The Shape Parameter of a Two-Variable Graph

WILLIAM S. CLEVELAND, MARYLYN E. McGILL, and ROBERT McGILL*

The shape parameter of a two-variable graph is the ratio of the horizontal and vertical distances spanned by the data. For at least 70 years this parameter has received much attention in writings on data display, because it is a critical factor on two-variable graphs that show how one variable depends on the other. But despite the attention, there has been little systematic study. In this article the shape parameter and its effect on the visual decoding of slope information are studied through historical, empirical, theoretical, and experimental investigations. These investigations lead to a method for choosing the shape that maximizes the accuracy of slope judgments.

KEY WORDS: Statistical graphics; Aspect ratio; Graphical perception; Visual perception.

1. INTRODUCTION

Figure 1 gives data on the amount of solar radiation penetrating sea water at different depths: the filled circles are actual measurements and the open circles are estimates (Littler, Littler, Blair, and Norris 1985). The dashed rectangle, which is the *data rectangle*, shows the maximum and minimum values of the data along both the vertical scale and the horizontal scale. Suppose the height of the data rectangle of a graph is h centimeters (cm) and the width is w cm. The *shape parameter*, or *shape*, of the graph is h/w. In Figure 1 the shape is .85.

Shape is a critical factor for two-variable graphs that show the dependence of y on x. Figures 2 and 3 show the Canadian lynx data (Elton and Nicholson 1942), a time series of substantial historical interest because of the many analyses it has inspired (Campbell and Walker 1977). In Figure 2 the shape parameter is 1, and in Figure 3 the shape is .074. In Figure 2 it is impossible to see a critical property of the data that can be seen in Figure 3—the number of lynx trappings rises more slowly than it declines. As we shall explain, this phenomenon of graphical perception—the better perception of the lynx rise and fall in Figure 3—is a result of the effect that the change in shape has on our judgments of slopes.

It is hard to find a statistical-graphics topic more universally discussed than shape. But despite the importance of this parameter and the ubiquity of comments on it, there has been almost no systematic study. In this article we study shape in several ways. Section 2 defines objects critical to the study. In Section 3, comments and recommendations from a sample of twentieth-century writings on graphical data display are reviewed. In Section 4, measurements of the shapes of 481 graphs are analyzed. In Section 5, observations are made about our processing of slope information on a graph; the discussion is critical to the subsequent study of shape and slope, because it determines in a fundamental way how the problem is approached. Section 6 contains theory; we hypothesize that the accuracy of slope judgments depends on what is called *orientation resolution*. The dependence of orientation resolution on other quantities is then investigated. Section 7 describes an experiment that probes the hypothesis of Section 6. In Section 8, the results of the theory and experimentation are used to develop an algorithm for choosing the value of the shape parameter that maximizes both the resolution and accuracy of slope judgments. Section 9 concludes the article with a discussion of the general applicability of our methods, including the significance (for the analysis of the lynx data) of the phenomenon observed in Figure 3.

2. DEFINITIONS

Suppose we have a two-variable graph showing how y depends on x. In such a case, the decoding of quantitative information encoded by the slopes of line segments is a fundamental visual task that we perform. The reason, of course, is that the slopes encode the rate of change of y as a function of x, and decoding the rate of change is important for understanding the dependence of y on x.

2.1 Actual Line Segments and Virtual Line Segments

The line segments on a graph that encode slope information might be *actual* line segments drawn on the graph. For example, the lynx numbers in Figures 2 and 3 are graphed by connecting successive yearly values by line segments. We visually decode the slopes of these segments to infer the local rate of change of the lynx numbers through time. The line segments can also be superimposed on the graph by our visual system. For example, in Figure 1 we can visually superimpose segments connecting successive points on the graph to judge the local rate of change of log radiation as a function of depth. We will follow Marr (1982) and Stevens (1978) and refer to these superimposed segments as *virtual* line segments.

2.2 Coordinate Systems

Consider a point (x, y) inside the data rectangle of a graph. Now, x and y can have the units shown on the two

* William S. Cleveland and Robert McGill are statisticians, AT&T Bell Laboratories, Murray Hill, NJ 07974. Marylyn E. McGill is President, MEM Research, Inc., Murray Hill, NJ 07974. John Chambers, Colin Mallows, and Daryl Pregibon made helpful comments on an earlier draft of this article. We are indebted to the referees, whose comments led to a substantial improvement in the exposition. Joseph Follettie convinced us, from his comments on our earlier experiments in graphical perception, that showing stimuli for short time intervals is an important noise-reduction technique.

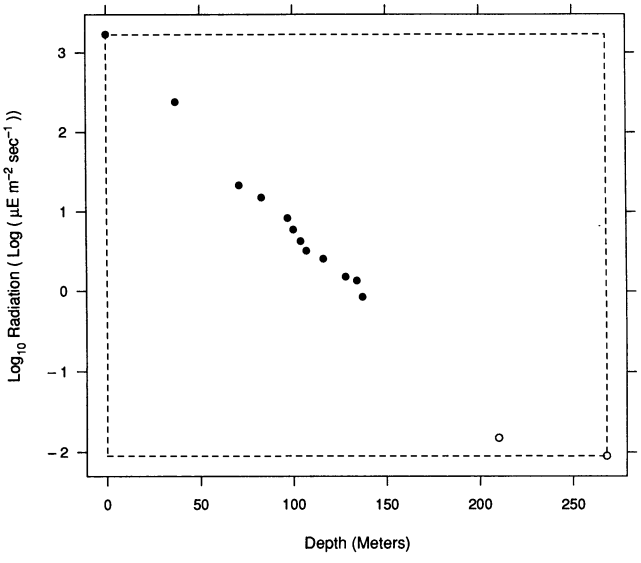

Figure 1. Shape Parameter. The data rectangle, shown by the dashed lines, is determined by the maxima and minima of the data along the two scales. The shape parameter is the height of the data rectangle divided by its width.

scale lines of the graph; this is the *scale-line coordinate system.* Or, x can be the distance in cm from the left side of the data rectangle and y can be the distance from the bottom. This is the *physical coordinate system.* Unless stated otherwise, notation will refer to the physical coordinate system. Note that the shape parameter is the slope in the physical coordinate system of a line segment connecting the lower left and upper right corners of the data rectangle.

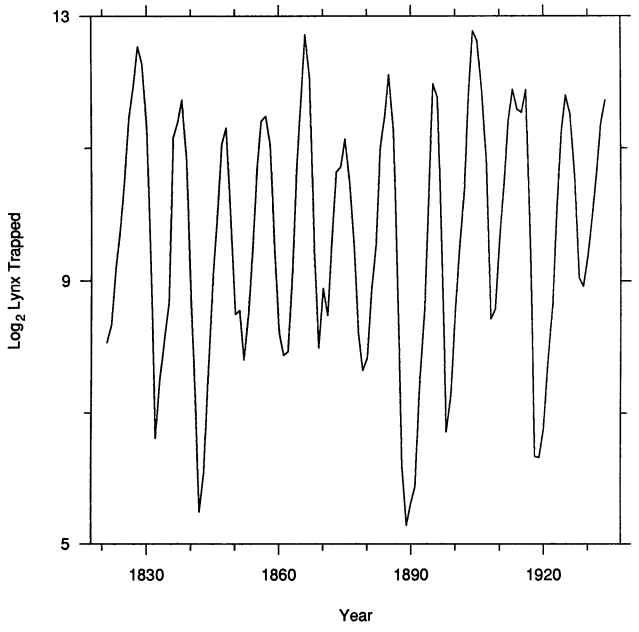

Figure 2. The Effect of Shape. The data are the Canadian lynx trappings from 1821 to 1934. The shape parameter is 1. The orientations of the line segments connecting successive data points are too close to 90° and −90° to allow us to see an important property of the data.

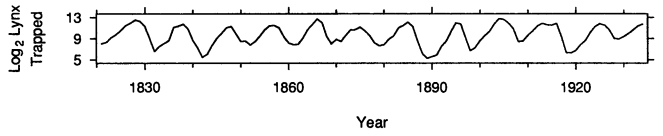

Figure 3. The Effect of Shape. The shape parameter of the graph is .074. The orientations of the line segments are in a range that allows better visual decoding of the slopes. We can now see what we could not see in Figure 2—the numbers tend to rise more slowly than they fall.

2.3 The Effect of the Shape Parameter on Slope and Orientation

Let s_i $(i = 1, \ldots, n)$ be a collection of physical slopes of line segments on a graph with shape parameter 1. The slopes are γs_i when the shape parameter is γ and the *orientations* of the segments are $\arctan(\gamma s_i)$, where the range of arctan is from $-90°$ to $90°$.

One shape parameter can lead to orientations of line segments that make visually decoding slope information easy, and another shape can result in orientations that make it very difficult. We have already seen this in Figures 2 and 3; unlike in Figure 3, however, the value of the shape parameter in Figure 2 puts the orientation of local segments in a region that greatly reduces the accuracy of local slope judgments and thus makes it difficult to detect the slower rise than fall of the lynx numbers. This article aims to determine how to choose the shape parameter of a graph to put the orientations in a region that allows effective slope judgments.

3. HISTORICAL STUDY

Major twentieth-century writings on graphical data display were sampled to see if previous studies on shape—either theoretical or experimental—had been undertaken, and to see if there was a consensus about how to choose the shape parameter. No studies were found in the survey. Costigan-Eaves (1984) carried out an extensive general survey of writings on data display, and also failed to uncover shape studies. Researchers in visual perception (Baird and Noma 1978; Julesz 1981) and computer vision (Marr 1982) have investigated related topics, but not shape directly.

Although there have been no systematic studies, there has been no lack of opinion; comments on shape go back decades and are widespread. There have been many different recommendations on how to choose the shape parameter. One category is that the ratio of the lengths of the scale lines be some fixed number, independent of the data, but sometimes dependent on the medium in which the graph is drawn (American National Standards Institute 1979a; American Standards Association 1938; Karsten 1923; Lefferts 1981; Schmid and Schmid 1979). The fixed ratio varies from one source to the next and takes on the value $\frac{2}{3}$, $1/\sqrt{2}$, or $\frac{3}{4}$. For example, Schmid and Schmid (1979) recommended, "Root-two dimensions, representing a ratio of 1 (short side) to 1.414 (long side), are particularly appropriate for statistical charts . . ." (p. 6).

In another category of recommendations, the authors suggest a particular shape parameter but argue that it should be modified to accommodate the data (American National Standards Institute 1979b; Lutz 1949; Tufte 1983; U.S. Department of the Army 1966). For example, the U.S. Department of the Army (1966) recommended the following: "The 'normal' grid proportion of the average time series chart is about 2 high by 3 wide. The regular exceptions to this 2 to 3 grid proportion are: (a) a long time series with a great many plottings, which require a wider than normal chart, (b) a long series of bars, which require a taller than normal chart" (p. 78).

In another category, authors state that general rules are not possible and that the shape should be determined solely by the data, on a case-by-case basis (Brinton 1914; MacGregor 1979; Meyers 1970). For example, Meyers (1970) wrote: "Many graphs . . . are drawn in the approximate proportions of $1:\sqrt{2}$. . . . However, the informational requirements of the graph should determine the proportions of the data rectangle, rather than a somewhat arbitrary ratio" (p. 40).

In the final category, authors state explicitly that it is the orientations of line segments that are affected by shape (see Sec. 2.3). There is, however, no unanimity about the optimum orientation. Von Huhn (1931) stated: "The writer believes that this angle probably is located somewhere between 30 and 45" (p. 320). Weld (1947) gave a range of 35°–45°, Hall (1958) gives a range of 30°–60°, and Bertin (1967, 1983) stated (with high accuracy but no evidence) that it is 70°.

Clearly, opinions on choosing the shape parameter vary widely and are contradictory. This is not surprising, since with no facts arising from either convincing theoretical arguments or careful experiments, there has been no scientific body of material around which opinions could coalesce. This problem is pervasive in the graph-construction area of statistical graphics (Cleveland 1985).

4. SHAPE PARAMETERS IN PRACTICE

The shapes of the 481 two-variable graph panels in Volume 207 of *Science* were measured. The graphs were divided into three categories: (a) *uncontrolled* (34 graphs)—neither variable is controlled, that is, both are just observed; (b) *controlled* (226 graphs)—one variable has values that are specified, for example, in an experiment, and the other variable is observed for the specified values of the first variable; (c) *time series* (221 graphs)—a variable is graphed against time. Since the time variable for the time series category is essentially a controlled variable, (c) is actually a subset of (b).

Figure 4 shows notched box plots (McGill, Tukey, and Larsen 1978) of the shape parameters of (a)–(c). (Since no pairs of notches overlap, each pair of medians is significantly different at the .05 level.) To the right of the graph, a rectangle is drawn by each of the four tick marks; the rectangle has a shape corresponding to the value at the tick mark. The median shape for the uncontrolled category is 1.00, corresponding to a square data rectangle; thus 50% of the graphs in this category are wider than tall. In the controlled category the median shape is .72, and 69% of the graphs in this category have data rectangles that are wider than tall. The time series graphs have data rectangles that are even more oblong than those in the controlled category; the median shape is .53, and 83% of the graphs are wider than tall.

Figure 4 shows several things. First, none of the fixed-shape rules from Section 3 ($\frac{2}{3}$, $1/\sqrt{2}$, $\frac{3}{4}$) is being universally followed. Second, in practice shape parameters vary enormously; for example, in the time series category the shapes range from .03 to 9.24. Finally, it is interesting that the location of the distribution of shape is about 1 (a square graph) for the uncontrolled category, that the distribution for the controlled category is shifted toward lower values, and that the distribution for the time series category is shifted even lower. (See Sec. 8.2 for further discussion.)

5. VISUAL PROCESSING AND ANALYTIC TASKS

Before the shape problem can be attacked, we must have as precise a delineation as possible of what needs to be studied. In this section, we focus on visual processing and analytic tasks. The discussion is fundamental in that it determines the directions taken in the theoretical and experimental work of Sections 6 and 7.

5.1 Visual Information Processing

Several forms of visual information processing are used to extract quantitative information from graphical displays. There is a highly perceptual processing in which we (among other things) rapidly judge magnitudes associated with geometric aspects such as position, length, and area. This processing is almost unconscious; in fact, it falls in a

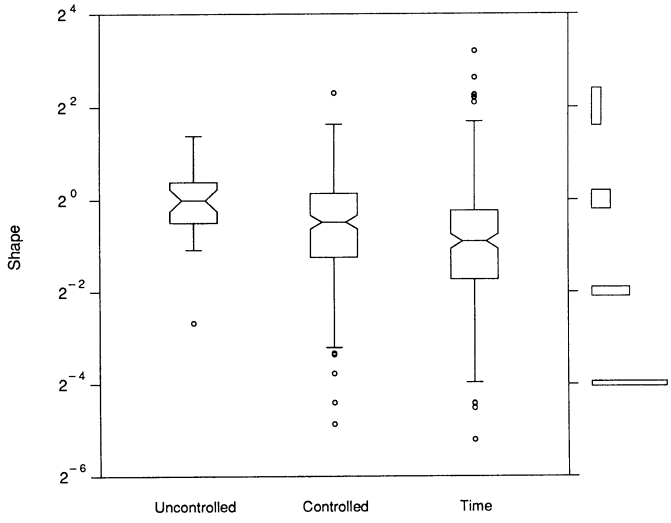

Figure 4. Shape Distributions of 481 Graphs. Notched box plots show the distributions of the shape parameters of three categories of graphs from Volume 207 of "Science." The categories are time (graphs of time series), controlled (one variable has values that are selected rather than observed), and uncontrolled (both variables are observed). The rectangles to the right of the graph have shapes corresponding to the values at the tick marks.

category of vision that Hermann Helmholtz called "unconscious inference" (Rock 1984). There is also scale reading, beginning with the perceptual process of scanning horizontally or vertically to visually project a plotting location onto a scale line and concluding with the more highly cognitive process of noting the value of a tick-mark label close to the projection; this gives us a value for a plotted point in the units of the data. The quantitative information decoded by these various forms of visual processing is combined with our knowledge of the data and the mechanism generating them to draw conclusions. For example, from Figure 1 we conclude that from sea level down to 150 meters, solar radiation is attenuated, to a good approximation, by a constant factor of about 1 order of magnitude per 40 meters. We do not mean to imply that there is an orderly progression of perception, scale reading, and contemplation; rather, we jump back and forth among them in a highly complex way that is not yet understood.

The perceptual processing involves eye movement and attentive search, and thus consists of more than what Julesz (1981) called "pre-attentive vision," which can occur in as few as 5 milliseconds. Nevertheless, graphical perceptual processing, such as identifying two line segments and judging their relative physical slopes, can be carried out in a second or even less time.

5.2 Analytic Tasks

Follettie (1986) has isolated three analytic tasks that we carry out in looking at graphical displays—discrimination, measurement, and comparative estimation. Discrimination involves a decision about whether two magnitudes are equal or not. For example, does the slope of one line segment appear to be the same as that of another? Measurement concerns the extraction of absolute magnitudes, such as determining (in the units of the data) the maximum value of a graphed function. Comparative estimation involves the decoding of relative magnitudes, such as judging the ratio of two physical slopes on a graph.

Measurement can be achieved in some cases by perceptual processes such as extracting the percentages from pie charts, but it usually requires scale reading. Discrimination and comparative estimation are carried out almost exclusively by perceptual processing. For most graphs, comparative estimation is the goal. Discrimination is important only insofar as it contributes to comparative estimation, because we are rarely interested only in whether two positive magnitudes m_1 and m_2 are exactly equal or not. If a discrimination task leads to the conclusion that they are equal, then the comparative estimation task has automatically been performed, since we have concluded that $m_1/m_2 = 1$, or nearly so. If the magnitudes are judged not equal, then we almost always judge the magnitude of the inequality. (See Sec. 6.1 for an example.)

5.3 Processing of Slope Information

Let s_i $(i = 1, \ldots, n)$ be the physical slopes of a collection of line segments, and let s_i^* be the slopes in the scale-line coordinate system. Then $s_i^* = cs_i$. The rapid perceptual processing discussed in previous sections yields information about s_i but not about c, so we get information only about ratios of the s_i^*. In other words, the perceptual processing of slope information results in comparative estimation and not measurement. To get a value of s_i^* from a line segment on a graph, we must employ scale reading and nonvisual cognitive processing in a complicated way: (a) Project the endpoints of the segment onto the vertical scale line, note the scale values, and do a mental subtraction to get the vertical change in the units of the data. (b) Carry out a similar process to get a horizontal change. (c) Do a mental division. Because this is tedious, we do little of it. Most of our extraction of slope information comes from a perceptual processing of physical slopes, the comparative estimation of slope information. For this reason, perceptual processing and comparative estimation are studied in Sections 6 and 7.

6. THEORY

In this section we use theory to consider slope judgments and how shape affects them. We first discuss the interplay between orientation and the judgment of slope ratios, then we present an important fact about the resolution of orientations, and finally we discuss two important perceptual issues.

6.1 A Hypothesis About Orientation Resolution and the Judgment of Slope Ratios

Suppose we have n line segments on a graph with positive slopes. The segments' orientations span a certain range. The *orientation resolution* of the n segments is the maximum orientation minus the minimum orientation. The *orientation midangle* is the average of the maximum and minimum orientations.

For the moment, suppose that there are two segments, that s_1 and s_2 are the physical slopes when the shape parameter is 1, and that $s_1 > s_2$. Thus when the shape parameter is γ, the slopes are γs_1 and γs_2, the orientation resolution is $\arctan(\gamma s_1) - \arctan(\gamma s_2)$, and the orientation midangle is $[\arctan(\gamma s_1) + \arctan(\gamma s_2)]/2$. Our ability to detect a difference in γs_1 and γs_2 depends on the orientation resolution: If this resolution falls below a certain discrimination limit, we judge the ratio $(\gamma s_2)/(\gamma s_1) = s_2/s_1$ to be 1, and the absolute error will be $1 - s_2/s_1$. As γ goes to 0, the orientation resolution goes to 0° because the orientations of both line segments go to 0°. Thus once γ is sufficiently small, we are guaranteed that our error will be $1 - s_2/s_1$, no matter how much bigger s_1 is than s_2. A similar statement holds as γ tends to ∞, since the orientations go to 90°.

We take this discussion one step further and hypothesize the following: Our judgment of the ratio of two slopes is most accurate when the orientation resolution is the greatest, and the accuracy decreases as the resolution decreases. (In Sec. 7 we treat this hypothesis experimentally.)

6.2 Orientation Resolution and Midangle

Since our working hypothesis is that accuracy increases with orientation resolution, it makes sense to ask when resolution is a maximum. Let us continue with the two line segments of Section 6.1. Let $f = s_2/s_1$. To make formulas simpler, we define $\lambda = \gamma(s_1 s_2)^{1/2}$ (λ is defined so that when $\lambda = 1$, the midangle is 45°). When the shape parameter is γ, the physical slope of the first line segment is $\gamma s_1 = \lambda/\sqrt{f}$, the physical slope of the second line segment is $\gamma s_2 = \lambda\sqrt{f}$, the resolution is

$$r(\lambda) = \arctan(\lambda/\sqrt{f}) - \arctan(\lambda\sqrt{f}),$$

and the midangle is

$$a(\lambda) = \frac{\arctan(\lambda/\sqrt{f}) + \arctan(\lambda\sqrt{f})}{2}.$$

From these formulas we can see the following:

1. As λ goes from 0 to ∞, $a(\lambda)$ increases monotonically from 0° to 90°.
2. $r(\lambda)$ has a maximum of $2[45° - \arctan(\sqrt{f})]$ at $\lambda = 1$, and at this value $a(1) = 45°$. In other words, the resolution is maximized when the midangle is 45°. Also, $r(\lambda)$ increases monotonically as λ goes from 0 to 1 and decreases monotonically as λ goes from 1 to ∞.
3. $r(\lambda) = r(\lambda^{-1})$ and $|a(\lambda) - 45°| = |a(\lambda^{-1}) - 45°|$. In other words, the resolution when the midangle is $45° - \theta$ ($0° \le \theta \le 45°$) is the same as when the midangle is $45° + \theta$.

The properties described in the previous paragraph are illustrated in Figure 5. The same data are graphed in all three panels. The midangle is 45° in Panel A, 6° in Panel B, and 84° = 90° − 6° in Panel C; in B and C the orientation resolutions are equal because of the symmetry in $r(\lambda)$ and are less than in the top panel because $r(\lambda)$ is maximized at $a(\lambda) = 45°$. Since the orientation resolution is much less in Panels B and C, we can barely detect the change in slope that is readily apparent in Panel A; consequently, in Panels B and C it is harder to see the curvature apparent in Panel A.

6.3 Orientation Distortion

In Section 6.2 we showed that the physical orientation resolution of two line segments with positive slopes is greatest when the midangle is 45°. Application of this fact assumes that our ability to discriminate depends only on the angular resolution and not otherwise on the orientations of the two segments. But this is not quite true. Suppose we have two segments whose orientation resolution is small and fixed. Our ability to discriminate between the orientations is somewhat less when the segments are near 45° than when they are near the horizontal or vertical (Fisher 1974). If we choose the shape parameter to make midangles closer to 45° we increase discrimination in principle by increasing the angular resolution on the graph, but the actual increase in discrimination might be offset somewhat by the decreased discrimination ability at 45°. Nevertheless, the variation in our discrimination ability is typically small compared with the change in physical resolution due to changing midangle, so we do not believe the variation in discrimination ability has a major effect.

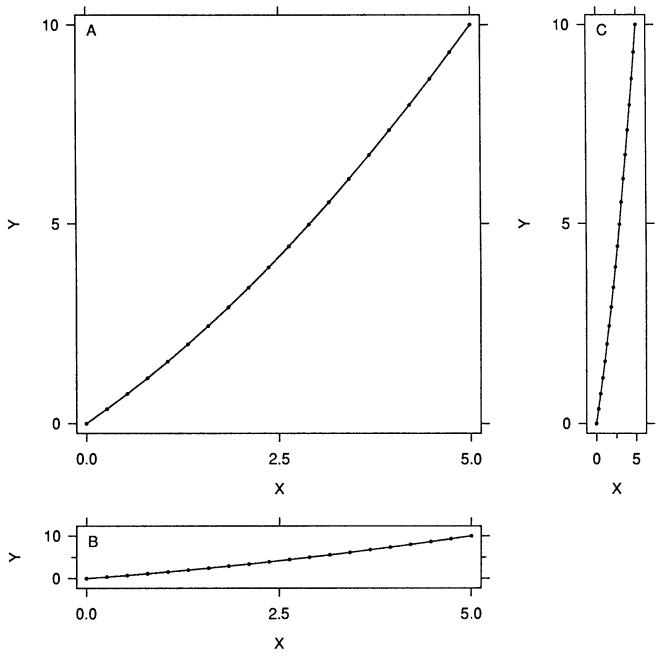

Figure 5. Orientation Resolution. The same data are graphed in all three panels. The orientation resolution is 18.4° in Panel A, 4.0° in Panel B, and 4.0° in Panel C. In Panels B and C it is difficult to perceive the curvature that is apparent in Panel A, because the resolutions are so much smaller. Orientation resolution is maximized when the midangle is 45°.

6.4 The Horizontal–Vertical Distortion

One well-known visual distortion is the horizontal–vertical illusion: Vertical extents are judged to be longer than horizontal ones (Coren and Girgus 1978). It is tempting to suppose that this will affect slope judgments, but analysis suggests that it is unlikely to have an appreciable effect. Actually, we know of no experimental evidence demonstrating the *virtual* vertical lines associated with the vertical extents of oblique actual lines are judged to be longer than virtual horizontal lines of the same lengths. Nevertheless, let us suppose that the illusion occurs as it does for actual horizontal and vertical lines. This means that the lengths of the virtual vertical lines would appear to be inflated by a factor of $1 + r$; Künnapas (1957) reported a value of $r = .071$ for an experiment in well-lighted viewing conditions. Whatever the value of r, as long as it is relatively constant over the line segments whose slopes are being judged, the inflation of the vertical distances does not affect our visual decoding of slope ratios, since the factor $1 + r$ cancels when a ratio is taken.

7. AN EXPERIMENT ON SLOPE PERCEPTION

We ran an experiment to test the hypothesis that accuracy increases with orientation resolution. In Section 6, we argued that rapid perceptual processing is used in decoding slope information. Thus in the experiment we limited subjects to a 2.25-second view to deter them from

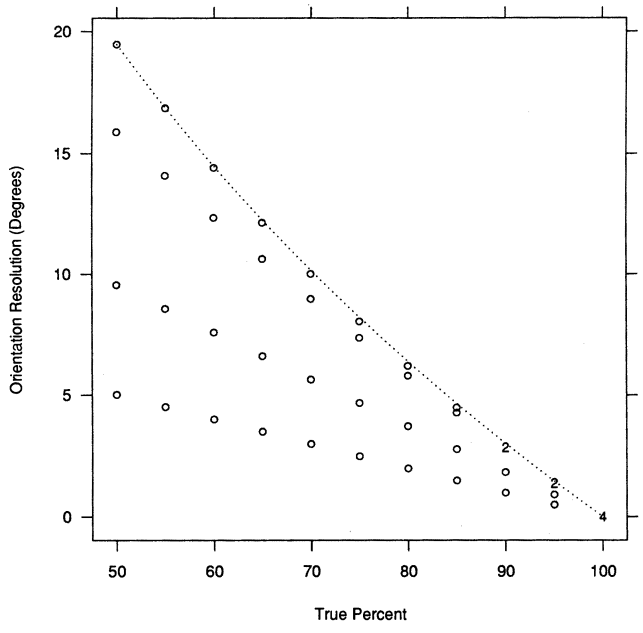

Figure 6. Design Configuration. An experiment was run to probe the hypothesis that the accuracy of slope judgments increases with increasing orientation resolution. Subjects judged 44 line-segment pairs. The circles and numerals show the orientation resolutions and slope ratios (expressed as a percent) for the 44 segments; a numeral indicates the number of values at or near its plotting location. The curve is the maximum resolution possible for each percent.

using highly cognitive processing. In Section 6 we also argued that the relevant task in slope judgments is comparative estimation, that is, judging slope ratios. Thus we had subjects judge ratios. Previous experiments showed that the accuracy of comparative estimation tasks depends on the value of the ratio being judged (Cleveland and McGill 1984, 1985). Hence we systematically varied ratio as well as resolution.

7.1 Design

Pairs of line segments with positive slopes were drawn on a computer-terminal screen. One member of each pair was designated as the standard, and subjects were asked to judge what percent the slope of the other segment was of the slope of the standard. Subjects were informed the true percents varied between 50 and 100. Judgments were entered through the terminal keyboard. The horizontal extents of the two line segments in each pair were equal. Thus the slopes of the line segments were proportional to the vertical extents. Subjects were asked to judge, specifically, the vertical extents of the segments. Horizontal extents varied, however, from one pair to another.

The standards had one of four different orientations: Standard 1, 55°; Standard 2, 35.5°; Standard 3, 19.7°; Standard 4, 10.1°. The standard always appeared in the upper left corner of the screen and the other segment in the lower right. For each of the standards there were 11 line-segment pairs, with the true percents ranging from 50 to 100 in steps of 5. Thus the stimulus set that subjects judged consisted of 44 line-segment pairs. The orientation resolutions ranged from 0° to 19.5°. Figure 6 portrays the design configuration; the circles and numerals show the orientation resolutions and the true percents. The curve is a boundary of the design space, showing the maximum resolution that one can get for given true percent. We know from Section 6.2 that this maximum is achieved when the midangle is 45° and is equal to $2[45° - \arctan(\sqrt{f})]$, where f is the true percent divided by 100.

Subjects judged the complete set 12 times; the order of the 44 line-segment pairs was randomized differently within each set. At the end of each set, each subject was given the average absolute error for the set, where the absolute error of a judgment is |true percent − judged percent|. The score was reported to encourage the subject to work hard at the task by making the session like a game. (Because there was sufficient training before the recorded trials, no consistent learning effect was observed.)

Sixteen subjects participated in the experiment. Backgrounds varied—there were statisticians, engineers, and technical typists, but all had considerable experience using computer keyboards.

The experiment considered two factors: orientation resolution and true percent. In a first experiment it seemed important to make the number of factors explored as small as possible and devote as much of the finite resources as possible to replication. The placement of the standard in the upper left and the other segment in the lower right was done to keep proximity and overall configuration about the same. The horizontal extents for the two segments of each pair were made the same to help keep the method of judgment consistent throughout the experiment. Because experiments of this type are probing processes that the human visual system uses in its algorithms for seeing in general, we would not expect level of technical training or experience in graph reading to be an important factor. We have found this to be the case in other experiments (Cleveland and McGill 1986) and thus we did not control for training and experience in this experiment.

7.2 Data Analysis

The data from the experiment consist of 8,448 subject judgments that form a four-way array: 16 subjects × 11 true percents × 4 standards × 12 replications. For each judgment, the absolute error was computed. A mean across the 12 replications was computed for each of the 44 × 16 = 704 combinations of line-segment pairs and subjects. Let m_{ijk} ($i = 1, \ldots, 11$, $j = 1, \ldots, 4$, $k = 1, \ldots, 16$) be the means. These means were further averaged across subjects to form 44 means for the line-segment pairs

$$m_{ij} = \frac{1}{16} \sum_{k=1}^{16} m_{ijk},$$

and a standard error was computed for each m_{ij} by

$$s_{ij} = \frac{1}{4} \sqrt{(1/15) \sum_{k=1}^{16} (m_{ijk} - m_{ij})^2}.$$

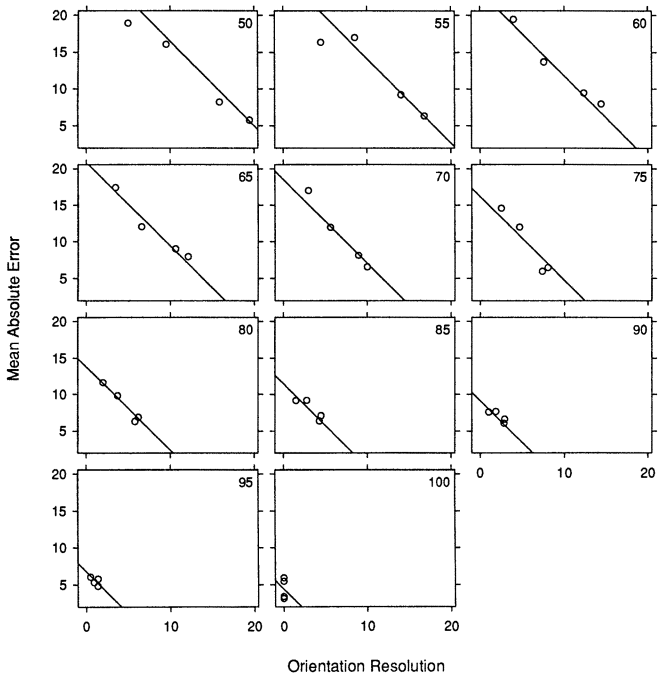

Figure 7. Experiment. In the experiment, subjects judged what percent the slope of one line segment was of the slope of another segment. (All slopes were positive.) Means (across replications and subjects) of the absolute errors for judgments of the 44 pairs of line segments are graphed against the orientation resolutions of the pairs. For each panel, the four values that are graphed are for pairs where the true percent is equal to the value shown in the upper right of the panel. The experiment supported the hypothesis. Judgment error was modeled as a linear function of orientation resolution and true percent. The line segments show the fitted model.

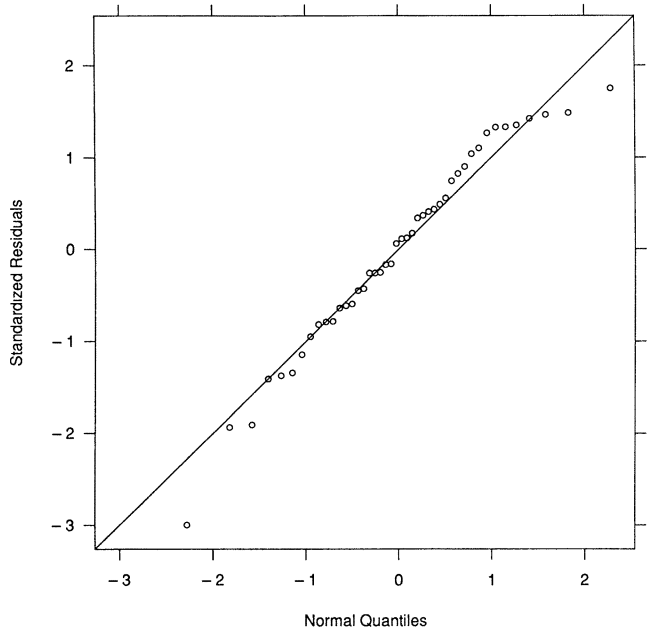

Figure 8. Residual Plots: Normal Probability Plot of Residuals Divided by Standard Deviations Across Subjects. Since the distribution of the standardized residuals is well approximated by an N(0, 1) distribution, the model has explained as much variation as possible.

Thus the sampling frame to which the s_{ij} are applicable is one in which the 16 subjects are a sample from a homogeneous population of potential subjects capable of understanding and carrying out the instructions of the experiment.

In the following, we let r_{ij} ($i = 1, \ldots, 11, j = 1, \ldots, 4$) be the orientation resolutions of the 44 line segment pair. We could denote the true percents by p_{ij}, but since for each i the four values of p_{ij} ($j = 1, \ldots, 4$) are equal, we use p_i instead. Figure 7 shows the dependence of m_{ij} on p_i and r_{ij}. The circles of the ith panel are a graph of m_{ij} against r_{ij} ($j = 1, \ldots, 4$); the true percent, p_i, is shown on the panel. (The line segments on the panels are explained in the next paragraph.) The hypothesis that error decreases with increasing resolutions is strongly supported by the data.

Exploratory plots such as Figure 7 suggest that m_{ij} is linear in p_i and r_{ij}. Thus we fit the model

$$m_{ij} = \beta + \beta_p(p_i - 100) + \beta_r r_{ij} + \varepsilon_{ij}$$

using weighted least squares with weight $1/s_{ij}^2$ for the (i, j)th observation. The value 100 was subtracted from p_i in the model to make β the error in judging the slope ratio of two parallel line segments; Figure 7 shows the errors are smallest under this condition. The estimated values of the parameters are $\hat{\beta} = 4.39$, $\hat{\beta}_p = -.47$, and $\hat{\beta}_r = -1.14$. The line segments in Figure 7 show the fitted model; on the ith panel, a line with intercept $\hat{\beta} + \hat{\beta}_p p_i$ and slope $\hat{\beta}_r$ is portrayed. We can see that the model provides an excellent fit to the data. Figure 8 is a normal probability plot of the residuals divided by the s_{ij}. The overall pattern of the points is well described by the line $y = x$. Hence the model accounts for as much of the variation in the m_{ij} as is possible, since the residual variation is commensurate with the variation across subjects.

8. A PROCEDURE FOR CHOOSING THE SHAPE PARAMETER

Suppose that a two-variable graph is made to determine how y depends on x. In most cases, judgments of slopes of various line segments—either actual or virtual—contribute to our understanding of the dependence. In this section we describe a method for choosing the shape parameter to enhance the visual decoding of the slopes, and then investigate its performance and properties.

8.1 The Median-Absolute-Slope Procedure

The experiment of Section 7 supports the following hypothesis: For segments with positive slopes, the accuracy of judgments of slope ratios increases with increasing orientation resolution. It seems entirely reasonable to suppose that the same is true for segments with negative slopes. In Section 6.2, it was shown that for segments with positive slopes the orientation resolution is a maximum when the midangle is 45°. A similar statement, with 45° replaced by $-45°$, holds for segments with negative slopes.

Consider a set of n line segments on a graph whose slopes we want to compare visually. Let s_i ($i = 1, \ldots, n$) be the physical slopes when the shape parameter is 1. As we have seen, the slopes are γs_i when the shape pa-

rameter is γ. We cannot, of course, choose a single γ so that all pairwise comparisons of positive s_i involve 45° midangles and all pairwise comparisons of negative s_i involve midangles of $-45°$. But we can attempt to center the segments with positive slopes on 45° and those with negative slopes on $-45°$. We do this by choosing γ so that median $|\gamma s_i| = 1$. This is called the *median-absolute-slope* procedure.

8.2 General Comments

Before discussing applications of the median-absolute-slope procedure, it is important that several general issues be made explicit. (Other important general comments are given in Sec. 9.)

An application of the median-absolute-slope procedure requires that there be a collection of line segments with slopes whose pairwise ratios we want to visually decode. The relevance of segments on a particular graph needs careful thought and depends on the application. In some cases we want to focus on segments connecting successive observations and in other cases we want to focus on the slopes of a smooth curve through the data. The choice of segments can have a substantial effect on the resulting shape.

The median-absolute-slope procedure cannot be used in a fully automated way. (No data-analytic procedure can be.) It needs to be tempered with judgment. As we shall see, its application can lead to more than one answer or to an infeasible answer (see Secs. 8.4 and 8.5).

Slope judgments are typically important for graphs that show how y depends on x. But slope judgments are typically not important for graphs that show how points (x_i, y_i) are distributed in the plane, and for which dependence is not an issue. Figure 6 is an example. Thus neither the median-absolute-slope procedure nor anything else in this investigation provides guidance for choosing the shape parameters of graphs such as Figure 6.

8.3 Applications

The median-absolute-slope procedure was used to choose the shape parameter in Figure 1. The line segments used as input for the procedure are the virtual segments connecting the radiation values at successive depths.

The median-absolute-slope procedure was also used to choose the shape parameter for Figure 3, where $\gamma = .074$. The segments used as input for the procedure are the actual line segments connecting successive observations. In Figure 2 the shape was simply set to 1, the default for many statistical graphics software routines. Figure 9 shows why Figure 3 works and Figure 2 does not. The top panel of Figure 9 is a quantile plot (i.e., sample-distribution function plot with the axes reversed) of the orientations of the line segments joining successive values in Figure 2, and the bottom panel is a quantile plot of the orientations in Figure 3. The orientations in Figure 2 are generally near 90° and $-90°$, making visually decoding the slopes difficult. The orientations in Figure 3 are spread out more uniformly, and many more fall in the range that leads to higher-accuracy slope judgments. For this reason, the slower rise than fall of the lynx numbers is apparent in Figure 3.

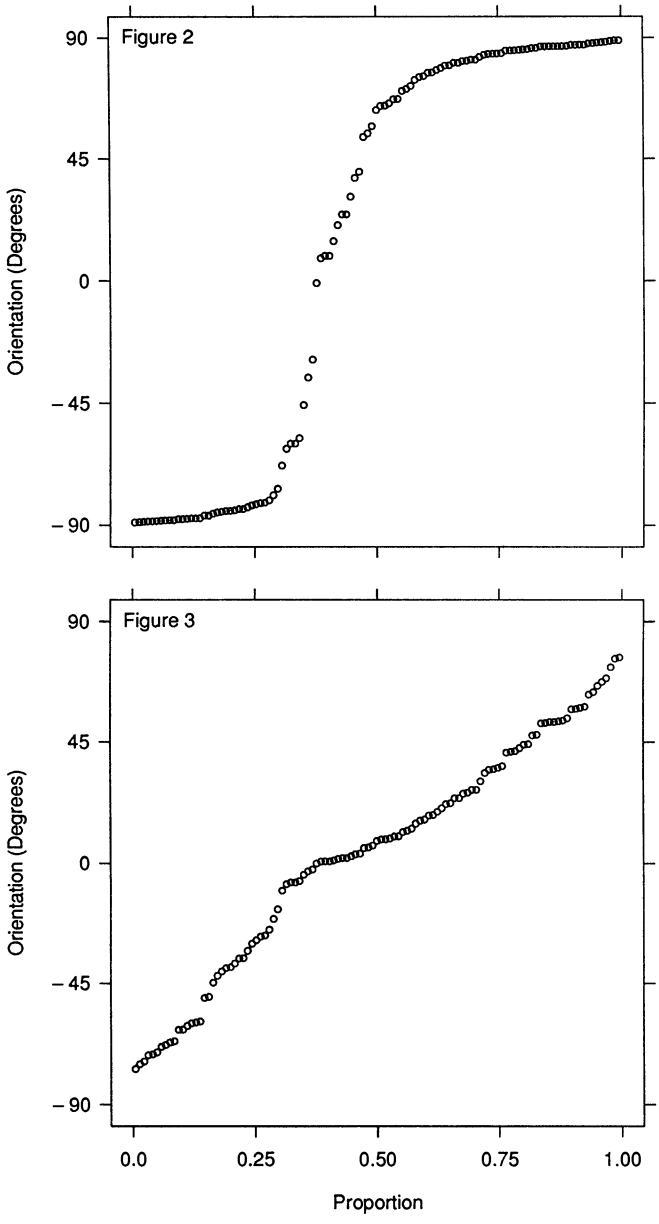

Figure 9. Centering Orientations on ±45°. The top panel is a quantile plot of the orientations of line segments in Figure 2. The bottom panel is the same for Figure 3. In Figure 3 the shape parameter was chosen so that the median absolute slope of line segments connecting successive points is equal to 1; this results in orientations that allow more effective slope judgments than in Figure 2.

In the two previous examples we focused on the line segments connecting successive observations, because the goal was to study the very local behavior of the rate of change of y as a function of x. For example, seeing that the slope of the rightmost virtual line segment in Figure 1 is considerably less than the rest is important: It suggests that the prediction of radiation at the lowest depth is incorrect. In Figure 3, the judgment of local slopes lets us perceive the shapes of the 10-year cycles and determine that they rise more slowly than they fall.

But consider Figure 10. The circles are a scatterplot of atmospheric temperature and ground-level ozone, an air pollutant. The purpose of the graph is to assess how ozone depends on temperature. In this case we are not interested in the local slopes, that is, those of line segments connecting successive observations, because the local slopes are dominated by the noise in the data. Rather, we are interested in the global slope change. This global change is portrayed by the regression curve, a nonparametric estimate using robust locally weighted regression (Cleveland 1979). The estimate, $\hat{y}(x_i)$, was computed at the distinct values of temperature, x_i, and graphed by connecting successive values of $(x_i, \hat{y}(x_i))$ by line segments; because the slopes of the line segments change smoothly with temperature, the visual appearance is a smooth curve. These segments were the input for the median-absolute-slope procedure, which was used to choose the shape.

8.4 Shapes That Are Too Large or Too Small

We must ask what else on the graph can be affected by having the shape optimize the visual decoding of slopes. The only major problem that can arise is a loss of resolution if the optimum shape is too large or too small. If the vertical or horizontal height of the data rectangle is too small, we lose resolution along the horizontal or vertical scale because the plotting device is not infinitely accurate and because the heights of plotting characters and the widths of actual line segments are not infinitesimal. For example, this resolution problem commonly arises when a long time series is graphed and the local slopes are used as input to the median-absolute-slope procedure. (In Sec. 8.6 we see that the optimum shape in such a case can fall off like n^{-1}.) In Figure 3, the optimum shape parameter of .074 is used. This value does not ruin the vertical resolution, but we are quite close to the limit.

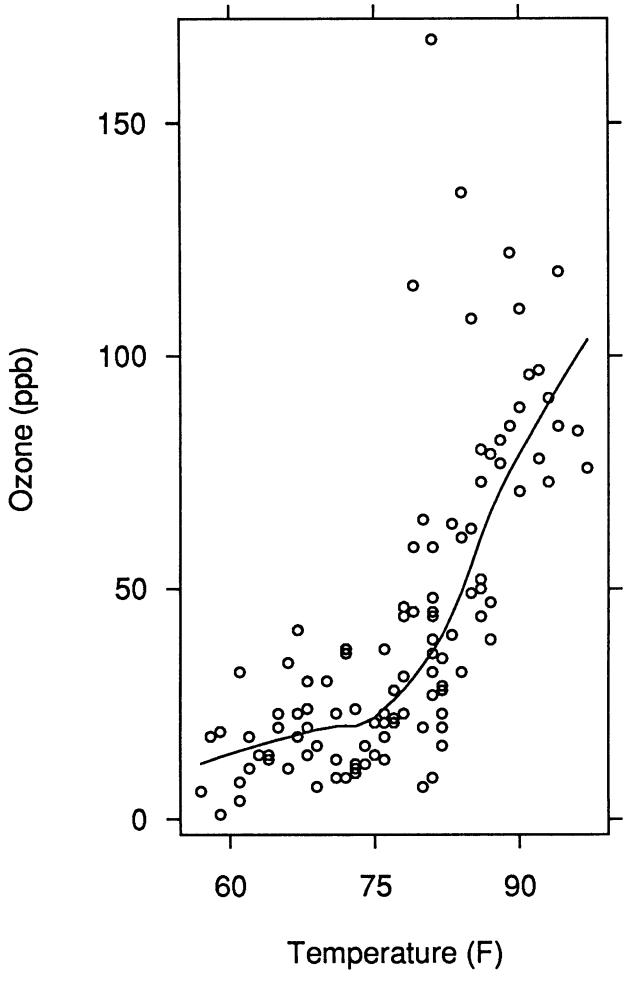

Figure 10. Optimizing the Slopes of a Smooth Curve. The curve, a nonparametric regression estimate using robust locally weighted regression, is graphed by connecting successive estimated values by line segments. The shape parameter was chosen by applying the median-absolute-slope procedure to these segments.

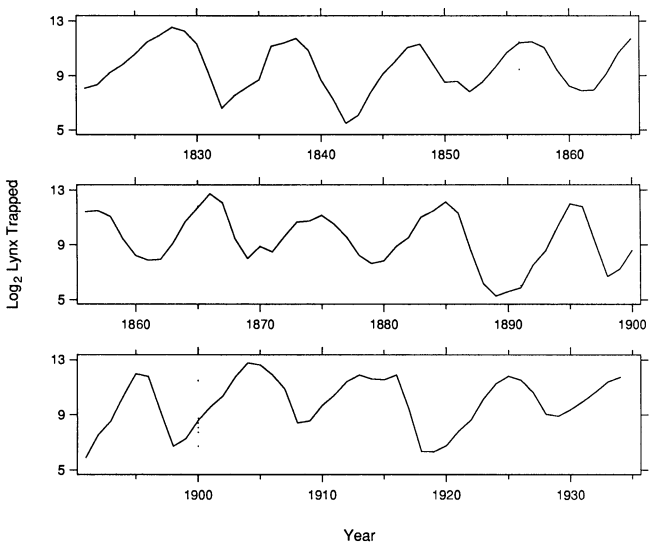

Figure 11. Multiple Panels. When the median-absolute-slope procedure is applied to a long time series, the resulting shape parameter is often too small to allow good vertical resolution. One solution is multiple panels, as illustrated here.

One solution when the optimum value is too small is to break up the horizontal scale into several intervals and graph the subsets of the data on juxtaposed panels. (Similarly, when the value is too large, we can break up the vertical scale.) This is illustrated in Figure 11, which shows the lynx numbers. The vertical scales of the three panels in the figure are the same, and the horizontal scales have the same number of units per cm. The common shape parameter of the panels is chosen by applying the median-absolute-slope procedure to the collection of slopes on all panels. It is sometimes helpful (as in Fig. 11) to have the intervals overlap somewhat, to ease the visual transition from one panel to the next; the overlap is indicated by the dotted vertical lines.

8.5 More Than One Set of Slopes

A graph may have more than one set of slopes that are of interest. In such a case, the different sets can lead to quite different optimum shapes. In Figure 3, the segments used as input for the median-absolute-slope procedure are those connecting successive observations. But another interesting set of segments are the virtual line segments connecting the peaks of the cycles and those connecting the troughs of the cycles, because the changes through time

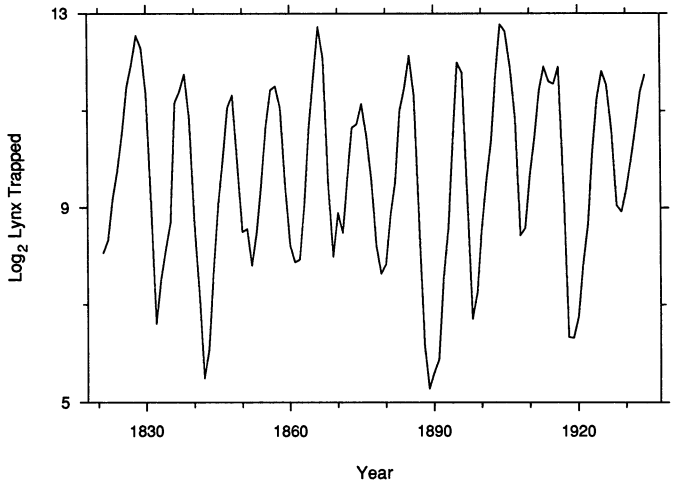

Figure 12. Two Sets of Slopes. In Figure 3 the shape was chosen by the median-absolute-slope procedure, with the line segments connecting successive observations as input. Here the shape was chosen by the same procedure but with the input changed to the virtual segments connecting successive peaks and the virtual segments connecting successive troughs.

in the extremes of the cycles are important and a challenge in modeling the data (Campbell and Walker 1977). In Figure 3 it is difficult to judge the changes in the peaks and troughs through time, however, because the orientation of these virtual segments are too close to 0. The solution is to make another graph, with the shape chosen using the virtual segments as input for the median-absolute-slope criterion. This is done in Figure 12, where the shape parameter is .67.

A dynamic graphics environment is particularly well suited to changing the shape of a graph. The graphs in this article are static in that they remain constant through time. But with a computer graphics terminal that has direct access to a computer processor (a bitmap display terminal), graphs can be changed nearly instantaneously (Becker, Cleveland, and Wilks 1987). Using an input device such as a "mouse," a data analyst can vary the shape virtually continuously and see the graph change on the screen. Buja, Asimov, Hurley, and McDonald (in press) implemented a system in which the user can dynamically change the height or width of the graph: If the graph becomes wider than the screen, the data analyst can scan horizontally back and forth to show the entire graph, which means we can get a small shape parameter without having a too-small height.

8.6 The Algorithm for a Special Case: A Connected Sequence of Line Segments

Suppose that the coordinates of a collection of values on a graph are (x_i, y_i) $(i = 1, \ldots, n + 1)$. Suppose also that $x_{i+1} > x_i$ and that the line segments whose slopes we want to compare are those with endpoints (x_i, y_i) and (x_{i+1}, y_{i+1}) $(i = 1, \ldots, n)$. That is, we want to study the very local behavior of the dependence of y on x. Finally, we suppose that no other graphical elements appear on the graph except the (x_i, y_i) and their connecting line segments. Two examples are the radiation data in Figure 1 and the lynx data in Figure 3. Let R_y be the range of the y_i, and let R_x be the range of the x_i. Also, let M be the median of the values $|y_{i+1} - y_i|/|x_{i+1} - x_i|$ $(i = 1, \ldots, n)$. Then the value of γ that satisfies the median-absolute-slope criterion is $M^{-1}R_y/R_x$. Note that this dimensionless expression holds whether we express (x_i, y_i) in scale-line coordinates or physical coordinates.

Suppose that the (x_i, y_i) are the plotted points of a time series measured at equally spaced points in time, as in Figure 3. Thus we can take $x_i = i$ $(i = 1, \ldots, n + 1)$. Then, the optimum shape parameter is

$$\gamma_n^* = \frac{\text{range } \{y_i\}}{n \text{ median } \{|y_{i+1} - y_i|\}}.$$

Suppose that y_i is stationary. Furthermore, suppose that y_i is a bounded random variable. In practice, this is a realistic assumption for many time series, such as the sunspot numbers. As n increases, $n\gamma_n^*$ converges with probability 1 to a constant. Thus as n increases, γ_n^* can be expected to get small, eventually falling off like n^{-1}. For this reason a long time series will tend to have a small optimum shape parameter.

The behavior of the optimum shape for a time series may explain why in Figure 4 the distributions of shape ratios for the time series graphs were shifted toward lower values compared with the other categories: People making the graphs may have found, in an ad hoc way using a purely visual process, that smaller shape parameters do a better job of showing the local variation of time series.

Let us also consider the plotting of deterministic functions. Let $f(x)$ be a function with a derivative $f'(x)$ on the interval $[a, b]$. Suppose that the x_i $(i = 1, \ldots, n + 1)$ are equally spaced points from a to b. Let R be the range of the function over the interval, and let M be the median of the distribution of the random variable $|f'(X)|$, where X is uniform on $[a, b]$. As n gets large, the optimum shape parameter converges to $\gamma^* = M^{-1}R/(b - a)$. If f is linear, then $\gamma^* = 1$. If $g(x) = uf(x) + v$, then f and g have the same value of γ^*. Suppose that f' is monotone; then $M = |f'((b - a)/2)|$, since the values of f' are already ordered. Suppose that $a = 0$, $b = 1$, and $f(x) = x^k$; then $\gamma^* = k/2^{k-1}$, when $k = 1$ or 2, $\gamma^* = 1$, but then γ^* decreases monotonically as k increases. Suppose that $f(x)$ is periodic with period p, and suppose that $b - a = mp$, where m is an integer; if $\gamma^* = \gamma_0$ when $m = 1$, then $\gamma^* = \gamma_0/m$ for general m.

8.7 The Algorithm for the General Case

The algorithm for the general case is actually quite similar to that for the special case of Section 8.6. Suppose that the endpoints of the ith of n line segments to be used as the input for the median-absolute-slope procedure are (x_{i1}, y_{i1}) and (x_{i2}, y_{i2}), where $x_{i1} < x_{i2}$. Let R_x be the range of all $2n$ values of x_{i1} and x_{i2}. Define R_y similarly. Let

$$M = \text{median}|(y_{i2} - y_{i1})/(x_{i2} - x_{i1})|.$$

Let c_y be the number of data units per cm on the vertical scale of the graph, and let c_x be defined similarly. To satisfy the median-absolute-slope criterion, choose c_x and c_y so that the shape parameter of the data rectangle just enclosing the $2n$ points (x_{i1}, y_{i1}) and (x_{i2}, y_{i2}) is $M^{-1}R_x/R_y$. Of course, if there are other data points on the graph that are not involved in defining the slopes for the procedure, then the horizontal and vertical scales must be extended, if necessary, to include the other points.

Figure 10 illustrates this general case. The (x_{i1}, y_{i2}) and (x_{i2}, y_{i2}) describe the endpoints of the line segments that make up the smooth curve. The shape of the data rectangle that just encloses these endpoints is $M^{-1}R_y/R_x = 1.10$. The horizontal and vertical scales must be expanded beyond this rectangle to accommodate the original data, plotted by the circles. The shape parameter of all of the graphical elements, and thus of the graph, is 2.02.

9. DISCUSSION

9.1 The General Principle and the Median-Absolute-Slope Criterion

Theory and experimentation led to a general principle, that judging the ratio of two positive slopes is optimized by a midangle of 45°, and judging the ratio of two negative slopes is optimized by a midangle of $-45°$. This principle led to the median-absolute-slope criterion. *But the principle is far more important than the criterion.* Any heuristic that brings orientations closer to $\pm 45°$ will enhance slope judgments. It probably would even suffice to make a rough choice visually as the shape is varied, say in a dynamic graphs environment. Nevertheless, one desirable aspect of an explicit criterion—such as the median-absolute-slope procedure—is that it provides something whose properties can be studied.

The examples we have used show that the set of slopes one chooses to enhance by the choice of the shape parameter is enormously varied, and much depends on the application. A careful selection of the slopes is, of course, an important part of a successful application of the method.

9.2 Graphing Slopes Directly

The examples and the experimental results show that at best we can only get a rough visual estimate of rate of change from slope judgments, even in the best of circumstances (i.e., when the orientations are optimized). In examples where studying rate of change with more accuracy is needed, it is important to graph rate of change directly so that the values can be visually decoded by the more accurate judgments of position along a common scale (Cleveland 1985; Cleveland and McGill 1985).

9.3 The Importance of Slope Assessments

The importance of accurate slope judgments should not be underestimated. Inaccurate slope judgments can lead to inappropriate descriptions of data and inappropriate statistical models. One example is the lynx data. The phenomenon observed in Figure 3—the slower rise than fall of the lynx trappings—is of substantial importance. It means that the series is not time-reversible. Such a time series cannot be both stationary and Gaussian. Unfortunately, most attempts at modeling the data depend on an assumption of these two properties (Campbell and Walker 1977). Apparently, the slower rise than fall has not been fully appreciated, possibly because the shape parameters of graphs of the data were too large to show the phenomenon. For example, in the article by Elton and Nicholson (1942) the untransformed data were graphed with $\gamma = .31$, and the slower rise than fall could not be seen. The optimum shape for the untransformed data is .11 when the line segments connecting successive observations are used as input. Bulmer (1974), Campbell and Walker (1977), and Priestley (1981) graphed the logs with shape parameters of .46, .23, and .99, respectively. In none of these graphs is the shape small enough for the phenomenon to be convincingly seen. As we have seen, the optimum shape for the logs is .074.

[*Received October 1986. Revised October 1987.*]

REFERENCES

American National Standards Institute (1979a), *Illustrations for Publication and Projection* (ANSI Y15.1M), New York: American Society of Mechanical Engineers.

——— (1979b), *Time-Series Charts* (ANSI Y15.2M), New York: American Society of Mechanical Engineers.

American Standards Association (1938), *Time Series Charts: A Manual of Design and Construction* (ASA Z15.2), New York: American Society of Mechanical Engineers.

Baird, J. C., and Noma, E. (1978), *Fundamentals of Scaling and Psychophysics,* New York: John Wiley.

Becker, R. A., Cleveland, W. S., and Wilks, A. R. (1987), "Dynamic Graphics for Data Analysis," *Statistical Science,* 2, 355–395.

Bertin, J. (1967), *Semiologie Graphique,* Paris: Gauthier-Villars.

——— (1983), *Semiology of Graphics,* Madison: University of Wisconsin Press. (English translation of *Semiologie Graphique.*)

Brinton, W. C. (1914), *Graphical Methods for Presenting Facts,* New York: Engineering Magazine Co.

Buja, A., Asimov, D., Hurley, J. A., and McDonald, J. A. (in press), "Elements of a Viewing Pipeline for Data Analysis," in *Dynamic Graphics for Statistics,* eds. M. E. McGill and W. S. Cleveland, Monterey, CA: Wadsworth.

Bulmer, M. G. (1974), "A Statistical Analysis of the 10-Year Cycle in Canada," *Journal of Animal Ecology,* 43, 701–715.

Campbell, J. J., and Walker, A. M. (1977), "A Survey of Statistical Work on the Mackenzie River Series of Annual Canadian Lynx Trappings for the Years 1821–1934 and a New Analysis," *Journal of the Royal Statistical Society,* Ser. A, 140, 411–431.

Cleveland, W. S. (1979), "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association,* 74, 829–836.

——— (1985), *The Elements of Graphing Data,* Monterey, CA: Wadsworth.

Cleveland, W. S., and McGill, R. (1984), "Graphical Perception: Theory, Experimentation, and Application in the Development of Graphical Methods," *Journal of the American Statistical Association,* 79, 531–534.

——— (1985), "Graphical Perception and Graphical Methods for Analyzing and Presenting Scientific Data," *Science,* 229, 828–833.

——— (1986), "An Experiment in Graphical Perception," *International Journal of Man-Machine Studies,* 25, 491–500.

Coren, S., and Girgus, J. S. (1978), *Seeing Is Deceiving,* Hillsdale, NJ: Lawrence Erlbaum.

Costigan-Eaves, P. J. (1984), "Data Graphics in the 20th Century: A Comparative and Analytic Survey," unpublished doctoral dissertation, Rutgers University, Dept. of Education (University Microfilms 8424036).

Elton, C., and Nicholson, M. (1942), "The Ten-Year Cycle in Numbers of Lynx in Canada," *Journal of Animal Ecology,* 11, 215–244.
Fisher, G. H. (1974), "An Experimental Study of Linear Inclination," *Quarterly Journal of Experimental Psychology,* 26, 52–62.
Follettie, J. F. (1986), "User Tasks of Statistical Graph-Using and Analytic Microtasks of Graphics Research," unpublished manuscript.
Hall, A. S. (1958), *The Construction of Graphs and Charts,* London: Pitman.
Julesz, B. (1981), "Textons, the Elements of Perception, and Their Interactions," *Nature,* 290, 91–97.
Karsten, K. G. (1923), *Charts and Graphs,* New York: Prentice-Hall.
Künnapas, T. M. (1957), "The Vertical–Horizontal Illusion and the Visual Field," *Journal of Experimental Psychology,* 54, 405–407.
Lefferts, R. (1981), *Elements of Graphics,* New York: Harper & Row.
Littler, M. M., Littler, D. S., Blair, S. M., and Norris, J. N. (1985), "Deepest Known Plant Life Discovered on an Uncharted Seamount," *Science,* 55, 57–59.
Lutz, R. R. (1949), *Graphic Presentation Simplified,* New York: Funk & Wagnalls.
MacGregor, A. J. (1979), *Graphics Simplified,* Toronto: University of Toronto Press.
Marr, D. (1982), *Vision,* San Francisco: W. H. Freeman.
McGill, R., Tukey, J. W., and Larsen, W. A. (1978), "Variations of Box Plots," *The American Statistician,* 32, 12–16.
Meyers, C. H. (1970), *Handbook of Basic Graphs: A Modern Approach,* Belmont, CA: Dickenson.
Priestley, M. B. (1981), *Spectral Analysis and Time Series,* New York: Academic Press.
Rock, I. (1984), *Perception,* New York: W. H. Freeman.
Schmid, C. F., and Schmid, S. E. (1979), *Handbook of Graphic Presentation,* New York: John Wiley.
Stevens, K. A. (1978), "Computation of Locally Parallel Structure," *Biological Cybernetics,* 29, 19–28.
Tufte, E. R. (1983), *The Visual Display of Quantitative Information,* Cheshire, CT: Graphics Press.
U.S. Department of the Army (1966), *Standards of Statistical Presentation,* Washington, DC: U.S. War Office.
Von Huhn, R. (1931), "A Trigonometrical Method for Computing the Scales of Statistical Charts to Improve Visualization," *Journal of the American Statistical Association,* 26, 319–324.
Weld, W. E. (1947), *How to Chart Facts From Figures With Graphs,* Norwood, MA: Codex Book Company.