Conceptual Limitations in Comprehending Line Graphs

Article *in* Journal of Experimental Psychology General · March 1995

DOI: 10.1037/0096-3445.124.1.43

Priti Shah
University of Michigan

Journal of Experimental Psychology: General
1995, Vol. 124, No. 1, 43–61

0096-3445/95/$3.00

Conceptual Limitations in Comprehending Line Graphs

Priti Shah and Patricia A. Carpenter
Carnegie Mellon University

This article examines the conceptual limitations that influence a viewer's mental representation of a data set of 3 continuous variables. Such data can be graphed in 2 ways, depending on which variable is depicted on the *x* axis and which is made a parameter on the curve (the *z* variable). The results of several studies indicate that a viewer's internal representation of a graph expresses quantitative functional information about the *x–y* relations (such as "*y* decreases exponentially as *x* increases"). By contrast, the *z–y* information is represented ordinally or even nominally. This difference in the mental representations of the variables was evident when viewers described or interpreted simple line graphs or judged the equivalence of successively presented graphs. The difference persisted in spite of the viewer's accurate encoding of the graph's visual pattern, as indicated by accuracy in reproducing them from memory. Viewers seldom formed an integrated representation of the 3 variables, even when explicitly instructed to encode them. The *x–y* line patterns may automatically activate quantitative concepts, whereas viewers may be unfamiliar with the quantitative interpretation of *z–y* patterns.

The display of numerical data in a graphic form is an invention of the late 18th century that takes advantage of highly developed human capabilities to recognize visual patterns (Bertin, 1983; Fienberg, 1979; Lewandowsky & Spence, 1989). The visual patterns revealed by graphic displays of data have been linked to several scientific discoveries, including the patterns of stellar evolution (Lewandowsky & Spence, 1989) and chaos in weather patterns (Gleick, 1987). Nevertheless, underlying the apparently effortless comprehension of graphs are a complex set of perceptual and conceptual processes, according to several theories of graph comprehension (Cleveland & McGill, 1985: Kosslyn, 1989; Lohse, 1991; Pinker, 1990). The present research specifies the processes underlying the comprehension of one commonly used graphic display: line graphs.

The current experiments investigate the comprehension of line graphs depicting three continuous variables, such as the line graphs in Figures 1a and 1b. The lines in a two-dimensional line graph of three variables are orthographic projections of a three-dimensional data set in which the variable on the *x* axis is depicted as continuous, and the other independent variable is depicted as discrete, whether or not the underlying scales are discrete or continuous. The graph in Figure 1c depicts the three-dimensional data set at an intermediate orientation lying between the two conventional line graphs in Figure 1a and 1b. As Figures 1a and 1b illustrate, two perspectives of the data set can be generated by selecting different independent variables to be graphed along the *x* axis. (For convenience, we will refer to the *x* and *z* variables of two-dimensional graphs as independent variables, although they need not be experimentally manipulatable.) The two perspectives, although based on the same data set, are not psychologically equivalent. For each perspective, some information about the *x–y* relations is automatically retrieved, whereas information about the *z–y* relations must be inferred by comparing the relative *y* values for a given *x* value for different lines. For example, one of the most salient features of Figure 1a is the negative effect of increasing noise on test performance; for Figure 1b, this effect must be inferred by comparing the *y* values associated with the two lines. Hence, such information is less likely to be available to the graph reader of Figure 1b. Similarly, a salient feature of Figure 1b, the differential effect of room temperature on achievement test scores for low and high noise levels, must be inferred from Figure 1a by comparing the *y* values of the two lines at different points for *x* (at a low noise level and also at a high noise level). The differing representations that a person internalizes from the two perspectives of the same data can be used to investigate the characteristics and limitations of the processes in interpreting line graphs.

A priori, there are three possible formats for the internal representation of three-variable line graphs: three-dimensional spatial, visual feature, and propositional. A three-dimensional spatial representation includes a representation of the data set's dimensionality, and is similar to the internal representation of a three-dimensional object. The subse-

Priti Shah and Patricia A. Carpenter, Department of Psychology, Carnegie Mellon University.

We thank Larry Wasserman for statistical advice and Susan Chipman, Jim Hoeffner, Marcel Just, Stephan Lewandosky, Akira Miyake, and Yuko Munakata for their helpful comments on earlier versions of this article. We also thank Paulette Williams for assistance in transcribing verbal descriptions and Rory Marler for his assistance in testing experimental participants for Experiment 2. This research was supported by grants from the Mellon Foundation and Office of Naval Research Grant N00014-92-J-1209.

Correspondence concerning this article should be addressed to Patricia A. Carpenter, Department of Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213.

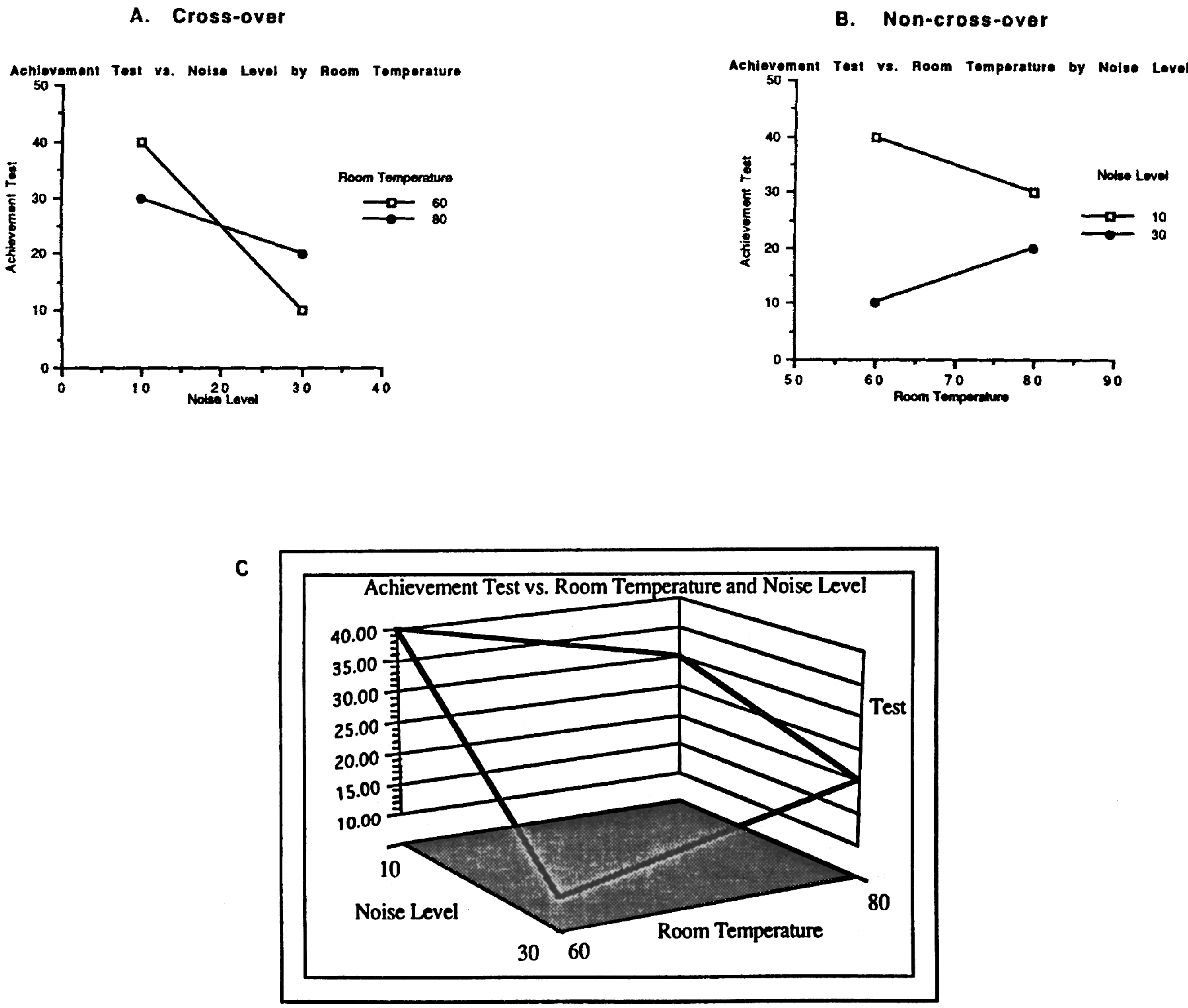

Figure 1. Two perspectives of an exemplar depicting pronounced interactions in which the main visual features of the graphs differ greatly. The graph depicted in Figure 1a is the *cross-over* graph; the graph in Figure 1b, the *non-cross-over*. See the text for further explanation of the types of interaction. The bottom graph depicts the three-dimensional data space depicted by the two graphs above. Note that we refer to the dependent variable (achievement test scores) as the y variable and the two independent variables as the x variable (room temperature in Figures 1b and 1c) and the z variable (noise level in Figures 1b and 1c; the parameter on the curves, or in the case of the three-dimensional graph, the variable displayed in depth).

quent experiments show little evidence that viewers spontaneously form a three-dimensional spatial representation, and indeed, do not even do so when given explicit instructions. By contrast, a visual feature representation has no representation of the dimensionality of three variables, but rather contains information about the two-dimensional visual pattern. A third possible format, a propositional representation, consists of a limited number of propositions describing the major functional relationships, such as the direction and possibly the magnitude of x–y relations for at least some levels of the z variable. The propositional format appears to be the dominant format, although a visual feature representation may coexist with it. In summary, we propose that the dominant internal representation of a line graph consists of propositions expressing the major functional x–y relations in the graph, and perhaps some representation of salient visual features.

To illustrate the possible representation of a line graph, consider Figure 1a (which will be called a *cross-over* interaction). A propositional representation of this graph might contain the following information, which is based on the verbal descriptions that viewers typically give such a graph:

> Achievement test scores decrease as noise level increases when it is 60 degrees.
>
> Achievement test scores decrease as noise level increases when it is 80 degrees.
>
> Achievement test scores decrease more when it is 60 degrees than when it is 80 degrees.

In addition to these propositions, there may be some internal representation of the salient visual features, such as the general steepness of the slopes or the fact that the lines cross in the middle of the graph.

In contrast to the description of Figure 1a, the graph in Figure 1b typically elicits a verbal description that is superficially quite different:

> Achievement test scores decrease with room temperature for low (10 db) noise levels.
>
> Achievement test scores increase with room temperature for high (30 db) noise levels.
>
> Achievement test scores for low (10 db) noise levels are higher than for high (30 db) noise levels.

In Figure 1a, both of the functional relations of the *x–y* lines are decreasing; by contrast, in Figure 1b, one *x–y* function is increasing. Although the two sets of propositions are obviously compatible, when the two graphs are displayed successively, viewers often erroneously judge that they are depicting different data sets. We claim that the two graphs are interpreted as different because the two perspectives tend to evoke different functional relations, and it is difficult to infer one set of *z–y* relations from the other set of *x–y* relations. This research examines the processes in interpreting such graphs by analyzing how the interpretation of a data set is influenced by the graphic perspective in which it is presented.

Recent analyses of graph comprehension have described three types of processes that are particularly relevant to the interpretation of line graphs (Bertin, 1983; Cleveland & McGill, 1985; Lohse, 1991; Pinker, 1990; Simkin & Hastie, 1987). One process is the encoding of the major visual patterns, such as whether there is a straight or jagged line, and if there are multiple lines, whether they are parallel, converging, intersecting, and so on. By focusing on simple line graphs, the current experiments minimize the likelihood that this process is a source of error. Nevertheless, accurate encoding is a necessary component of comprehending line graphs, and its role will be evaluated in Experiment 2.

A second set of processes is the translation of visual features to the conceptual relations that are represented by those features (Kosslyn, 1989; Pinker, 1990). These processes include the retrieval of quantitative knowledge associated with the visual pattern, such as the knowledge that an upwardly curved line represents an accelerating function. When visual patterns readily evoke the appropriate quantitative concept, comprehension appears to be relatively effortless. However, certain quantitative or functional information that a graph viewer has not previously associated with visual features may require complex inferential processes (Pinker, 1990). Specifically, the *z–y* functions are indicated by the pattern of spaces (relative *y* values) for different *z* values that share the same *x* value, and these patterns tend not to trigger the retrieval of an associated quantitative label. Consequently, in Figure 1a, the relationship between room temperature (the *z* variable) and achievement test scores (the *y* variable) must be explicitly inferred by noting the *y* values first at the lower noise levels and then at the higher noise levels. Such inferential processes may be less accurate and less likely to be executed than the automatic retrieval of quantitative functions that are associated with *x–y* patterns. The current research begins to specify what types of concepts are or are not associated with various visual patterns in line graphs, particularly contrasting between the knowledge evoked by patterns in the *x–y* plane compared to the *z–y* relations.

A third group of processes includes determining the referent of the concepts being quantified and associating those referents to the encoded functions. One aspect of the referential processes, which we will call *labeling processes*, involves reading the legends and the quantitative values on the axes and relating these to the corresponding lines and data points. Labeling processes are likely to interact with both pattern recognition and the retrieval of associated conceptual knowledge, if, for example, viewers look up the numerical values of salient points or features that indicate important mathematical relations (such as a cross-over point or a point of inflection). These labeling processes will be shown to be particularly error prone if the values of the third variable are in a legend, as they are displayed in many technical articles and in the output of several computer graphics packages (Schmid, 1983).

The characteristics and limitations of the major types of processes (pattern encoding, retrieving and inferring functional relations, and labeling) can be analyzed by determining when they result in an incorrect or incomplete representation of a data set. Specific errors or types of incompleteness can be predicted by two hypotheses about the properties of the internal representation. The first hypothesis is that the internal representation of the *z* variable is, in general, much less likely to have metric information about the absolute or relative quantitative differences among the *y* values, compared with the representation of the *x* variable. This hypothesis follows from the claim that, whereas *x–y* patterns may automatically evoke quantitative labels, the *z–y* patterns usually require inferences to be computed. Without such inferences, the representation of the *z* variable may be more similar to an ordinal representation, with information about the rank ordering of the values (first, second, third), or to a nominal scale in which values are distinguished but not ranked. For example, an ordinal representation of an accelerating function would contain information that the *y* values increase but would not contain information about the relative differences in the amount of increase as a function of changes in the independent variable.

The second hypothesis is that there is an asymmetry in the internal representation of the alternative perspectives of three-variable line graphs. This hypothesis also arises from the claim that *x–y* patterns trigger the retrieval of the associated functional relations, whereas the representation of z–y patterns requires inferences. Thus, the x–y relations should be more frequent and more salient in the viewer's representation than the z–y relations. It is this asymmetry that is exploited in the comparison of Figures 1a and 1b. In each graph, the more automatically retrieved x–y relations may differ from predicates describing the z–y relations. Consequently, the x–y relations generated for each perspective seem to describe incompatible features.

These two hypotheses provide the rationale for the methodology we used to study the processing of line graphs. We constructed data sets with specific characteristics that would illuminate each hypothesis. The limitation in representing metric information was examined by constructing two types of *metric* data sets that each had specific metric properties (acceleration or inflection) that we hypothesized would more frequently be a source of error in encoding, retrieval, or labeling processes when it was depicted on the z axis than when it was on the x axis. The neglect of z–y relations relative to x–y relations was examined by constructing two *interaction* data sets in which the x–y relations differed from the z–y relations and determining which relations were more frequently described. In the next section, we describe the two metric data sets and then the two interaction data sets.

1. *Acceleration data set: curve and intercurve spacing perspectives.* This type of metric data set contains an accelerating (or decelerating) function to test the hypothesis that its metric property would be more accurately represented by the viewer when it is plotted on the x axis than when it is plotted on the z axis. On the x axis, the function is realized as an accelerating curve, as shown in Figure 2a; this perspective will be referred to as the *curve* perspective. By contrast, when the function is plotted on the z axis, the acceleration is manifested as unequal spaces (y increments) between successive lines that represent increments in the z dimension. This perspective, exemplified in Figure 2b, will be referred to as the *intercurve spacing* perspective. Inaccuracy in comprehending graphs such as Figure 2b may arise from two sources. First, the inequality of the successive spaces between functions may be less visually salient than the acceleration in the curve perspective, so that viewers may not be as likely to encode that visual feature. Second, viewers may not have associated a change in intercurve spacing with an accelerating mathematical function. In this case the intercurve spacing pattern may be encoded, but it may not evoke the retrieval of the associated quantitative property.

2. *Inflection data set: V-pattern and legend perspectives.* This type of metric data set has an inflected function, in which the y values for one variable decrease and then increase (or vice versa). It was predicted that the inflection is more likely to be accurately represented when it is plotted as the x–y relation than as the z–y relation. On the x axis, the low (or high) value is visually salient, as in Figure 3a; this will be called the *V-pattern* perspective. By contrast, if the inflected function is the z variable, as in Figure 3b, the inflection must be inferred from the legend (hence, the *legend* perspective), using labeling processes to encode the values in the legend and associate them with the appropriate lines. These labeling processes, if errorful or incomplete, could result in the incorrect inference that the successive z values of lines in the legend graph increase monotonically.

The remaining two data sets were *interaction* data sets, which were constructed to reveal the asymmetry in the internal representation that we hypothesized neglects z–y relations and emphasizes the x–y relations. In an interaction, the effect of the x variable on the y variable depends on the level of the z variable. Logically, an interaction is dependent on both the x and the z variables, but psychologically, the two variables are not equivalent because the x–y relation-

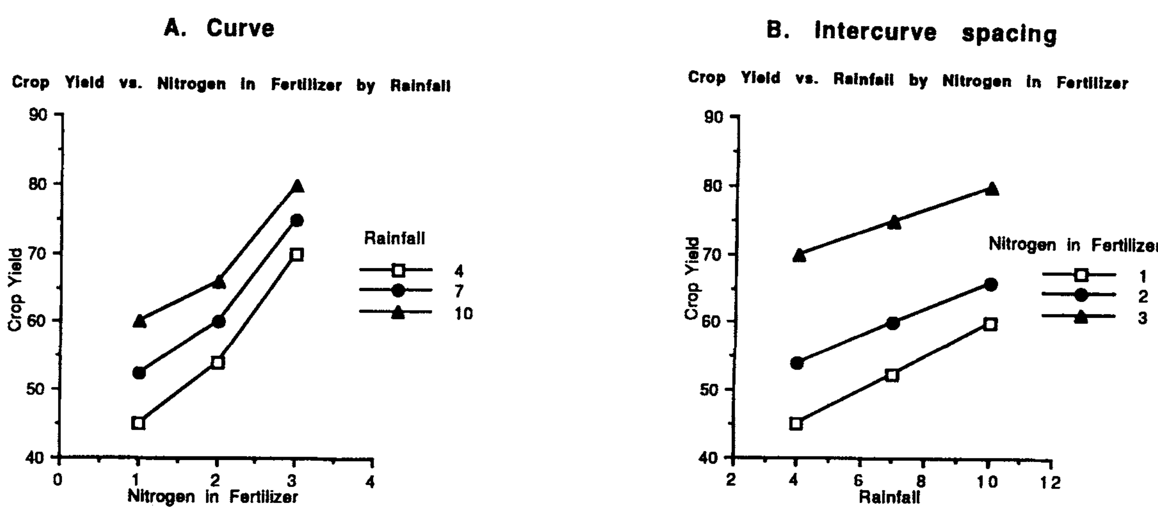

Figure 2. Two perspectives of an exemplar used to investigate the linearity assumption in encoding the third variable. A: An example of a *curve* graph; the x variable is curvilinear and hence, the nonlinearity is salient and familiar. B: An example of a *intercurve spacing* graph; the nonlinearity in the z variable can be inferred from the unequal spacing between equally spaced values of the third variable.

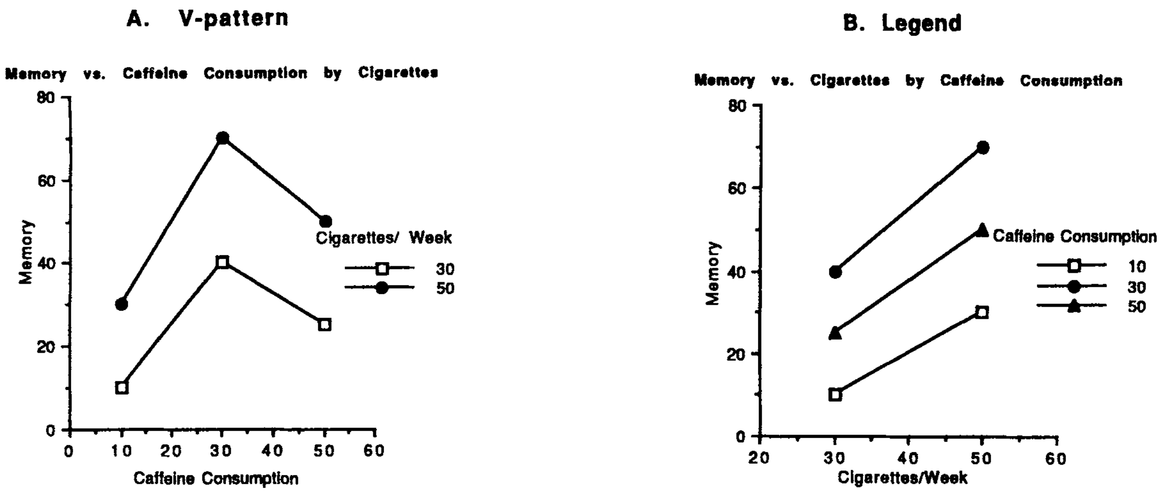

Figure 3. Two perspectives of an exemplar used to investigate the difficulty in the labeling process of the lines on the graph and the monotonicity assumption in encoding the third variable. A: An example of the *V-pattern* graph; the inflection is placed in the x–y plane. B: A *legend* graph; the inflection can be inferred by comparing the legend to the lines on the graph.

ship is directly retrievable, whereas the z–y relationship must be inferred from the directed differences between the z values at various points along the x dimension.

3. *Common value interaction data set: Line and point perspectives.* In a "common value" interaction, one independent variable has no effect on the y variable for one value of the second independent variable. However, the lack of effect may be internally represented differently for the two perspectives. For the *line* perspective (see Figure 4a), one of the salient properties is the horizontal line, which may automatically evoke the retrieval of the associated conceptual relation "no effect." For the alternative *point* perspective (see Figure 4b), the salient properties may include increasing or decreasing lines but may not include the lack of effect which is depicted by the two (or more) functions meeting at a point. The hypothesis is that a graph viewer will be more likely to represent the x–y lines as decreasing, and perhaps less likely to retrieve the knowledge that a common point indicates a lack of effect. Hence, the prediction is that the lack of effect is less likely to be part of the interpretation for the point perspective than for the line perspective.

4. *Cross-over interaction data set: Cross-over and non-cross-over perspectives.* In a *cross-over* interaction, one variable has a greater or lesser effect, depending on the value of the second independent variable, as illustrated earlier in Figures 1a and 1b. The hypothesis is that the alternative perspectives of this interaction may generate

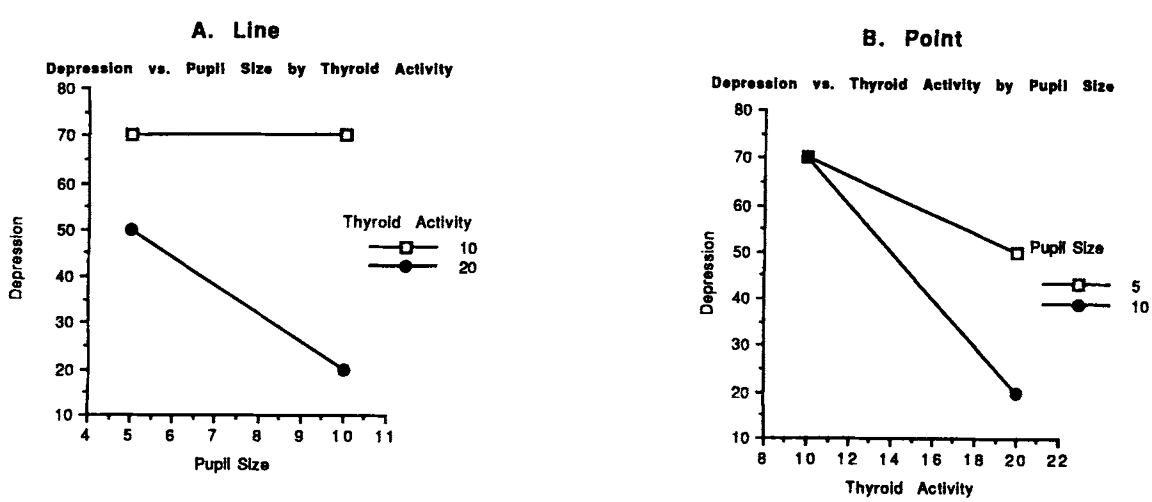

Figure 4. Two perspectives of an exemplar that depicts an interaction in which there is no effect of one of the variables at one value of the other variable. In the *line* graph (A), this lack of effect is less visually salient than the *point* graph (B).

different representations because the graphs differ in which aspects of the interaction they make salient, through the slopes of the x–y lines. To represent the z–y relations may require more complex inferences. In the cross-over graph in Figure 5a, the salient features are the main x–y relations, in this case, that both functions increase with the x dimension, and that one increases at a greater rate. By contrast, in the *non-cross-over* perspective (as in Figure 5b), the salient x–y relations include the fact that the direction of the effect of x on y reverses for the two functions. The hypothesis is that the tendency to retrieve automatically the x–y relations for each perspective will result in different internal representations of the two types of graphs.

In summary, the current studies examine the processes in interpreting three-variable line graphs with an emphasis on the format and limitations of the resulting representation. We propose that a graph viewer's representation of three-variable data consists primarily of a set of propositions about the x–y relations and does not directly represent the dimensionality of the data. The internal representation typically has little metric information about the z–y relations, although as Experiment 2 will indicate, the representation may include information about the major visual features in the graph. Experiment 1 establishes the basic limitations in the comprehension of these simple line graphs; Experiment 2 helps to localize the processing sources of the errors; and Experiment 3 demonstrates that the same limitations constrain experts (who have had years of experience in empirical research and statistical training) and novices (college undergraduates).

Experiment 1

A major hypothesis of this study is that viewers are less likely to represent accurately the metric information from the z variable than from the x variable. Instead viewers may represent the z variable ordinally or nominally, due to the relative difficulty of inferring z–y relations compared to retrieving quantitative labels associated with the x–y pattern. To test this hypothesis, participants were asked to describe briefly a series of individually presented graphs that varied the patterns that were displayed on the x and z axes. The viewers' descriptions were classified according to the type of information that they provided about the z variable—whether the values of the lines were only named (a nominal encoding) or whether the description included ordinal relations, incorrect ordinal relations, or metric properties and relations. The prediction was that metric properties would be less frequently and less accurately described if the function containing those properties were on the z axis than if it were on the x axis.

To provide additional evidence about what information was preserved in the internal representation, the description task was followed by a perspective comparison task. The initial graph was removed, and the viewer judged the comparability of the underlying data to that depicted in the newly presented graph of the alternative perspective. The alternative perspective depicted the same data on half of the trials, but it was altered to be a plausible foil on the other half.

Theoretically, several alternative strategies could be used to perform the perspective comparison task, and these strategies may be related to the format of the internal representation of the graph. One strategy is to generate a three-dimensional representation of the underlying data set from the initial perspective, and then to rotate it mentally into the alternative perspective to perform the comparison task (Shepard & Cooper, 1982). The accuracy of the comparison would be expected to correlate with spatial ability (Just & Carpenter, 1985). In fact, there was no evidence of a three-dimensional representation either from subject report or from the pattern of correlations between comparison task performance and spatial ability measures. Intuitively, this strategy seems to be effortful, and in Experiment 2, we demonstrate that even architecture and design students find

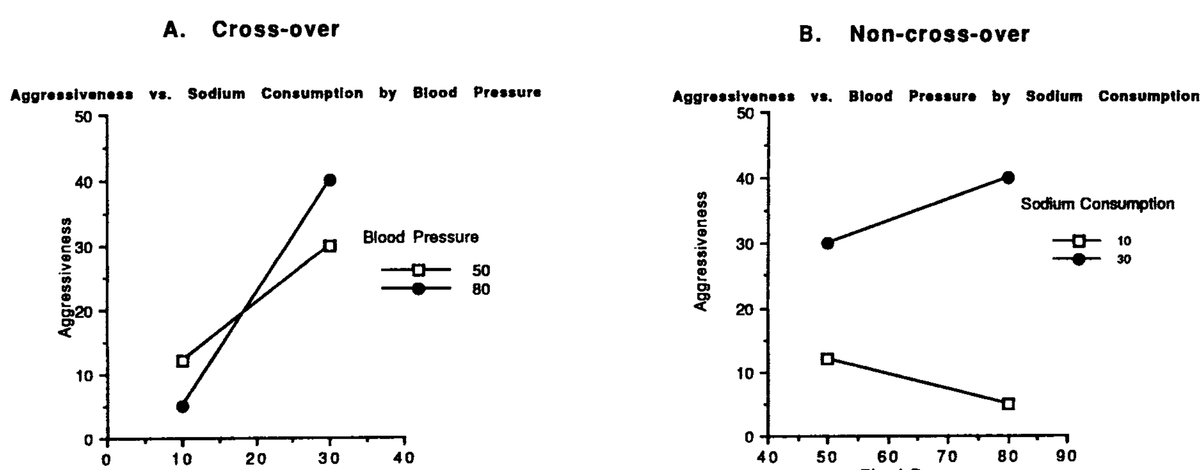

Figure 5. Two perspectives of a second exemplar of the cross-over interaction data set. A: the cross-over graph; B: the non-cross-over graph.

it difficult to generate a three-dimensional representation of a data set and rotate it.

Another strategy for the comparison task is very commonly reported, and it is consistent with a representation of the first graph that includes the major x–y functional relations. In this strategy, individuals check if the x–y functions they had coded from the initial graph are compatible with the z–y properties of the comparison graph. One prediction arising from this strategy is that the number of distinct functional relations depicted in the initial graph should influence the number of relations to be checked; consequently, it might correlate with individual's accuracy in the comparison task. The number of functional relations should be related to the number of lines with different slopes or shapes. In addition, the type of information that is represented about the z–y relations in the initial description may correlate with the accuracy of the comparison task. Thus, trials for which metric information about the z–y relations is described may generally result in more accurate comparisons than trials on which only nominal or ordinal information is described.

Finally, there is some evidence that viewers represent some of the visual features of the graph; visual features alone are not a sufficient basis for accurate performance in the comparison task. Nevertheless, the role of the visual feature representation will be examined in Experiment 2.

Method

Participants. Students from Carnegie Mellon University participated in the study ($N = 24$; 20 undergraduate and 4 graduate).

Stimuli. Figures 2 through 5 show examples of the eight graph types that viewers described, although the labels at the top of each graph (such as "cross-over") are just for current explanatory purposes and were not presented in the experiment. Each viewer described a total of 16 graphs, consisting of two exemplars of each of the two perspectives of the four data sets.

In all cases, the three labels for a graph represented variables that are plausibly related to one another (e.g., achievement test, noise level, and room temperature), but the relation is not commonly known.

To generate the 16 graphs, each of the four data types (i.e., acceleration, inflection, common-value interaction, cross-over interaction) was instantiated in two distinguishable versions that contained the main property of that data type. For example, for the cross-over data set, in one version y sharply increased as a function of one of the independent variables (as in Figure 5a and 5b), and in the other version, y sharply decreased (as in Figure 1a and 1b). Two versions of each data set were used so that in the comparison task, the same type of perspective graph could be presented twice to a subject, once for a *same* and once for a *different* comparison trial.

For each version of a data set, two variants were constructed that contained identical graphic lines but different labels and whose numerical values were linear transformations of each other. Thus for each data set, a total of four variants were created: two different versions (with distinguishably different lines) and within each version, two sets of labels and values.

For the comparison task, a graph was paired with either a same or different foil. For the same trials (half of the trials), the viewers saw the identical data points plotted from the alternative perspective. For the different trials (the other half of the trials), the alternative perspective data was systematically altered. For the two metric data sets (acceleration and inflection), the different foils varied only in this metric property. For example, in a different foil version of the curve graph, the x–y function had no acceleration. For the interaction data sets, the different foils depicted functional relations that were similar to those in the initial graph, but qualitatively dissimilar from those in the accurate alternative perspective. In general, errors in judging the comparability of the second graph would support the hypothesis that the subjects were relying primarily on their memory of the ordinal or nominal properties of the initial graph.

In the comparison task, each graph was randomly paired with a same or different foil, with the constraint that a same and a different foil occur once for each perspective type. The same and different response to each initially described graph was balanced by forming a second condition with the opposite assignment of same and different foils to graphs. A third and four condition were formed by presenting a different variant of each of the 16 graphs, and balancing the same and different foils across these. Thus, there were four different groups, to which viewers were randomly assigned with the constraint that there be an equal number in each.

In addition to the 16 graphs each viewer described, another 12 served as fillers. For six fillers, the axes of the comparison graph were the same as the axes of the initial graph, to encourage participants to check the axes. For other fillers, the graphs for both perspectives were visually similar, as in graphs with parallel functions from both perspectives. The filler trials were easier than the experimental trials and were included to minimize frustration with the comparison task.

The physical layout of the graphs, such as the characteristics and location of labels and the legend, were typical of line graphs, and are standard options of the graphing program that was used to generate the graphs (Cricket graph for the Macintosh). As illustrated in Figures 1–5, the x and y variables were clearly labeled near the axes. The label and values of the z variable were given by labeled line segments in a legend on the right side of the graph. A title was written above the graph that specified all three variables, following the conventional order of mention: y as a function of x by z. Each graph measured approximately (4.5 × 7 in.).

Procedure. Viewers briefly described the data in the first graph. It was removed, and they were then shown a second graph and were asked to judge whether the data in the second graph were similar in relationships and values to those in the first. They were told that individual values need not be identical, although the values were identical in the same condition. However, the purpose of the instructions and the structure of the sequential comparison task was to encourage viewers to attend to the gist of the pattern, and not to compare values point by point. Before the experiment began, subjects were shown three examples to give them practice with the general procedure. The order of 16 experimental and 12 filler trials was randomized for each viewer with the constraint that the two versions of a single perspective occurred in different halves of the experiment. In addition, viewers were administered a psychometric test of spatial ability, the Vandenberg Mental Rotations Test (Vandenberg, 1971), and they completed a questionnaire that asked about their mathematics and verbal Scholastic Aptitude Test (SAT; Educational Testing Service) scores.

Several features of the experiment were designed to instruct and remind viewers that the z variable of the initial graph was continuous, rather than nominal or ordinal, so that a failure to encode the metric properties of the z variable would not reflect the incorrect inference that the z variables were truly ordinal. The continuous nature of the z variable was described in the task instructions.

Moreover, this property was implicit in features of the graphs, including the title of the z variable, the values in the legends, and in the perspective comparison task itself, in which the x and z variables were interchanged.

Classifying verbal descriptions. The viewers' descriptions of the initial graph were classified according to the information given about a particular variable, specifically, whether it was described nominally, ordinally, metrically, or not mentioned at all. One analysis concerned the z variables of the two alternative perspectives, and so, involved different mathematical properties. Another analysis contrasted the description of the same property on different graphs, as a function of whether the property was depicted on the z axis or on the x axis.[1]

Nominal descriptions included the names of the z values but did not provide any ordinal or metric information about the z–y relation. For example, for the non-cross-over perspective graph (shown in Figure 5b) that related the blood pressure (the x variable) and sodium consumption (the z variable) to aggressiveness levels (the y variable), a nominal description would be, "For a sodium consumption of 30, an increase in blood pressure shows an increase in aggressiveness, while for a sodium consumption of 10 an increase in blood pressure shows a decrease in aggressiveness." The description is classified as nominal because the relation between the performance at the 10 and 30 sodium consumption (which happens to be higher at 30 than 10) is not explicitly described. Obviously, the nominal classification does not preclude the possibility that the viewer encoded the order information.

Ordinal descriptions included the explicit mention of the relation between the third variable and the dependent variable. In the example above, the order information might have been described as: "And, the aggressiveness levels for high sodium consumption are greater than for low sodium consumption." If the order was incorrectly described, it was classified as an "incorrect ordinal" description.

Metric descriptions included a more detailed quantitative explanation of the relations between the third variable and the dependent variable. For graphs depicting the metric data sets (the acceleration and inflection data sets), a description was classified as metric if it included interval or ratio detail of the relationship between the third variable and the dependent variable. That is, a metric description included information about changing influence of the z–y relations as a function of the z variable. For example, a metric description of the z variable for the intercurve spacing graph in Figure 2b would have included a statement about the increasing influence of nitrogen on crop yield. For graphs depicting the interactions (the cross-over interaction and common value interaction data sets), a description was classified as metric if it mentioned differential effects of the third variable on the dependent variable as a function of the other independent variable. That is, a metric description included information about the changing influence of the x variable on the z–y relations. In the example of the non-cross-over graph above, a metric description might include the phrase, "There is a greater increase in aggressiveness levels as sodium consumption increases when blood pressure is high."

Results and Discussion

Descriptions. The typical graph descriptions that viewers gave reflected the relative ease of retrieving the explicitly represented x–y relations and the relative difficulty of inferring the z–y relations. The descriptions included the major x–y relations, often labeled with the individual z values, but typically did not include information about the main effect of the z variable, or included either incomplete or incorrect information about it. Consistent with the first hypothesis, the proportion of trials on which subjects described the specific metric property of the acceleration and inflection data sets was greater when the property was displayed on the x axis than when it was displayed on the z axis. The property of acceleration was twice as likely to be described when it was displayed on the x axis (48% of the descriptions of the *curve* perspective) than when it was on the z axis (23% of the descriptions of the intercurve spacing perspective), $t(23) = 2.93$, $p < .01$. These descriptions are classified as metric in Table 1. Similarly, inflection was more likely to be described when it was displayed on the x axis (92% of the descriptions for the V-pattern perspective) than when it was displayed on the z axis (53% of the descriptions for the legend perspective), $t(23) = 3.4$, $p < .01$. (These numbers are the sum of the ordinal and metric categories in Table 1). In fact, for 28% of the descriptions given to the legend graph, the ordinal information about the values of the z variable were incorrect, presumably because viewers attended to the x–y lines and assumed that the z values were monotonically ordered. There were no such errors for the descriptions of the V-pattern perspective because the nonmonotonicity is part of the x–y pattern. Hence, the types of descriptions that were generated were consistent with the hypothesis that metric information is more likely to be internally represented, and hence described, if it is a property of the x variable than if it is a property of the z variable.

Consistent with the second major hypothesis, viewers' descriptions of the graphs depicting the interaction data sets emphasized the x–y relations and resulted in different types of descriptions for the alternative perspectives of the same data set. They referred to the x–y functions (typically, the sign of the slopes and relative steepness) in 94% of the descriptions of the graphs depicting the interaction data sets. By contrast, they did not typically describe the z–y functions. Instead, on 55% of the trials, they used the z values as labels (nominal description) to distinguish among the x–y lines. (The Appendix presents the frequency of various types of descriptions of the z variable for all the graph types; it differs from Table 1, which gives the descriptions of specific metric properties that were either on the x axis or on the z axis).[2]

Overall, viewers' descriptions tended to consist of the major x–y functions, with the z values used as labels to distinguish qualitatively different functions. Specifically, the frequency with which viewers gave nominal descrip-

[1] Forty-eight (13%) of the descriptions were coded by a second judge, who agreed with the first judge on 88% of the descriptions.

[2] The frequencies of different types of descriptions of the z variable were consistent across the two exemplars and perspectives of a data set with the exception of the two exemplars of the common value interaction data set. One exemplar elicited nominal descriptions 55% (for the point graph) and 78% of the time (for the line graph); the lines in the graphs were rated less similar to each other than those in the other exemplar, which elicited nominal descriptions 33% and 35% of the time, respectively.

Table 1
Experiment 1: Descriptions of Property Depending on Whether the Function was Displayed on the z or x Axis (in Percent)

Property	Classification			
	Nominal	Incorrect ordinal	Ordinal	Metric
Acceleration				
z axis (intercurve spacing)	11	0	64	23
x axis (curve graph)	4	0	48	48
Inflection				
z axis (legend graph)	15	28	51	2
x axis (V-pattern graph)	2	0	63	29
M	8	7	57	25

Note. The proportion of trials in which the relevant variable was not mentioned (an average of 3% of the trials) can be calculated by subtracting the sum of the remaining columns in a row from 100.

tions of the z variable correlated with the rated dissimilarity of the x–y lines in a graph, $r(14) = .68$, $p < .01$. The dissimilarity rating was obtained by asking an independent group of individuals to rate the similarity of the unlabeled x–y lines within a graph on a scale from 1 (*very similar*) to 10 (*dissimilar*). Graphs for which the lines were rated as dissimilar, such as the non-cross-over graph shown in Figure 5b, depict multiple distinguishable x–y functions. In these cases, the z-values were more likely to be used as labels to distinguish among lines (resulting in a nominal description). For such graphs, inferring the z–y relations would require keeping track of multiple relationships which depend on values of the x variable. Graphs for which x–y lines were rated as similar, the z values were less likely to be used as nominal labels and instead, the descriptions were more likely to contain ordinal or metric information about the z variable. An example is the *intercurve spacing* graph in Figure 2b, which depicts nearly parallel lines. Deriving an ordinal representation of the z–y relationships in such cases only requires the identification of a single set of relations that is consistent throughout the range of the x variable. In summary, the types of descriptions viewers generated were consistent with the claim that the internal representation was primarily propositional, and that z–y relations have little metric detail and are less salient than the x–y relations.

Comparison task. Performance on the perspective comparison task was also consistent with the hypothesis that the internal representation of these line graphs was a propositional representation rather than a three-dimensional spatial representation. Viewers never spontaneously reported imaging a three-dimensional representation or mentally rotating their representation in order to perform the perspective comparison task. Moreover, there was no correlation between the error rates in the perspective comparison task and spatial ability. The only significant predictor of the proportion of correct comparison judgments was the Mathematics SAT, when these scores were analyzed using a logistic regression as a function of the Vandenberg Mental Rotation (MRT) score ($M = 21$, $SD = 9$) as a test of spatial skill, Verbal SAT ($M = 588$, $SD = 119$), and Mathematics SAT ($M = 715$, $SD = 83$), logit(p) = .0029 (MSAT), $p < .05$.

The accuracy in the perspective comparison task (62%) was better than chance, $t(23) = 4.85$, $p < .01$, but low enough to support the argument that the internal representation of the initial graph is incomplete and that the comparison task is difficult. Subjects' comments during the perspective comparison task suggested that they often (74% of trials) checked whether relations in their initial description were consistent with data presented in the comparison graph.[3] Not surprisingly, less detailed descriptions of the z-variable in the initial graph correlated with poorer performance in the comparison task, $\chi^2(2, N = 180) = 14.2$, $p < .01$.[4] Specifically, comparison task accuracy was 54% for trials on which viewers gave nominal descriptions of the z variable in the initial graph, compared to 65% for ordinal descriptions and 82% for metric descriptions. Incomplete descriptions of the z–y relations imply an internal representation that does not contain the information necessary to accurately recognize the alternative perspective. Of course, such incomplete descriptions were more frequent for the interaction data sets for which the z–y relations were complex because the x–y lines had dissimilar slopes.

The pattern of errors in the comparison task provides further evidence of the relative ease of retrieving x–y relations and the relative difficulty of inferring the z–y relations. We will first describe these errors for the metric data sets and then for the interaction data sets. For the graphs that depicted metric properties in the z variable, the intercurve spacing and legend perspectives (Figure 2b and 3b), most of the errors were caused by subjects incorrectly responding "same" to a different foil. The different foils did not contain the acceleration or inflection of the original z–y relations, suggesting that those metric properties were not salient in the subjects' internal representations of the data.

Comparison task performance also supported the conclusion that inflection was more salient when it was a property of the x variable (in the V-pattern perspective) than when it was a property of the z variable (in the legend perspective) because performance was much better for the former case than the latter case, 75% versus 49% correct, $t(23) = 2.5$, $p < .05$. Comparison task performance was similar for curve and intercurve spacing graphs (49% vs. 48% correct) despite the difference in describing the acceleration property in the two graphs; the similarity suggests that subjects had difficulty identifying acceleration in the intercurve spacing perspective even after encoding the property in the curve perspective, an interpretation that is supported in Experiment 2.

Comparison task performance for the interaction data sets reflected the predicted asymmetries in the internal represen-

[3] On 17% of the trials, the viewers' comments suggested that they based their judgment on finding a specific visual feature of the first graph, such as a salient data point. Only 5% of the time did they report simply guessing, and another 5% of the comments were unclassifiable.

[4] Because the observations were not independent, this test technically violated one of the assumptions of a chi-square equation.

tations of alternative perspectives of a data set. For the common value data set (Figure 4), comparison task performance was greater for the line perspective, which explicitly depicted lack of effect for one value of the z variable (74% accuracy), than the point perspective (54% accuracy), $t(23) = 2.1, p < .05$. For the cross-over interaction data set (Figures 1 and 5), viewers incorrectly considered the same data to be different on nearly half of the trials (48%), suggesting that their internal representations of the two perspectives often seemed incompatible.

In summary, both the descriptions and comparison task performance supported the hypothesis that the internal representation of a line graph primarily contains information about the x–y relations. The viewer's relative neglect of the z variable might reflect the typical uses of line graphs, in which the more important variable is frequently on the x axis. It might also indicate that the z variable can be (and often is) a nominal or ordinal variable. However, it is important to note that in this experiment, viewers knew that the z variable was metric and that because of the perspective comparison task, there was some value to encoding that metric detail. Despite this, both the descriptions and the specific errors on the perspective comparison task suggest that viewers generally neglected the metric properties of the z variable. Presumably, some properties of the z variable could be made more visually salient in the graph, and hence, be more likely to be prominent in the viewer's internal representation of the data. For example, a large graphic distance between successive z values might encourage subjects to give more prominence to its effect (although the representation might still be primarily ordinal and have little metric detail). The argument is not that the z variable is inevitably or entirely neglected in the internal representation of three-variable line graphs, but rather that the metric properties of the z–y relations are relatively neglected because they are not automatically retrieved and instead require effortful inferences to identify. In the next study, we show that even when viewers are specifically instructed to represent the metric structure of three-variable graphs, most are unable to do so.

Experiment 2

One purpose of this experiment was to further pinpoint the sources of the incomplete or incorrect representation of the metric properties of the third variable in terms of the processes of retrieving the conceptual relations associated with a particular visual pattern, encoding a visual pattern, or mentally associating the lines with the designated values (called *labeling processes*). For these types of line graphs, we postulate that a major source of error and incompleteness is a failure to retrieve the conceptual knowledge about the quantitative interpretation of an encoded pattern. For example, an individual might note that there was an unequal separation between successive lines in the intercurve spacing graph (Figure 2b), but not retrieve the fact that this pattern represents an accelerating function. By contrast, encoding processes are not expected to be a major source of error because the visual patterns are very discriminable. Nevertheless, encoding could be considered a source of error if a viewer failed to note some significant graphic feature, such as the non-uniform increase in the spacing between successive lines in the intercurve spacing graph. A third source of error is the labeling processes used to associate information in the legend with the visual patterns. Such processes are expected to be particularly problematic for the legend graph (Figure 3b) if viewers incorrectly assume or misremember that the z values increase monotonically.

To distinguish among the retrieval, encoding, and labeling processes as sources of error, the current experiment employed a drawing task. After the initial graph was described, the graph was removed and viewers drew the data in either the same or the alternative perspective. If the critical visual feature (such as the relative space between successive lines for the intercurve spacing graph) were drawn correctly in the same perspective, it would suggest that encoding was accurate and the viewer had an accurate visual feature representation. However, if the viewer made errors for the comparable alternative perspective, it would suggest that the internal representation did not support an accurate translation to that perspective. In particular, it would suggest that the z–y relations had been not been accurately retrieved or inferred from the original graph. More specific processing sources of difficulty could be indicated by the specific types of errors.

A second goal of the study was to contrast a propositional representation with a three-dimensional spatial representation of the graphs. This was done in a subsidiary experiment, in which we used explicit instructions to increase the use of three-dimensional spatial representations. We recruited a special group of participants, undergraduates who were majoring in design and architecture, and instructed them to use the graph to generate a three-dimensional representation of the underlying data set. It was thought that they might be more skillful than other college students at generating a three-dimensional spatial representation from a line graph because of their experience in interpreting orthographic projections, such as those used in architectural diagrams. If they were successful at generating a spatial representation, their descriptions and drawings should have more metric information about the z variable than those of viewers who generated a propositional representation.

Method

Participants. Undergraduate students ($N = 24$) participated in the main experiment, and another group of architecture and design students ($N = 13$) participated in the subsidiary experiment. One additional architecture student was not included in the final analysis because of his persistent failure to understand the instructions that resulted in consistent and gross errors.

Stimuli. The design and stimuli were identical to those in Experiment 1, with two exceptions. First, in place of same and different foils, viewers were presented with labeled but blank graphs that indicated they should draw the data from either the same or the alternative perspective. Second, minor changes were made to two versions of a data set to make them more similar to each other.

Procedure. The instructions and procedure for the main group of viewers were similar to those used in Experiment 1, except that a drawing task replaced the perspective comparison task. The viewer first described a graph, then it was removed and the viewer was given a graph with a set of labeled axes but no data. On half of the trials, the axes were in the same configuration as the initial graph and the viewer's task was to reproduce the original drawing. On the other half of the trials, the x and z axes were interchanged and the viewer's task was to draw the data from the alternative perspective. Within each condition, graphs were presented in random order for each viewer with the constraint that the two perspectives of a single data set occurred in different halves of the experiment.

The architecture and design students were instructed to generate a spatial representation of the three-dimensional structure. They were also given an example that visually depicted how a line graph was an orthographic projection of a three-dimensional data structure. The students were told to represent the underlying data structure spatially and to use that spatial representation to generate the alternative perspective, when that perspective was required in the drawing task.

As in Experiment 1, all subjects were given the Vandenberg Mental Rotations Test and asked about their mathematics and verbal SAT scores.

Scoring the drawings. Two scoring systems were used to analyze separately the nature of the drawings for the same and alternative conditions. The same perspective drawings were used to assess the accuracy of the original encoding of the two-dimensional pattern. The same drawings were first sorted according to whether they were qualitatively accurate in preserving the correct ordering of conditions. (This sorting was particularly important for the inflection data set because if viewers ignored the values in the legend, their drawing would be highly inaccurate). For those graphs that were qualitatively accurate, the quantitative accuracy was calculated by measuring the percentage of deviation of the depicted value from the correct value for each point, squaring that percentage, and then calculating the root mean squared deviation of each y value across viewers, and taking the average across drawings. For the alternative perspective condition, drawings were considered correct if they accurately depicted the relative magnitudes and directions of the functional relations, even if the drawn values of the points differed quantitatively from the correct answer. The large number of qualitatively incorrect drawings for the alternative perspective precluded any further analysis of the quantitative accuracy. Qualitative errors were of two kinds. The first was clearly related to the specific property designed to be investigated (such as failing to note the order of conditions in the legend graph). The second included general gross mistakes, such as drawing the data lines vertically, rather than horizontally. Typically such mistakes occurred when viewers had difficulty translating a function from one perspective into the alternative perspective.

Results and Discussion

Overall, the drawings suggested that viewers usually encoded the major x–y relations and visual features in the x–y plane, but that they did not represent the z–y relations in a way that allowed them to accurately draw the alternative perspective. Over 90% of the drawings for the same trial performance were qualitatively accurate, with the exception of some specific errors associated with legend and intercurve spacing graphs that will be discussed below (see Table 2). A quantitative analysis of the drawn points in the same trials indicated that average root mean squared deviation was only 8%. The deviation primarily reflects simplifications of minor features that were made more symmetrical, flatter or more parallel than the original; such simplifications are consistent with the findings of Tversky and Schiano (1989) on memory for graphic information. The accuracy of the same drawings is consistent with the hypothesis that the internal representation includes a set of properties about the major x–y relations, perhaps in addition to some two-dimensional visual feature information. The few consistent errors in the same drawings, described below, are related to incomplete information about the metric properties of the z variable.

Failure to encode a visual feature was identified as a non-negligible source of error on 50% of the same trials for the intercurve spacing graph when the lines on the response graph were drawn equally spaced, rather than unequally spaced to represent the acceleration (or deceleration) of the z variable. Encoding may be a much more frequent source

Table 2
Experiment 2: Unselected Subjects' Classification of Drawings (in Percent)

Data space	Same perspective, correctly depicted main property	Alternative perspective		
		Correct	Specific error	General confusion
Acceleration				
Graph: Intercurve spacing	50	27	36	36
Graph: Curve	100	13	42	45
Inflection				
Graph: Legend	43	21	79	0
Graph: V-pattern	100	25	35	39
Common value interaction				
Graph: Point	96	17	33	50
Graph: Line	95	38	21	42
Cross-over interaction				
Graph: Cross-over	92	22	57	22
Graph: Non-cross-over	96	33	33	33
M	84	24	42	33

of error when the visual feature information is not visually salient, the graphs are extremely complex, or the display triggers visual illusions (e.g., Cleveland & McGill, 1985; Poulton, 1985). However, as the general accuracy on other same perspective drawings indicated, the failure to encode a major graphic feature is not a common source of difficulty for these perceptually simple graphs.

The labeling processes were a source of errors when subjects incorrectly assumed a monotonic ordering of z values, either because they did not encode or remember how the values in the legend corresponded to the individual lines. For the legend graph, which had nonmonotonically ordered z values, the order of the z values was described incorrectly on 45% of the trials (a rate that is higher than the 28% in Experiment 1); the same-perspective drawing was subsequently drawn incorrectly 87% of the time. When the order was described correctly or not mentioned, the order of z values on the same drawing was inaccurate only 40% of the time. Thus, labeling can be a source of error, particularly when the z values are in a legend and are in a nonmonotonic order.

In contrast to the high levels of accuracy for the same-perspective drawings, accuracy for the alternative-perspective drawings was low (24%) for the eight data sets, as indicated in Table 2. In general, the errors suggest that viewers had difficulty inferring the quantitative properties of the z variable or drawing those properties as a function of the z variable in order to generate the alternative perspective. For example, after an obviously accelerating function was displayed on the x axis of the curve graph, 42% of the drawings on the alternative-perspective trials had no representation of the acceleration for the z variable but were otherwise correct. Similarly, although the x variable was obviously inflected for the V-pattern graph, only 26% of the viewers accurately translated that property into the z dimension when drawing the alternative perspective.

An analysis of the types of specific errors that subjects made in drawing alternative perspectives supports the hypothesis that the internal representation of the initial data set may consist primarily of a propositional representation of the x–y relations. An error was called *specific* if it could have resulted from not encoding a specific metric property of the original z variable or by incorrectly generalizing the x–y relations to the z–y relations (but correctly depicting the x–y relations) (see Table 2). For example, the middle panel of Figure 6 depicts a specific error for the alternative perspective when the initial graph was a cross-over graph (shown in the left-most panel and the accurate alternative perspective is shown in the right-most panel). The viewer's drawing accurately reflects two important properties of the original x–y relation (sodium consumption–aggressiveness): First, as sodium consumption increases, so does aggressiveness; second, the magnitude of the increase is greater when blood pressure (the z variable) is higher than when it is lower. However, the increasing x–y relations are incorrectly generalized to the z–y relations (blood pressure–aggressiveness). This particular specific error, which occurred on 57% of cross-over trials, is consistent with a nominal representation of the third variable in which the effect of the original z variable on the magnitude of the x–y slope is drawn correctly, but the correct z–y relationship is not drawn.

Another common type of error (33%), called *general errors* in Table 2, reflected more general difficulties in the translation to the alternative perspective. On about half of these trials, the viewers drew lines that were oriented and positioned identically to the lines in the initial graph, either because the viewers incorrectly inferred the functions were the same in the alternative perspective or as a default strategy that they used when they were unsure of the properties of the third variable. In the other half of these trials, they grossly violated graphic conventions by drawing lines of extremely different lengths, attaching multiple labels to a single line, or by drawing vertical lines, which likely reflected viewers' difficulty in translating the discrete values encoded from the z axis into the continuous values of the x

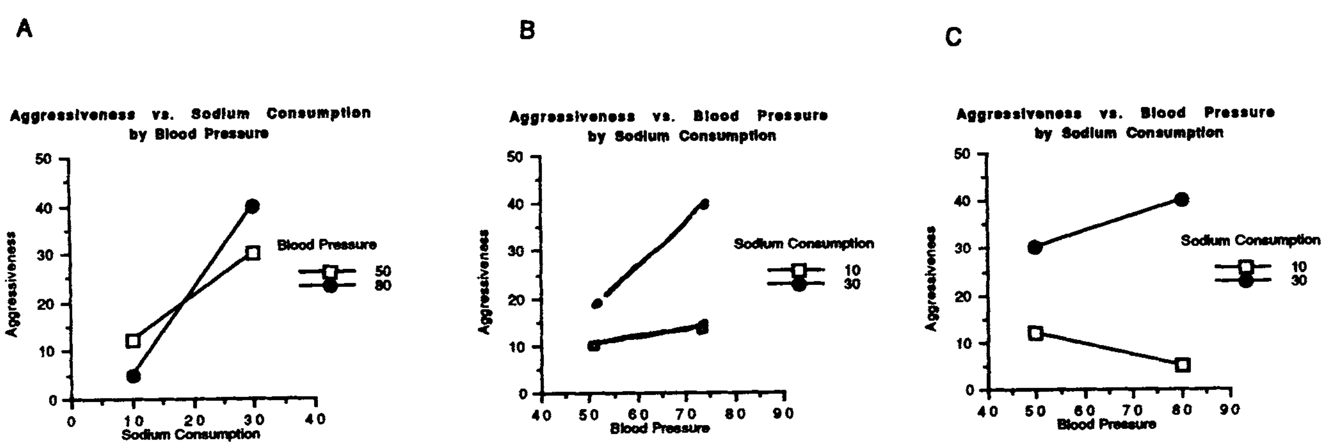

Figure 6. An example of a specific error in drawing the alternative perspective. From left to right, the figure shows the initially presented cross-over graph, the viewer's incorrect drawing, and the correct alternative-perspective graph. A typical error was to replicate the main effect of the initial x variable (age) but not capture the effect of the initial z variable.

axis. Overall, the errors on the alternative-perspective drawings indicate that viewers had difficulty translating surface features of the x–y lines in the initial graph to another perspective and that viewers may have inappropriately generalized the major x–y relations in the initial perspective.

Description task. The types of descriptions given by most viewers were similar to those given in Experiment 1. Moreover, the frequencies of nominal descriptions of the z variables for the eight graph types were similar in Experiment 1 and 2. As shown in Table 3, nominal descriptions of the z values were more frequent for graphs containing lines that were less visually similar, specifically, the pronounced interaction data sets, as was true in Experiment 1.[5]

Architecture and design students. The attempt to induce architecture and design students to construct a three-dimensional spatial representation of the data set was successful with 3 architecture students who indicated that they used a spatial representation and an explicit rotation strategy on over 50% of the trials. Overall, these 3 viewers were more accurate in drawing the alternative perspectives than the other architecture students who seemed to use a propositional representation (42% correct vs. 16%), although with so few participants this large difference was not statistically reliable, $t(11) = 2.12$, *ns.* All three frequently drew sketches, indicating they were attempting to construct a three-dimensional spatial representation while computing the alternative perspective, suggesting that some external aid might be necessary to accurately compute the alternative orthographic projection. Two of these viewers lightly drew diagonal lines connecting the x–y lines of the initial graph, indicating the depth dimension. One even drew three-dimensional sketches, an example of which is shown by Figure 7c. Figure 7d depicts the components which the viewer drew first (adding dashed lines to complete the drawing from that orientation), showing a position similar to the perspective of the initial graph (Figure 7a). Figure 7e depicts the portion of the drawing of the rotated perspective. These drawings show how the comprehension of the dimensionality of the data space could aid the comprehension of the relationships among the three variables in the graphs. One additional observation is that these three viewers sometimes reversed values on a dimension (as in Figure 7), in spite of accurately encoding the three-dimensional shape of the data space. Such confusions suggest that the three-dimensional spatial representation of data and its numerical interpretation can be dissociated. Moreover, the fact that so few viewers used this spatial strategy suggests that this type of representation is difficult to generate. Consequently, it is consistent with the hypothesis that line graphs are typically (but not inevitably) represented as an ordered set of propositions that expresses the prominent x–y functional relations.

The remaining 10 students gave descriptions and sketched drawings that were indistinguishable from those of the unselected viewers and did not sketch intermediate drawings. They may not have followed the instructions because the data spaces were relatively impoverished structures from an architectural viewpoint. In addition, the maintenance and comprehension of quantitative information may have competed with the generation and maintenance of a spatial representation. Given this result, it is not surprising that as a group, the architecture students were no more accurate in the alternative perspective condition than the unselected viewers (21% vs. 24% correct). Moreover, the failure of the instructed viewers to represent the metric information strongly supports the claim that spatial representations are a difficult and infrequent format in this graph interpretation task.

In summary, the drawing task suggests that viewers accurately encode and represent the main visual features and have a propositional representation of the x–y relations. However, this information is insufficient to allow them to retrieve or easily infer the z–y relations, resulting in numerous errors when they sketch the alternative perspective.

Table 3
Experiment 2: Unselected Subjects' Descriptions of the Third Variable of Initial Graph (in Percent)

	Classification			
Data space	Nominal	Incorrect ordinal	Ordinal	Metric
Acceleration				
Graph: Intercurve spacing	14	0	57	23
Graph: Curve	12	0	73	7
Curvilinearity				
Graph: Legend	16	45	16	16
Graph: V-pattern	36	4	49	11
Common value interaction				
Graph: Point	35	0	49	14
Graph: Line	58	0	30	9
Cross-over interaction				
Graph: Cross-over	63	0	26	11
Graph: Non-cross-over	76	0	7	13
M	38	6	38	13

Note. The proportion of trials in which the z variable was not mentioned (an average of 5% of the trials) can be calculated by subtracting the sum of the columns in a row from 100%.

Experiment 3

Experiment 3 examined possible expert–novice differences in interpreting graphs, particularly the contribution

[5] The correlations between spatial ability and drawing accuracy revealed a very small but consistent correlation, in contrast to the results of Experiment 1, $\text{logit}(p) = .06(\text{MRT}) + .03(\text{MSAT}) - .05(\text{VSAT})$, $p < .01$, for MRT (Vandenberg Mental Rotation task). This suggests that viewers sometimes coded and manipulated spatial relations, and on those trials viewers were more likely to be accurate in their drawing. However, the fact that the correlation was low and that the verbal descriptions resembled those in the preceding experiments, suggests that the strategy was infrequently used. The psychometric scores of the unselected viewers in Experiment 2 were similar to those of the viewers in Experiment 1, MRT = 21 (SD = 9); MSAT = 684 (SD = 54) and VSAT = 575 (SD = 71). The architecture and design students had similar scores, except for slightly lower Mathematics SAT scores: MRT = 20 (SD = 7); MSAT = 598 (SD = 101); VSAT = 576 (SD = 84).

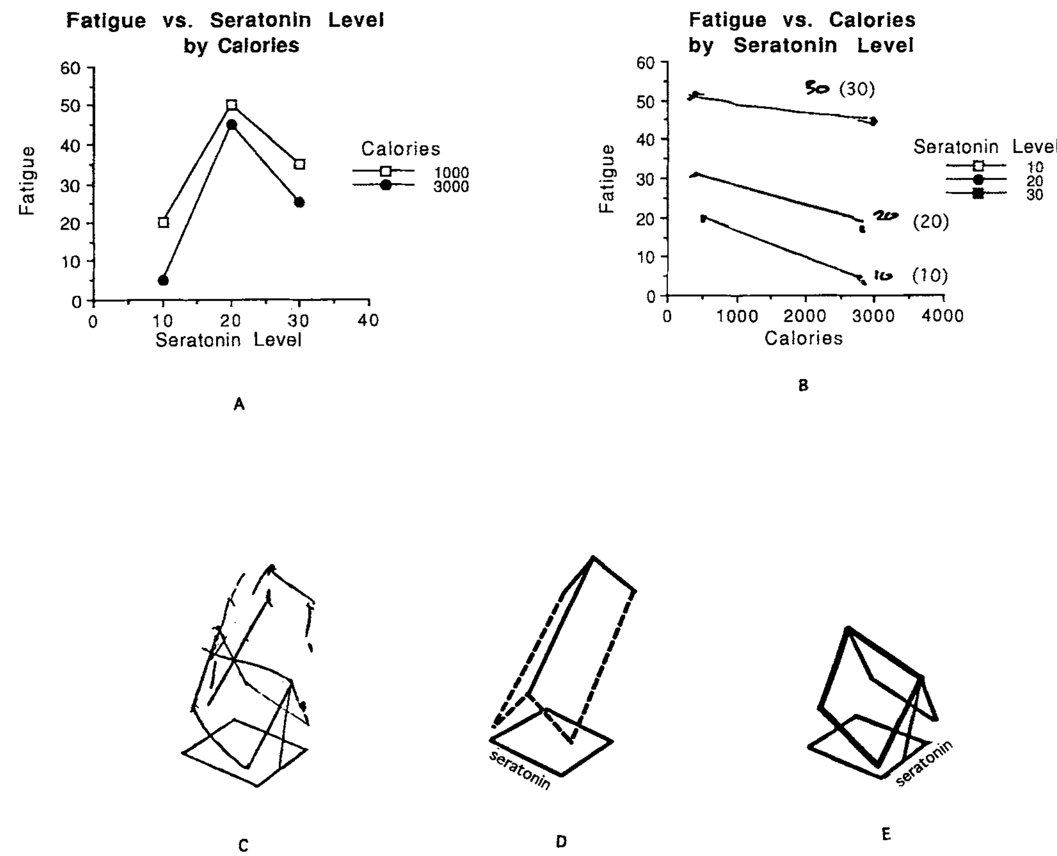

Figure 7. An example of a drawing by an architecture student, which indicated use of a spatial strategy in drawing the alternative perspective of the graph. This individual was able to draw the legend graph correctly except for mislabeling the lines. The initially described graph and the viewer's drawing of the alternative perspective are on the first row. The second row shows the viewer's three-dimensional drawing on the left. The middle drawing is an extension (dashed lines) of the viewer's original drawing showing the initial perspective of the data. The viewer then drew the data from the alternative perspective three-dimensionally, and this part of his drawing is emphasized in the right-hand drawing.

of expertise in data analysis and graphic interpretation. Psychology graduate students with extensive experience with data interpretation were given the same task as the undergraduate students. If the difficulty in the comprehension of the third variable is based on general limitations in the encoding, labeling, and retrieval processes found in Experiments 1 and 2, even the experienced graph reader may not be able to encode and remember the multiple relationships represented in these line graphs. Alternatively, experts' greater familiarity with data interpretation may mitigate the limitations that beset the novice, as has been found in content domains such as physics problem solving (Chi, Glaser, & Farr, 1988). In that case, experts may encode more metric information about the third variable.

A second focus of this study was the effect of the viewer's goals on the comprehension of line graphs. The type of information that a person extracts from a graph usually depends on the reason for looking at the graph (Carswell & Wickens, 1987). Therefore, we wanted a task that would be commonly associated with the interpretation of data, but which might also encourage attention to the entire pattern. Consequently, instead of asking viewers to describe the data (as in Experiment 1), we asked them to generate an explanation of the data. The explanation task could invite more attention to the neglected metric properties of the z variable, as the viewer attempts to account for the three variables. If so, viewers may increase the frequency of metric descriptions of the z variable and therefore have greater accuracy in the perspective comparison task. An alternative hypothesis concerning the task is that a viewer develops explanations from internal descriptions of the graph—the same type of incomplete (and sometimes inaccurate) internal representations that were elicited in Experiments 1 and 2. In that case, the explanations would also be incomplete and sometimes inaccurate. Moreover, accuracy in the perspective comparison task would not be better than in the description task condition of Experiment 1.

Method

Participants. Carnegie Mellon University undergraduate students ($N = 24$) participated for course credit. The experts were an additional 7 psychology graduate students who had taken an average of 5.6 statistics courses, had conducted empirical research of their own, and reported regularly reading empirical journals that contained graphs of data. By a variety of indexes, the graduate students had considerable experience in statistics and the interpretation of graphic information.

Stimuli. The graphs were fundamentally similar to those in Experiment 1. One change is that in some cases, more common labels were substituted for the variables used in Experiment 1, to ensure that the three variables would be familiar to the viewers and to make the possible causal relations among them more salient. Groups of three names were composed on the basis of familiar phenomena (e.g., heart attacks, sodium intake, cholesterol intake; car accidents, alcohol consumption, traffic density). The label sets were randomly assigned to graphs. Each graph appeared in two versions; each version had the same y variable label, but the x and z labels were balanced across versions of the graph. For half of the viewers, one of the independent variables was plotted on the x axis; for the other half, it was plotted on the z axis.

Procedure. The procedure was the same as in Experiment 1, except that the viewers were asked to generate a possible mechanism or explanation to account for the data. The instructions provided examples of mechanistic accounts of two-dimensional graphs. Also, each viewer was given only 8 graphs to describe, rather than the 16 that were used in Experiment 1, because generating an explanation took more time than describing a graph.

Results and Discussion

Expertise. The performance on the perspective comparison task indicated no substantial benefit was conferred by the graduate students' extensive experience in data analysis and graph interpretation. The 7 psychology graduate students averaged 53% accuracy, which is similar to the 50% accuracy of the unselected undergraduate students, $t(30) = .39$, *ns.* Moreover, the types of explanations and descriptions generated by the graduate students were very similar to those by the less experienced undergraduates, suggesting that both groups are vulnerable to the same processing limitations. These results do not preclude the possibility that viewers can be taught to comprehend graphs more accurately, but they suggest that such skill does not automatically emerge as a result of the experiences of these students. These experienced graph viewers have not learned to retrieve automatically the z–y relations from the graphic features in line graphs; thus, like the inexperienced undergraduates, they would need to infer these functions.

Explanation task. In general, the types of explanations given by the undergraduate students suggested that the interpretations built on the incomplete and sometimes incorrect internal representations like those generated by the viewers in Experiment 1. In fact, there were slightly fewer metric descriptions of the z variable (15%) than in Experiment 1 (21%) for these graphs, as shown in Table 4, although the difference is not statistically significant, $t(46) = 1.22$. Viewers provided explanations on 66% of the trials, and descriptions on other trials. The explanations were frequently incomplete, and they sometimes contained inaccurate inferences about the data. For example, on 23% of the trials, viewers reversed the direction of the causal or correlational relations by describing how variation in the x variable led to the observed z values. Overall, the descriptions and explanations focused on the x–y relationship and contained little metric information about the z variable.

Consistent with the hypothesis that explanations are built on incomplete representations of the data set, the accuracy in the perspective comparison task was only 50%, a rate that is even lower than the accuracy for the comparable graphs in Experiment 1 (61%), $t(46) = 2.1, p < .05$ (see Table 4).[6] Clearly, the task of generating a "deeper" interpretation of the data did not improve the viewer's ability to judge the data's similarity to another perspective, even though viewers spent much more time on each graph in the initial phase in the current study in comparison with the time spent in Experiment 1. The demands imposed by the explanation task may have contributed to the less detailed descriptions and the poorer comparison task performance in this experiment. Viewers tried to generate a causal path between the x, z, and y variables, and this may have distracted them from attending to the precise metric properties of the variables.

The results of this study further support our hypothesis that the difficulty in comprehending graphs may arise from fundamental processing limitations that differentially affect the z and x variables. These limitations affect even the most experienced graph users. A second important conclusion from this study is that the process of generating an explanation did not improve the accuracy or quality of the internal representation of the data. Moreover, this study suggests that the difficulty of inferring the metric properties of the z variable also influences the types of explanations that are generated for a data set. Explanations may be developed on the basis of incomplete and biased internal descriptions. Hence, different perspectives of a data set that give rise to different descriptions may also give rise to different explanations.

General Discussion

Interpreting or describing a line graph to oneself is a process of representing the major x–y functional relations; very rarely do viewers construct an integrated representation of the metric properties of the three variables. The incompleteness and errors in the representation of simple line graphs primarily result from limitations in conceptual processes rather than inaccuracy in perceptual processes. The drawing task demonstrated that for the most part, viewers had an accurate representation of the visual features of the original graph. Moreover, the x–y lines in the present studies did evoke the associated mathematical functions, as indicated by the viewers' frequent use of specific mathematical terms, such as *accelerating, linear,* or *exponential.*

[6] The viewers' psychometric scores in Experiment 3 were similar to those of viewers in Experiments 1 and 2, suggesting that all the groups were from the same population (MRT = 21; $SD = 8$, MSAT = 702; $SD = 59$, and VSAT = 551; $SD = 80$).

Table 4
Experiment 3: Descriptions of the Third Variable of Initial Graph (in Percent) and Performance on the Perspective Comparison Task

Data space	Classification				Comparison task (% correct)	
	Nominal	Incorrect ordinal	Ordinal	Metric	Experiment 2	Experiment 1
Inflection						
Graph: Intercurve spacing	0	0	62	38	62	49
Graph: Curve	0	0	94	2	44	48
Cross-over interaction						
Graph: Cross-over	36	29	18	11	51	69
Graph: Non-cross-over	42	0	47	9	47	70
M	19	7	55	15	50	61

Note. The proportion of trials in which the z variable was not mentioned (an average of 4% of the trials) can be calculated by subtracting the sum of the columns in a row from 100%.

By contrast, they did not use such terms for the comparable z–y relations, suggesting that graph viewers lack the knowledge of the associated quantitative functions. If the z–y relations are to be inferred, it is by an effortful comparison of the y values of successive lines at particular values of x. For example, to infer that the z–y relation is accelerating in the graph in Figure 2b, the viewer would compare the distances of the three lines which depict different values of the z variable (nitrogen) at different values of the x variable (rainfall). The difficulty of inferring z–y relations leads to an asymmetry in the internal representation of line graphs, in which the z variable is more likely to have an ordinal or even nominal internal representation than is the x variable.

One reason why viewers may be unfamiliar with the metric properties of the third variable may be related to the typical uses of line graphs, in which the more important metric information is typically plotted as the x variable and the less important variable is the parameter on the curves. Moreover, the third variable need not be continuous and often is only nominal or ordinal. Hence, correlations between importance and depiction may cause viewers to achieve some cognitive economy by primarily attending to the x–y features. However, as we argued in Experiment 1, viewers were aware of the continuous nature of the z variable through the instructions, labels, and the task itself. This suggests that viewers may not have developed schemas for interpreting metric information in the third variable rather than that they simply did not choose to use such schemas. Consistent with this claim, even explicit instructions (in Experiment 2) were generally unable to elicit the representation of the metric properties of the z variable.

The limitation and biases revealed in these studies may not be entirely attributable to familiarity, but may also reflect a limitation on the number and type of relations that can be comprehended and simultaneously retained. The hypothesis that only a limited amount of information is represented is supported by the finding that there was a negative correlation between the metric detail about the z variable in the descriptions generated in Experiment 1 and the rated complexity of the graphs. Thus, viewers gave less information about the z variable when there was more to say about the x–y relations. Also compatible with this capacity view is the fact that the descriptions generated by graduate students (who had a great deal of experience in interpreting graphic data) were as limited as the descriptions generated by undergraduates. Certainly, it seems uncontroversial that there are limitations in the amount of information that individuals can extract from a graph. If so, what may be most surprising about the current research is that the limitation is evident even with these relatively simple three-variable data sets.

The use of a three-dimensional spatial representation, which architecture students were instructed to use in Experiment 2, provides one potential way to circumvent the capacity limitation. A spatial representation may permit multiple relations and interactions to be coded because the relations are distinguished by their differing locations and directions in three-dimensional space. However, using a three-dimensional spatial format was apparently effortful and only successfully used by the 3 viewers who generated external diagrams to support the translation between perspectives. But in a more propositional–visual feature representation, viewers may only code and maintain information about a subset of functional relations and graphic features, and these tend to be the x–y relations.

Verbal Descriptions

Verbally describing and interpreting graphs are tasks that are representative of the ways that graphs are used in many reading and comprehension situations. Indeed, it is difficult to imagine a graph comprehension task that relies solely on pattern perception. Even though some type of verbal description may be intrinsic to most situations in which graphs are used, this feature of the task could influence the content of the representation. For example, the described relations are presumably more easily retrieved, but their description could also further contribute to their accessibility and influence the comparison and reproduction tasks. Such an influence would explain the fairly consistent correlation between the type of description subjects gave of the third variable and their accuracy in subsequent perspective comparison task. Support for this possibility comes from research that has shown that viewers who verbally describe a face later show exaggerated reliance on the features in their description, and their memory performance may be less accurate than that of viewers who do not verbally describe the face and presumably are less biased toward the verbalizable features (Schooler & Engstler-Schooler, 1990). Such an influence also would be consistent with earlier research on memory for visual information that indicated viewers' memory for an ambiguous line drawing was influenced by its label (Carmichael, Hogan, & Walter, 1932). The analogy to such studies with nongraphic stimuli may be illuminating in so far as line graphs also have many visual features and their interpretation requires selecting certain features that are important; as we argued earlier, these features typically are associated with the x–y relations. In summary, the description of the graph's meaning probably does bias the information maintenance. However, some description process, such as reading a description or generating one, is an integral part of most tasks involving the interpretation of line graphs.

Alternative Formats for Graphic Data

The current studies have implications for those who use line graphs to explore the interpretation of data. Even with a simple three-variable data set, the placement of the variables in the display influences its interpretation. Alternative formats that explicitly represent all the important functional relations might encourage a more complete and accurate representation of data. One alternative format is to present three-variable data from both perspectives; another alternative is to use linear perspective to encourage the viewer to encode a spatial representation of data which includes metric relations about the third variable (as in Figure 1c; Wickens, 1992). Perhaps it is obvious that neither solution has been adopted in psychology and cognitive science journals. An examination of the graphs in four major journals for the last decade reveals that none presented higher dimensional data in alternative perspectives or with a linear perspective, although a linear perspective format was sometimes used to represent the predictions of hypothetical data on the basis of theoretical models.[7]

Although a linear perspective format may seem to be a plausible way to ameliorate interpretation biases revealed in the current studies, several difficulties may limit its usefulness. These difficulties arise from three sources. First, the use of linear perspective may facilitate understanding the general three-dimensionality of the data, but it degrades the specific metric information (Wickens, 1992). For example, parallel lines in the third dimension are not necessarily parallel. Second, because such formats are rare, viewers are not familiar with the interpretation various features such as the "tilted plane" depicted in Figure 1c. Third, the format can make it difficult to perceive or distinguish some lines, making it hard to visually encode complex relationships, such as interactions. These difficulties may be exacerbated when the graphs depict sparse data sets. Some types of denser data sets may be more amenable to a three-dimensional format because they have fewer ambiguities, in the same way that a landscape is less ambiguously rendered in perspective than is a sparse stimulus, like a Necker cube.

Another possible format for representing three-variable data is the use of animation to rotate among the alternative perspectives, as is permitted in animation programs such as Macspin (Donoho, Donoho, & Gasko, 1988). Although not applicable to conventional print media, such animation might help the viewer form a three-dimensional spatial representation. However, it is not currently known whether such animation would eliminate the conceptual constraints that dominate the interpretation that viewers give to more conventional formats.

The conceptual processes that constrain the internal representation of three-variable data undoubtedly also constrain the representation of higher dimensional data (four or more variables). The current studies suggest what biases might arise in interpreting a four-variable data set that is displayed through several two-dimensional, three-variable line graphs, where each graph represents one value of a fourth variable. Specifically, the representation of the fourth variable would be even more likely to be nominal than the representation of the third variable. Relations among values of the fourth variable must be inferred by comparing the set of graphs; these relations are not visually salient and would likely require more cognitive effort to encode than the relations in the third variable. It is also likely that the relations in the fourth variable will be considered as they relate to the x–y plane but not as they relate to the third variable, the parameter on the curves. Finally, the sheer burden of maintaining the representation of one graph may limit the amount of additional information that is encoded and maintained from another graph or graphs. As limited as the representation is for three dimensional data sets, it is likely to be proportionally less detailed for higher dimensional information.

The tasks of describing and explaining graphs are representative of real-world graph comprehension tasks and are useful to understanding the comprehension and representation of graphic displays. Earlier studies and theories of graphic presentation have emphasized the tremendous potential of harnessing the abilities of visual pattern recognition processes in the detection of empirical relations. Undoubtedly, the perceptual features of graphs can either enhance or detract from the detection of such relations, and an analysis of these features can benefit both theories of perception as well as graphic design (Cleveland & McGill, 1985). The studies presented here are part of a recent trend in display research to use comprehension and description tasks that begin to examine the conceptual, in addition to the

[7] The graphs from 1980 through 1991 in four psychology journals were examined: *Journal of Experimental Psychology: General, Journal of Experimental Social Psychology, Journal of Experimental Psychology: Learning, Memory and Cognition* (called *Journal of Experimental Psychology: Human Learning and Memory* through 1981), and *Psychological Review.*

perceptual, processes in graphic displays (Carswell, Emery, & Lonon, 1993; Howell, 1993). Viewers fail to make the inferences needed to represent the metric properties of the z–y relations. Consequently, even relatively simple line graphs, such as those used in the current studies, are incompletely (and sometimes incorrectly) internally represented.

References

Bertin, J. (1983). *Semiology of graphics: Diagrams networks maps* (W. Berg, Trans.). Madison: University of Wisconsin Press.

Carmichael, L., Hogan, H. P., & Walter, A. A. (1932). An experimental study of the effect of language on the reproduction of visually perceived form. *Journal of Experimental Psychology, 15,* 73–86.

Carswell, C. M., Emery, C., & Lonon, A. M. (1993). Stimulus complexity and information integration in the spontaneous interpretation of line graphs, *Applied Cognitive Psychology, 7,* 341–357.

Carswell, C. M., & Wickens, C. D. (1987). Information integration and the object display: An interaction of task demands and display superiority, *Ergonomics, 30,* 511–527.

Chi, M., Glaser, R., & Farr, M. J. (1988). *The nature of expertise.* Hillsdale, NJ: Erlbaum.

Cleveland, W. S., & McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. *Science, 229,* 828–833.

Donoho, A. W., Donoho, D. L., & Gasko, M. (1988). MacSpin: Dynamic graphics on a desktop computer. In W. S. Cleveland & M. E. McGill (Eds.), *Dynamic graphics for statistics.* (pp. 331–351). Monterey, CA: Wadsworth.

Fienberg, S. E. (1979). Graphical methods in statistics. *The American Statistician, 33,* 165–178.

Gleick, J. (1987). *Chaos: Making a new science.* New York: Penguin Books.

Howell, W. C. (1993). Engineering psychology in a changing world, *Annual Review of Psychology,* 231–263.

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial abilities. *Psychological Review, 92,* 137–172.

Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology, 3,* 185–226.

Lewandowsky, S., & Spence, I. (1989). The perception of statistical graphs. *Sociological Methods and Research, 18,* 200–242.

Lohse, J. (1991). A cognitive model for the perception and understanding of graphs. In S. P. Robertson, G. M. Olson, & J. S. Olson, (Eds.), *Reaching through technology: CHI '91 Conference proceedings,* New Orleans.

Pinker, S. (1990). A theory of graph comprehension. In R. Freedle (Ed.), *Artificial intelligence and the future of testing* (pp. 73–126). Hillsdale, NJ: Erlbaum.

Poulton, E. C. (1985). Geometric illusions in reading graphs. *Perception and Psychophysics, 37,* 543–548.

Schmid, C. F. (1983). *Statistical graphics: Design principles and practices.* New York: Wiley.

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22,* 36–71.

Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations.* Cambridge, MA: MIT Press.

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

Simkin, D., & Hastie, R. (1987). An information-processing analysis of graph perception. *Journal of the American Statistical Association, 82,* 454–465.

Tversky, B., & Schiano, D. J. (1989). Perceptual and conceptual factors in distortions in memory for graphs and maps. *Journal of Experimental Psychology: General, 118,* 387–398.

Wickens, C. D. (1992). *Engineering psychology and human performance.* New York: Harper Collins.

Vandenberg, S. G. (1971). *Mental rotations test.* Boulder: University of Colorado.

Appendix

Descriptions of the Third Variable of the Initial Graph (in Percent) and Performance on the Perspective Comparison Task

Data space	Classification: Nominal	Incorrect ordinal	Ordinal	Metric	Comparison task (% correct)
Acceleration					
Graph: Intercurve spacing	11	0	64	23	49
Graph: Curve	13	0	67	19	48
Inflection					
Graph: Legend	15	28	51	2	49
Graph: V-pattern	27	0	63	3	75
Common value interaction					
Graph: Point	42	0	26	28	54
Graph: Line	49	0	36	9	74
Cross-over interaction					
Graph: Cross-over	63	0	11	26	69
Graph: Non-cross-over	65	0	15	19	70
M	35	3	41	16	62

Note. The proportion of trials in which the z variable was not mentioned (an average of 3% of the trials) can be calculated by subtracting the sum of the columns in a row from 100%. Note that it is not possible to compare comparison task performance across data space types because of potentially nonequivalent false foils.

Received March 14, 1994
Revision received August 18, 1994
Accepted August 25, 1994 ■