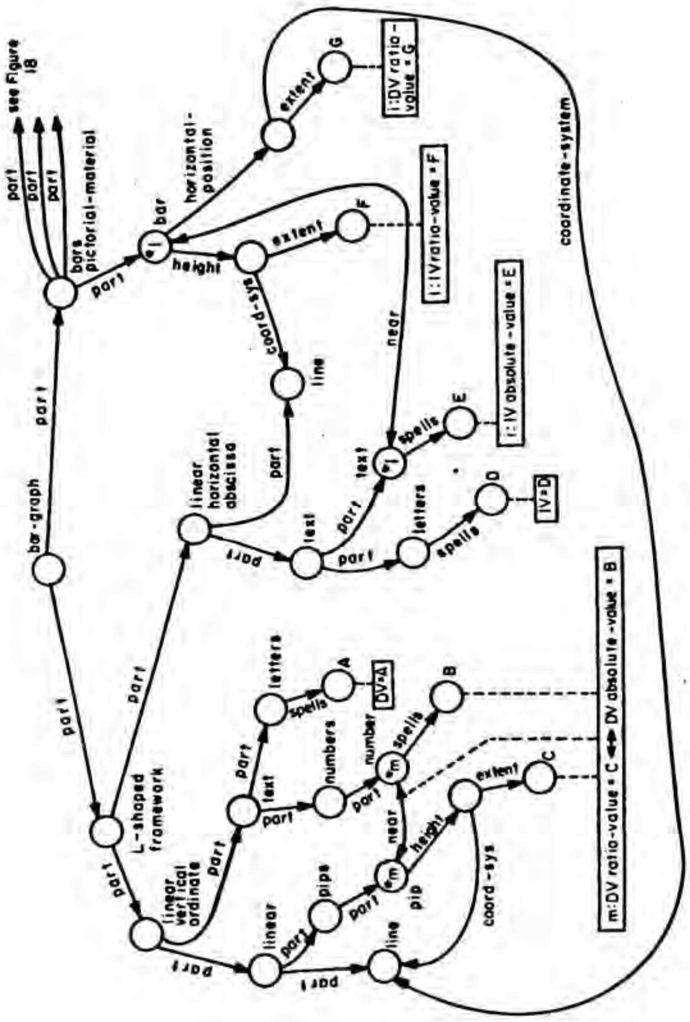


FIG. 4.17.

"DV" or Dependent Variable) with the bar's height with respect to the ordinate. In addition, the absolute value of the independent variable for an entry will be equated with the meaning of whatever label is printed below it along the abscissa. Finally, the referents of each variable will be equated with the meaning of the text printed alongside its respective axis.

In devising these formalisms, I was at one point distressed that there was no straightforward way to derive absolute values for the dependent variable. The ratio value of each bar, corresponding to its height, could easily be specified, but since the absolute values are specified in equal increments along the ordinate, far from most of the bars, and specific to none of them, no simple substitution process will do. However, a simple glance at a bar graph should convince the reader, as it convinced me, that this is not a liability but an asset. The absolute value of the dependent variable at a given level of the independent variable is indeed *not* immediately available from a bar graph. Instead, one seems to assess the height of a bar in terms of some arbitrary perceptual or cognitive scale, and then search for the pip along the ordinate whose vertical position is closest to that height or mentally extrapolate a horizontal line until it hits the ordinate (see Finke & Pinker, 1983). The number printed next to the nearest pip, or a number interpolated between the numbers printed next to the two nearest pips, is deduced to be its absolute value. In contrast, the absolute value of a given level of the independent variable (i.e., which month it is), or the relative values of the dependent variable (e.g., its maximum and minimum values, its trends, or differences between adjacent values) seem available with far less mental effort. The most natural mechanism for representing absolute values of the dependent variable within the bar graph schema, and the one that happens to be in accord with the actual difficulty of perceiving these values, is to add to the conceptual message special entries asserting an equivalence between a certain level of the referent's absolute value and a certain level of the referent's ratio value, each entry derived from a labeled pip on the ordinate. The leftmost message flag in Fig. 4.17 sets up these entries; the symbol "=" indicates that the two equations are equivalent. Presumably, higher-level inferential processes, unspecified here, can use these equivalence entries to convert ratio values to absolute value within other entries in the conceptual message, calculating interpolated values when necessary.[3]

Earlier, we mentioned that the visual system can encode predicates that

[3] The schema presented here perhaps unfairly anticipates that the bar-graph example will have individual labels for each bar along the absicssa and a graduated scale along the ordinate. In fact, graduated scales often appear along the abscissas of bar graphs as well. In a more realistic bar-graph schema, the subschemas for the pips of a graduated scale would be appended to the abscissa as well as to the ordinate.

stand for well-defined groups of objects, and also that conceptual messages can contain entries specifying a trend of one variable over the range of another. An implication of the theory, then, is that graph readers (or at least experienced graph readers) should be able to translate directly a higher-order perceptual pattern, such as a group of bars comprising a staircase, into the quantitative trend that it symbolizes, without having to compute the trend by successively examining each element. Furthermore, the difference in height between a pair of adjacent bars might be encodable into a single predicate, which should be directly translatable into an entry expressing a difference in the symbolized values. Also, a salient perceptual entity might be encoded as extreme (independently of the encoding of its precise extent on a ratio scale), and this should be directly translatable into an entry expressing the extremeness of its corresponding variable value, again without the mediation of ratio scale values. These direct translations at several levels of globality, we shall see, play an important role in predicting the difficulty of a graph or the effectiveness of a graph reader. In the theory, the translations are accomplished by the message flags in Fig. 4.18 (which should actually be part of Fig. 4.17, but is depicted separately for the sake of clarity). Fig. 4.18 shows that bars in a graph can be described not only in terms of their heights and horizontal positions, but also in terms of being extremely tall or short, in terms of differences between the heights of adjacent pairs, or in terms of groups that constitute a perceptual whole. In each case the appropriate equation is attached to the predicate which encodes the attribute. Two additional notational conventions are introduced in the figure: the location of a pattern that occupies an extended region of the array is specified by its endpoints along a ratio scale (i.e., H–I), both in the visual description and in the conceptual message. In addition, one of the equation flags for a pair of bars makes reference to nodes standing for the bars themselves, p_j and q_j, rather than for an attribute like horizontal position. It is assumed that when a pair of bars is encoded as a pair, some information about each bar is encoded as well. This information, be it ratio value, absolute value, or level, can then be linked with or substituted for appropriate symbols for the bars (p_j or q_j) within the entry for the pair.

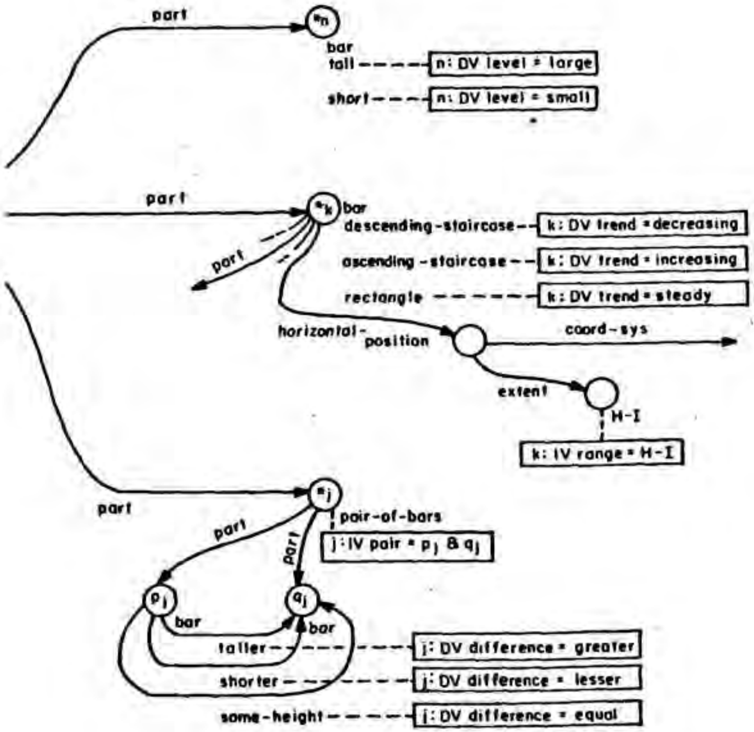


FIG. 4.18.

VIII. PROCESSES

In the account so far, I have relied upon the intelligence and cooperativeness of the reader to deduce how the various structures are manipulated and read during graph comprehension. In order to use the theory to make predictions, it will be necessary to define explicitly the procedures that access the structures representing graphic information. Four procedures will be defined: a *MATCH* process that recognizes individual graphs as belonging to a particular type, a *message assembly* process that creates a conceptual message out of the instantiated graph schema, an *interrogation* process that retrieves or encodes new information on the basis of conceptual questions, and a set of *inferential processes* that apply mathematical and logical inference rules to the entries of the conceptual message.

A. The MATCH Process

The term is borrowed from Anderson and Bower's (1973) theory of long-term memory. This process compares a visual description in parallel with every memory schema for a visual scene, computes a goodness-of-fit measure for each schema (perhaps the ratio or difference between the number of matching nodes and predicates and the number of mismatching nodes and predicates), and selects the schema with the highest goodness-of-fit measure. This schema, or rather, the subset of the schema that the limited capacity processes can keep activated, is then instantiated (i.e., the parameters in the schema are replaced by the appropriate constants found in the visual description). This is the procedure, alluded to in vague terms before, that uses the graph schema to recognize a graph as being of a certain type (e.g., bar graph, pie graph).[4]

B. Message Assembly

This process accomplishes the translation from visual information to conceptual information, also alluded to in previous sections. It searches over the instantiated graph schema, and when it encounters a message flag, it adds the message it contains to the conceptual message, combining into a single entry all equations sharing a given prefix (i.e., all those beginning with the same *i*:). It is assumed that at the time that the MATCH process instantiated the parameters of the graph schema, the parameters within the message flags were instantiated as well.

Memory and processing limitations imply that not every message flag in the graph schema is converted into an entry into the conceptual message. Some may not be instantiated because the visual description was reduced or because the default encoding likelihood of the predicate was low; some may not be instantiated because of noise in the MATCH process; and some may be skipped over or lost because of noise in the message assembly process. For these reasons, we need a process that adds information to the conceptual message in response to higher-level demands.

C. Interrogation

This process is called into play when the reader needs some piece of information that is not currently in the conceptual message (e.g., the difference between two values of the dependent variable corresponding to a

[4] This process has been oversimplified in several ways, in accordance with certain oversimplications in the graph schema itself. For one thing, conceptual labels such as "abscissa" do not appear in visual descriptions, and so should not count in the goodness of fit calculations. This could be accomplished by distinguishing the conceptual or graph-specific predicates from the rest, perhaps by listing them, too, as message flags, which are "read off" the schema, but not used to instantiate it. The second complication is that different nodes and predicates should count differently in the recognition process. Some might be mandatory, some might be mandatorily absent, some might be characteristic to various degrees, some might occur in sets from which one member must occur, and so on. There are several ways of accomplishing this, such as the introduction of logical operators into schemas, or the use of a Bayesian recognition procedure, but limited space prevents me from outlining them here (see Anderson, 1976; Anderson & Bower, 1973; Minsky, 1975; Smith, Shoben, & Rips, 1974; Winston, 1974).

given pair of independent variable values). As mentioned, each such request can be expressed as a conceptual message entry with a "?" replacing one of the equation values. The interrogation process works as follows: The message flag within the graph schema that matches the conceptual question (i.e., is identical to it except for a constant or parameter in the place of the "?") is activated. If it already contains a constant (i.e., if the equation it contains is instantiated, and thus, complete), the equation is simply added to the conceptual message. If it contains a parameter (i.e., is incomplete), the part of the visual description that corresponds to that branch of the schema is checked to see if it contains the desired constant (e.g., if a certain ratio-value of the dependent variable is desired, the visual description is checked for the presence of a constant attached to the node representing the bar's height). If this constant is absent from the visual description, the encoding process for the relevant predicate (e.g., the process that encodes height) is commanded to retrieve the desired information for the relevant part in the visual array. It can do so by using the retinal coordinates attached to the node for the part, which are assumed to be present in the visual description (though they have been omitted from the diagrams in this chapter). Often, however, these coordinates will have decayed, and the coordinates of an associated part together with the degree and direction of the association will be used to direct the encoding process to the correct location in the visual array. In other words, the conceptual question can initiate a top–down search for the desired part or part parameter in the array. Once the desired information is encoded into the visual description, it can be instantiated in the schema and its message flags, and the instantiated equation within the flag can be added to the conceptual message.

D. Inferential Processes

Human intelligence consists of more than the ability to read graphs. In the category *inferential processes*, I include the ability to perform arithmetic operations on the quantitative information listed in the conceptual message (e.g., calculating the rate of increase of a variable by subtracting one value from another value and dividing by a third value), to infer from the context of the graph (e.g., the paragraph in which it is embedded) what information should be extracted from the graph, to draw qualitative conclusions relevant to some domain of knowledge based on the information in the graph, and so on. Naturally, I have little to say about these abilities here; they are part of the study of cognition in general and not the study of graph comprehensions. However, I mention them here because many types of information can be obtained either directly from a conceptual message or indirectly from inferential processes operating on the conceptual message.

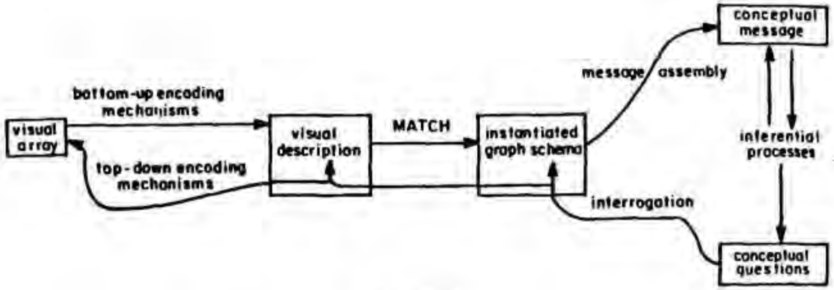


FIG. 4.19.

Which method is used, we shall see, affects the difficulty of a graph and the efficiency of a graph reader.

The flow of information specified by the current theory is summarized in Fig. 4.19, where blocks represent information structures and arrows represent processes that transfer information among them.

IX. WHERE DO GRAPH SCHEMAS COME FROM?

The graph schema discussed so far embodies knowledge of bar graphs (in fact, a subset of bar graphs). Clearly, the theory must also account for people's ability to read other common types of graphs (line graphs, pie graphs, pictograms, etc.) and to understand completely novel forms of graphs as well (e.g., one in which the length of a ray of light emitted from a disk represents the price of gold in a given month). I propose that people create schemas for specific types of graphs using a *general graph schema*, embodying their knowledge of what graphs are for and how they are interpreted in general. A plausible general graph schema is shown in Fig. 4.20. There are three key pieces of information contained in the schema. First, some objects, or parts of objects (i.e., a display's pictorial content) are described in terms of several visual attributes. Each visual attribute symbolizes a conceptual variable, and the set of values of the n visual attributes encoded for an object or object part corresponds to a particular n-tuple of associated values of the respective conceptual variables for a given conceptual entity. Second, the ratio magnitudes of attributes are usually to be specified in terms of a coordinate system centered upon a part of the graph framework. Third, textual material perceptually grouped with an object specifies the absolute value of the object; textual material perceptually grouped with the framework specifies the real-world referent of the attribute that the coordinate system centered on the framework

helps to specify; textual material associated with specific local regions of the framework specifies pairings of absolute and ratio values of the attribute specified by the associated coordinate system. Note that for maximum generality, text is linked to perceptual entities by the predicate "associated," which can symbolize proximity, similarity, continuity, and so on. This helps to encompass graphs with parts directly labeled and graphs exploiting common colors or shapes in keys and legends. Similarly, the predicate "attribute" is meant to encompass length, width, orientation, lightness, color, and so on. However, the indispensability of visual space motivates "geometric shapes" as opposed to arbitrary visual predicates being specified as typical frameworks, and spatially localizable "parts" being specified as the units over which attributes are defined.

In encountering a certain type of graph for the first time, a reader will generate a specific graph schema for it using the general graph schema. The reader will have to replace the predicates "pictorial content," "associated," "attribute," "geometric figure," and so on, by the actual visual predicates found in the visual description of the novel graph. This will be possible when the visual description has a structure similar to that of the general graph schema, with objects described in terms of attributes defined with respect to a framework, and textual labels associated with each. In addition, an astute graph reader will add to the new specific graph schema higher-order predicates (e.g., "descending-staircase") that can be taken to symbolize global trends (e.g., a decrease in the dependent variable). However, the availability of these higher-order predicates, and how transparently they symbolize their trends, will differ arbitrarily from graph type to graph type, and so these predicates cannot be included in any simple way within the general graph schema but must be created case by case. This process will be discussed in more detail in the section describing what makes a graph reader efficient.

Pushing the question back a step, we may ask, "Where does the general graph schema come from?" This question is more profound, and the answer to it is correspondingly murkier. In one sense, one could answer that people are explicitly taught how to read certain types of graphs. But, this still leads one to wonder how people can generalize from the small set of graph types that they are exposed to in school (basically, bar graphs, line graphs, pie graphs, and pictograms) to the myriad exotic forms that are created and easily understood in popular magazines or areas of expertise. This is especially problematic given that formal instruction in graph reading does not teach the abstract concepts such as "attribute," "extent," "ratio value," and so forth, that in fact define what all graphs have in common. A deeper answer to this question is that a great many abstract concepts seem to be mentally represented by structures originally dedicated to the representation of space and the movement of objects within it, a phenomenon that

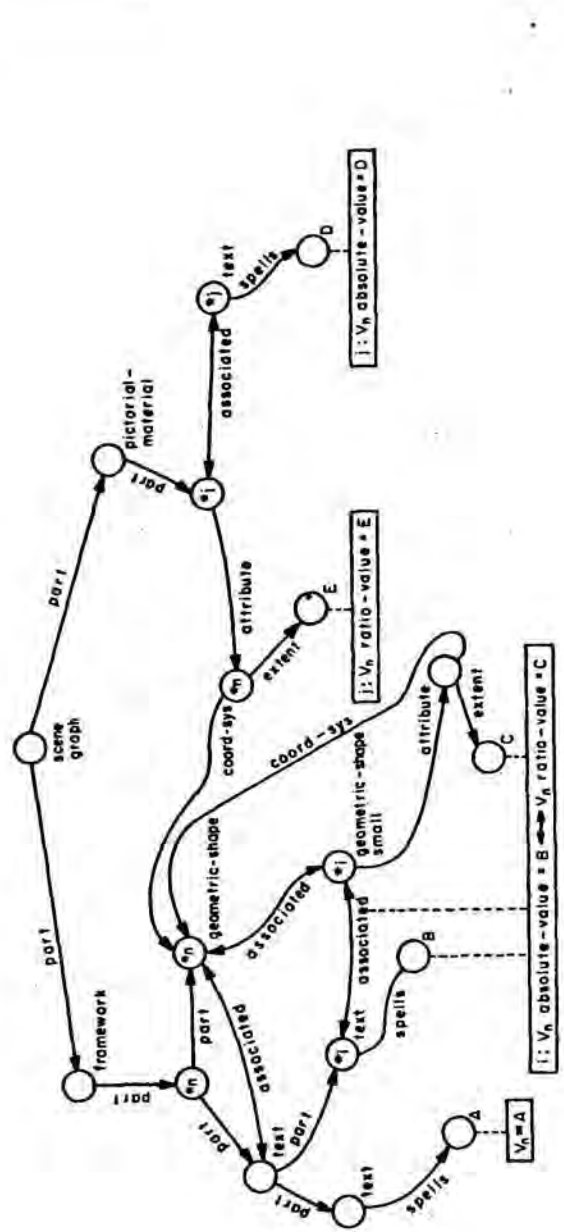


FIG. 4.20.

manifests itself in language in many ways (see Clark, 1973; Cooper & Ross, 1975; Jackendoff, 1978; Lakoff & Johnson, 1980; Talmy, 1978). In particular, abstract quantities seem to be treated mentally as if they were locations on a spatially extended scale (as can be seen in expressions such as *The temperature is rising*, *John weighed in at 200 lbs.*, and many others), or more generally, as corresponding to virtually any other abstract continuum as long as the "positive" and "negative" ends of the two continua are put into correspondence (see Cooper & Ross, 1975; Pinker & Birdsong, 1979). Thus, the use of continuous spatial predicates to represent abstract variables is part of a larger cognitive pattern of using spatial properties to symbolize nonspatial ones; beyond this informal observation there is, unfortunately, little that can be added with any precision.

X. THE DIFFICULTY OF COMPREHENDING A GRAPH

In this section, I consider what makes different types of graphs easy or difficult when particular types of information have to be extracted (by "type of information," I am referring to different conceptual questions, such as ones referring to ratio values vs. differences vs. trends).

Aside from the limitations of the peripheral encoding mechanisms (i.e., limits on detectability, discriminability, and the accuracy of encoding magnitudes), the structures and processes described here permit *any* quantitative information whatsoever to be extractable in principle from a graph. This is because no information is necessarily lost from the visual description "upward" and there are no constraints on what the inferential processes can do with the information in the conceptual message.

In practice, though, limits on short-term memory and on processing resources will make specific sorts of information easier or more difficult to extract. I have assumed that the visual description that is encoded is, in fact, a small subset of the complete visual description, and that noise in the MATCH and message assembly processes causes only a subset of that reduced visual description to be translated into conceptual message information. The remaining conceptual message entries will contain the information that is "easily extracted" from a graph, since a simple look-up procedure suffices to retrieve the information. On the other hand, if the desired information is not already in the conceptual message, it will have to be generated either by the top–down interrogation process, which adds entries to the conceptual message, or by the inferential processes, which perform computations on existing entries. Each of these processes can involve a chain of (presumably) capacity-limited computations, and each process properly includes the look-up of information from the conceptual message. Therefore, they are necessarily more time consuming and memory consuming (since the results of intermediate computations must be temporarily stored) than the look-up of existing information in the conceptual message. And, in a limited-capacity, noisy system such as the human mind, greater time and memory requirements imply increased chances of errors or breakdowns, hence, increased difficulty. This conclusion can be called the *Graph Difficulty Principle*: A particular type of information will be harder to extract from a given graph to the extent that inferential processes and top–down encoding processes, as opposed to conceptual message look-up, must be used.

There will, in turn, be two factors influencing whether a desired type of information (i.e., the answer to a given conceptual question) will be present in a conceptual message. First, a message entry will be assembled only if there are message flags specific to that entry appended to the graph schema. That, in turn, will depend on whether the visual system encodes a single visual predicate that corresponds to that quantitative information. For example, I have assumed that because of the nature of visual encoding, a bar graph schema appends message flags to predicates for height, horizontal position, extremeness in height, extreme differences in height between adjacent objects, and extended increases or decreases in height. This respectively makes ratio values of the dependent and independent variables, extremeness in value, extreme differences in values, and global trends easily extractable. On the other hand, our visual systems do not supply a visual predicate for an object being a given number of ordinate scale units high, or for one bar's height to be a precise ratio of the height of another, or for the leftmost and rightmost bars to be of the same height, and so on. Therefore, there can be message flags and no conceptual message entries for the absolute value of the dependent variable, the exact ratio of dependent variable values corresponding to successive values of the independent variable, or the equality of dependent variable values corresponding to the most extreme independent variable values. If a reader wishes the graph to answer these conceptual questions, he or she can expect more difficulty than for the conceptual questions discussed previously.

The second factor influencing whether a conceptual message entry will be assembled is the encoding likelihoods of the predicates attached to the corresponding equation flags in the graph schema. In the example we have been using, if the predicate "descending-staircase" has a very low default-encoding likelihood, and hence is absent from the visual description on most occasions, the entry specifying a decreasing trend will not find its way into the conceptual message until interrogated explicitly. Incidentally, apart from innateness and automaticity factors, it is conceivable that the encoding likelihood of a predicate is also influenced by "priming": When a graph schema is activated (i.e., when the graph is recognized as being of a particular type), the encoding likelihoods of the visual predicates may

be temporarily enhanced or "primed" (see Morton, 1969). In other words, when a graph is recognized on the basis of partial recognition, the schema may make the rest of the information more likely to be encoded for as long as the schema is activated.

As simple as the Graph Difficulty Principle is, it helps to explain a wide variety of phenomena concerning the appropriateness of different types of graphs for conveying different types of information. Consider Cartesian line graphs, for example. The English language has a variety of words to describe the shapes of lines: straight, curved, wiggly, V-shaped, bent, steep, flat, jagged, scalloped, convex, smooth, and many more. It also has words to describe pairs of lines: parallel, intersecting, converging, diverging, intertwined, touching, X-shaped, and so on. It is safe to assume that the diverse vocabulary reflects an equally or more diverse mental vocabulary of visual predicates for lines, especially since the indispensability of visual space (see Section IIIA) implies that predicates for configural spatial properties such as shape should be readily available. The availability of these predicates affords the possibility of a line graph schema with a rich set of message flags for trends. For example, if "x" and "y" are nodes representing lines on a graph, with V_1 the abscissa, V_2 the ordinate, and V_3 the parameter, the propositions on the left side of Table 4.1 can be flagged with the conceptual message equations on the right side of the table: This makes line graphs especially suited to representing functions of one variable over a range of a second, the covariation versus independence of two variables, and the additive versus interactive effects of two variables on a third, and so on. In contrast, the mental vocabulary for the shapes implicit in the tops of a set of grouped bars is poor, perhaps confined to "ascending-staircase," "descending staircase," and "rectangular," as implied in Fig.

TABLE 4.1
Some Quantitative Trends Associated With Visual Patterns

Predicate	*Equation Flag*
Flat (x)	V_2 trend = unchanging
Steep (x)	V_2 trend = increasing rapidly
Inverted U-shape (x)	V_2 trend = quadratic
U-shape (x)	V_2 trend = quadratic
Jagged (x)	V_2 trend = random
Undulating (x)	V_2 trend = fluctuating
Straight (x)	V_2 trend = linear
S-shape (x)	V_2 trend = cubic
Rectilinear (x)	V_2 trend = abruptly changing
Not flat (x)	V_1 affects V_2
Parallel (x, y)	V_1, V_3 additively affects V_2
Converging (x, y)	V_1, V_3 interactively affects V_2

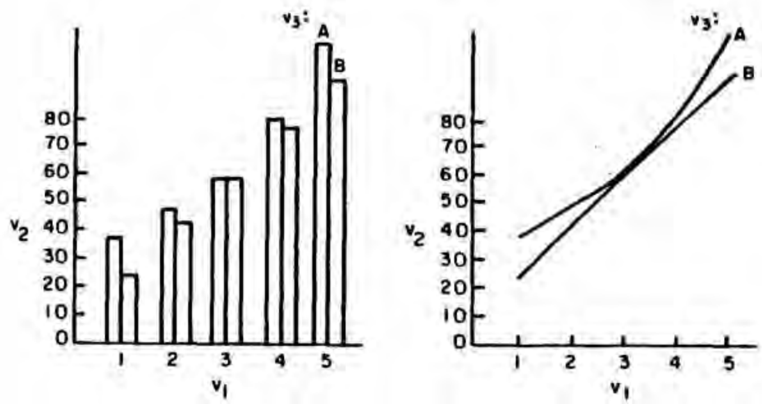
TABLE 4.2
Data Illustrating the Relative Efficacy of Reading Trends From Tables, Line Graphs, and Bar Graphs

V_1:		V_2: 1	2	3	4	5
V_3:	A	30.0	35.0	45.0	60.0	80.0
	B	20.0	32.0	45.0	57.5	70.0

4.18. Correspondingly, there will be fewer possibilities for specifying trends in a schema for bar graphs, and less likelihood of assembling specific "trend" and "affects" entries in the conceptual message when a bar graph is processed. And the predicates for a *pair* of shapes implicit in the respective tops of two integrated *groups* of bars will be even scarcer, preventing "additively affects" and "interactively affects" entries from being encoded. Small wonder, then, that line graphs are the preferred method of displaying multidimensional scientific data, where cause-and-effect relations, quantitative trends, and interactions among variables are at stake. To convince yourself of the appropriateness of line graphs for these purposes, try to determine the nature of the trend of V_2 over the range of V_1, and the nature of the interaction of V_1 and V_3 (a variable with two levels, A and B) on V_2, from Table 4.2, Fig. 4.21a and Fig. 4.21b. It should be easy to see from the line graph in Fig. 4.21b that at level A of Variable 3, Variable 2 is increasing and positively accelerating, whereas at B, it is increasing linearly. Similarly, one can see that Variables 1 and 3 interact in their effects on Variable 2. This is because the "straight" and "concave-up predicates, corresponding to "linear" and "positively accelerating" trends are readily encodable. In contrast, the like-colored bars in Fig. 4.21a do not form a group where relative heights can be described by a single predicate, and so inferring the trend necessitates a top–down, bar-by-bar height comparison, a difficult chore because it is hard to keep the heights of all the bars in mind (i.e., activated in the visual description) at once. It is even more difficult to extract the trends from the table, because not only is a number-by-number comparison necessary, but the process of encoding a multidigit numeral's magnitude seems to be intuitively slower and more effortful than the encoding of a bar's height.[5]

[5] Incidentally, though a line graph is better than other forms of data presentation for illustrating trends, typically only one way of constructing the line graph will illustrate a given trend optimally. For example, a line graph that used Variable 3 (i.e., A vs. B) as the abscissa and Variable 1 as the parameter would not illustrate the linear and accelerating trends as transparently as the graph in Fig. 4.21b, since these trends no longer correspond to single attributes of a distinct perceptual entity, but must be inferred from the successive intervals separating the left end points of the five lines and those separating the right end points of those lines, respectively.

However, try to answer the following question by examining the table, bar graph, and line graph just considered: What is the exact value of Variable 2 at level B of Variable 3 and level 4 of Variable 1? Most people I have asked find the question easiest to answer with reference to Table 4.1, a bit harder with reference to the bar graph, and hardest of all with reference to the line graph. This illustrates the purpose-specificity of graphs, which has frequently been noted in the graph comprehension literature, and which is an inescapable consequence of the present theory: Different types of graphs are not easier or more difficult across the board, but are easier or more difficult depending on the particular class of information that is to be extracted. In this case, we have already seen that absolute values of the dependent variable in a bar graph cannot be directly entered into the conceptual message since there are no visual predicates that correspond to them. Rather, specific ratio values of the dependent variables can be encoded, as can pairings between arbitrary absolute values and ratio values (from the numbers printed along the ordinate), but the absolute value of a particular entry must be computed by effortful inferential processes using these two kinds of information. The line graph is harder still, because the Gestalt principles cause each entire line to be encoded as a single node rather than being broken up into a set of nodes, each corresponding to a level of Variable 1. Thus, when the conceptual question addresses the absolute value of Variable 2 corresponding to a particular value of Variable 1, there is no visual description node specific to the part of the line signifying that value, and one must be created by a top–down encoding process focused on a perceptually arbitrary point along the line.



(a) (b)

FIG. 4.21.

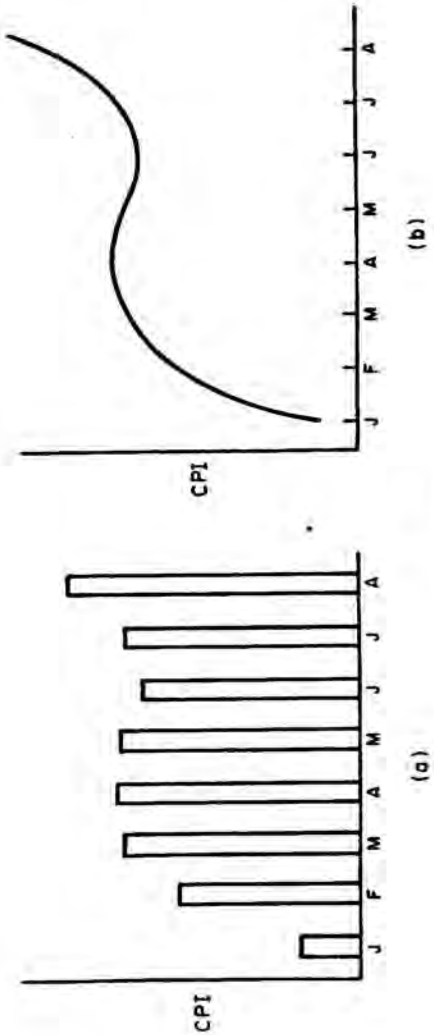


FIG. 4.22.

That is also why it is sometimes easier to use a bar graph than a line graph to determine the difference between two levels of one variable corresponding to a pair of values on another (e.g., whether the Consumer Price Index is higher for March or June in Figs. 4.22a and 4.22b).

In sum, we have seen that extracting information from a graph is easiest when the visual description contains predicates linked to message flags displaying equations that answer the conceptual question (less technically, when the information is conveyed by an easily perceivable visual pattern in the graph and when the reader knows that that pattern encodes desired information). As a consequence, (a) Line graphs should be best for illustrating trends and interactions (since there exist many visual predicates for line shapes); (b) Tables should be best for illustrating absolute values of the dependent variable (since there is no way to specify absolute values for particular levels of the independent variable in line or bar graph visual descriptions and graph schemas); and (c) Bar graphs may be better than line graphs or tables for illustrating differences between dependent variable values corresponding to specific independent variable values (since the desired values are specified individually in the bar but not the line graph, and since it seems to be easier to encode a bar's height than to read a multidigit number). It is comforting to know that these three conclusions have been borne out many times in the empirical literature on graph comprehension (Carter, 1947; Culbertson & Powers, 1959; Schutz, 1961a, 1961b; Washburne, 1927), scanty though that literature is (see Wainer & Thissen, 1981).

Some Further Determinants of Graph difficulty

In general, a graph maker will do best if he or she designs the graph so that the visual system parses it into units whose attributes correspond to the quantitative information that he or she wishes to communicate. In the previous section, we saw how this principle favors either line graphs, bar graphs, or tables, depending on the type of question the reader is to answer. Of course, these are not the only choices that face a graph designer. In this section, I briefly show how often design choices might be resolved by the Graph Difficulty Principle.

1. One Graph with Two Lines or Two Graphs with One Line? As mentioned, the visual system has predicates describing groups of nearby lines (e.g., *Parallel (x, y)*, *Fan-shaped (x, y, z)*, *Intersecting (x, y)*, etc.). These correspond to specific types of interactions between variables (e.g., additive, multiplicative, inversely multiplicative, etc.). Thus, questions about interactions can be answered quickly if the lines are in close enough proximity to the predicate describing them as a group to be encoded. However, if the lines are in different graphs, they will be encoded as units and their interactions must be extracted by interrogating their slopes separately and inferring the interaction from these slopes. Thus, when interactions are of interest, lines should be plotted on a single graph (unless, of course, the number of overlapping lines is large, which may lead to spurious groupings of line segments belonging to different lines). Schutz (1961b) indeed found that graphs with multiple lines were easier to understand than multiple graphs, if the number of lines is small.

2. Legends or Labeled Lines? As noted earlier, the visual system groups stimuli that are in close proximity. A graph schema can exploit this fact by specifying that a label near a graph element signals the absolute value of a variable for the conceptual message entry specified by that element (e.g., in the bar graph schema we examined previously). If the correspondence is specified instead in an insert or legend (i.e., with a label next to a small patch sharing the color, shading, or internal cross-hatching or stippling of the lines or bars), that correspondence must be extracted by the inferential processes, using one entry specifying the distinguishing feature of the bar or line in the graph, and a second entry linking that distinguishing feature to the appropriate absolute value, based on the legend or insert. Therefore, labeled lines should be better (again, assuming the number of elements is not so large that spurious groupings arise).

In fact, Parkin (1983) has conducted an experiment deliberately designed to test this prediction of the theory. He composed five methods of labeling for line graphs that exploited varying numbers of Gestalt principles to associate lines with their labels. He had each label next to its corresponding line somewhere along its length (proximity), next to each line and aligned with its end (proximity, good continuation), aligned with the end of each line but separated from it by white space (good continuation), in a legend (no Gestalt principle) or in the caption (no Gestalt principle). In addition, lines and labels were sometimes printed in the same respective colors (similarity) and sometimes not. As expected, the greater the number of Gestalt principles associating lines with labels (and the fewer the number of principles leading to a competing organization of labels with labels), the faster subjects were able to answer questions about relative heights and slopes of the different lines at given points on the X-axis. As expected, this effect interacted with geometric complexity as measured by the number of times the lines crossed over one another. For simple graphs, labels close to the lines fostered quicker answers than labels next to and aligned to the ends of lines, which in turn were quicker to read than labels simply aligned with line ends. However, the three labeling methods were equally difficult for complex graphs. Similarly, captions and legends were much harder than labeled lines for simple graphs, but only somewhat harder for complex graphs. Finally, Parkin found that associating lines with their labels via a common color led to quicker responding. Thus, the experiments stand as a confirmation of the prediction that the processes used by the visual system to associate objects with one another exert effects on how quickly graph elements can be associated with their labels, and of the prediction that these effects might be weakened or nullified when the graph displays complex patterns of line intersection.

XI. THE EFFICIENCY OF A GRAPH READER

Though I have referred to a single idealized "graph reader," flesh-and-blood graph readers will differ from one another in significant ways. For example, some people may have swifter elementary information processes, or a larger short-term memory capacity, or more powerful inferential processes. Though these factors may spell extreme differences in how easily different people comprehend graphs, they are not specific to graph comprehension, and I will not discuss them further. Instead, I will focus on possible differences among people in their abilities to read graphs per se.

A natural way of determining what makes a person good at reading graphs is to examine what makes the graph-reading process more or less easy (i.e., the considerations in the preceding section) and to predict that individual differences in the nature of the structures and processes involved will spell differences in the general ease with which individuals read graphs.

Recall that in the last section I predicted that a given type of information was easy to extract from a given type of graph if there were message flags in the graph schema specific to that information, and if the visual predicates to which the flag was attached were presented in the activated visual description of the graph. Each factor allows for individual differences. First, a person's graph schema may lack important message flags. Thus, he or she may not know that parallel lines in a line graph signal the additivity of the effects of two variables on a third. When pressed to determine whether additivity holds in a certain graph, such a person would have to resort to costly inferential processes operating on a set of entries for ratio or absolute values. In general, the theory predicts that the presence or absence of message flags in a person's schema will have dramatic effects on how easily that person can extract the information specified by the flag. Second, the predicates that trigger the process whereby message flags are assembled into conceptual message entries may be more or less likely to appear in the visual descriptions of different people. The needed predicates, because of lack of practice at encoding them, may not yet be automatic, and hence may have low default encoding likelihoods. Furthermore, the links between those predicates and the rest of the graph schema may be weak, dissipating the "priming effect" which assists the encoding of missing predicates once a graph has been recognized.

Of course, this begs the questions of what in fact determines whether people have the necessary equation flags in their schemas, and whether the encoding likelihoods and links among predicates in a schema will be sufficiently strong. As to the first question, there are probably three routes to enriching graph schemas with useful flags:

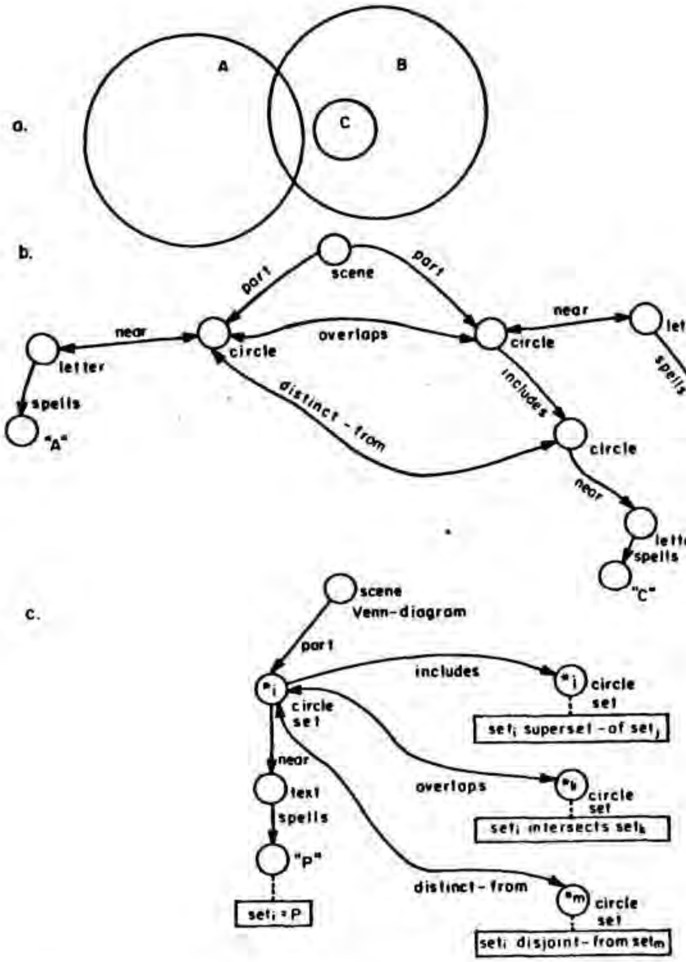
1. *Being told*. It is common for formal instruction in mathematics and science to spell out what to look for in a graph when faced with a particular question. For example, students learning statistical procedures such as the Analysis of Variance are usually told that nonflat lines indicate main effects, nonparallel sets of lines indicate interactions, U-shaped lines indicate quadratic trends, and so on.
2. *Induction*. An insightful reader or graph maker might notice that quantitative trends of a given sort always come out as graphs with particular visual attributes (e.g., quadratic functions yield U-shaped lines). He or she could then append the message flag expressing the trend to the predicate symbolizing the visual attribute in the graph schema.
3. *Deduction*. Still more insightful readers could infer that owing to the nature of the mapping between quantitative scales and visual dimensions in a given type of graph, a certain quantitative trend *must* translate to a certain visual property. For example, a person could realize that the successive doublings of a variable by a particular exponential function must lead to a curve that becomes increasingly steep from left to right.

Taken together, these principles suggest that improvements in the ability to read graphs of a given sort will come (a) With explicit instructions concerning the equivalences holding between quantitative trends and visual attributes (so as to enrich the graph schema); (b) With instruction as how to "see" the graph (i.e., how to parse it perceptually into the right units, yielding the appropriate visual description), and with practice at doing so (making the encoding process automatic and thereby increasing the encoding likelihoods and associative strengths of the relevant visual predicates); and (c) With experience at physically plotting different quantitative relationships on graph paper (affording opportunities for the induction and deduction of further correspondences between visual attributes and quantitative trends, to be added as message flags to the graph schema).

XII. EMPIRICAL TESTS OF THE THEORY

As Wainer and Thissen (1981) note, there has been very little systematic research on the psychology of graph comprehension. Experiments cited herein on the relative ease of extracting information from line graphs, bar graphs, and tables are consistent with the general claim that graph readers can translate visual patterns directly into trend information when possible (via message flags appended to visual predicates in graph schemas), that readers require that the visual marks signifying values or pairs of values form good Gestalts, and that absolute value information cannot be perceived directly from a graph. However, these data alone are not optimal tests of the theory, or even of parts of the theory. First, in most cases, we have no independent evidence for the perceptual phenomena that figure into the explanations of graph-reading difficulty. For example, we do not in fact know that the shape of a line is easier to perceive than the shape implicit in the tops of a set of upright rectangles, or that a segment within a smooth curve is harder to isolate perceptually than a closed rectangle in the set of stimuli used in these experiments. Without independent evidence concerning the perceptual properties of the display, the theory's explanations risk becoming circular. Second, whenever familiar graph formats are used, there is a risk that subjects' explicit training with that format prior to entering the lab may influence their responses in a way that may undermine attempts to interpret their responses in terms of perceptual factors. For example, graphic style manuals advise designers to use line graphs to convey trends. If designers heed this advice, they may on the average use line graphs more often in contexts demanding the extraction of trends, giving readers more practice reading trends from line graphs. In turn, they may come to execute the sequence of mental operations necessary to verify trend information from line graphs more quickly, even if the operations themselves were at first intrinsically difficult, the style manuals notwithstanding.

There are a small number of experiments that I and my collaborators have performed which are explicitly designed to test the theory proposed here while avoiding the problems described. In Pinker (1983), I report three experiments which tested the hypotheses that graph readers have knowledge of the correspondences between trends and visual patterns when these visual patterns are readily encodable, and that they are then able to translate the perceived pattern into the desired trend directly without having to examine more local units one by one. A novel graph format was invented, consisting of a chain of line segments joined end to end corresponding to the months of the year. The length of a segment (greater than or less than an inch) represented the rainfall for that month relative to a



d.

$set_1 = A$
$set_2 = B$
$set_3 = C$
set_1 intersects set_2
set_2 superset-of set_3
set_1 disjoint-from set_3

FIG. 4.23.

reference level, and its angle with respect to the previous segment (greater or less than 180°) represented its temperature relative to a reference level (or vice versa). It was expected that single values for temperature or rainfall would be easier to extract when encoded as segment length than when encoded as segment angle, because the perception of segment angle requires attention to a pair of segments and normalization of the orientation of the first (since we must mentally rotate a pattern into a standard orientation in order to determine its handedness—corresponding here to the sign of its angle; Cooper & Shepard, 1973, 1975). In contrast, the detection of whether temperature or rainfall was consistently above or below the reference level, versus sometimes being above it and sometimes below, and a similar discrimination involving the detection of alternation, were predicted to be easier when the variable was encoded by segment angle. This is because for the angle variable, consistent years yield uniformly convex curves and inconsistent years yield curves with a concave region, a discrimination our visual systems are adept at making (Hoffman & Richards, 1985), whereas the length variable does not yield curves with recognizable shape differences contingent on the consistency or alternation of the variable. Thus, subjects should be able to create a graph schema in which there are message flags signifying consistency of one of the dependent variables appended to the predicate for convexity and message flags signifying inconsistency appended to the predicates for concave regions. With no such shape predicates for length, there can be no message flags for consistency of the other dependent variable.

In the first experiment, subjects were shown the stimuli described as visual patterns, not as graphs, and answered questions about the lengths and angles of particular segments or the consistency and alternation of the lengths or angles of the entire sequence. As predicted, single lengths were recognized more quickly and accurately than single angles, but consistent sequences of lengths were recognized more slowly and less accurately than consistent sequences of angles. Presumably, this was because consistency of angle translated into convexity, and convexity is an easily encodable visual predicate. Thus, the subjects did not have to encode the angle of each pair of line segments individually, and as a result, they took less time (rather than more) to verify the consistency of the entire sequence of angles than to verify a single angle. This provides independent motivation for predictions about graph-reading difficulty in two subsequent experiments. In the second experiment, the stimuli were introduced as graphs and subjects were told how segment angle and length conveyed information about temperature and rainfall, and were also told that consistency of the variable conveyed by angle translated into convexity and alternation into zigzags (i.e., I tried to induce them to form a graph schema with the right message flags by telling them about the appropriate trend–shape correspondences directly. As a control for attention differences, they were also told how consistency and alternation of the other variable translated into patterns of lengths, though based on the results of the first experiment it is unlikely that visual predicates exist for consistently long or short sequences of line segments). When answering questions about rainfall and temperature, subjects showed the same pattern of reaction times as did their counterparts in the first experiment when answering questions about the corresponding geometric properties of the stimulus. That is, they were faster and more accurate at detecting single values of the variable conveyed by length than of the variable conveyed by angle, but showed the reverse pattern when detecting the consistency of the global sequence. This indicates that the subjects, as predicted, were able to exploit the correspondences between trends and shapes that would be encoded explicitly in a graph schema, allowing them to recognize trends directly without examining individual point values.

In a third experiment, subjects were only told how the graphs conveyed information about temperature and rainfall for individual months, not about how shapes translated into trends. Nonetheless they, too, showed the interaction found in the first experiment. (However, they did show the interaction less strongly, suggesting that being told about shape–trend correspondences for a graph may facilitate the extraction of those trends, as predicted by the discussion in Section XI.) In general, then, these experiments confirm that a given sort of conceptual information is easily extractable from a graph to the extent that the graph encodes the desired conceptual information as an easily perceivable visual predicate and to the extent that the correspondence between the two is represented in the mind of the reader.

Simcox (1983) reports three experiments that are more narrowly addressed to the application of these principles to the issue of line versus bar graphs per se. Specifically, Simcox wanted to see whether there is independent perceptual evidence that in the default case we are more likely to encode a line in terms of its slope and overall height, whereas we are likely to encode pairs of bars into the individual height of each one (this putative perceptual phenomenon is at the heart of my account of the respective superiority of line and bar graphs at displaying trend and point information). He reasoned that whatever attributes we do encode by default should be available in "pure" form to discrimination processes and irrelevant information should not interfere with such judgments. In contrast, if we have to discriminate stimuli on the basis of some attribute that is not part of the default encoding, then the discrimination would require an internal transformation of the encoded information into the desired attributes, which should result in more time and errors, and in interference from values of irrelevant attributes conflated with the desired one in the default encoding.

In his first experiment, Simcox found that people could sort a deck of cards with bar graph-type stimuli into two piles on the basis of the height of one of the bars, and their sorting times were not significantly affected by whether the height of the other bar was constant or varied randomly. However, when asked to sort the deck into two piles on the basis of the slope or average height defined by the two bars, they were significantly slower when the irrelevant attribute (average height or slope) varied randomly than when it was held constant. These two patterns are diagnostic of "separable" and "integral" stimuli, respectively (Garner, 1974) and indicate that people naturally encode pairs of bars into a representation in which the height of each one is stated explicitly; overall height or slope must be inferred from that representation. Precisely the opposite pattern was found when subjects sorted analogous decks of cards depicting line graph-like stimuli. When sorting according to the height of one of the endpoints of the lines, the subjects were slowed down when the height of the other endpoint varied randomly compared with when it was constant. However, when sorting according to the slope of the line or its overall height, the irrelevant attribute (overall height or slope, respectively) did not affect the sorting speed. Thus, a line segment depicted within an L-shaped framework is easily encoded into a representation in which its overall height and slope are stated explicity; the height of individual endpoints must be inferred from that representation or encoded in a second look at the stimulus.

In a second experiment, Simcox found that when subjects are simply asked to classify the overall height, height of one point, or slope of a line or pair of bars, they were faster at classifying the height of bars than the height of one of a line's endpoints, but faster at classifying the slope or overall height of a line than those defined by a pair of bar heights. Finally, in a third experiment involving the speeded sorting of a graph-like stimulus with two lines, Simcox found that sorting by either height or by slope was slowed down when the irrelevant factor (slope or height, respectively) varied randomly and when the variation of that irrelevant factor yielded intersecting versus nonintersecting lines. However, when the stimuli were varied so that intersection was not a concomitant of varying the slope or height, each of the two attributes could be attended to selectively without interference from the other. This suggests that the global property of line intersection finds its way directly into the default encoding of a pair of lines, rather than it being a property derived from heights and slopes of the component lines.

Taken together, the Pinker and Simcox studies offer strong support for the hypothesis that is at the core of the present theory: that graphs will be easy to comprehend when the visual system naturally encodes the geometric features of the graph with visual predicates that stand in one-to-one correspondence (via the graph schema) with the conceptual message that the reader is seeking. The Pinker (1983) study showed that readers indeed can infer global quantitative trends directly from global geometric features representing them so long as those global features are ones that our visual systems perceive easily. Simcox (1983) showed in particular that this sort of explanation can account for the widely observed superiority of line graphs at conveying trends and of bar graphs at conveying point values. He did so by showing that, as the theory would require, uninterpreted stimuli resembling line graphs are more naturally encoded in terms of the overall height, slope, and intersection of lines, whereas uninterpreted stimuli resembling bar graphs are more naturally encoded in terms of the heights of the individual bars.

XIII. EXTENSION OF THE THEORY TO CHARTS AND DIAGRAMS

Quantitative information is not the only kind that is transmitted by visual displays, and it would be surprising if the charts and diagrams used to express qualitative information were comprehended according to principles radically different from those governing graph comprehension. In fact, the theory described in these pages can be extended virtually intact to the domain of charts and diagrams. Again, a visual description of the diagram would be encoded, obeying the principles of grouping, the indispensability of space, and so on, and again, there would be a "chart schema" for a particular species of chart, which specified (a) The constituents of the visual description that identify the graph as being of the appropriate sort (e.g., a flow chart vs. a Venn diagram); and (b) The correspondences between visual predicates and conceptual message entries. The conceptual message entries would be of a form appropriate to the qualitative information represented, and conceptual questions would consist of conceptual message entries with the "?" symbol replacing one of the constants. The MATCH, message assembly, interrogation, and inferential processes would play the same roles as before. Charts would be easier or more difficult depending on whether the visual system encoded them into units corresponding to important chunks of conceptual information, and chart readers would be more fluent to the extent that their chart schemas specified useful correspondences between conceptual information and visual attributes, and to the extent that those visual attributes were encoded reliably. A brief example follows.

Venn diagrams, used in set theory, consist of interlocking circles, each of which represents a mathematical set. Presumably, they are effective because the visual system can easily encode patterns of overlap (which will translate into set intersection), inclusion (translating into the subset–

superset relation), nonoverlap (translating into disjointness), and so on (see Ullman, 1984, for a discussion of how some of these patterns might be recognized). Simplified Visual Array, Visual Description, Chart Schema, and Conceptual Message representations specific to Venn diagrams appear in Figs. 4.23a through 4.23d.

Even from these simplified examples, one can see that, as before, the difficulty of retrieving a given type of information will depend on what is in the visual description and graph schema and not simply what is on the page. For example, here the reader would have to infer the fact that Set C is a subset of Set B from the conceptual message entry stating that Set B is a superset of Set C. A more efficient diagram reader might have a richer schema, containing the predicate "included-in" together with a message flag stating that one set is a subset of the other. This would spare that reader from having to rely on inferential processes.

Other sorts of diagrams and charts use other visual predicates to convey their messages efficiently: For example, flow charts use shape predicates to signify the type of operation (e.g., action vs. test), they use the contiguity of shapes with lines to indicate the flow of control, and they use the orientation of arrowheads to indicate the direction of that flow. The linguist's tree diagrams for the phrase structure of sentences use horizontal position to signify precedence relations among constituents, proximity to common line segments to signify dominance (inclusion) relations, and above/below predicates to signify the direction of the dominance relations. For each type of diagram, there would be a specific schema spelling out the correspondence between visual predicates and conceptual messages.

XIV. CONCLUSIONS

This chapter began with a warning that our understanding of graph comprehension would advance in proportion to our degree of understanding of general perceptual and cognitive faculties. As we have seen, the theory outlined here indeed borrows heavily from perceptual and cognitive theory, adopting, among others, the following assumptions: the importance of propositional or structural descriptions at certain levels of representation; the indispensability of space as it relates to visual predicates, selective attention, creation of perceptual units, and accuracy of encoding; the limited capacity of short-term visual representations; the use of distributed coordinate systems for encoding shape and position; the use of schemas to mediate between perception and memory; the effects of physical salience on encoding likelihood; conceptually driven or top–down encoding of visual attributes; a MATCH process for recognition; "priming" of visual predicates; and strengthening of associative links with practice. I hope that this enterprise is not totally parasitic, though, since in developing the theory, significant gaps in our knowledge of visual cognition came to light; for example, the exact constraints on which physical attributes can serve as visual predicates, the determinants of their likelihood of being encoded, the relative strengths of the Gestalt principles, the format in which the groupings they impose should be represented in structural descriptions, the constraints that determine how message flags can be appended to predicates in schemas, the ways that descriptions can guide top–down encoding processes, and how general the information in a general schema (like the general graph schema) can be. Perhaps the most salient conclusion of this exposition, then, is that our understanding of basic cognitive processes will be the rate-limiting step in our understanding of applied cognitive domains, and that unexplained but pervasive phenomena in applied domains can be very effective diagnostics of important gaps in that basic knowledge.

ACKNOWLEDGMENTS

This research was supported by NSF grants BN5 80–24337, 81–14916, 82–16546, 82–09450, 85–18774, and NIE grant 400–79–006. I thank Stephen Kosslyn, Bill Simcox, Michael Kubovy, and Barbara Tversky for helpful discussion. An earlier draft appeared as Technical Report No. 10 in the MIT Center for Cognitive Science Occasional Paper Series in 1981.

REFERENCES

Anderson, J. R. (1976). *Language memory and thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. New York: V. H. Winston.

Bertin, J. (1967). *Semiologie graphique: Les diagrammes—les reseaux–les cartes*. The Hague: Mouton.

Bregman, A. S. (1977). Perception and behavior as compositions of ideals. *Cognitive Psychology*, *9*, 250–292.

Carter, L. F. (1947). The relative effectiveness of presenting numerical data by the use of tables and graphs. In P. F. Fitts (Ed.), *Psychological research on equipment design*. Washington, DC: U. S. Government Printing Office.

Clark, H. H. (1973). Space, time, semantics, and the child. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*. New York: Academic Press.

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

Cooper, L. A., & Shepard, R. N. (1975). Mental transformations in the identification of right and left hands. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 48–56.

Cooper, W. E., & Ross, J. R. (1975). *World order*. Notes from the Parassession on functionalism. Chicago: Chicago Linguistics Society.

Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Culbertson, H. M., & Powers, R. D. (1959). A study of graph comprehension difficulties. *Audio Visual Communication Review*, *7*, 97–100.

Finke, R. A., & Pinker, S. (1983). Directed scanning of remembered visual patterns. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 142–147.

Garner, W. P. (1974). *The processing of information and structure*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gibson, E. J., Bergman, R., & Purdy, J. (1955). The effect of prior training with a scale of distance on absolute and relative judgements of distance over ground. *Journal of Experimental Psychology*, *50*, 97–104.

Hinton, G. E. (1979). Some demonstrations of the effects of structural descriptions in mental imagery. *Cognitive Science*, *3*, 231–250.

Hoffman, D. D., & Richards, W. (1984). Parts of recognition. In S. Pinker (Ed.), *Visual cognition*. Cambridge, MA: MIT Press.

Jackendoff, R. (1978). Grammar as evidence for conceptual structure. In M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic theory and psychological reality*. Cambridge, MA: MIT Press.

Kosslyn, S. M., Murphy, G. L., Bemesderfer, M. E., & Feinstein, K. J. (1977). Category and continuum in mental comparisons. *Journal of Experimental Psychology: General*, *106*, 341–375.

Kosslyn, S. M., Pinker, S., Smith, G. E., & Shwartz, S. P. (1979). On the demystification of mental imagery. *Behavioral and Brain Sciences*, *2*, 535–548.

Kubovy, M. (1981). Concurrent pitch segregation and the theory of indispensable attributes. In M. Kubovy & J. Pomerantz, (Eds.), *Perceptual organization*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three dimensional shapes. *Proceedings of the Royal Society*, *200*, 269–294.

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

Miller, G. A., & Johnson–Laird, P. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.

Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision*. New York: McGraw–Hill.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*, 165–178.

Newell, A., & Simon, H. (1973). *Human problem solving*. Englewood Cliffs, NJ: Prentice–Hall.

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

Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory and Cognition*, *3*, 519–527.

Parkin, L. (1983). *A comparison of various graph labelling methods in the context of gestalt organizing principles*. Unpublished manuscript, Consulting Statisticians, Inc., Wellesley, MA.

Pinker, S., Birdsong, D. (1979). Speakers' sensitivity to rules of frozen word order. *Journal of Verbal Learning and Verbal Behavior*, *18*, 497–508.

Pinker, S. (1983). *Pattern perception and the comprehension of graphs*. Unpublished manuscript.

Posner, M. I., Snyder, C. R., & Davidson, B. S. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109*, 160–174.

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

Schank, R. A., & Abelson, P. E. (1977). *Scripts, plans, goals and understanding*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1–66.

Schutz, H. (1961a). An evaluation of formats for graphic trend displays: Experiment II. *Human Factors, 3*, 237–246.

Schutz, H. (1961b). An evaluation of methods for presentation of graphic multiple trends: Experiment III. *Human Factors, 3*, 108–119.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127–190.

Simcox, W. A. (1983). *A perceptual analysis of graphic information processing*. Unpublished doctoral dissertation, Tufts University, Medford, MA.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decision. *Psychological Review, 81*, 214–241.

Stevens, S. S. (1961). To honor Fechner and repeal his law. *Science, 133*, 80–86.

Talmy, L. (1978). The relation of grammar to cognition—A synopsis. In D. Waltz, (Ed.), *Proceedings of TINLAP–II* (Theoretical Issues in Natural Language Processing). Urbana: University of Illinois Press.

Ullman, S. (1984). Visual routines. In S. Pinker (Ed.), *Visual Cognition*. Cambridge, MA: MIT Press.

Wainer, H., & Thissen, D. (1981). Graphical data analysis. *Annual Review of Psychology, 32*, 191–241.

Washburne, J. N. (1927). An experimental study of various graphic, tabular, and textual methods of presenting quantitative materials. *Journal of Educational Psychology, 18*, 361–376, 465–476.

Wertheimer, M. (1938). Laws of organization in perceptual forms. In W. D. Ellis (Ed.), *A source book of Gestalt psychology*. London: Routledge & Kegan Paul.

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