DOCUMENT RESUME

ED 222 191 IR 010 427

AUTHOR Simcox, William
TITLE Cognitive Considerations in Display Design.
INSTITUTION Consulting Statisticians, Inc., Wellesley, MA.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE 81
CONTRACT 400-79-006
NOTE 51p.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Cognitive Processes; Computer Graphics; *Design Requirements; Display Aids; *Display Systems; Flow Charts; *Graphs; Models; Visual Aids; *Visual Perception
IDENTIFIERS Graphic Communication; *Selective Attention

ABSTRACT

This investigation of the encoding features of graphs begins with a description of a cognitive framework which allows designers to factor into the process of designing displays how people interpret the information found and what display properties are responsible for this interpretation. The framework also provides a performance measure for use in assessing alternative designs and cognitive effort. Properties of displays--features, dimensions, and configurations--that can be mentally represented and serve as a basis for response are also discussed, as well as the strategy used to determine which display property is actually being represented and a processing justification for its use. An experiment is then described in detail which used a speeded classification task to test whether any of the display properties discussed were actually the encoding features and to assess the degree of interference in two univariate and two orthogonal conditions. It was found that, if the display format was not compatible with the information to be presented, classification errors increased, which in turn would increase the likelihood of miscommunication at higher levels of processing. A 38-item bibliography, 7 figures, and 4 tables are attached. (LMM)

COGNITIVE CONSIDERATIONS IN DISPLAY DESIGN

Contract No. 400-79-006

Prepared by:

William Simcox
Senior Engineer

Submitted to:

Joseph Psotka, Ph.D.
The National Institute of Education
Contracts & Grants Management Division
1200 19th Street, N.W., Room 808
Washington, D.C. 20208

Consulting Statisticians, Inc

Cognitive Considerations in Display Design

Introduction

Human information processing of complex visual displays has emerged as a problem to be reckoned with in recent years. A large variety of displays are encountered daily by people, products of our technological revolution. As examples of such displays we can think of computer graphics, flow diagrams, radar and sonar scopes and the like.

The simple graph is probably the most commonly encountered example of such visual displays yet it has received only intermittent attention by the human factors community (cf. MacDonald-Ross, 1977; Schutz, 1961a, 1961B; Vernon, 1952, Washburne, 1927). Graphs are found in most of the materials we read today, spanning technical books and journals to recreational magazines. This abundance has probably made most people take graphs for granted, but they are indeed displays which must be designed. They must be constructed using principles and tools common to all display design problems.

Traditional display design has focused mainly on the properties of the display which makes the information more perceptually accessible to the person processing it. Properties such as brightness, contrast, texture, size and the like are all display variables affecting how the information is detected and discriminated by the visual system. For example, without adequate contrast a person could not separate displayed material from background. Within a psychophysical framework it is possible to investigate systematically the effects of these display properties on our ability to identify the symbols of the display (cf. Shurtleff, 1980 for a nice consolidation of this literature).

There is another aspect to the display design problem, apart from legibility considerations, which must be addressed, and as yet has not been to any substantial degree. The problem is that once the information gets inside the person, so to speak, it must be interpreted to effect a response, whether it be overt, such as answering a question, or simply an integration into existing cognitive structures. Legibility, per se, does not guarantee that information will be successfully communicated. While the form of the symbols may be quite clear, their meaning may remain obscure and liable to misinterpretation. The source of the misinterpretation must lie in the properties of the display, as read by a person with specific knowledge and goals. Traditional psychophysical research, which views the person as a measuring instrument, does not provide an adequate framework for systematically investigating the interaction between these higher-order variables.

Instead we must turn to a more cognitive framework which views the person as a processor of information. Within this framework, the display properties are mentally represented in data structures and processes operate on these structures, transforming and manipulating them so that eventually we come to understand the communicated information. Each process takes time and resources to perform its appropriate function, and a fundamental assumption of cognition is that increases in processing times implicate increased cognitive effort. Since presumably there is a finite amount of resources available to accomplish these functions, processes needing additional effort will reduce the amount available for other operations. This in turn will increase the likelihood of error, and consequently, miscommunication. Thus it is possible to investigate certain design alternatives within this framework using cognitive effort as a design criterion.

There will, in turn, be two factors that influence the amount of effort necessary to extract a given piece of information from a display. The first such factor is the correspondence between the display properties used to represent the information and the properties encoded by the person. There are many different ways of representing the same information graphically. However, there may be only a few ways in which correct interpretation follows naturally. Any other design alternatives will result in greater processing effort. The second factor is a person's knowledge of the correspondence. Thus a given piece of information can be correctly represented in the display but the person may lack the appropriate knowledge structures to detect the correspondence and consequently, process it efficiently. This lack of efficiency results in increased effort. In this investigation, I assume that the second factor is nonoperative, i.e., the appropriate structures exist. Consequently the investigation addresses the first factor only. That is, we want to choose the display properties in such a way that the least amount of effort is needed to process the desired information.

How do we accomplish this objective? I think that we must begin by considering the graphical representation of information from the observer's point of view. That is, when we encounter graphic displays, we seem to look at them in certain ways and not others. There seems to be some sort of 'natural' perceptual representation driven by the display properties themselves. It is this 'data-driven' description that we, as designers, must determine since, in all likelihood, the conceptual information it carries is the information extracted most easily. Thus one objective of this investigation is to determine this 'data-driven' description and to look at some processing consequences associated with it.

Discovering the Display Properties

I will refer to the display properties that are mentally represented by an observer as <u>encoding features</u> of the display. This is the perceptual information found in the data structures alluded to earlier. Garner (1978) has identified two general types of display properties which can serve as encoding features: <u>component properties</u> and <u>wholistic properties</u>.

Component properties are called <u>attributes</u>. Garner defines attributes as properties which help define a stimulus but do not denote the stimulus. For example the attributes of the following object (■) are form, size, brightness and position. There are two major types of attributes: dimensions and features.

A <u>dimension</u> is a variable attribute of a stimulus such that if the dimension exists for the stimulus it exists at some positive level and these alternative levels are mutually exclusive. For instance if hue is a dimension of a stimulus then the stimulus must exist at some particular hue, such as red or green and if the stimulus is red then it cannot be green and vice versa. Note that a zero level can be a positive level on a dimension as with zero brightness.

A <u>feature</u> is an attribute of a stimulus that either exists or does not exist, but if it exists it is defined on a single level only, the alternative level being the absence of the feature. An important property of a feature is that it can be removed from a stimulus without otherwise affecting the rest of the stimulus. For example the bottom horizontal line of an upper case E is feature and it can be removed (F) without affecting the rest of the stimulus. Note that the zero level of a feature implies its absence.

Wholistic properties are those aspects of a stimulus which are not perceived analytically (i.e., in terms of component parts) but nonetheless are perceived directly. One particular type of wholistic property defined by Garner which we will find useful as a display property is a configural property or <u>configuration</u>. A configuration has properties that emerge from the stimulus as a consequence of interactions and interrelationships between the components parts. These 'emergent' properties exist in addition to, not instead of, component properties but results indicate that when they exist emergent properties dominate perception, overriding the component properties which create them (Pomerantz and Garner, 1973). Configural properties are also different from the component properties which define them. For example the following stimulus:

is defined by relations between the horizontal and vertical position of a sequence of dots (component parts) but, as is intuitively obvious, configures so as to form the appearance of a line having the property of 'curvedness' which is certainly different from the properties of the component parts. Other examples of configural properties include symmetry (Λ), repetition (//////////), intersection (X) as well as other higher order properties.

<u>Methodology</u>

It is possible to analyze a simple display for its encoding features, be they dimensions, features or configurations using measures obtained from the selective attention paradigm (Egeth, 1967; Garner and Felfoldy, 1970). In this paradigm the subject is presented with stimuli varying along two attributes (usually dimensions) and is required to classify (or respond to) each stimulus according to one attribute while ignoring the other. Figure 1 shows a typical set of stimuli generated from two dimensions each existing at two levels. One type of task requires the subject to classify stimuli varying along one dimension, the other dimension held constant at a particular level. Any row or column of Figure 1 would satisfy this task requirement. This is called the univariate or discrimination task since the subject must simply discriminate between the levels of the relevent dimension. A second type of task called the gating or filtering task (Posner, 1964) must be performed when both dimensions are operative and the subject must classify on the basis of one of the dimensions (a1 vrs a2 or b1 vrs b2), ignoring the other. Of interest is the classification performance (time or errors) of this orthogonal condition relative to performance on the univariate condition. If stimuli are classified more slowly when the irrelevant dimension is orthogonally varied than when it is held constant, orthogonal interference results. This means that the irrelevant attribute interferes with our ability to selectively attend to the relevent attributes. Another way of interpreting this is to say that it is not possible to 'filter' out the irrelevent dimension. On the other hand if no difference in classification speed between orthogonal and univariate tasks is found, it is concluded that selective attention to the relevent attribute is possible.

INSERT FIGURE 1 HERE

Using the selective attention paradigm, then, I will operationally define an encoding feature of a display as that display property which allows selective attention. A justification for this, based on the cognitive framework introduced, is the following. In classification tasks, the experimenter, in telling a person how to respond, sets a <u>response basis</u>. That is, the person is told to represent the stimulus in a certain way and thus must try to look at the set of stimuli in terms of this stored representation. When a stimulus is encountered it must be perceptually represented. If this representation is the same as the response basis then simply locating the correct value on the relevent basis is sufficient for a correct response (classification). If on the other hand this representation is different from the response basis then additional resources are needed to 'transform' it so its correct value can be obtained. Presumably this increased effort will require greater processing time resulting in a failure of selective attention. In the univariate condition such a reparsing is unnecessary since any distinguishing display property can be used. Therefore I will equate success of selective attention with minimum cognitive effort for this task.

Empirical evidence in support of this claim is provided by research on the processing of multidimensional stimuli. This research has demonstrated quite convincingly that the experimenter defined aspects of a stimulus are not necessarily the same aspects represented by the subject (Garner, 1974; Pomerantz, 1981; Garner and Felfoldy, 1970). For example Garner and Felfoldy (1970) had subjects sort decks of stimulus cards containing circular Munsell patches varying along the dimensions of value and chroma. They observed orthogonal interference for this pair of dimensions. The assumption here is that these are the attributes represented by the observer. Suppose these attributes are not the encoding features of this stimulus, but instead the encoding features are shape and the integrated percept color. Using these attributes as the basis for classification, Gottwald & Garner (1972) found no evidence of orthogonal interference. Additionally, the perceptual grouping research (Pomerantz, 1981, Pomerantz and Garner, 1973) has shown that if component properties configure then it is the configural properties which become the encoding features, not the component parts. For example, using pairs of elements differing in the direction of curvature of each element, e.g. (), ((,)),)(, Pomerantz and Garner (1973) found failure of selective attention to the individual elements. On the other hand, the following stimuli, (◡, (⌒, ·)◡, ⌒) show no such failure. Note that while both sets of stimuli have the same elemental properties, the relationships between these properties in the first set produce additional configural properties such as symmetry, closure etc which seems to dominate. If classification were based on such properties then evidence supporting selective attention may exist.

Investigating the Encoding Features of Graphs

Thus far, I have introduced a cognitive framework which I feel can be used to design more communicatively effective displays. This framework allows us to factor into the design process how people interpret the information found in graphic displays and what display properties are responsible for this interpretation. Additionally the framework provides us with a performance measure useful in assessing alternative designs, cognitive effort. Drawing upon concepts from the psychology of perception, I then describe those properties of displays, features, dimensions and configurations, which can be mentally represented and thus serve as a basis for response. Finally, I introduce a methodology that can be used to determine which display properties are actually being represented, using cognitive effort as a criterion, as well as a processing justification for its use.

Now this methodology is directed towards the study of a specific display, the simple graph. Graphs are an integral part of the communication process today, and hence a poorly designed graph could have dire consequences in a wide range of fields. Thus there is a real need for design with respect to these displays. Additionally the properties of graphs are simple extensions of the basic stimulus concepts found in the perceptual literature allowing us to build upon existing theory and methodology. Finally, since the properties of these displays do not change over time, their properties can be efficiently controlled and manipulated. This makes them ideal tools for study.

Suppose we are given the four basic graphic displays shown in Figure 2. The first display is defined by pairs of lines on two L-shaped frameworks; the second display is defined by a pair of lines on a single L-shaped framework; the third and fourth displays are defined by pairs of unconnected and connected points respectively on L-shaped frameworks. Note that the L-shaped framework constitutes a feature of the display since its removal does not affect the remaining aspects which are the dimensions on which the objects are defined. Formally we can define these dimensions as the height of each point of the pair. Thus we should be able to classify stimuli generated from different values of these dimensions on the basis of either. Using these dimensions as the response basis, then, if success of selective attention is evidenced for any of these displays then the height of each point would constitute an encoding feature.

INSERT FIGURE 2 HERE

However, the heights of each point are not the only properties that can perceptually define the four displays. In Figures 2A and 2B the objects may not be the points at all, but instead the points may be ignored in favor of the lines connecting them to the framework. In this case we could define the displays in terms of the height of the lines rather than the height of the points. Similarly, as we move from left to right across the displays the defining properties may begin to shift from the height of each line or point, depending on the display, to properties reflecting the relationships between the pair, culminating in slope and overall height of the pair (see Figure 2D). Again using the selective attention task, we can establish whether these properties are in fact encoding features by using them as a response basis and measuring the amount of interference for each display.

Processing Consequences

The strategy for determining the encoding features of a given display should be fairly evident from the discussion thus far. First, the display is analyzed for its potential information conveying properties. For example, using the displays of Figure 2, the properties chosen for investigation are the heights of the individual objects (be they points or lines) and the slope and overall height of the pair. Second, these properties serve as the response basis for a classification task. That is, each display is classified with respect to both sets of dimensions. Then, using the selective attention paradigm, the magnitude of interference between orthogonal and univariate conditions is determined. If there is no difference in performance between these conditions we conclude that the display properties chosen are in fact the encoding features.

What happens if the display properties we choose to represent the information are not the encoding features? For example, suppose that slope and height are not the encoding features for any of the displays shown in Figure 2. Would we expect the amount of interference (reflecting the amount of additional effort needed) to be the same for each display with respect to this response basis? The justification for using the selective attention task, which I presented, earlier would imply that the amount of interference will vary depending on how similar the perceptual representation is to the response basis.

A proposed processing model which could account for varying amounts of interference is shown in Figure 3. When a visual stimulus is encountered it is represented perceptually by its encoding features, whether they be dimensions, features or configurations. This perceptual representation forms one input to a similarity based 'match' process. The other input comes from a stored representation of the response basis set by the experimental context. Similarity is being used in the sense of Tversky (1977). That is, the similarity between 'a' and 'b' is some 'matching function' of both common and distinctive properties of a and b. The matching function measures the degree to which two objects overlap and will vary monotonically as a function of common properties shared by a and b as well as the distinctive properties specific to each. For example E would be more similar to F than to I because E and F have more common features than E and I. Thus the match is based on some function of the number of properties the two representations have in common and by how much they differ. If the two representations share many properties a relatively quick decision can be made and a response generation process can begin. On the other hand if the two representations share relatively few properties then additional processes are necessary to reparse the perceptual representation to effect a match. This reparsing will result in more processing resources and hence greater processing times. This increase in time will be evidenced as orthogonal interference. Note that this model puts the locus of interference in post-perceptual processing which is more consistent with late mode structural attention theories (Deutsch & Deutsch, 1963; Norman, 1968).

INSERT FIGURE 3 HERE

Therefore, using both height of each object and slope and overall height of the pair as the response basis it is possible, using the selective attention task, to assess this degree of interference hypothesis. As we move from left to right across the displays it seems intuitively apparent that the displays start sharing properties more common to the response of basis slope and overall height then height of each object. Similarly, as we move from right to left the opposite occurs, that is more properties common to the heights of each object and less properties common to slope and overall height of the pair. Since greater mismatch between encoding features and response basis means greater reparsing is necessary, we should expect greater amounts of interference evidenced as we move from left to right with classification is based on heights of each point or line while the same would be true as we moved right to left with classification based on slope and overall height of the pair.

Using a speeded classification task, a task commonly used in selective attention research, the first experiment will test whether any of the above display properties are in fact the encoding features. Additionally the degree of interference hypothesis will be assessed using the resulting interference magnitudes.

EXPERIMENT 1

Method

Subjects

Sixteen coworkers at CSI/Datacrown served as subjects in this experiment. Performance measures on one subject were not included in subsequent analyses because, under some experimental conditions, these measures were excessively large (greater than ± 3σ) relative to the treatment means. As a consequence they were considered outliers and discarded.

Stimuli

Two sets of stimuli, corresponding to two response sets, were used in this experiment. They are shown in Figure 4 and Figure 5. Each set consisted of four different types of stimuli, 1) Pairs of vertical lines, centered on separate L-shaped frameworks, spaced 2 cm apart, 2) pairs of vertical lines, centered on a single L-shaped framework, spaced 5 mm apart, 3) pairs of unconnected points, centered on a single L-shaped framework, spaced 5 mm apart, and 4) pairs of connected points, centered on a single L-shaped framework, spaced 5 mm apart.

INSERT FIGURES 4 and 5 HERE

The dimensions used to generate the first set of stimuli were defined in terms of the pair of lines or points. The dimensions were the slope of the pair (0.25 or 0.45) and the overall height of the pair (1.27 or 1.52 cm). The overall height was defined in terms of the midpoint between the pair. The dimensions used to generate the second set of stimuli were defined in terms of the component properties of the objects. These dimensions are the height of the left line or point (4.95 or 5.97 mm) and the height of the right line or point (7.24 or 8.72 mm).

The two values of slope used in this experiment were chosen so as not to make the length of the left and right lines of stimulus type 1 in Figure 4 unduly large or small. Such conditions could occur for this stimulus type because the spacing between lines is much greater than spacing for the other stimulus types. As a result a much greater change in vertical length is necessary to generate the same slope as the other stimulus types. The heights used in the second stimulus set were obtained by constraining the major diagonal of Figure 1 (a1b1 and a2b2) to have a slope of 0.45, one minor diagonal element (a2b1) to have a slope of 0.25 and constraining the difference in levels along both height dimensions to be 5 just-noticeable-differences (JNDs) using Ono's (1967) differential sensitivity index.

For both response sets and within each of the four stimulus types there were decks of 32 cards generated in accordance with four experimental conditions; two univariate conditions and two orthogonal conditions. Each deck contained an equal number of cards relevant to that condition. For example when slope was the relevant dimension the univariate condition consisted of a deck of 32 cards, 16 of which defined the pair of lines or points at an orientation of 0.25 and another 16 which defined the pair at an orientation of 0.45. The overall height dimension was held constant at 1.52 cm for this condition. Similiarly when overall height was the relevent dimension, 16 cards defined the pair at a height of 1.27 cm and 16 cards defined the pair at 1.52 cm, slope held constant at 0.25. When the defining dimensions were height of left object and height of right object, the irrelevant dimension was held constant at 7.24 and 5.97 mm respectively for the univariate condition. In the orthogonal condition a deck was made up of 8 cards each from the four stimuli which could exist as a result of the combination of two dimensions each having two levels (each cell in Figure 1).

The process of generating the stimulus cards was semi-automated using a Tektronix 4662 interactive digital plotter driven by PLOT 10 software. Standard sheets of 27.9 x 35.6 cm (11x14 in) Bristol Board were partitioned into a 6x3 grid pattern, each grid being 5.6 cm wide x 8.3 cm high. The plotter was scaled such that 2.54 cm (1 in) mapped into 10 display units. Centered within each grid, the particular stimulus alternative was drawn in black ink. Eighteen such stimuli were drawn per sheet. The stimuli were then manually cut out, the top left corners clipped to avoid any potential upside down mixup (and additionally to facilitate error checking), and became part of the appropriate deck. Two decks per condition were generated in this way.

Procedure

Subjects were required to sort the deck of 32 cards into two piles corresponding to the two levels of the relevent dimension. Exemplars of each classification level were placed on the table in front of the subject. Each subject was told the purpose of the experiment, handed a deck of cards and told to sort them into two piles consistent with the targets as quickly as possible but without making errors. Time to sort each deck was measured to the nearest one hundreth second using a Model 54035 (Lafayette Instrument Co.) clock/counter. The clock was started by the experimenter when the first card left the deck and stopped when the last card was placed on one of the piles. Upon completion of the task the subject was told the sorting time. The sorted cards were then set aside for purposes of determining errors, and the next deck was given to the subject. Within each condition subjects sorted the deck of cards two times, the first sort representing practice and not considered in the analyses.

Design

Thirty-two different conditions were generated by combinations of two response sets (slope/height, height of left/height of right), four types of graphs (lines on two frameworks, lines on one framework, points unconnected, points connected) and four tasks (two univariate, two orthogonal). Each subject participated under all conditions. Subjects sorted conditions corresponding to one response set first, were given a five-minute break and then sorted conditions corresponding to the other set. Response set one occurred first for half of the subjects while response set two occurred first for the other half. Within each set order of presentation of the sixteen conditions was balanced over the sixteen subjects in a Latin square arrangement. An experimental session lasted about 1.5 hours.

Data Analysis

Of particular interest in this experiment, were two sets of a prior planned comparisons. One family of sixteen orthogonal contrasts tested the magnitude of interference, i.e. differences in means between orthogonal and univariate tasks for each pair of dimensions under each graph type and response set were tested. The other family of twelve nonorthogonal contrasts tested the degree of interference effect. Differences in means, averaged over both dimensions, between orthogonal and univariate tasks for one graph type were contrasted with equivalent differences for the other graph types within the specific response set. The results of these tests will be couched in terms of confidence intervals since it is felt that this approach not only allows one to test the specific hypothesis under consideration but also to gauge the strength of any inference made as a result by noting the width of the interval.

In addition, note that as the number of comparisons increase the likelihood of at least one Type I error (i.e. reject Ho when it is true) increases. For example, if we did sixteen independent significance tests, each at the 0.05 level , the probability of at least one Type I error would be $1-(.95)^{16} = 0.56$. This value is termed the error rate per family. Since inferences drawn from results of this experiment are based on the family as a whole, using p = .05 for each individual test, would make the findings somewhat tenuous and not readily reliable since spurious results would be expected in over half the replications. For this experiment the error rate per family was chosen to be 0.20 resulting in a family confidence coefficient of 0.80. This confidence coefficient means that if the experiment is repeated, say, 100 times then we would expect the same sixteen confidence intervals in 80 out of the 100 replications. This family confidence coefficient results in a significance level per comparison of about 0.02 for the twelve nonorthogonal contrasts. For the family of sixteen orthogonal contrasts, however the significance tests, and hence the confidence intervals, need not be equally weighted (Myers, 1979 Chapter 2). It is only necessary that the sum of the individual α level per comparison add to the error rate per family or 0.20. Preliminary results from a pilot study using the four graph types with the response set consisting of height of left vrs. height of the right classification showed that, of the graph types showing potential interference effect, the greatest power was needed in assessing effects for bars on one framework. Differences in means for this graph type showed significance at approximately p = .03 while the other graph types showed significant effects at $p < .001$. For this family of contrasts, then, the α level used to test interference effects for lines on one framework for both response sets was set at 0.03. The remaining graph types all were tested at a significance level of 0.007 such that the overall error weight was 0.20.

Results and Discussion

The positive correlation coefficient ($r=0.18$, $p<.001$) between time and errors indicates an absence of any speed-accuracy tradeoff which would otherwise qualify interpretation of the of the data. However plots of variances against means for both sorting times and errors indicated variance stabilizing transformations were necessary for both these measures. It was found that a logorithmic transformation for times and a square root transformation for errors were sufficient to induce spherical distributions for both measures. Thus all times and errors to be presented will be expressed in logs and square roots respectively.

Times

The logarithm of sorting times in seconds averaged over subjects is shown in Table 1 for each graph type, response set and task.[1] Also shown is the magnitude of interference, Δ, defined as the difference between the orthogonal and univariate sorting task.

---

[1] For a repeated measures design there is no appropriate error term from which the significance of the four way interaction of Response Set x Graph Type x Task x Subject can be tested since each cell has a sample of size one. However a single df test developed by Tukey (1949) provides a means of evaluating a particular form of this interaction. This test involves further analyzing the interaction sums of squares into two components, one which represents the interaction component distributed on 1 df, the other representing error distributed on the remaining dfs. For this data $F(1,125)=0.83$, $p>.25$ indicates the interaction component is not significant implying the mean square provides an independent estimate of the error variance.

INSERT TABLE 1 HERE

Table 2 shows confidence intervals for the preplanned orthogonal comparisons testing the sixteen interference magnitudes found in Table 1. Any interval not enclosing zero implies a reliable difference between the means of the two sorting tasks. The tabled results show that for the response set height of left/height of right, selective attention (i.e., no interference) was possible only for the graph type symbolized by the pair of vertical lines on two L-shaped frameworks. Thus it appears that the encoding features for this stimulus are in terms of its component parts (the individual heights) while for the other graph types, these parts interact to produce other defining properties.

INSERT TABLE 2 HERE

The success of selective attention for the pair of lines on two frameworks and the failure of selective attention for the pair of lines on one framework may be a consequence of the proximity of these lines to one another. Pomerantz and Schwaitzberg (1975) showed that the magnitude of interference diminished as the spacing between two elements (left and right parentheses) was increased and disappeared entirely at a 4-deg of visual angle separation. The separation between the pair of lines on two frameworks, at a viewing distance of 28 cm was 4.1-deg while that for the lines on one framework was about 1-deg. Thus it is quite likely that the proximity of the lines on one framework induces an encoding in terms of the pair as a perceptual unit. Further support for this likelihood is provided by the success of selective attention for the pair of lines on one framework when classification is based on the overall height of the pair (contrast L12 in Table 2).

In a similar vein the encoding of pairs of unconnected points the pairs of connected points is in terms of some relationship between the pair. These properties of these graph types seem to configure such that slope and overall height become the dominant properties. This is evidenced by the success of selective attention when classification was based on these properties, while failure of selective attention resulted when classification was based on the individual heights of the points. For both of these displays, then, we can conclude that the encoding is in terms of the slope and height of the pair of points.

Finally, note that within the response set of slope and overall height, the confidence intervals show an interference asymmetry (Garner, 1974) for two graph types. This means that selective attention was evidenced for one dimension of the pair but not the other. For the pair of lines on two frameworks, when classification was based on overall height failure of selective attention was evidenced but not so for classification based on slope. However inspection of the stimuli used for this graph type, shown in Figure 4, suggest that the source of the asymmetry may be the particular dimensional values chosen to represent this graph type. Note that the two stimuli having the greatest slope also have the smallest height of the left line and largest height of the right line. Contrasts L1 & L2 imply that we represent this graph type in terms of the individual heights of each line. This means that when we look at this graph type it is in terms of the individual heights initially. This makes these extreme values quickly noticeable since presumably they keep calling attention to themselves. Since the response basis makes ambiguous the attention allocation policy, both extreme values should be noticed equally often. Thus, for example in the univariate condition where classification is based on stimulus 1c or 1d (see Figure 4) a basis of, say, extreme (left or right)/other would be sufficient. Similarly such a basis would be sufficient to classify stimulus 1a/1c vrs stimulus 1b/1d (the orthogonal condition) and no interference would be evidenced as is the case. This suggests that there may be a tendency to use the encoding features as a response basis whenever possible.

Just the opposite asymmetry was evidenced for the pair of lines on one framework, i.e. failure of selective attention when classification was based on slope and success of selective attention when classification was based on height. Since the confidence interval (contrast L11) encloses a value whose true magnitude is much greater than zero (relative to the marginal lower bounds of other contrasts), this interference effect is not tenuous and the height of the lines confounds our perception of slope with this graph type. This in turn implies that we may not mentally represent this graph type in terms of the slope of the lines but instead, for example, by some aspect associated with the differences in heights of the lines.

Post hoc comparisons (Scheffe, 1959) of the interference magnitudes within each graph type showed no reliable difference in means between either of the two component properties or the properties defined in terms of the pair. For example, there were no difference between the height of left and height of right response for any particular graph type nor were there any differences between slope and overall height. Thus for each graph type within the appropriate response set the interference magnitudes were averaged together and these average magnitudes are plotted in Figure 6. The figure clearly shows a monotonic increase in the amount of interference as we move from pairs of lines on two frameworks, having the greatest match to the response basis, to connected points, having the greatest mismatch to the response basis, when classification is based on the component properties. Conversely a monotonic decrease is apparent when classification is on the basis of properties defined in terms of the pair, which produces the opposite matching characteristics.

INSERT FIGURE 6 HERE

Table 3 shows the confidence intervals of the preplanned comparisons testing this degree of interference effect [Garner, (1970) uses the term degrees of integrality]. Considering the response set height of left/height of right, the average interference for the connected points is higher than averages for any of the remaining graph types. Contrasts involving differences between the other graph types were not significant with respect to the family, but less conservative tests (p=.05) show that both lines on one framework and unconnected points result in greater amounts of interference than lines on two separate frameworks. For the response set slope/height, contrasts show that lines on two separate frameworks result in greater amounts of interference than either connected or unconnected points, while lines on one framework yield greater average amounts of interference than connected points.

INSERT TABLE 3 HERE

The monotonicity of interference magnitudes as the degree of presumed mismatch between encoding features and response basis increases suggests that it is possible in principle to assess different design alternatives for a given piece of information using cognitive effort as a criterion. For example, if we want to convey, conceptually, the 'trend' of one variable (increasing or decreasing) over some discrete 'range' of another variable, this information should be represented graphically by connected points. As we have shown, such a graphical representation results in the least amount of effort for its perceptual processing. The least preferred design using a single framework representation would be the bar graph. On the other hand, if we want to convey, conceptually, the 'level' of one variable (high or low) for a certain value of the other variable using a single framework graph, the bar graph is the preferred design from a processing viewpoint.

Selective Attention Considerations

Using the selective attention paradigm, I stated earlier that it might be possible to use the data from this experiment to further elaborate on some of the theories and models on which the paradigm is based. For instance, the fact that some of the contrasts shown in Table 3 are reliably different indicates a very important consequence concerning selective attention. That is, selective attention is not an all-or-none affair. The monotonic relationship between interference magnitude and degree of mismatch seems to be more consistent with shared capacity theories of attention (Kahneman, 1973; Norman and Bobrow, 1975). These theories posit the prime determinant of task difficulty is governed by the allocation policy of resources (attention). One model, based on this theory, has beenproposed by Dykes and Cooper (1978). They suggest that orthogonal interference could be due to the misallocation of attention. Assuming a fixed amount of attention that must be shared amongst all incoming signals, if attention is misallocated to the irrelevant attribute then less attention can be directed to the relevent attribute. This, then, is evidence by orthogonal interference.

I feel that while this assumption can account for processing errors (as we see in the next section), it is somewhat hard pressed, as it stands, to account for increases in times. As Norman & Bobrow (1975) suggest, attention is a resource to be allocated to a cognitive task. As more resources are applied to a task then presumably better performance results. However the allocation of more processing resources will require more processing time. Thus the allocation of less attention (resource) to the relevent dimension as Dykes and Cooper suggest will not produce an increase in time. Instead, as I have suggested, orthogonal interference arises from the increased attention necessary to reparse the perceptual representation, the resultant of a post-perceptual match process.

Errors

Table 4 presents the results of the experiment in terms of the transformed error scores. That is, square root errors averaged over subjects are presented for graph type, response set and task. However, in contrast to sorting times, these mean values are tempered by individual differences.[2]

INSERT TABLE 4 HERE

The significance of this high order Subject X Treatments interaction means that while the group of subjects showed the overall pattern of results found in Table 4, this relative pattern differs markedly for some individuals. The data shown in this table, though,

---

[2]Tukey's test showed a significant component of the four way interaction involving subjects, and the three other variables, $F(1,125) = 3.9$, $p < .25$. In testing this effect we want to guard against making a Type II error, i.e., we do not want to conclude no interaction when in fact it exists. This protection is accomplished by increasing the probability of making a Type I error, or level of significance. Tukey suggests an α level of 0.25.

suggests a potential source of this interaction lies in the discriminability difference between levels defining the component properties left/height of right). Scheffe's (1959) post hoc comparison of the univariate tasks for this response set showed a reliable difference between the mean transformed errors for sorting based on the height of the left (1.47) and those for the height of the right (0.583). No such difference was found between univariate tasks for the other response set. Additionally, visual inspection of the subject data for the graph type symbolized by the unconnected point, for example, showed that, while no subject made more errors in the univariate task than the orthogonal tasks when classification is based on the height of the right attribute, four subjects did so when classification was based on the height of the left attribute. Similar effects were not atypical among other graph types as well. Thus differences in discriminative ability among subjects to the attributes for the different graph type and reponse sets may be the potential source of the interaction.

Data-limited errors. In this experiment, errors may be more of a consequence of what Norman and Bobrow (1975) refer to as data-limited processes. Data-limited processes result in performance which is independent of processing resources such as attention. They depend, instead, only on the quality of the input data signal or the quality of the representation stored in memory. Consider the type of task we are asking subjects to do: absolute judgements. Absolute judgement tasks not only require the subject to mentally represent the information but also to compare this representation with a representation stored in memory. The quality of the input data, such as its contrast, proximal size, duration, etc. can affect our representation of the input. However, these factors are not responsible in this case. What is responsible seems to be the quality of the memory trace which depends on both the number of items held in memory and their discriminability, and as we have suggested, discrimination is exceedingly difficult. Further support for data-limited considerations is evidenced by the magnitude of the correlation coefficient between time and errors, r=0.18. Note that while this coefficient is statistically significant[3], its magnitude implies only 3% of the variance in errors is accounted for by differences in times. If, as we assume in the previous section, differences in times are due to differences in processing resources, the lack of a meaningful correlation coefficient suggests processes affecting accuracy are different from those affecting latencies. Data-limited processes bear important considerations in display design, and it is these types of processes to which traditional display design has addressed itself.

Resource-limited errors. However, the pattern of errors in Table 4 appear to be somewhat similar to the pattern found for sorting times, i.e. an increasing amount of interference as the degree of mismatch between encoding features and response basis increases. Since we have assumed sorting times reflect a resource-limited process (i.e., the amount of attention allocated), we can attribute some resource consequence to errors as well. It might be that errors are resource-limited up to a point, and then become data-limited. For example, suppose less than sufficient resources than are necessary are allocated to the correct input in an absolute judgement task as Dykes and Cooper (1978) suggest.

---

[3]The significance of this coefficient is based on 478 degrees of freedom. This large number of degrees of freedom provide relatively large power to detect even a very small change from zero.

Since the activation and maintenance of the stored representation and match process are all resource demanding operations associated with this input, there is less resources to allocate among them. Thus, we might have only a partial match or a match based on a more degraded representation, but in either case the likelihood of error increases and this likelihood is independent of data quality.

In this experiment, we assume a fixed amount of attention (Kahneman, 1973) and further that all the attention which can be allocated to the task is allocated.[4] One way less attention can be allocated to the correct response basis, then, is if more of it is allocated to mental operations involving some other aspect of the task. These other operations could involve a reparsing of the input description so that the appropriate input can be compared with the stored representation. Such a reparsing would be necessary if the encoding features are in fact different from the response basis.

Further evidence suggesting that the encoding features of a stimulus are not necessarily the response basis is provided by the pattern of errors. Multidimensional scaling studies (Shepard, 1964; Handel & Imai, 1972, Somers & Pachella, 1978) demonstrate that stimuli which do not differ in terms of their encoding features are best described in terms of their overall similarity structure (i.e., Euclidean metric). This means that perceived differences between any two stimuli depend on relations between levels of the underlying dimensions. In other words, the dimensions interact. On the other hand, stimuli

---

[4]The relative discriminabilities of the stimuli used in this experiment impose a very capacity demanding operation on the subjects. Debriefings at the end of each session indicated that all subjects found the task very difficult and had to really concentrate to see the difference.

differing in terms of their encoding feature show no such interaction (i.e., are characterized by a city-block metric). Note that the stimuli used in our experiments differ on levels of two dimensions (i.e. height of left/height of right or slope/overall height). Thus if the stimuli to be compared, albeit a mental comparison, do not differ in terms of their encoding features then we should see some evidence of interaction between the dimensions. A procedure called logit analysis (Goodman, 1972; Theil, 1970) allows us to test for interaction effects using the confusion matrices[5]. It was found that for the eight graph types showing orthogonal interference (see Table 2), only one (lines on one framework, classification by height of right) failed to show a significant interaction effect, $\chi^2(1)=0.830$, $p<.36$. However, this condition resulted in the lowest frequency of errors relative to all others, decreasing the power to detect this interaction. Conversely, it was found that for the eight graph types not showing interference only one (unconnected points, classification by orientation) showed a

---

[5]Specifically we want to know, for a given level (high or low) of the relevent dimension whether the pattern of errors (i.e., cell frequencies) is influenced by the level of the irrelevent dimension (high or low). The dependent variable was dichotomized to correct classification/incorrect classification resulting in a 2 x 2 x 2 contingency table. Thus the usual ANOVA framework, which assumes that the dependent variable is continuous and normally distributed is not applicable in this case. However, a class of procedures developed by Goodman (1970) and others (Bishop, et al., 1975) allow qualitative information to be analyzed with the same degree of sophistication that was once reserved for quantitative data. Specifically, the procedures examine the effects of a set of categorical predictor variables on the log odds of a binary dependent variable similar to the way in which predictor effects are evaluated in traditional ANOVA designs. Just as in ANOVA designs, the dependent variable is partitioned into a set of main effects and interactions, but the partition is in terms of the natural logarithm of success to failure (log odds) for each level of the independent variable rather than the mean. A likelihood Chi-square statistic is then used to evaluate any specific effects in the model.

significant interaction, $\chi^2(1)=11.5$, $p<.001$. This condition, though, was the only one which had no errors for a particular combination of relevant/irrelevant dimensions. Thus, apart from a pair of matrices, orthogonal interference did show dimensional interactions, while no interference showed no such interactions as expected.

To elucidate the nature of the interactions, log-odds plots[6] are presented next to the confusion matrices associated with graph types showing interference. These matrices and plots are represented in Figure 7. The dimensional interactions are clearly evident from the plots. Consider for example Figure 7A showing misclassifications for lines on one framework when classification is based on the height of the left line. When this dimension exists at a low level, greater misclassification results when the irrelevant dimensions exist at a high level than when it exists at a low level. However, just the opposite is true for a high level defining the relevant dimension (hence the interaction). Comparing the four stimuli associated with

---

[6]Log-odds plots are obtained by plotting the natural logarithm of the odds being in one category of the dependent variable. Thus for any combination of the two dimensions the odds of misclassification are calculated by dividing the number of errors by the number of correct responses. For example, for Figure 7A, if the relevant dimension (height of left line) is at a low level and the irrelevant dimension (height of the right line) is also at a low level, the odds of misclassification is 9/111=0.081 and the log-odds is -2.512. Log-odds are computed for each combination in the same fashion and when these values are displayed, Figure 7A results.

The graph is easily interpretable if several points are remembered. First, when two values are equally likely the odds are 1 and the log-odds are 0. Thus, any log-odds above 0 indicates that the probability of the numerator, in this case misclassification, is greater than that for the denominator, correct classifications, while log-odds less than zero imply the converse. Secondly, the higher the point on the graph, the more likely misclassification results. Moreover, since points of equal heights imply equal log-odds, changes in the slope of the line or equivalently when the line is not parallel to the abscissa, an effect is identified. Similarly an interaction between dimensions occurs when the pairs of lines are not parallel.

this graph type (see Figure 5), the pattern of misclassification imply that stimulus G2A is most confused with stimulus G2C and stimulus G2B is most confused with stimulus G2D. Suppose some aspect associated with differences in heights of the lines using the left line as a reference point represents an encoding feature, the overall height of the pair representing another. Since these aspects will keep calling attention to themselves, and as we have already seen discrimination on the basis of the height of the left line is exceeding difficult, there may be a tendency to use them for classification if possible. For example, it appears quite easy to contrast stimulus G2A and stimulus G2D on the basis of overall height and similarly stimulus G2B and G2C on the basis of the difference between heights. However, it is not as easy to distinguish stimulus G2B and G2D or stimulus G2A and G2C on the basis of either of these features under forced pace conditions.[7] In this case a reparsing of the perceptual description is likely and this requires attention, some of which was initially allocated to maintaining the trace associated with these two stimuli. This reduction in attention should result in a less then optimal comparison involving these stimuli. Thus more confusion should result between these pairs as is the case. Note, too, that the same type of interaction occurs for this graph type when classification is based on orientation which is not an encoding feature (see Figure 7G).

INSERT FIGURE 7 HERE

[7]Not coincidently these pair of stimuli differ from each other in terms of one dimension only, while the less confused stimuli differ on both dimensions. Both Eriksen & Hake (1955) and later Lockhead (1966) have shown that stimuli differing on two or more dimensions are more accurately identified than stimuli differing on one dimension provided the defining dimensions are not the encoding features (Garner & Lee, 1962).

The other graph types in this response set (with the exception of one) show a pattern of errors that is markedly different. For example, consider both height of left/height of right classification for the graph type symbolized by the connected points. The log-odds plots show that the greater likelihood of errors occurs when the relevant and irrelevant components exist either at both low levels or at both high levels. In terms of the stimuli comprising this group (see Figure 5) this means that stimuli G4A and G4D are most often confused. Suppose the slope of the line keeps calling attention to itself, i.e., an encoding feature. With respect to such an encoding feature stimuli G4A and G4D differ on only one dimension, height, whereas any other pair differ on both height and slope. Using the same considerations as before, more confusions should result between the pair. We might also expect the same pattern for the unconnected points since our results showed no reliable interference effects when classification was based on orientation and height. However, while Figure 6d shows a similar pattern to the connected point, Figure 6c is not consistent with this expectation. In fact, Figure 6c shows a pattern similar to that for the lines on one framework. The fact that the amounts of interference between these graph types is in fact different (see Table 3), though, suggests that how we represent the pair of unconnected points may be somewhat different than our representation of the connected points. It may be that the pair of unconnected points is represented in a way that is intermediate between the lines on one framework and connected points and may be modified depending on the demands of the task. This flexibility may thus manifest itself in how we use the representation accounting for the disparate interaction results between the connected and unconnected points.

DISCUSSION

The results of this experiment provide initial justification for considering display design issues using a cognitive framework. Simply considering those properties of the display that make the information accessible, while necessary, is not sufficient for effective graphic communication. As we have seen, how a person perceptually interprets the information represented graphically is determined, in part, by the display format we choose. If this format is not compatible with the to be presented information, results show that classification errors increase which in turn means greater likelihood of miscommunication at higher levels of processing.

Thus the properties of the display which influence its format must be chosen with respect to the observer's point of view. That is, when a person looks at a graph, for instance, there seems to be a 'natural' encoding of certain of its properties, whether they be dimensions, features, or configural properties. These properties, called encoding features, are mentally represented with little cognitive effort on the part of the person. If we assume that the perceptual representation leads directly to a conceptual interpretation at a deeper level of processing (Craik and Lockhart, 1972), then certain types of conceptual information should likewise follow naturally. Turning this problem around, then, for a given conceptual message we wish to communicate, there may exist a graphical format that leads 'naturally' to its extraction. Any other format will require a reparsing or elaboration of the perceptual representation corresponding to this format, and, as the results imply, requires greater processing effort and more time.

Thus it becomes possible in principle to 'maximize' the likelihood of correct interpretation for any message we wish to communicate by representing the specific conceptual message to be communicated using encoding features which correspond to it.

In the next experiment this optimality assumption is tested using the display types shown in Figure 2 and a discrete reaction time (RT) task. For each display type, mathematical scales and values on these scales will be added. This will result in a set of paired observations, where the first member can be either a particular value of the independent variable or a pair of values. For example if the graph symbolizes the price of gold over months, months constitutes the independent variable. We can then talk about a particular month, say January, or a pair of months, January vs. February. The second member of each pair can be a ratio value (e.g., "twice as high"), an absolute value (e.g., "$150/ounce"), a level (e.g., "high"), or a trend (e.g., "increasing") defined on the dependent variable. Thus it is possible to 'set' the subject conceptually by asking a particular 'conceptual question', which contain certain of these pairs. The subject then extracts from the graph the information necessary to answer the question and the time it takes to answer is recorded. Presumably the conceptual question sets a response basis at the perceptual level of analysis. Therefore, for a given display, if the conceptual question induces a representation other than that defined by the encoding features (i.e., a default visual description), increased RT to the question answering task can be expected for the reasons outlined earlier.

REFERENCES

Bishop, Y., Feinberg, S., and Holland, P. Discrete Multivariate Analysis: Theory and Practice. Cambridge, MA: MIT Press, 1975.

Craik, F.I.M. and Lockhart, R.S. Level of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 671-684.

Deutsch, J.A. and Deutsch, D. Attention: Some theoretical considerations. Psychological Review, 1963, 70(1), 80-90.

Dykes, J.R. and Cooper, R.G. An investigation of the perceptual basis of redundancy gain and orthogonal interference for integral dimensions. Perception and Psychophysics, 1978, 23, 36-42.

Egeth, H. Selective attention. Psychological Bulletin, 1967, 67(1), 41-57.

Erikson, C.W. and Hake, H.W. Multidimensional stimulus differences and accuracy of discrimination. Journal of Experimental Psychology, 1955, 50(3), 153-160.

Garner, W.R. The stimulus in information processing. American Psychologist, 1970, 25, 350-358.

Garner, W.R. The Processing of Information and Structure. Potomac, MD: Lawrence Erlbaum Associates, 1974.

Garner, W.R. Aspects of a stimulus: Features, dimensions, and configurations. In E. Rosch and B.B. Lloyd (Eds.), Cognition and Categorization. Hillsdale, NJ: Erlbaum, 1978.

Garner, W.R. and Felfoldy, C.L. Integrality of stimulus dimensions in various types of information processing. Cognitive Psychology, 1970, 1, 225-241.

Garner, W.R. and Lee, W. Analysis of redundancy in perceptual discrimination. Perceptual and Motor Skills, 1962, 15, 307-388.

Goodman, L.A. The multivariate analysis of qualitative data: Interactions among multiple classifications. Journal of the American Statistical Association, 1970, 65, 226-256.

Goodman, L.A. A modified multiple regression approach to the analysis of dichotomous variables. American Sociological Review, 1972, 37, 28-46.

Gottwald, R.L. and Garner, W.R. Effects of focusing strategy on speeded classification with grouping, filtering, and condensation tasks. Perception and Psychophysics, 1972, 11, 179-182.

Handel, S. and Imai, S. The free classification of analyzable and unanalyzable stimuli. Perception and Psychophysics, 1972, 12(1B), 108-110.

Kahneman, D. Attention and Effort. Englewood Cliffs, NJ: Prentice-Hall, 1973.

Lockhead, C.R. Effects of dimensional redundancy on visual discrimination. Journal of Experimental Psychology, 1966, 72(1), 95-104.

MacDonald-Ross, M. How numbers are shown. AV Communication Review, 1977, 25(4), 359-409.

Myers, J.L. Fundamentals of Experimental Design. Boston, MA: Allyn and Bacon, 1979.

Norman, D.A. Toward a theory of memory and attention. Psychological Review, 1968, 75(6), 522-536.

Norman, D.A. and Bobrow, D.J. On data-limited and resource-limited processes. Cognitive Psychology, 1975, 7, 44-61.

Ono, H. Difference threshold for stimulus length under simultaneous and nonsimultaneous viewing conditions. Perception and Psychophysics, 1967, 2, 201-207.

Pinker, S. Theory of Graph Comprehension. National Institute of Education Contract Report No. 400-79-0066, 1980.

Pomerantz, J.R. Perceptual organization in information processing. In M. Kubovy and J.R. Pomerantz (Eds.) Perceptual Organization, Hillsdale, NJ: Erlbaum, 1981.

Pomerantz, J.R. and Garner, W.R. Stimulus configuration in selective attention task. Perception and Psychophysics, 1973, 14, 565-569.

Pomerantz, J.R. and Schwaitzberg, S.D. Grouping by proximity: selective attention measures. Perception and Psychophysics, 1975, 18, 355-361.

Posner, M.I. Information reduction in the analysis of sequential tasks. Psychological Review, 1964, 71, 491-504.

Scheffe, H. The Analysis of Variance. New York: John Wiley and Sons, 1959.

Schutz, H.G. An evaluation of formats for graphic trend displays. Human Factors, 1961A, 3, 99-107.

Schutz, H.G. An evaluation of methods for presentation of graphic multiple trends. Human Factors, 1961B, 3, 108-119.

Shepard, R.N. Attention and the metric structure of stimulus space. Journal of Mathematical Psychology, 1964, 1, 54-87.

Shurtloff, D.A. How to make displays legible. La Mirada, CA: Human Interface Design, 1980.

Somers, P. and Pachella, R.G. Psychophysical compatibility and information reduction tasks. Proceedings of the Human Factors Society - 22nd Annual Meeting, 1978, 69-73.

Thiel, H. On the estimation of relationships involving qualitative variables. American Journal of Sociology, 1970, 76, 103-154.

Tukey, J.W. One degree of freedom for nonadditivity. Biometrics, 1949, 5, 232-242.

Tversky, A. Features of similarity. Psychological Review, 1977, 84, 327-352.

Vernon, M.D. The use and value of graphical material in presenting quantitative data. Occupational Psychology, 1952, 26, 22-34.

Washburn, J.N. An experimental study of various graphic, tabular and textural methods of presenting quantitative material. Journal of Educational Psychology, 1927, 18, 361-376.

LIST OF FIGURES AND TABLES

Figure 1	Method of Generating an Orthogonal Set of Stimuli
Figure 2	Four Symbolic Displays
Figure 3	Information Processing Account of Orthogonal Interference
Figure 4	Stimulus Set Corresponding to Response Set Orientation/Overall Height
Figure 5	Stimulus Set Corresponding to Response Set Height of Left/Height of Right
Figure 6	Magnitude of Interference (Experimental Sorting Time--Univariate Sorting Time) Averaged Over Defining Properties of Each Graph Type
Figure 7	Confusion Matrices and Log-odds Plots of Graph Types Showing Orthogonal Interference
Table 1	Transformed Sorting Times (sec) as a Function of Graph Type, Response Set and Task
Table 2	Confidence Intervals, t-statistics, and Significance Levels for Preplanned Paired Comparisons of Selected Means (LOGTIME) Involving the Three Way Interaction of Graph Type x Response Set x Task
Table 3	Confidence Intervals for Preplanned Comparisons of Selected Contrasts Testing the Degree of Interference for Graph Type Response Set
Table 4	Transformed Errors as a Function of Graph Type, Response Set and Task

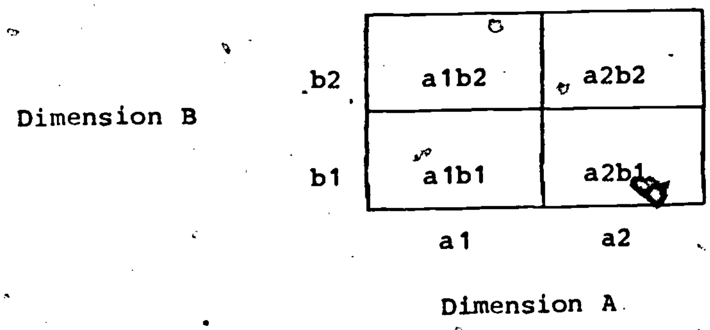

Figure 1 Method of Generating an Orthogonal Set of Stimuli

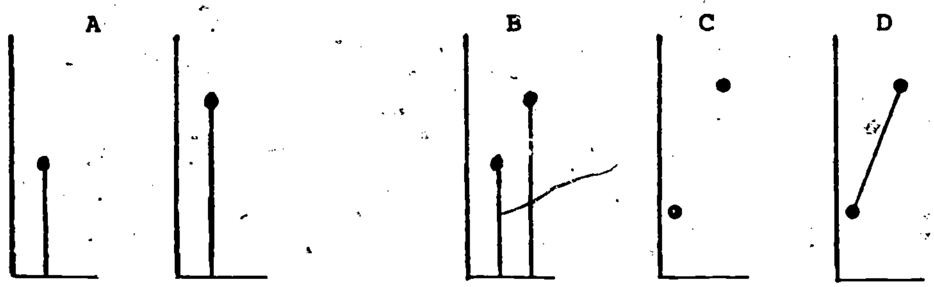

Figure 2 Four Symbolic Displays

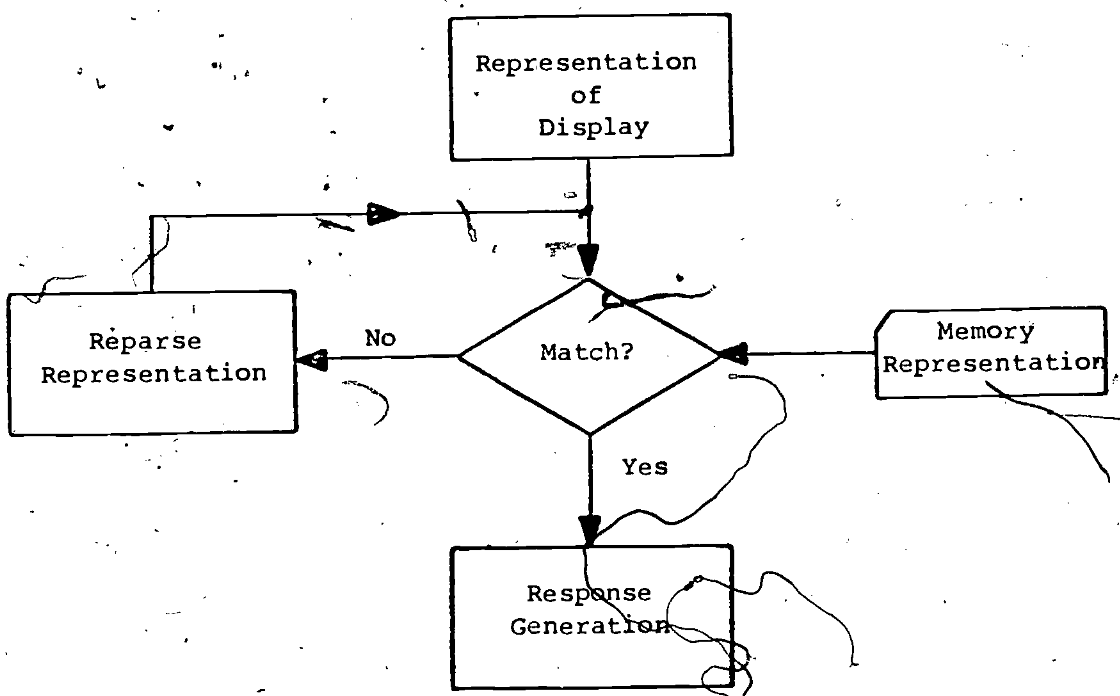

Figure 3. Information Processing Account of Orthogonal Interference

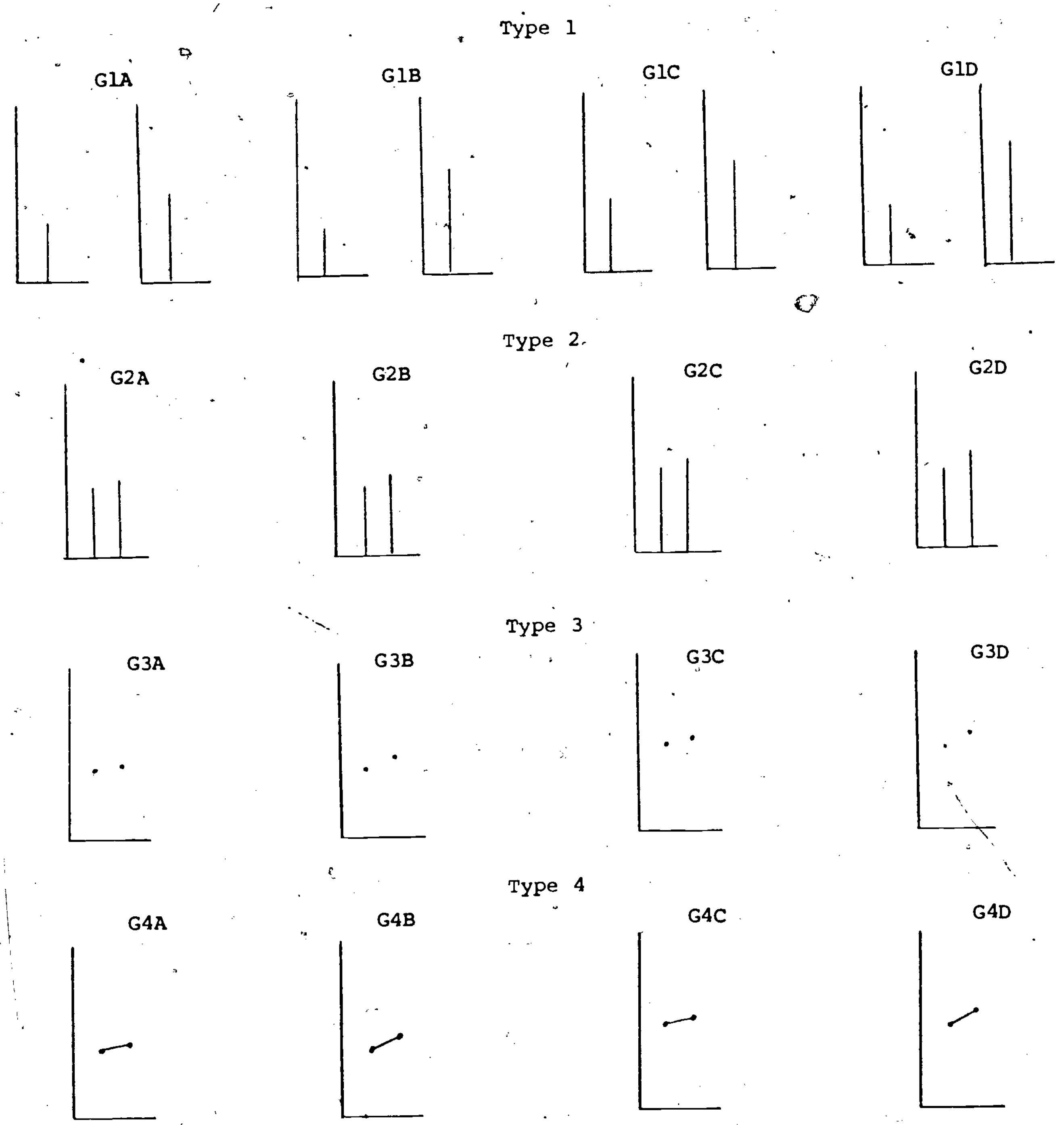

Figure 4: Stimulus set corresponding to response set orientation/overall height

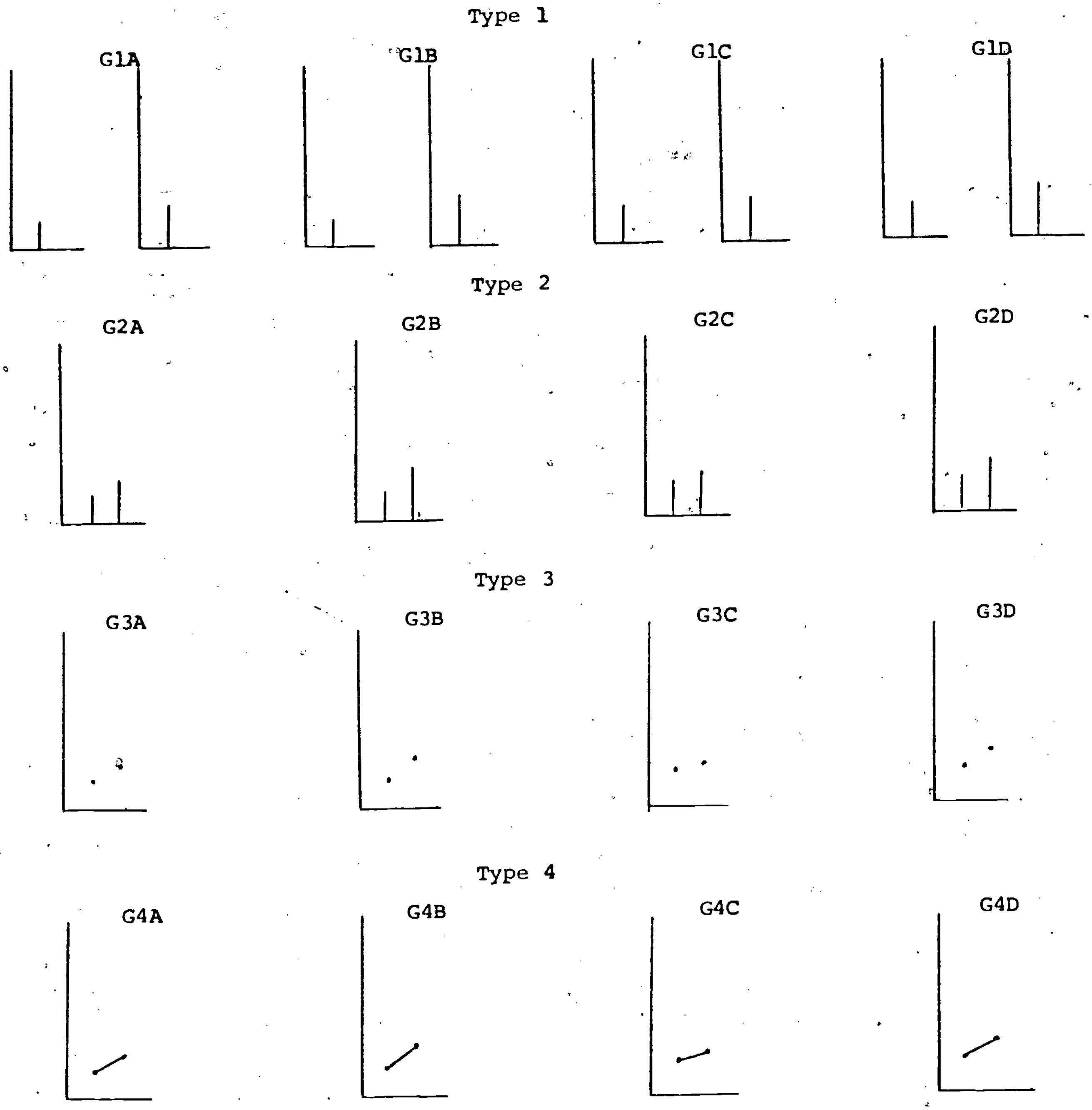

Figure 5: Stimulus set corresponding to response set height of left/height of right

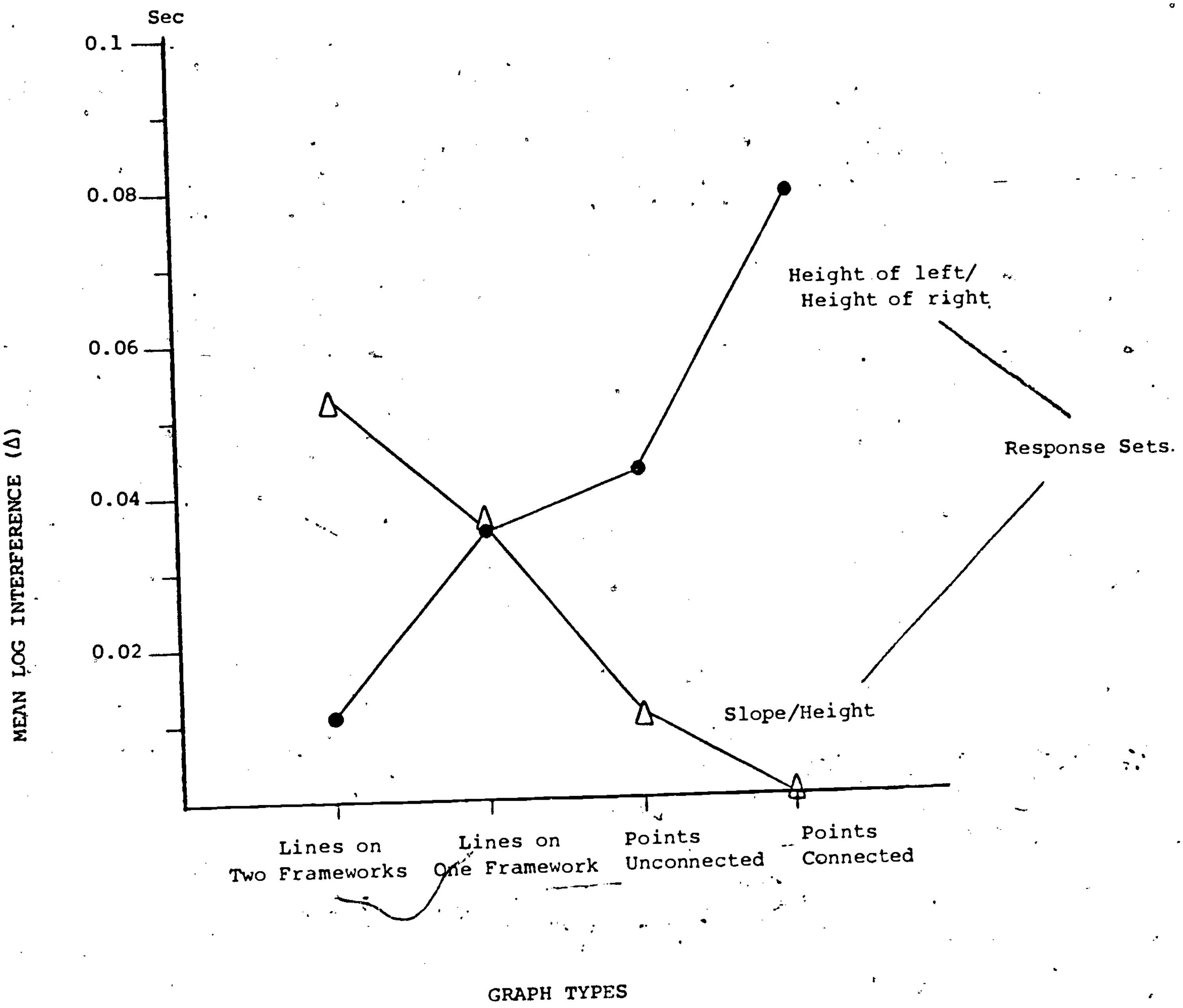

Figure 6 - Magnitude of interference (Experimental sorting time - Univariate sorting time) averaged over defining properties of each graph type.

Height of Left	Height of Right	Misclassification Yes	Misclassification No
Low	Low	9 (.075)	111 (.925)
	High	22 (.183)	98 (.817)
High	Low	11 (.092)	109 (.908)
	High	5 (.042)	115 (.958)

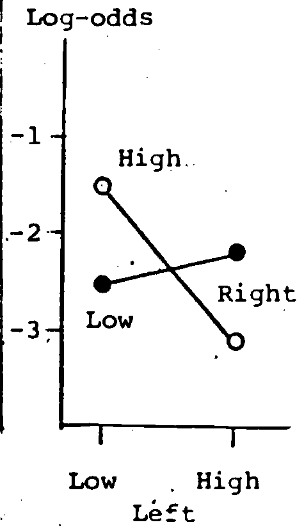

Fig. 7A - Lines on one framework, classification by height of left.

Height of Right	Height of Left	Misclassification Yes	Misclassification No
Low	Low	2 (.017)	118 (.983)
	High	4 (.033)	116 (.967)
High	Low	2 (.017)	118 (.983)
	High	11 (.092)	109 (.908)

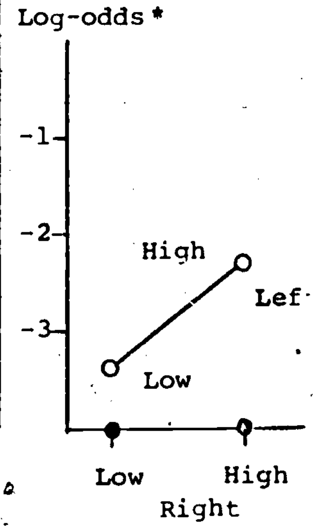

Fig. 7B - Lines on one framework, classification by height of right.

Height of Left	Height of Right	Misclassification Yes	Misclassification No
Low	Low	8 (.067)	112 (.933)
	High	24 (.200)	96 (.800)
High	Low	26 (.217)	94 (.783)
	High	24 (.200)	96 (.800)

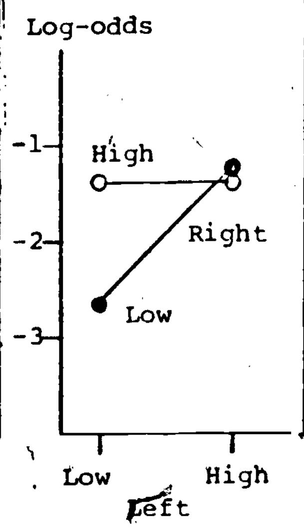

Fig. 7C - Unconnected points, classification by height of left.

Height of Right	Height of Left	Misclassification Yes	Misclassification No
Low	Low	10 (.083)	110 (.917)
	High	5 (.042)	115 (.958)
High	Low	3 (.025)	117 (.975)
	High	17 (.142)	103 (.858)

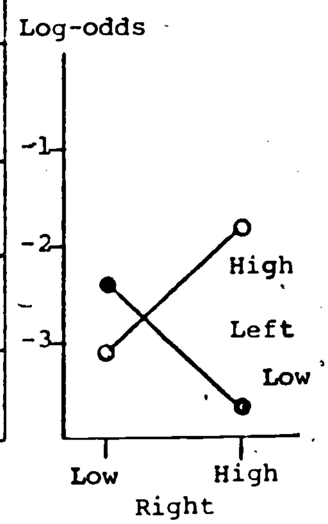

Fig. 7D - Unconnected points, classification by height of right.

* Not significant.

FIGURE 7 - CONFUSION MATRICES AND LOG-ODDS PLOTS OF GRAPH TYPES SHOWING ORTHOGONAL INTERFERENCE

Height of Left	Height of Right	Misclassification Yes	Misclassification No
Low	Low	26 (.217)	94 (.783)
	High	15 (.125)	105 (.875)
High	Low	15 (.125)	105 (.875)
	High	23 (.192)	97 (.808)

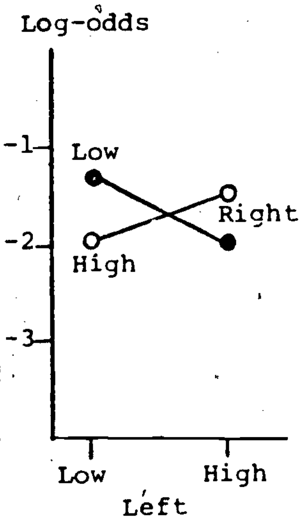

Fig. 7E - Connected points, classification by height of left.

Height of Right	Height of Left	Misclassification Yes	Misclassification No
Low	Low	20 (.167)	100 (.833)
	High	8 (.067)	112 (.933)
High	Low	1 (.008)	119 (.992)
	High	10 (.083)	110 (.917)

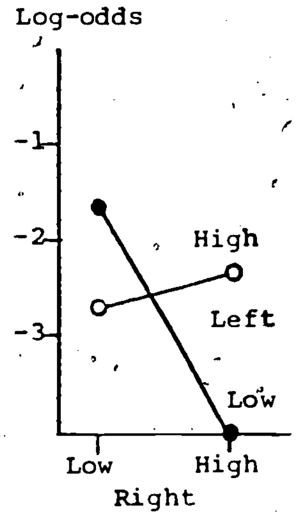

Fig. 7F - Connected points, classification by height of right.

Orientation	Height	Misclassification Yes	Misclassification No
Low	Low	16 (.133)	104 (.867)
	High	22 (.183)	98 (.817)
High	Low	4 (.033)	116 (.967)
	High	1 (.008)	119 (.992)

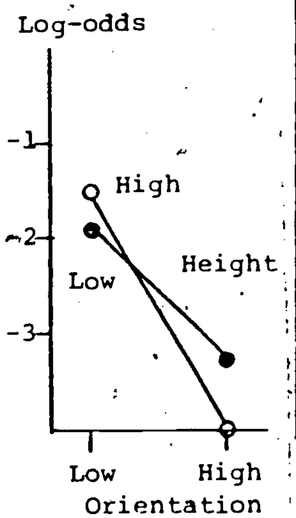

Fig. 7G - Lines on one framework, classification by orientation.

Height	Orientation	Misclassification Yes	Misclassification No
Low	Low	7 (.058)	113 (.942)
	High	34 (.283)	86 (.717)
High	Low	16 (.133)	104 (.867)
	High	24 (.200)	96 (.800)

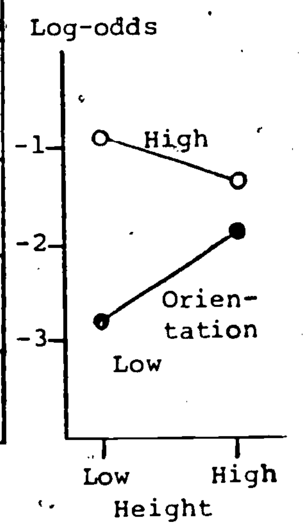

Fig. 7H - Lines on two frameworks, classification by overall height.

FIGURE 7 - CONFUSION MATRICES AND LOG-ODDS PLOTS OF GRAPH TYPES SHOWING ORTHOGONAL INTERFERENCE (CONT.)

TABLE 1

Transformed sorting times (sec) as a function of graph type, response set and task

		Task			
Graph Type	Response	Univariate	Orthogonal	Δ	Contrast
Lines on two frameworks	Height of left	1.393	1.403	0.01	L1
	Height of right	1.376	1.389	0.013	L2
Lines on one framework	Height of left	1.385	1.421	0.036	L3
	Height of right	1.344	1.378	0.034	L4
Points unconnected	Height of left	1.443	1.489	0.046	L5
	Height of right	1.395	1.436	0.041	L6
Points connected	Height of left	1.413	1.494	0.081	L7
	Height of right	1.344	1.421	0.077	L8
Lines on two frameworks	Slope	1.414	1.427	0.013	L9
	Height	1.421	1.513	0.092	L10
Lines on one framework	Slope	1.399	1.450	0.051	L11
	Height	1.347	1.371	0.024	L12
Points unconnected	Slope	1.393	1.396	0.003	L13
	Height	1.388	1.406	0.018	L14
Points connected	Slope	1.396	1.389	-0.007	L15
	Height	1.380	1.355	-0.025	L16

TABLE 2

Confidence intervals, t-statistics and significance levels for preplanned paired comparisons of selected means (LOGTIME) involving the three way interaction of graph type x response set x task

Contrast	Δ	Confidence Interval*	t-statistics	α
L1	0.01	-0.028<L<0.048	~2.7	.0067
L2	0.013	-0.025<L<0.051	2.7	.0067
L3	0.036	0.005<L<0.065	~2.2	.03
L4	0.034	0.003<L<0.065	2.2	.03
L5	0.046	0.008<L<0.084	~2.7	.0067
L6	0.041	0.003<L<0.079	2.7	.0067
L7	0.081	0.043<L<0.119	~2.7	.0067
L8	0.077	0.039<L<0.115	2.7	.0067
L9	0.013	-0.025<L<0.051	~2.7	.0067
L10	0.092	0.054<L<0.130	2.7	.0067
L11	0.051	0.020<L<0.082	~2.2	.03
L12	0.024	-0.007<L<0.055	2.2	.03
L13	0.003	-0.035<L<0.041	~2.7	.0067
L14	0.018	-0.035<L<0.041	2.7	.0067
L15	-0.007	-0.045<L<0.031	~2.7	.0067
L16	-0.025	-0.063<L<0.013	2.7	.0067

Ms Graph Type X Response Set X Task X Subject = 0.0015

*If interval encloses zero, then nonsignificant.

TABLE 3

Confidence intervals,* for preplanned comparisons of selected contrasts testing the degree of interference for graph type respnse set

Contrast	Δ	Confidence Interval
1/2(L3+L4)-1/2(L1+L2)	0.024	-0.009<L<0.057
1/2(L5+L6)-1/2(L1+L2)	0.032	-0.001<L<0.065
1/2(L7+L8)-1/2(L1+L2)	0.067	0.034<L<0.100
1/2(L5+L6)-1/2(L3+L4)	0.008	-0.025<L<0.041
1/2(L7+L8)-1/2(L3+L4)	0.044	0.011<L<077
1/2(L7+L8)-1/2(L5+L6)	0.035	0.002<L<0.068
1/2(L9+L10)-1/2(L11+L12)	0.015	-0.018<L<0.048
1/2(L9+L10)-1/2(L13+L14)	0.042	0.009<L<0.075
1/2(L9+L10)-1/2(L15+L16)	0.068	0.035<L<0.101
1/2(L11+L12)-1/2(L13+L14)	0.027	0.006<L<0.060
1/2(L11+L12)-1/2(L15+L16)	0.053	0.020<L<0.086
1/2(L13+L14)-1/2(L15+L16)	0.026	-0.007<L<0.059

$t(\hat{L}) = 2.356$
$s(\hat{L}) = 0.014$
$p = 0.02$

TABLE 4

Transformed errors as a function of graph type, response set and task

		Task			
Graph Type	Response	Univariate	Orthogonal	Δ	Contrast
Lines on two frameworks	Height of left	1.426	1.426	0.000	L1
	Height of right	0.856	0.839	-0.017	L2
Lines on one framework	Height of left	1.093	1.456	0.362	L3
	Height of right	0.476	0.831	0.355	L4
Points unconnected	Height of left	1.876	2.172	0.296	L5
	Height of right	0.522	1.153	0.631	L6
Points connected	Height of left	1.471	2.060	0.539	L7
	Height of right	0.476	1.348	0.972	L8
Lines on two frameworks	Orientation	0.821	0.820	-0.001	L9
	Height	0.990	2.299	1.309	L10
Lines on one framework	Orientation	0.986	1.577	0.591	L11
	Height	0.255	1.047	0.792	L12
Points unconnected	Orientation	0.910	1.283	0.373	L13
	Height	0.731	0.938	0.207	L14
Points connected	Orientation	0.899	0.932	0.033	L15
	Height	0.824	1.136	0.312	L16