Graphical Perception: The Visual Decoding of Quantitative Information on Graphical Displays of Data

Author(s): William S. Cleveland and Robert McGill

Source: *Journal of the Royal Statistical Society. Series A (General)*, Vol. 150, No. 3 (1987), pp. 192-229
Published by: Wiley for the Royal Statistical Society
Stable URL: http://www.jstor.org/stable/2981473
Accessed: 11/07/2014 05:17

J. R. Statist. Soc. A (1987)
150, *Part* 3, *pp.* 192–229

Graphical Perception: The Visual Decoding of Quantitative Information on Graphical Displays of Data

By WILLIAM S. CLEVELAND† and ROBERT MCGILL

AT&T Bell Laboratories, USA

[Read before the Royal Statistical Society on Wednesday, February 11th, 1987, the President Professor J. Durbin in the Chair]

SUMMARY

Studies in graphical perception, both theoretical and experimental, provide a scientific foundation for the construction area of statistical graphics. From these studies a paradigm that has important applications for practice has begun to emerge. The paradigm is based on elementary codes: Basic geometric and textural aspects of a graph that encode the quantitative information. The methodology that can be invoked to study graphical perception is illustrated by an investigation of the shape parameter of a two-variable graph, a topic that has had much discussion, but little scientific study, for at least 70 years.

Keywords:

1. INTRODUCTION

When a graph is made, information is *encoded* on the graph by a variety of aspects such as positions of plotting symbols, lengths and slopes of line segments, areas of planar regions, texture, and colour. When a graph is studied, the encoded information is *visually decoded.* This decoding process, which is called *graphical perception*, is a controlling factor in the ability of a graph to convey information; this is illustrated in the following example.

1.1. *An Example: The Shape of a Graph and Slope Judgments*

Fig. 1 shows data on world track records as of 1984 for 13 metric distances from 100 m to 30 km. The vertical scale is the distance divided by the record time and the horizontal scale is the logarithm of distance. The dashed rectangle, which is the *data rectangle*, shows the maximum and minimum values along the vertical scale and along the horizontal scale. Suppose the height of the data rectangle of a graph is h cm and suppose the width is w cm. The *shape parameter*, or *shape*, of the graph is h/w, which is the slope of a line segment joining the lower left corner and the upper right corner of the data rectangle. In Fig. 1 the shape parameter is 1.2.

When a two-variable graph is made to see how y depends on x, we judge the slopes of line segments to determine the rate of change of y as a function of z. The shape parameter of the graph is important because it has a substantial effect on our ability to visually decode the slopes. This is illustrated in Figs 2 and 3, which show the yearly sunspot numbers from 1749 to 1924 that were analyzed by Yule (1927) in his landmark paper on autoregression. In Fig. 2 the shape parameter is 0.065 and in Fig. 3 it is 1. In Fig. 3 it is impossible to see a critical property of the data that is quite clear in Fig. 2—the sunspot numbers rise more rapidly than they decline. The reason for the failure of Fig. 3, as we shall see later in the paper, is that the shape of the graph has made certain slope judgments difficult.

1.2. *Summary of the Paper*

This paper contains a general discussion of graphical perception. In Section 2, three areas of statistical graphics are delineated—computing, methodology, and construction. Graphical perception is of fundamental importance for the construction area because its study provides a scientific foundation for many issues that arise in graph construction. Section 3 describes

† *Address for correspondence*: AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, N.J. 07974, USA.

 0035–9246/87/150192 $2.00

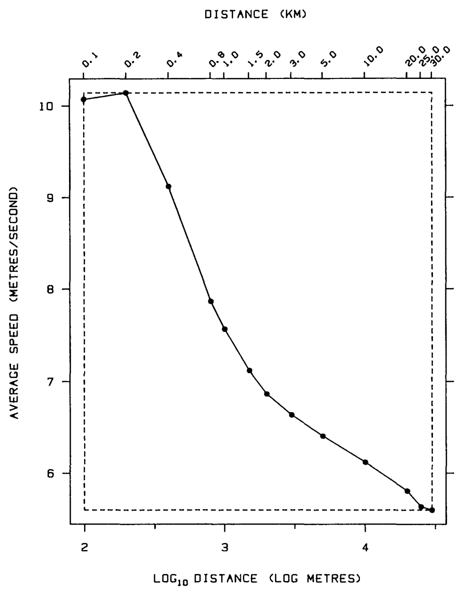

Fig. 1. The shape parameter of a graph. The data rectangle is the dashed rectangle on the graph. The shape parameter is the height of the data rectangle divided by its width, where the height and width are measured in the same units of length. For this graph the shape parameter is 1.2.

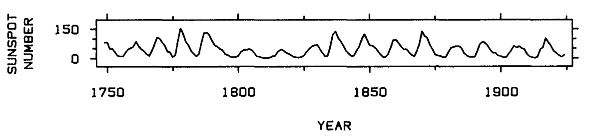

Fig. 2. Sunspot numbers. The shape parameter is 0.065. We can see from this graph that the sunspot numbers rise more rapidly than they fall.

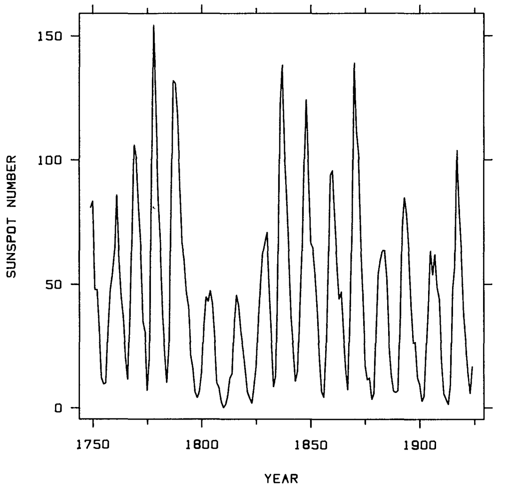

Fig. 3. Sunspot numbers. The shape parameter is 1. We can no longer see the more rapid rise than fall because the change in the shape parameter has made our visual decoding of slopes less accurate than in Fig. 2.

the methodology that can be used to study graphical perception—informal experimentation with graph constructions, formal experiments with human subjects, invoking the theory of vision, and the development of paradigms for graphical perception. In Section 4 one such paradigm is described. Elementary codes of graphs are identified. The codes are then ranked, on the basis of theory and experimentation, according to how accurately we judge the quantitative information that they encode. The ranking has an important application to graph construction; data should be plotted so that the elementary codes employed are as high in the ranking as possible. In Section 5 many general points made in earlier sections are illustrated by a brief description of a study of the shape parameter of a graph and its effect on our visual decoding of slope information. In the past, a lack of systematic study of shape has led to widely-varying, weakly-justified opinions on choosing the shape parameter. In Section 6 the paper concludes with an argument that graphical perception can benefit from an interdisciplinary approach combining knowledge from research in graphical methods and knowledge from research in visual perception.

2. THE ROLE OF GRAPHICAL PERCEPTION IN STATISTICAL GRAPHICS

2.1. *Three Areas of Statistical Graphics—Computing, Methods, and Construction*

Statistical graphics is a newly activated area of statistics because of the computer graphics revolution. High-quality hardware systems are now available at low cost for generating graphs by computer (McDonald and Pedersen, 1985a and 1985b), and the development of software for graphing data is today a particularly vibrant area of statistical graphics (Becker and Chambers, 1984; Donoho, Donoho, and Gasko, 1985; Bentley and Kernighan, 1986). An important trend in this area has been a movement away from batch processing on mainframes toward highly interactive graph production on personal workstations.

A second area of statistical graphics is graphical methods for data analysis. Such a method consists, essentially, of a choice of certain quantitative information to show on a graph to help the analyst understand the data or understand the performance or properties of a statistical model fit to the data (Tukey, 1977; Nelson, 1979; Cox, 1978; Barnett, 1981; Chambers *et al.*, 1983; Tufte, 1983). A particularly promising part of this area is dynamic, high-interaction graphical methods (Becker, Cleveland, and Wilks, 1986) in which a data analyst interacts with a display of data on a computer graphics screen through the use of a control device such as a mouse. The analyst can specify in a visual way points or regions on the display and cause nearly instantaneous changes in the graph. These capabilities provide a new medium for the invention of graphical methods. One such invention is brushing a scatterplot matrix (Becker and Cleveland, 1987). Graphical methods, both static and dynamic, are an important part of the collection of techniques that have come to be known as exploratory data analysis (Tukey, 1977; Chatfield, 1985).

Having decided what quantitative information it is useful to display, one needs to construct a graph. This is a third area of statistical graphics—graph construction. We must decide what geometric aspects of the graph will encode the quantitative information, choose the scales, choose the shape parameter, and so forth. Many of the decisions are controlled by conventions for graph construction that have arisen through an informal process that has led to common usages. But a large amount of the detail of graph construction is nevertheless in the hands of the person making the graph.

2.2. *Graphical Perception: A Scientific Foundation for Graph Construction*

Graphical perception is of fundamental importance for the construction area of statistical graphics. The details of construction of a graph determine what visual processes we must employ to decode the information. The construction is successful only if our visual systems perform this graphical perception with accuracy and efficiency.

By studying graphical perception, both theoretically and experimentally, and then applying the results to making graphs, we are giving graph construction a scientific foundation (MacDonald-Ross, 1977; Kruskal, 1982; Cleveland and McGill, 1984 and 1985). This is badly needed. Writings on graph construction consist of a plethora of how-to books and journal articles that have appeared in the 20th century. Costigan-Eaves (1984), who has carried out a very thorough historical survey, cites Brinton (1914) for one of the early and influential such treatises. The problem with this how-to tradition is that there has been a reign of pure opinion and little scientific study. One symptom of this problem is that without experiments and theories there has been little convergence of opinion on many issues; consequently, one finds a wide range of contradictory advice on how to deal with these issues in constructing graphs. One example is the shape parameter, as we will show in Section 5.2.

3. RESEARCH METHODOLOGY

There are a number of ways that one can probe aspects of graphical perception. However, to make meaningful progress on a particular issue, one almost always has to invoke all of them.

3.1. *Informal Experimentation*

One way to study graphical perception is to change an aspect of a graph and compare the new with the old. Such a process is helpful for building intuition and can often answer questions about the ease of detection, that is, whether it is easier or harder to detect certain behaviour in the data as a result of the change in the graph. Figs 2 and 3 are an example; by changing the shape parameter from 1 to 0.065 we can detect the more rapid rise than fall of the sunspot cycles.

The reason why this informal process works at all is that there is a reasonable amount of uniformity in the human visual system. Since we (the authors) were able to perceive the faster rise in Fig. 2 and not in Fig. 3, it was reasonable to conclude that others would have the same perception. One must, of course, be extremely careful in drawing conclusions by such informal processes, and soliciting the responses of others, even in an informal way, is a good practice.

3.2. *Formal Experimentation*

Informal experimentation can carry us just so far. Many issues, particularly those having to do with the accuracy or efficiency of different methods of visual decoding, need controlled experiments with subjects judging graphical objects and recording responses (Macdonald-Ross, 1977; Wilkinson, 1982; Cleveland and McGill, 1984 and 1985; De Soete and De Corte, 1985; Dickson, DeSanctis, and McBride, 1986). One such experiment is described in Section 5.4. As with any experimental science, there are many pitfalls in experiments in graphical perception. Fortunately, however, expertise already exists that can serve as a guide: experimental techniques from the field of visual perception. The following are two examples.

The first is timed displays. In later sections the perceptual tasks that will be investigated are the rapid visual scans we perform to detect the geometric patterns that lead us to insights about the behaviour of the data; these scans are distinct from more highly cognitive tasks such as scale reading. Thus in experiments that probe these scans, it is important to ensure that subjects make their judgments rapidly without resorting to cognitive judgments. One technique, an old one in vision research, is to show displays for a short time period; for example, in the experiment described in Section 5.4, displays appeared on a computer screen for about $2\frac{1}{4}$ seconds. In experiments in visual perception that probe preattentive vision, displays appear in some cases for as little as 5 ms (Sagi and Julesz, 1985).

A second example is subject motivation. A subject in a typical experiment in graphical perception records a value as a result of perceiving a magnitude on a visual display, for example, the percent that the length of a shorter line is of a longer one. It is a mistake to suppose that what the subject records is what he or she sees. The recording introduces noise. An experimental protocol that provides careful instructions and thorough training to subjects and that motivates them to concentrate on the task can reduce the noise substantially. For example, in the experiment described in Section 5.4, subjects were motivated by conveying a sense of importance of the experiment to them, rewarding them for participation, and introducing a competitive aspect of the tasks in which each subject competed against his or her earlier performance in the experiment.

3.3. *The Theory of Vision*

More general studies of vision have given us more than just techniques of experimentation. The classical theory of visual perception sheds light on processes of graphical perception. An example is Weber's Law, one of the oldest and most basic laws of perception (Baird and Noma, 1978). Let us consider the law in terms of line length. Suppose we are judging the lengths of two lines and the goal is detection, that is, to determine if the lengths are different. Weber's Law states that detection depends only on the ratio of the lengths and not on the overall sizes of the two lines; that is, our ability to detect a difference in two lines of lengths 5 mm and 5.1 mm is the same as that for two lines that are 50 mm and 51 mm. Experimentation has shown that the law is quite accurate except for very small line lengths that are near the limits of what our visual systems can see. Weber's Law suggests that in experiments in graphical perception in which subjects judge the relative lengths of line segments, we need to control only for the ratio of the lengths and not for overall size. One experiment (Cleveland and McGill, 1985) suggests that this deduction is true. Note that if we were forced to control for overall length we would also have to control for the distance of the judged object to the viewer's eyes.

Concepts from the rapidly-evolving fields of cognitive science and computational vision are also relevant to graphical perception. Pinker (1982), Haber and Wilkinson (1982), Kosslyn (1985), Follettie (1986), and Simkin and Hastie (to appear) have employed recent ideas of visual information processing in quite interesting ways to study graphical perception.

3.4. *Paradigms for Graphical Perception*

Another important methodological tool for making progress in graphical perception is the development of paradigms, or frameworks for thinking about the subject. Paradigms are important for graphical perception for the same reasons that they are important for other scientific subjects. They serve as a guide for experimentation; their theoretical components provide predictions that can then be tested, leading to a confirmation or revision of the components. One problem with the small amount of experimentation in graphical perception that had been carried out until a few years go was that it was wholly empirical and unguided by attempts to develop basic concepts. Experiments tended to be of a form in which whole graph forms, such as pie charts and divided bar charts, were compared, rather than breaking these complex structures into simpler pieces, attempting to understand the pieces, and then inferring the properties of the graph forms from an understanding of the pieces and their interactions. Were we to continue only comparing whole graph forms and never developing basic concepts, every issue would have to be resolved by an experiment and deduction, an important tool of science, would not be possible. In the next section we describe a paradigm that helps us to make deductions.

4. A PARADIGM FOR GRAPHICAL PERCEPTION

4.1. *Elementary Codes*

Cleveland and McGill (1984 and 1985) have developed a paradigm for graphical perception that begins with the isolation of elementary codes of graphs, which are listed in Table 1. These are fundamental geometric, colour, and textural aspects that encode the *quantitative* information on a graph. The judgements of the codes make up the rapid processing that we perform to extract information visually about the *relative magnitudes* of quantities shown on the graph. They are not meant to address highly cognitive tasks such as scale reading.

TABLE 1
Elementary codes and an order

Rank	*Code*
1	Positions along a common scale
2	Positions along identical, nonaligned scales
3	Lengths
4	Angles
4–10	Slopes
6	Areas
7	Volumes
8	Densities
9	Colour saturations
10	Colour hues

Fig. 4 illustrates four of the elementary codes. Panels I and II are dot charts (Cleveland, 1985). To compare visually the values in Panel I we make judgments of positions along a common scale. To compare a value in Panel I with a value in Panel II we make judgments of positions on identical but nonaligned scales. Panel III is a divided bar chart. To compare the three values of Item 1 or to compare the totals of groups *A*, *B*, and *C* we can judge positions along a common scale, but to compare any other set of values—for example, the values of Group *A* or the values of Item 2—we must make length judgments. Panel IV is a two-variable graph; the *x* values can be visually decoded by judgements of positions along a common scale; the same is true of the *y* values. The graph also allows us to study the

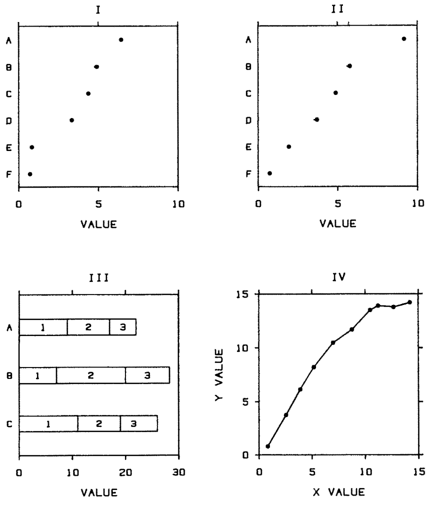

Fig. 4. Elementary judgments of graphical perception. Graphical perception is the visual decoding of information from a graph. To visually extract quantitative information we judge elementary codes. Four are illustrated here. (1) Judgments of positions along a common scale—comparing values in Panel I. (2) Judgments of positions along identical, nonaligned scales—comparing a value on Panel I with one on Panel II. (3) Length judgments—comparing Items 1, 2, and 3 in Group A on Panel III. (4) Slope judgments—comparing the local rates of change of *y* as a function of *x* on Panel IV.

relationship of x and y, for example, how y changes as a function of x. The local rate of change of y as a function of x can be visually decoded by judging the slopes of the line segments connecting successive points; the overall visual impression is that the slope tends to decrease as x increases.

4.2. *A Ranking*

The next aspect of the paradigm is a ranking of the elementary codes based on the accuracy with which we judge them. This was done by using the methodology described in Sections 3.1 to 3.3, and the result is the ordering shown in Table 1.

The ordering needs several qualifications. There has been no formal experimental verification of the codes below the dashed line, so their ranking relative to one another is conjectural. That they rank below the other codes can be convincingly established with informal experiments. The relative ranking of codes above the dashed line is consistent with the theory of vision and has a moderate amount of experimental verification. It was initially established by a series of three experiments (Cleveland and McGill, 1984 and 1985) and by drawing on an experiment of Croxton and Stein (1932) who had established the length-area-volume ordering. Recently, Simkin and Hastie (to appear) have corroborated the position-length-angle ordering. Furthermore, Cleveland, McGill, and McGill (1986) have investigated slope judgments in detail; this work is reported in Section 5.

Note that slope judgments span a range of positions in the ordering. Suppose m_1 and m_2 are the magnitudes of two slopes or of two line lengths or of some other geometric aspects. The order of the elementary codes is an ordering of the accuracy with which we judge m_1/m_2, that is, of the accuracy of *comparative estimation* (Follettie, 1986; Simkin and Hastie, to appear). Now for position, length, area, and volume judgments there is good evidence that accuracy does not further depend on m_1 and m_2 conditional on the ratio, m_1/m_2. (This was discussed for length judgments in Section 3.3.) For angle judgments this is true to a good approximation, although if m_1 or m_2 is 90° or 180°, special considerations are needed. However, for slope judgments, there is, as we shall see in Section 5.4, a strong dependence on m_1 and m_2. For this reason, slope judgments have a range of accuracies. At their most accurate, slope judgments have an accuracy similar to that of angle judgments, and at their worst, slope judgments carry no meaningful quantitative information at all.

4.3. *The Role of the Paradigm in Data Display*

The purpose of the paradigm is to improve data display. By our choices in constructing a graph, we can control, to some extent, the codes that are judged to decode quantitative information. Choosing codes that result in accurate judgments leads to more effective data display (Cleveland, 1985; Cleveland and McGill, 1985). We do not mean to imply that the direct goal of a graph is to show the data to as many decimal places as possible; the direct goal is to see patterns in the data and understand the overall behaviour, but the more accurately the data are visually decoded, the better our chance to detect and properly understand the patterns and behaviour of the data. This will be amply illustrated in Section 5.6.

4.4. *The Elementary Codes as Elementary Particles*

One person's elementary particle can be another's complex system. For most of the field of chemistry, the molecule is the elementary particle, but the particle physicist sees the molecule as made up of a large number of more basic constituents.

The elementary codes of Table 1 are one collection of elementary particles of graphical perception. Their virtue is that *they can be manipulated by the maker of a graph.* However, from the viewpoint of vision theory they are not so elementary. The textons of Julesz (1981) are a more fundamental set of elementary particles of vision theory. The important point is that, like a chemist manipulating molecules to good purpose, the graph maker can manipulate the elementary codes to make better graphs.

5. AN INVESTIGATION OF SHAPE AND SLOPE JUDGMENTS

This section is a brief report on certain studies of shape and its effect on slope judgments. A full description is given by Cleveland, McGill, and McGill (1986). One purpose of the section is to provide some foundation for the more general remarks that have been made in previous sections. Another is to provide a method for choosing the shape parameters of two-variable graphs that are made to see how one variable depends on another.

5.1. *The Effect of Shape on Slope Judgments*

Slope judgments, which give us information about the rate of change of y as a function of x, are an essential ingredient of our graphical perceptions when we study a two-variable graph to see how y depends on x. Suppose s_i for $i = 1$ to n are the *physical slopes* of a collection of line segments on a graph. By *physical slopes* we mean the slopes when the vertical and horizontal coordinates are the physical distances from the left and bottom sides of the data rectangle (see Section 1.1), where both distances are measured in the same units. (Note that if s_i^* are the slopes in the units of the data, then $s_i^* = cs_i$.) Furthermore, we shall suppose that the s_i are the physical slopes when the shape parameter is 1. When the shape is γ, the physical slopes are γs_i and the *physical orientations*, or angles, of the line segments are arctan (γs_i), where we will take the range of arctan to be $-90°$ to $90°$. Note that as γ tends to 0, the physical orientations tend to 0; as γ ends to ∞ the physical orientation of the ith line segment tends to $-90°$ if $s_i < 0$, is always $0°$ if $s_i = 0$, and tends to $90°$ if $s_i > 0$.

It only takes a little informal experimenting with graphing a set of two-variable data to observe that one shape parameter can lead to orientations of line segments that make visually decoding slope information easy and another shape parameter can result in orientations that make it very difficult. An example of this has already been given in Section 1.1.

5.2. *Historical Study*

Shape has received much attention in writings on data display for at least 70 years (*e.g.*, Brinton, 1914). The syndrome described in Section 2.2—lack of consensus and no scientific study—pervades the writings. There appear to be three categories of recommendations. One is to choose the shape parameter to be a fixed ratio, independent of the data; suggested values are $\frac{2}{3}$ (Karsten, 1923), $1/\sqrt{2}$ (American Standards Association, 1938), and $\frac{3}{4}$ (American National Standards Institute, 1979). In a second category of recommendations, authors state that general rules are not possible and that the shape parameter should be determined solely by the data. For example, Meyers (1970) writes: "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." No authors in this category provide guidance on choosing shape.

In a third category of recommendations, authors recognize the phenomenon described in Section 5.1—different shapes lead to different orientations of line segments. With no experimentation, however, there is no consensus on an optimum orientation. Von Huhn (1931) writes: "The writer believes this angle probably is located somewhere between 30 and 45." But Weld (1947) gives a range of 35° to 45°, Hall (1958) states it is 30° to 60°, and Bertin (1967 and 1983) states, with certainty but with no evidence, that it is 70°. In Sections 5.3 and 5.4 we will show how systematic study, both theoretical and experimental, leads to a convincing specification of an optimum orientation.

5.3. *Theory*

It turns out that a theorem about the orientations of line segments with positive slopes points us strongly in the direction of determining an optimum orientation. Fig. 5 illustrates the basic idea of the theorem. In each panel the same two pairs of points are graphed and connected by line segments. The shape parameter is 1 in Panel A, 1/15 in Panel B, and 15 in Panel C. Let s_1 be the smaller slope in the middle panel and let s_2 be the larger. Suppose the

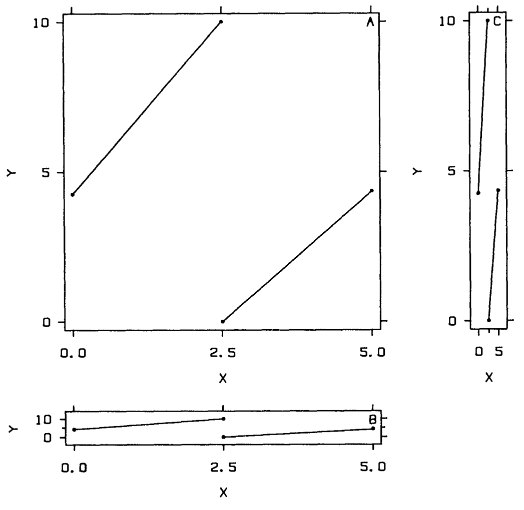

Fig. 5. Orientations of line segments. The same two pairs of points are graphed in each panel. The orientation resolution (difference in angles) of the two segments is greater in Panel *A* than in Panels *B* and *C*. Consequently we can see a slope difference in Panel *A* but not in *B* and *C*. A theorem shows that the orientation resolution of two line segments with positive slopes is maximized when the shape parameter is chosen to make the mid-angle of the two segments equal to 45°.

goal is to judge s_1/s_2. In Panel *B* the slopes are $s_1/15$ and $s_2/15$ and in Panel *C* the slopes are $15s_1$ and $15s_2$, so the ratio, of course, remains the same. The orientations of the line segments are shown in the second and third columns of Table 2. The absolute difference between the orientations of the two segments in each panel is the *orientation resolution*; the resolutions are shown in the fourth column of Table 2.

TABLE 2
Orientations of segments in Fig. 5

Panel	*Left Segment*	*Right Segment*	*Orientation Resolution*	*Mid-Angle*
A	49°	41°	8°	45°
B	4.39°	3.32°	1.07°	3.86°
C	86.68°	85.61°	1.07°	86.14°

We would like the orientation resolution of two segments with positive slopes to be as large as possible. This gives us the best chance of visually resolving differences in the slopes; if the orientation resolution is very small, the segments will appear to be parallel and we will judge the ratio of the physical slopes to be 1. For example, in Fig. 5 the ratios in Panels B and C appear to be 1 because the orientation resolution is so small; because the orientation resolution is greater in Panel A, we can perceive that the slope of the right segment is less. The question then is, when is the orientation resolution maximized? The following theorem shows maximization occurs when the average of the two orientations, which we call the *orientation mid-angle*, is 45°. The mid-angles of the three pairs of line segments in Fig. 5 are given in the fifth column of Table 2.

Maximum Resolution Theorem: Let $s_1 > 0$ and $s_2 > 0$ be the physical slopes of two line segments on a graph when the shape parameter is one. For $i = 1$ and 2 let

$$\theta_i(\gamma) = \arctan(\gamma s_i)$$

be the orientations when the shape parameter is γ. Let

$$r(\gamma) = |\theta_1(\gamma) - \theta_2(\gamma)|$$

be the orientation resolution and let

$$a(\gamma) = \frac{\theta_1(\gamma) + \theta_2(\gamma)}{2}$$

be the orientation mid-angle. Then

(i) $r(\gamma)$ is maximized when $\gamma = \gamma^* = 1/\surd(s_1 s_2)$
(ii) The mid-angle at the maximum is $a(\gamma^*) = 45°$
(iii) For $\delta > 0$, $r(\delta\gamma^*) = r(\delta^{-1}\gamma^*)$
(iv) $r(\delta\gamma^*)$ is monotonically decreasing for $\delta \geqslant 1$
(v) For $\delta > 0$, $a(\delta\gamma^*) = 90° - a(\delta^{-1}\gamma^*)$
(vi) $a(\delta\gamma^*)$ is monotonically increasing for $\delta \geqslant 1$. ■

The theorem, whose proof is straightforward, shows several things in addition to the optimum mid-angle of 45°. As γ increases or decreases away from the optimum value, the mid-angle moves monotonically away from 45° toward 0° or 90° and the orientation resolution monotonically decreases. Furthermore, the value of $r(\gamma)$ when the mid-angle is $\alpha(\gamma)$ is the same as that when the mid-angle is $90° - \alpha(\gamma)$. These facts are illustrated in Fig. 5 and Table 2. In Panel *A* the mid-angle is 45° and the resolution is 8°. In Panel *B* the mid-angle is 3.86° and the resolution is considerably less; in Panel *C* the mid-angle is $86.14° = 90° - 3.86°$ and the resolution is the same as in Panel *B*.

The Maximum Resolution Theorem *suggests* that our judgment of s_1/s_2 is most accurate when the mid-angle is 45° and that it decreases in accuracy as the mid-angle moves away from 45°. This hypothesis was investigated by an experiment that will be summarized in the next section.

5.4. *Experimentation*

An experiment was run to investigate the accuracy of slope judgments. Each *stimulus* that subjects judged consisted of two line segments with positive slopes that appeared on the screen of a microcomputer for $2\frac{1}{4}$ seconds. The segment in the upper left of the screen had the larger slope, and subjects judged what percent the slope of the other line segment was of the larger. Subjects recorded their answers by typing them on the keyboard.

There were 44 distinct stimuli, or line segment pairs. The line segment in the upper left corner had one of 4 orientations—55°, 35.5°, 19.7°, and 10.1°. For each of these orientations there were 11 stimuli for which the true percents to be judged varied from 50 to 100 in steps of 5. Subjects were told that the true percents lay between 50 and 100.

There were 12 replications of the 44 stimuli in the experiment, so each subject made $44 \times 12 = 528$ judgments. Within each replication the order was randomized and each replication had a different random order. Sixteen subjects participated in the experiment, which resulted in $16 \times 528 = 8448$ judgments. Since subjects were given sufficient instructions and training before the recorded trials began, there was no significant learning effect.

The absolute error of each judgment is

$$|\text{true percent} - \text{judged percent}|.$$

For each subject-by-stimulus combination, an average error across replications was computed.

This resulted in 704 mean absolute errors,

$$m_{ijk} \quad \text{for} \quad i=1 \text{ to } 11, \quad j=1 \text{ to } 4, \quad \text{and} \quad k=1 \text{ to } 16.$$

A stimulus mean error, m_{ij}, for each of the 44 stimuli was computed by averaging the m_{ijk} across subjects. That is,

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

Furthermore, an estimate of the variance of m_{ij} was taken to be

$$v_{ij}=\frac{1}{16}\left(\frac{1}{15}\sum_{k=1}^{16}(m_{ijk}-m_{ij})^2\right).$$

The v_{ij} provide information about the sampling variability of the m_{ij} where the 16 subjects are a sample from a homogeneous population of potential subjects: people who are capable of understanding and carrying out the instructions of the experiment.

The mid-angles of the 44 line segment pairs, which we will denote by a_{ij}, range from 7.6° to 55°. The Maximum Resolution Theorem suggests that the m_{ij} should decrease as a_{ij} gets closer to 45°. Previous experiments (Croxton and Stein, 1932; Cleveland and McGill, 1984 and 1985) suggest that the m_{ij} will also depend on the true value of the percent being judged. To be consistent with the previous notation—that is, m_{ij} and a_{ij}—we could denote the true percents of the 44 stimuli by p_{ij}; however, since for each i the four values of p_{ij} are equal, we will denote them by p_i.

The dependence of the m_{ij} on a_{ij} and p_i is shown in Fig. 6. In the i-th panel, m_{ij} is graphed against a_{ij} for $j=1$ to 4, and the value of p_i is shown in the lower left corner of the panel. The error bars portray plus and minus $1.96\sqrt{v_{ij}}$. The figure shows that the hypothesis suggested by the Maximum Resolution Theorem is strongly supported by the data. The general pattern for each panel is a decrease in error as the mid-angle gets closer to 45°; for true percents in the range 50 to 70, the rate of decrease in error is substantial, and as the true percent increases to 100 the rate is reduced to nearly zero. There is another informative pattern in Fig. 6. The 11 minimum values of m_{ij} on the panels are all reasonably close to the same value,

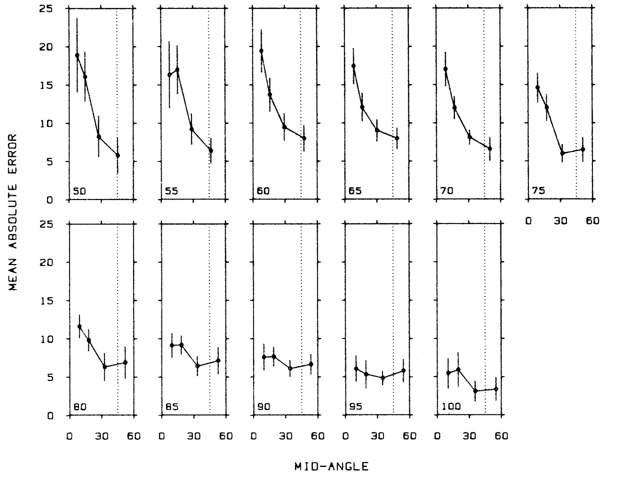

Fig. 6. Results of an experiment. Subjects judged what percent one slope was of another. Average error is graphed against mid-angle for 11 true percents, shown in the lower left corners of the panels. Errors are at a minimum when the mid-angle is in the vicinity of 45°, which is shown by the dotted lines.

approximately 5. This means that when the mid-angle is close to 45°, there is little dependence of error on the true percent. An error of 5 approaches the minimum error that one would expect in an experiment of this type where subjects tend to round answers to multiples of 5.

The dependence of m_{ij} on a_{ij} and p_i observed in Fig. 6 can be described by a simple model:

$$m_{ij} = \mu + \beta(100 - p_i^*) \mid a_{ij} - \theta \mid + \varepsilon_{ij}$$

where μ, β, and θ are unknown parameters and $p_i^* = \min(60, p_i)$. The parameters were estimated by weighted least-squares with weight v_{ij} for the (i, j)th observation. The second column of Table 3 shows the estimates. The sampling variability of the estimates was investigated by the jackknife (Quenouille, 1949; Tukey, 1958) and the bootstrap (Efron, 1982); the 16 subjects were subsampled and the weighted least-squares estimation was carried out for each subsample. Columns three and four of Table 3 show the estimated standard errors. The jackknife and the bootstrap estimates of variability are in good agreement.

TABLE 3
Inferences for parameters in model for absolute error

Parameter	*Estimate*	*Jackknife S.E.*	*Bootstrap S.E.*
θ	41.1	1.09	0.82
μ	4.59	0.53	0.46
β	0.0108	0.00142	0.00127

The model provides an excellent fit to the data; the median absolute residual is 0.62 and the maximum absolute residual is 2.58. A variety of diagnostic tests—including a normal probability plot of $\hat{\varepsilon}_{ij}/\sqrt{v_{ij}}$—showed that weighted least-squares is an appropriate method of estimation and that the model fits the data over the entire region of values of a_{ij} and p_i. The estimated optimum orientation—that is, the orientation that minimizes the error—is 41.1°. The estimated variability of $\hat{\theta}$ puts 45° slightly outside a probable range, however, it is important to point out that the above assessment of variability is conditional on the form of the model and does not take into account the aspects of the model that were chosen in an informal way on the basis of exploring the data. Consequently we will not regard the estimate of θ and its variability as providing strong evidence that the optimum orientation is slightly less than 45° rather than 45° itself. Actually, whether we use 45° or 41.1° is of little practical significance in applications.

5.5. *The Median Absolute Slope Procedure*

Suppose we have n line segments on a graph whose slopes we want to compare. Let s_i for $i = 1$ *to* n be the physical slopes of the segments when the shape parameter is one. The theory and experimentation of the previous two sections suggest that our visual comparison of two positive slopes is most accurate when the mid-angle is 45°. It is quite reasonable to suppose that the same is true of two negative slopes when the mid-angle is $-45°$ because a maximum resolution theorem, with 45° replaced by $-45°$, holds for negative slopes.

We cannot, except in a few very special cases, choose the shape parameter, γ, so that all pairwise comparisons of positive slopes will involve 45° mid-angles and all pairwise comparisons of negative slopes will involve $-45°$ mid-angles. We can, however, attempt to *centre* the orientations of segments with positive slopes on 45° and the orientations of segments with negative slopes on $-45°$ as a way of optimizing the collection of comparisons. One method for doing this, which we call the *median-absolute-slope procedure*, is to select γ so that

$$\text{median} \mid \gamma s_i \mid = 1.$$

The justification for this procedure is discussed in more detail by Cleveland, McGill, and McGill (1986).

5.6. *Applications of the Median-Absolute-Slope Procedure*

The median-absolute-slope procedure has been used in Fig. 2, and $\gamma = 0.065$. The line segments that were the input to the procedure are those connecting successive observations. Note that these are *actual line segments* that appear on the graph since the method of plotting was to connect successive observations by drawn segments. The difficulty of Fig. 3, where $\gamma = 1$, is that the orientations of the segments are too close to $\pm 90°$, a region of low accuracy for slope judgments.

Another example, again a time series, is shown in Fig. 7. The data are average monthly CO_2 concentrations measured at the Mauna Loa Observatory in Hawaii (Keeling, Bacastow, and Whorf, 1982). There is a yearly seasonal oscillation due to the waxing and waning of vegetation in the Northern Hemisphere and a trend due to man-made emissions of CO_2. The purpose in making the graph was to show the trend, which if continued will eventually produce the greenhouse effect (Hansen *et al.*, 1981). One major aspect of the graph that gives us visual information about the trend is the changes in the yearly maxima and the changes in the yearly minima. We judge the rate of change of the minima and the maxima by visually decoding the slopes of the *virtual line segments* that connect the successive peaks and the virtual segments that connect the successive troughs. Thus the shape parameter of the graph was chosen by the median-absolute-slope procedure with these virtual segments as the input. We can see clearly, principally by judging the virtual segments, that the trend is nonlinear; the rate of change of the CO_2 concentrations is increasing.

It is quite common in graphing time series to use a small shape parameter; in one study, the lower quartile for the shape parmeters of time series graphs was 0.30 (Cleveland, McGill, and McGill, 1986). In Fig. 8 the CO_2 concentrations are graphed again, but with the shape parameter equal to 0.3. Now the orientations of the virtual segments are in a range that makes comparing slopes difficult; we cannot as easily visually resolve differences in the slopes with the result that the trend now appears linear.

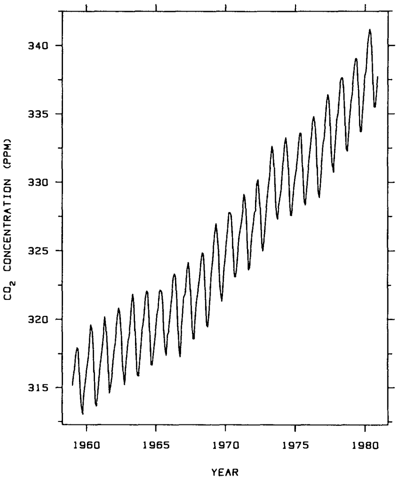

Fig. 7. CO_2 concentrations. By judging the slopes of the virtual line segments connecting peaks and the virtual segments connecting troughs we can see the nonlinearity in the trend. The shape parameter has been chosen so that the median-absolute-slope of the virtual segments is 1. In general, this median-absolute-slope procedure tends to center the orientations of segments with positive slopes on 45° and the orientations of segments with negative slopes on $-45°$.

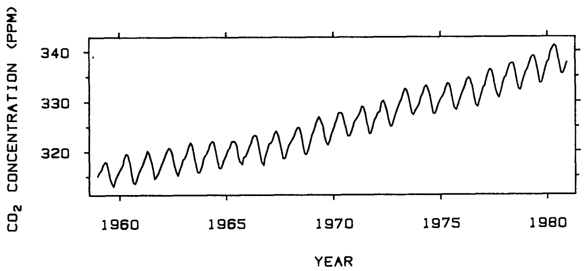

Fig. 8. CO_2 concentrations. The shape parameter is less than the optimal value in Fig. 7 and now the nonlinearity is more difficult to see because the slope judgements are less accurate.

The median-absolute-slope procedure can be used, of course, for data other than time series. For example, it is used in Figs 1 and 6. In Fig. 1 the line segments that serve as input to the procedure are the actual line segments that appear on the graph. In Fig. 6 the 11 panels have the same scales. The line segments used as input to the procedure are the actual segments connecting successive data values. The horizontal and vertical scales and the shape parameter were chosen as if all data were to be plotted on the same panel, and then the 11 data sets were graphed separately.

5.7. *Problems with the Median-Absolute-Slope Procedure and Some Solutions*

The median-absolute-slope procedure typically works quite well but it can fail; in attempting to compromise between centering positive slopes on 45° and negative slopes on −45° it can, of course, fall to do either. Also, it may well be that there is more than one interesting collection of slopes on a graph, and the γ that optimizes one collection might be quite different from that which optimizes another. One solution is to make more than one graph. The sunspot numbers in Figs 2 and 3, again, provide an example. The change in the peak values of the cycles through time is an important aspect of the data. However, it is difficult to assess this change in Fig. 2 because the orientations of the virtual segments connecting successive peaks are too close to 0°, a region of low accuracy. (Recall that the shape parameter of Fig. 2 was chosen to optimize the orientations of the actual segments connecting successive data values.) When the peak-to-peak segments are the input to the median-absolute-slope procedure, the result is Fig. 9, where $\gamma = 0.35$; the peak-to-peak change can be much more readily assessed than in Fig. 2.

Another problem with the median-absolute-slope procedure is that in some applications the optimal γ can be so small or so large that resolution in the horizontal or vertical direction

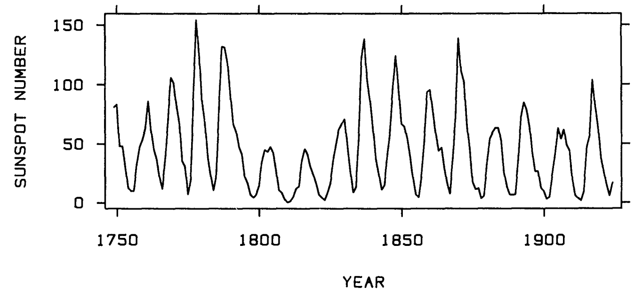

Fig. 9. Sunspot numbers. In the graph of the sunspot numbers in Fig. 2 the median-absolute-slope procedure was used to choose the shape parameter with the actual line segments connecting successive observations serving as the input. In this graph the procedure was used with the virtual segments connecting the successive peaks as the input.

is lost because the size of plotting symbols and the widths of drawn line segments are not infinitesimal. For example, for a long time series, γ is typically small when the input to the procedure is line segments connecting successive observations; if the series is stationary, bounded, and measured at equally spaced points in time, the optimal γ falls off like $O_p(n^{-1})$ where n is the number of observations (Cleveland, McGill, and McGill, 1986).

The general solution for a very small or very large γ is to break up the data into subsets, graph each subset on a separate panel, and use an appropriate coordination of the scales on the different panels. This will be illustrated by the winning times in four track events—100 m, 200 m, 400 m, and 800 m—in the 19 Olympic competitions from 1896 to 1984. The times for the first 17 competitions were analyzed in a paper by Chatterjee and Chatterjee (1982) and discussed further in several letters to the editor. Fig. 10, whose shape parameter has been set equal to 1, shows the problem in graphing these data; the decrease in the log times for each event is small compared with the differences between events so that when the shape parameter is 1, the orientations of line segments connecting successive observations are too close to 0° for slope judgments to be effective. Essentially, the graph provides information only about how running time varies with event and not about how it varies through time for each event. The median-absolute-slope procedure, with the line segments connecting successive observations as input, results in a γ of 284, which is much too large to be practical.

One solution to the problem of graphing the log running times is shown in Fig. 11. Each series is graphed on a separate panel together with a regression line fit to the data of the panel by robust regression with the Huber weight function (Huber, 1973). The horizontal scales of the four panels are the same. The vertical scales are not the same but they have a important property: the number of data units per cm is the same, which means that changes, slopes in particular, can be compared from one panel to the next. The common shape parameter of the four panels has been chosen by the median-absolute-slope procedure with the four regression lines as input. In effect, what we have done is to take four horizontal slices of Fig. 10, each with the same height and each containing one series, and to expand the vertical scales of the slices by the same amount and contract the horizontal scales by the same amount.

The manner in which Fig. 11 has been constructed gives us much insight into the structure of the data and into problems that arose in their analysis. First, the trend in each series of log times appears linear and the slopes of the 4 regression lines are similar. Thus, if w_{it} is the

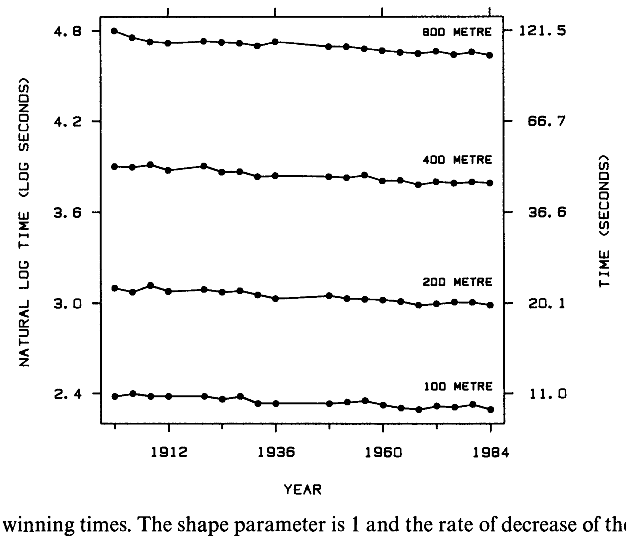

Fig. 10. Olympic winning times. The shape parameter is 1 and the rate of decrease of the log times cannot be accurately decoded.

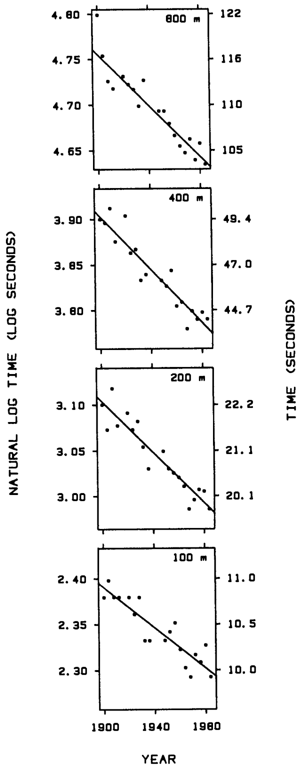

Fig. 11. Olympic winning times. The number of data units per cm is the same on the vertical scales of all panels so we can graphically compare slopes on different panels. A robust regression line is graphed on each panel. The median-absolute-slope procedure with the regression lines as input was used to choose the shape parameter.

winning time (in seconds) for the ith event at time t (in years), the model,

$$\log_e(\mathrm{w}_{it}) = \mu_i + \beta t + \varepsilon_{it},$$

provides a reasonable, rough description of the data. Note that the vertical scales are natural log time; since the range of the log times on each panel is at most about 0.15, we can interpret a change of c in the log times on a particular panel as a $100c$% change in the times themselves. Thus, since the estimate of β in the above model, again using robust regression, is -0.0013, the winning times for each race tend to decrease by about $\frac{1}{2}$% from one Olympic event to the next. The small change in the log times also means that for each series, taking logarithms is very nearly a linear transformation and the trend in w_{it} through time is also linear. That is, letting $\alpha_i = \exp(\mu_i)$, the trend for w_{it} is nearly

$$\exp(\mu_i + \beta t) \approx \alpha_i + \beta\alpha_i t.$$

Thus the advantage of logarithms is that the rate of decrease in the trend of $\log(w_{it})$ does not appreciably depend on i.

A data analyst who carried out an initial data analysis (Chatfield, 1985) by making Fig. 11 and who made the above observations would be justified in being highly pessimistic about reliably estimating curvature in the w_{it}. In fact, Wooton and Royston (1983) and Mayer (1983) have demonstrated that the attempt of Chatterjee and Chatterjee (1982) did not succeed. Another graph employing the median-absolute-slope procedure shows that a modelling attempt of Dewey (1984) was also unsuccessful. Dewey argues that the speeds, $v_{it} = d_i/w_{it}$,

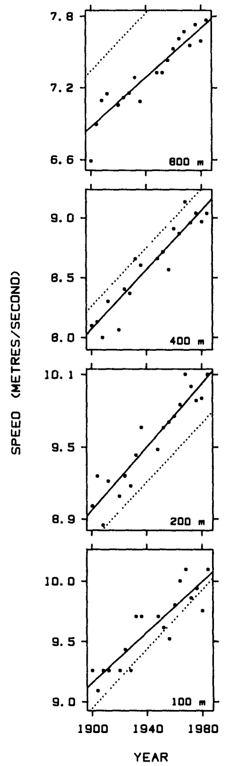

Fig. 12. Olympic speeds. The distances divided by the winning times are graphed in the same manner as Fig. 11. The dotted lines portray a model fit to the speeds by Dewey.

where d_i is the distance in meters of the i-th event, are linear in both time and distance, and fit the model

$$E(v_{it}) = -14.096 + .01225t - 0.0023d_i.$$

Fig. 12 shows the speeds, graphed in the same manner as the log times in Fig. 11; that is, the number of data units per cm is the same on the vertical scales, the solid line on each panel is a robust regression line fit to the data of the panel, and the common shape parameter of the four panels was chosen using the median-absolute-slope procedure with the regression lines as the input. The dotted lines portray the model fit by Dewey. The lack of fit occurs because the Olympic speeds, like the world record speeds in Figure 1, are not linear in distance.

The important point, for our purposes in this paper, is that the analysis of the running times from Figs 11 and 12 depends heavily on making slope judgments and that these judgments are enhanced by the choice of the scales and the shape parameter.

6. GRAPHICAL PERCEPTION: WHICH DISCIPLINE?

The thesis in the previous sections has been that to put graph construction on a scientific foundation we must study graphical perception. We will conclude with another thesis: statisticians, as well as researchers in visual perception, should be involved in this work in graphical perception; interdisciplinary studies are likely to lead to far more than if the research is contained within one field or the other.

As we have argued earlier, aspects of the general theory of visual perception are highly relevant to graphical perception. However, it is equally important to point out what vision theory cannot do. Equipped with theories of vision one can shed light on many issues of graphical perception but cannot unequivocally settle very many of them. The reason is that the general theories do not focus sufficiently incisively on issues that are relevant to graphical perception. Thus experimentation in graphical perception that focuses on these issues is needed. As we have also argued, experimental techniques from vision research are important for this needed experimentation in graphical perception; invoking them can greatly increase the chances of a successful experiment.

However, the input of ideas from research in graphical methodology is equally important. There are two reasons. Vision theory and experimental tecniques are tools, but the subject under study is statistical graphics. Thus intuition flowing from work in data display is an important input for developing paradigms for graphical perception. A second reason has to do with goals. For studies in graphical perception to be relevant for the statistician's need of improving data display, they must be targeted at factors over which the graph maker has control, even if these factors are not interesting objects for a vision theorist to study. Statistical computing is a good analogy. Developments, both conceptual and technological, in computer science and engineering are important for statistical computing. However, we cannot expect that computer scientists will develop systems tailored to the precise needs of statistics. Statisticians must be involved. The same is true of graphical perception.

REFERENCES

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

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

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

Barnett, V., editor (1981) *Interpreting Multivariate Data*. Chichester: Wiley.

Becker, R. A. and Chambers, J. M. (1984) S: An Interactive Environment for Data Analysis and Graphics. Monterey, California: Wadsworth.

Becker, R. A. and Cleveland, W. S. (1987) Brushing scatterplots. *Technometrics*, **29**, 127–142.

Becker, R. A., Cleveland, W. S. and Wilks, A. R. (1936) Dynamic graphics for data analysis. *Statistical Science*, to appear.

Bentley, J. L. and Kernighan, B. W. (1986) GRAP—A language for typesetting graphs. *Commun. Ass. Comput. Mach.*, **29**, 782–792.

Bertin, J. (1967) *Semiologie Graphique*. Paris: Gauthier-Villars. Paris-La Haye: Mouton et Cie.

Bertin, J. (1983) *Seminology of Graphics*. Madison, Wisconsin: University of Wisconsin Press. English translation of *Semiologie Graphique*.

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

Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A. (1983) *Graphical Methods for Data Analysis*. Monterey, California: Wadsworth.

Chatfield, C. (1985) The initial examination of data. *J. R. Statist. Soc.* A, **31**, 214–253.

Chatterjee, S. and Chatterjee, S. (1982) New lamps for old: an exploratory analysis of running times in Olympic games. *Appl. Statist.*, **31**, 14–22.

Cleveland, W. S. (1985) *The Elements of Graphing Data*. Monterey, California: Wadsworth.

Cleveland, W. S. and McGill, R. (1984) Graphical perception: theory, experimentation, and application to the development of graphical methods. *J. Amer. Statist. Ass.*, **79**, 531–554.

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

Cleveland, W. S., McGill, M. E. and McGill, R. (1986) The shape parameter of a two-variable graph. *Statistics Research Report No. 31*. Murray Hill, New Jersey: AT&T Bell Laboratories.

Costigan-Eaves, P. J. (1984) Data graphics in the 20th century: a comparative and analytic survey. Doctoral dissertation, Rutgers University, University Microfilms No. 8424036.

Cox, D. R. (1978) Some remarks on the role in statistics of graphical methods. *Appl. Statist.*, **27**, 4–9.

Croxton, F. E. and Stein, H. (1932) Graphic comparisons by bars, squares, circles, and cubes. *J. Amer. Statist. Ass.*, **27**, 54–60.

De Soete, G. and De Corte, W. (1985) On the perceptual salience of features of Chernoff faces representing multivariate data. *Appl. Psychol. Measur.*, **9**, 275–280.

Dewey, M. E. (1984) Letter to the editor. *Appl. Statist.*, **33**, 75–76.

Dickson, G. W., De Sanctis, G. and McBride, D. J. (1986) Understanding the effectiveness of computer graphics for decision support: a cumulative experimental approach. *Commun., Ass. Comput. Mach.*, **29**, 40–47.

Donoho, A. W., Donoho, D. L. and Gasko, M. (1985) *MACSPIN-TM Graphical Data Analysis Software*. Austin, Texas: D^2 Software.

Efron, B. (1982) *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia: Society for Industrial and Applied Mathematics.
Follettie, J. F. (1986) User tasks of statistical graph-using and analytic microtasks of graphics research. Unpublished manuscript.
Haber, R. N. and Wilkinson, L. (1982) Perceptual component of computer displays. *IEEE Computer Graphics and its Appl.*, **20**, 23–35.
Hall, A. S. (1958) *The Construction of Graphs and Charts*. London: Pitman.
Hansen, J., Johnson, D., Lacis, A., Lebedeff, S., Lee, P., Rind, D. and Russell, G. (1981) Climate impact of increasing atmospheric carbon dioxide. *Science*, **213**, 957–966.
Huber, P. J. (1973) Robust regression: asymptotics, conjectures, and Monte Carlo. *Ann. Statist.*, **1**, 799–821.
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.
Keeling, C. D., Bacastow, R. B. and Whorf, T. P. (1982) Measurements of the concentration of carbon dioxide at Mauna Loa Observatory, Hawaii, *Carbon Dioxide Review: 1982*, 377–385. (W. C. Clark, ed.) New York: Oxford University Press.
Kosslyn, S. M. (1985) Graphics and human information processing. *J. Amer. Statist. Ass.*, **80**, 499–512.
Kruskal, W. H. (1982) Criteria for judging statistical graphics. *Utilitas Mathematics, Ser. B*, **21B**, 283–310.
Macdonald-Ross, M. (1977) How numbers are shown. *AV Communication Rev.*, **25**, 359–409.
Mayer, M. J. (1983) Letter to the editor. *Appl. Statist.*, **32**, 313–314.
McDonald, J. A. and Pedersen, J. (1985a) Computing environments for data analysis I. Introduction. *SIAM J. Sci. Stat. Comput.*, **6**, 1004–1012.
McDonald, J. A. and Pedersen, J. (1985b) Computing environments for data analysis II. Hardware. *SIAM J. Sci. Stat. Comput.*, **6**, 1013–1021.
Meyers, C. H. (1970) *Handbook of Basic Graphs: A Modern Approach*. Belmont, California: Dickenson.
Nelson, W. (1979) *How to Analyze Data with Simple Plots*. Milwaukee, Wisconsin: American Society for Quality Control.
Pinker, S. (1982) A theory of graphic comprehension. *Occasional Paper No. 15*. Cambridge, Massachusetts: MIT Center for Cognitive Sciences.
Quenouille, M. (1949) Approximate test of correlation in time series. *J. R. Statist. Soc.* B, **11**, 18–84.
Sagi, D. and Julesz, B. (1985) Detection versus discrimination of visual orientation. *Perception*, **14**, 619–628.
Simkin, D. and Hastie, R. Towards an information processing view of graph perception. *J. Amer. Statist. Ass.*, (to appear).
Tufte, E. R. (1983) *The Visual Display of Quantitative Information*. Cheshire, Connecticut: Graphics Press.
Tukey, J. W. (1958) Bias and confidence in not quite large samples. *Ann. Math. Statist.*, **29**, 614.
Tukey, J. W. (1977) *Exploratory Data Analysis*. Reading, Massachusetts: Addison-Wesley.
von Huhn, R. (1931) A trigonometrical method for computing the scales of statistical charts to improve visualization. *J. Amer. Statist. Ass.*, **26**, 319–324.
Weld, W. E. (1947) *How to Chart Facts from Figures with Graphs*. Norwood, Massachusetts: Codex Book Company.
Wilkinson, L. (1982) An experimental evaluation of multivariate graphical point representations. *Proceedings: Human Factors in Computer Systems*. New York: ACM.
Wooton, R. and Royston, J. P. (1983) Letter to the editor. *Appl. Statist.*, **32**, 88–89.
Yule, G. U. (1927) On the method of investigating periodicity in disturbed series, with special reference to Wölfer's sunspot numbers. *Phil. Trans. Roy. Soc.* A, **226**, 267–298.

DISCUSSION OF THE PAPER BY DRS CLEVELAND AND MCGILL

Dr C. Chatfield (University of Bath): It is my privilege and pleasure to welcome Bill Cleveland to this country and thank the authors for bringing this important topic to the attention of the Society. We all use graphs regularly, and yet the general principles underlying their construction are only partially understood.

It is customary for discussants at read papers to concentrate on areas of disagreement. This is rather difficult as I agree with most of the paper. My main quibble is that the paper perhaps has a tendency to use rather too much jargon. For example the title of the paper might usefully be shortened to "Understanding Graphs" and the summary to "Graph construction should have a proper scientific basis. Some theoretical and experimental work is described and the two-variable graph is used as the main example".

So, rather than criticize the paper, I would like to discuss the general principles of graph construction in a broader context. A recent summary is given by Chapman (1986). Some apparently simple rules are as follows:

(1) Graphs should have a clear self-explanatory title. Units of measurement should be stated. All axes should be labelled.

(2) The scales on each axis need to be carefully chosen so as to avoid distorting or suppressing information. Use scale breaks for false origins.
(3) The mode of presentation needs to be chosen carefully. This includes the plotting symbol and the method of connecting points. An example illustrating the dramatic differences which can appear in different graphs of the same data is given by Chatfield (1977, Figure 1).
(4) Trial-and-error can be used to improve a graph. This, I suppose, is what the authors call "informal experimentation".

Do we need any further guidance? Surely the answer is "yes". In particular, rules (2) and (3) above are not straightforward to implement in practice. It is therefore surprising that I cannot recall a paper specifically on graphs in the Society's journals since Mahon (1977) some ten years ago. At the same meeting Ehrenberg (1977) read a paper to the Society on the kindred topic of tables, and how their presentation might be improved. Afterwards I heard various informal views on these papers ranging from "these topics are trivial" to "How refreshing to have intelligible papers on important practical topics". A similar range of views may also apply to today's paper so let me make clear that, in my view, these topics are important and they are not trivial. Ten statisticians will probably plot the same data in ten different ways and they can't all be "right". To reinforce this remark, I looked in the journals to find examples of poorly presented graphs and tables. Unfortunately, this was all too easy to do as I illustrated in my oral presentation, but I will not name names in this written version. However, these examples of unintelligible or incomplete graphs and tables indicate how important these matters are.

The authors have made a useful start in trying to give graph construction a more scientific foundation. In particular, as a "time-series person", I was interested in the suggestion that the choice of scales for a time plot depends on whether one wants to study any cyclic or seasonal variation, or if one wants to look at the trend in peak values. In summary, it is clear that I enjoyed today's paper and have very great pleasure in proposing the vote of thanks.

Dr Allan Seheult (University of Durham): I should like to add my welcome to Bill Cleveland and congratulate the authors on a very thoughtful and timely paper on graphical perception, an area which should be an important focus for graphics research. They rightly stress that together statisticians and psychologists can, and should, play a key role in the development of graphics for presentation of quantitative information. It is to be hoped that statisticians will initiate and support such collaboration. The ever-increasing availability of inexpensive fast computer graphics facilities for micros, minis and mainframes makes it imperative that statisticians should have a strong and active influence in graphics research and development.

However, it is a difficult area in which to do research: unfortunately, some will say it is a subject not worthy of serious study by statisticians. The authors of tonight's paper should therefore be applauded for suggesting a sensible and proper scientific framework within which to conduct research on graphical perception. Moreover, they rigorously follow their own recipe in the case of the shape parameter (or aspect ratio) for a two-variable display. One might say that the paper has everything: paradigms, theory, an experiment and its analysis to test the theory, and a straightforward, sensible algorithm (including a discussion of its shortcomings) to implement the theory on some substantive illustrative examples.

Cynics might suggest that the paper is a dressed-up display to exhibit a simple algorithm for solving an uninteresting problem. I believe such an opinion would be wholly wrong and very short-sighted: even within the narrow confines of the shape parameter of a two-variable graph the historical perspective given in the paper clearly indicates that this apparently simple problem has a long history during which no well-founded recommendation has emerged; until tonight, that is. However, to be convincing the framework should be widely applicable, and perhaps the authors in their reply (to the Discussion) might indicate other specific problems to which their approach might be successfully applied.

No doubt there will be some criticisms of the experiment described in the paper. My question is simply: is the conduct of the experiment wholly relevant to the intended use of the graphical display? I presume that the intended use is to aid informed textual or verbal commentary. Consider, for example, the statement "you will see from the graph that sunspot numbers rise more rapidly than they fall". What is required here is not that the reader or listener can make a judgement in a few seconds but rather that the graph accurately and efficiently exemplifies the statement. In practice the reader or listener will have much more than $2\frac{1}{4}$ seconds to contemplate and test the statement.

I will conclude with some comments and observations on the shape parameter. The median-absolute-slope algorithm has an interesting resistance-to-outliers property. Consider the special case of successive slopes; that is, the collection

$$s_i = \frac{y_{1+1} - y_i}{x_{i+1} - x_i},$$

where $x_{i+1} > x_i$ for $i = 1, \ldots, n$. According to the median-absolute-slope criterion the optimum shape parameter is then

$$\frac{\text{Range } y}{\text{Range } x} \text{ median} \left| \frac{x_{i+1} - x_i}{y_{i+1} - y_i} \right|.$$

A few outliers in x or y will not substantially affect the median but will, of course, affect the ranges. Thus, for example, outliers in y will increase the shape parameter of the data rectangle. However, the *effective* shape parameter for the outlier-free part of the data will be as required. The increased vertical aspect is there only to accommodate the outliers. Of course, there is a page-size restriction and the limits of Weber's law will apply to the shape parameter itself!

The notion that different shape parameters will emphasise different aspects of the data is a simple but an important idea to emerge from the paper, one which raises some questions. Space restrictions may force a compromise between two shape parameters γ_1 and γ_2; for example, Figs 2 and 9 of the sunspot data. In such cases, one possibility might be to use the geometric mean $\gamma = (\gamma_1 \gamma_2)^{\frac{1}{2}}$: in the example $\gamma = (.065 \times .35)^{\frac{1}{2}} = 0.15$. This compromise has the appealing property that $\gamma/\gamma_2 = \gamma_1/\gamma$, so that if, for example, the smaller shape parameter is doubled then the larger one will be halved. The selection of a particular set of slopes presupposes some insight into the data. In an exploratory mode it would be useful to try and detect sets of interesting slopes. Thus, for a time series, especially where there are seasonal or cyclical effects, it may be useful to examine, for a reasonable collection of lags l, the collection of shape parameters γ_l determined by the sets of slopes $(y_{i+l} - y_i)/l$, $i = 1, \ldots, n - l$. Any modes in the collection $\{\gamma_l\}$ may suggest interesting shape parameter values to consider.

However, perhaps such considerations will soon be rendered obsolete by forthcoming developments in dynamic computer graphics at AT&T Bell Laboratories, the University of Washington, and elsewhere. Nevertheless, this is a very welcome paper, and one which should help to stimulate further research. It is a pleasure for me to second the vote of thanks.

The vote of thanks was passed by acclamation.

Mr J. C. Gower (Rothamsted Experimental Station): How people perceive graphs and charts has not received the attention it deserves and I am glad that the authors have brought the topic before us tonight. Much depends on who the graphs are intended for – other statisticians, or the general public. Learning and training as well as visual perception play a role in interpretation; am I the only statistician who found the Andrews Plot a little less than immediately helpful?

Table 1 includes the 'code' *length* which implies, to me, the joining of two points, but omits *distance* which implies a relationship between two unjoined points (*length* seems to be synonymous with the term *actual line-segment* and *distance* with *virtual segment* used in Section 5.6).

These comments serve as an introduction to a good example of where background knowledge is vital in interpreting a graph. With ordination diagrams the natural thing is to use distance as the basis for interpretation. Often this is correct but sometimes (e.g. certain aspects of correspondence analysis) the inner-product is the proper interpretive tool. The locus of points equidistant from a fixed point is a circle and hence with distance-interpretations equal scaling of axes is obligatory and there is no freedom to select any shape parameter. With inner-products the corresponding locus is a straight line and although the inner-products themselves are affected by scaling, linear aspects are not. Both modes of interpretation are invariant to rotation of axes, so as well as the usual statistical criteria for deciding on optimal rotation, new visual criteria might also be worth considering. Criteria for optimal linearization too – but one would have to be careful not to encourage the well-known propensity for researchers to detect false linearities; indeed the whole field of visual false-perception needs exploring. I suppose that statisticians might be expected to be aware of these interpretive variants of what might seem to be the same kind of graphic display but I doubt if they all are and I am quite certain that few non-statistical research workers and even fewer members of the general public have any inkling of the problem. Not all graphs that look alike are alike, and I believe this is true of the x-y plots discussed in the paper. A further implication is that general-purpose automatic graphical scaling algorithms must be intelligent enough to decide the proper mode of interpretation for every graph plotted. I would be interested to hear of any work that has addressed the roles of education and learning in using graphs.

Dr D. J. Hand (London University Institute of Psychiatry): I suppose a topic such as tonight's, which is markedly less axiomatic and more subjective than much of statistics, invites discussion and disagreement. The authors are to be commended for having the courage to present their ideas and for starting this debate about an area which is clearly of fundamental importance to practical statistics – and which promises to become more so as computer graphics come into wider use.

Elementary statistics texts sometimes provide humorous representations of how advertisers can seek to deceive by means of subtle distortions of their graphs and diagrams. Perhaps now the tables are turned, for there must be a sizeable body of knowledge amongst graphics designers in art schools and advertising agencies on how to produce graphs which effectively convey information. I wonder if the authors have explored this.

I appreciate the more general nature of the paper, but I would like to make a few comments about simple slope comparisons between two lines. In Fig. 5 it is true that one can perceive in *A* that the orientations of the two line segments are different, but it is surely even more striking that the slopes are so similar. Surely the real question one asks when one looks at such a figure is "Can the slopes be treated as effectively equal?" Perhaps a more pertinent display would be one where *A* had very different slopes and *B* and *C* showed how much distortion had to be introduced in order for subjects to be uncertain that they differed.

I also wonder how many subjects in the experiments of section 5.4 attributed apparent slope differences, in the cases where the real differences were very small, to display distortion on the screen.

As I am certain the authors will agree, the figure should be chosen for its purpose. If detecting a small slope difference between two lines is of paramount importance then a diagram like Fig. 5.A is inappropriate. Instead the two lines should be superimposed as far as possible. I recognise, however, that the authors were not suggesting that their Fig. 5 was the ideal display for slope comparisons of two lines.

Mr G. J. S. Ross (Rothamsted Experimental Station): While the authors are to be thanked for raising, in an elementary way, the subject of graphical interpretation, they have omitted many topics which concern statisticians. One such topic is the recognition of families of response curves by analysis of changes of slope perceived in a scatter diagram. The authors have used such judgments, in commenting on the asymmetry of form in the sunspot data. In practical consulting statisticians are asked to suggest suitable curves to fit to data, which they attempt to do by looking for evidence of asymptotes, maxima or minima, inflexions or asymmetry, as well as trying to incorporate any relevant knowledge about the data, including the likely consequences of extrapolation.

Some years ago I performed a few simple experiments on my colleagues by presenting them with identical scatter plots on a blank sheet of paper and asking them to fit by eye a suitable smooth curve to each, and to indicate certain features such as the tangent at a point or the position of a maximum. The fitted values supplied by each individual were then analysed together with the least squares fits of various standard curves, by principal component analysis. The main results were as follows:

(1) Different statisticians see different curves in the same data
(2) Empirical curves resemble theoretical curves with more rather than fewer parameters (i.e. freehand curves give a better fit by following the data).
(3) Standard errors of fitted values, slopes and optima are similar to those obtained from least squares fits of suitable theoretical curves.

A more difficult visual task is the analysis of contour plots. The oval contours surrounding the optimum of a log likelihood function in two dimensions may appear to be ellipses, but we have little or no training in distinguishing different forms of ovals. Departures from quadratic form may have serious consequences for inference, so it is important to be able to measure the discrepancy in some way. A direct approach (which illustrates a well known principle in graphical methods: remove the obvious so that the less obvious can be revealed) is to subtract the fitted quadratic approximation and plot the contours of the remainder function. Two such plots are illustrated in Figs. D1 and D2. They correspond to different parameterisations of the same model fitted to the same data. In Fig. D1 the likelihood contours (broken lines) are slightly banana-shaped and the cubic remainder function shows six zones, alternately positive and negative. In Fig. D2 the contours are more nearly ellipses, and the remainder function is much flatter, revealing only a slight asymmetry in that the contours are closer together on the right hand side than on the left hand side of the diagram.

The message is that good graphics convey a clear interpretation while poor graphics impose too great a task on the viewer.

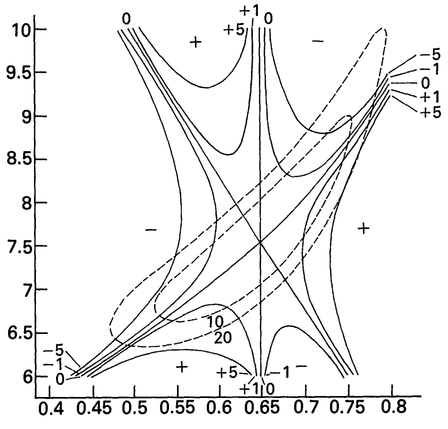

Fig. D1 Likelihood contours (– – – –) for two-parameter model, and remainder-function contours (————) after subtraction of quadratic approximation.

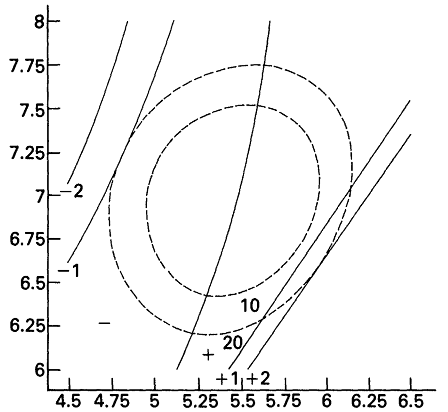

Fig. D2 The same model as in Fig. D1 with a different parameterisation.

Dr C. Jennison (University of Bath): I would like to join the other discussants in congratulating the authors on today's paper. My comments concern the experiment reported in Section 5.4. One notable feature is the absence of lines with slope greater than 55°, with the consequence that the diagrams in Fig. 6 are not as conclusive as they might have been. Although the proposed model fits the data well, so do other quite different models, indeed, it would not require great ingenuity to produce a plausible model with an optimal mid-angle of 70°, in line with Bertin's suggestion. Clearly, data points for larger values of the mid-angle would resolve this problem. An issue raised elsewhere in the paper is the subject's ability to judge which of two lines has the greater slope. The experiment does not tackle this question directly as subjects are told that the left hand line is at least as steep as the right hand line. Would it have been better to randomize the positions of the two lines and ask subjects to state which line had the greater slope as well as estimate the ratio of the slopes?

Experiments in graphical perception are relatively novel but some other studies have been reported in the statistical literature. Cleveland, Diaconis and McGill (1982) studied the effect of scale transformations on perceived correlation and concluded that subjects tend to rate a small plot as more "associated" than a large plot of the same data. However, Ehrenson (1982) noted that the smaller point cloud fell around the line $y = x$ and the observed results could be explained by subjects' unconciously adding the origin as a fixed point. Mosteller *et al.* (1985) describe an experiment in which subjects were asked to fit straight lines by eye to four data sets. One example contains an influential high leverage point which lies so close to an axis that it might easily be overlooked, this is also the only example with a negative slope and the confounding of all these factors can not be conducive to strong conclusions. My point is that the usual rules of good experimental design apply to studies conducted by statisticians too! No doubt, experience will produce a list of common pitfalls to be avoided and we can look forward to a greater understanding of graphical perception as future studies are conducted.

Dr D. A. Preece (Institute of Horticultural Research, East Malling): Western man reads from left to right and from top to bottom, and this affects our perception of pictures. Great artists of Western civilisation have testified to this in their paintings, with the composition often carrying the viewer's eye down a path close to what mathematicians might call the main diagonal of the picture. Also, as advertisements on financial pages of newspapers remind us, a graph that zigzags from top left to bottom right is *bad*, whereas one that zigzags from bottom left to top right is *good*. I conclude that there is not perceptual symmetry between positive and negative slopes. This conclusion seems to be confirmed by the square part of Fig. 5, when I view it either just as it is printed or after it has been rotated through 90°, 180° or 270°. To me, the angle between the two lines in the diagram seems greatest when Fig. 5 is viewed as printed. (I have tested this only whilst sober and whilst wearing spectacles that were intended to correct for my slight astigmatism.) I do not know where all this fits into the authors' story; I invite their comments.

Dr Peter Clifford (Oxford University): This is simply to remark that the effect of applying linear transformations to plotted data can be assessed by the expedient of tilting the paper away from the eye, possibly after rotation. I believe that I was taught this technique by Professor T. Lewis some 20 years ago. The trick is also helpful in revealing curvature in apparently linearly related data. Here the paper must be rotated so that the line of increase lies along the line of sight.

Dr A. J. Lawrance (University of Birmingham): Drs Cleveland and McGill have illustrated their paper with Yule's sunspots data, perhaps the most beloved numbers of time series analysts. The asymmetry property they refer to in Fig. 2 is the clearest depiction of the time irreversibility (directionality) of this series I have seen; it is cogent argument for the thoughtful use of graphics in time series. In this case there are also theoretical implications, because Gaussian linear autoregressive processes are reversible, Weiss (1975). Thus, if Yule in 1927 had realized this, and seen Fig. 2, today's world of time series might have been different.

The following contributions were received in writing, after the meeting.

Dr Andreas Buja (University of Washington and Bell Communications Research): The work of the authors on graphical perception deserves our admiration: it has evolved over the years into a substantial body of knowledge with great impact on how we think about the graphical representation of data. In developing new graphical methods we have often been surprised by the behaviour of human visual perception. Especially memorable was to see viewers of a graphics demonstration fail to recognize two plots (Fig. D3) of the same data in different scaling. In the demonstration the plots were transformed into each other in real-time motion by simultaneously stretching the horizontal axis and shrinking the vertical axis, resulting in a smooth deformation of the plot. The viewers could convince themselves that the motion changed only the scaling. Nevertheless, the subjective experience is that from a certain point on one is looking at a different object, and it takes a short moment of rationalization before one accepts that the two plots show indeed "the same object". The authors of the paper to be discussed here deal mainly with the problem of "how well" we can perceive certain aspects of data in different scalings of the axes, but the effect of scaling can be much more dramatic: not only do "aspects" change, but the whole object seems to change its identity. Interestingly, the possibility of using scale distortion for artistic purposes has been known to painters in the Renaissance: Hans Holbein's painting "The Ambassadors" shows in the foreground a strange elongated object which turns out to be a skull if viewed at a very flat angle to undo the distortion of scale.

Returning to our example of the effect of scale on perception, we mention that the seemingly "wrong" scaling which conceals the periodicity of the data actually lead us to an insight: the peak values seem separated from the bulk of the series. As it turns out, this makes sense since the vertical axis represents monthly averages of river flow of a river in Oregon. The upshot of this is twofold: as Cleveland and McGill state (beginning of Section 5.7), there is often more than one interesting scaling; second, there

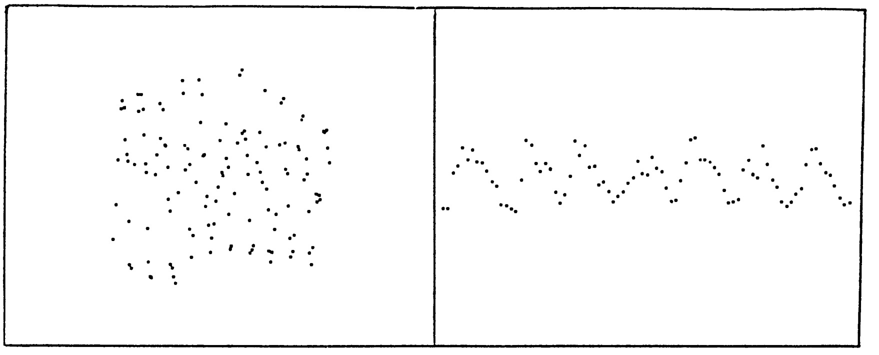

Fig. D3 Plots of monthly river flow versus time. Left: x compressed, y expanded. Right: x expanded (and partly clipped), y compressed.

are other types of scale dependent structure besides slopes. Another example for the latter are outliers in highly deterministic time series where often proper scaling together with strong magnification is needed to reveal them graphically.

At this point, we would like to distinguish between two purposes of graphics: presentation and exploration. For the former we know what kind of structure we wish to display, but for the latter we simply don't know. Therefore, for exploratory analysis we should have tools available which allow us to search all possible scalings of a plot. Since only two parameters are involved, we can afford an exhaustive search of the interesting ranges of scales, provided that we can do the search in real-time using motion graphics. We have such a system in place (called "Data Viewer", partially documented in Buja *et al.*, 1986). It convinced some of us that real-time scale control is more basic than some other better known techniques such as 3*d*-rotations.

We are thankful to the authors for bringing this important issue to the attention of a wider audience, and we feel we now understand a crucial aspect of graph perception better.

Dr P. Costigan-Eaves (Institute of Educational Technology, The Open University): We can think of the emergence of statistical graphics as a pasture in which the intellectual readiness for such methods began to ripen in the work of Lambert, Crome, Playfair and a host of others. By the end of the nineteenth century the pasture had hundreds of new graphical innovations sprouting in many academic arenas. In the twentieth century we find ourselves lost in a graphical jungle with patches of excellence, excursions into the exotic and plenty of what Tufte (1983) so humorously calls 'computer debris' and 'chartjunk'. I view Cleveland and McGill as a pair of courageous gardeners, trying to establish the beginnings of scientific order in what has largely remained a DIY world.

This is not say that the curious history of statistical graphics has not had its systematic thinkers. Over the last hundred years there have been a collection of analyses, outlines, taxonomies, schemes and working theories, all attempting to systematise the great range of graphical devices in existence. While this approach is worthwhile, it is limited. It is like glancing into a mirror and classifying every expression, forgetting that it is the person who generates the reflection. So it is with graphics. It is the user we need to know about and who Cleveland and McGill are inviting us to think about more intelligently.

In putting the spotlight on the user we need to prove more deeply than 'elementary codes' as Cleveland and McGill are probably well aware. We can think of a host of other issues, such as the user's purpose, background, interest etc. A thing that does not figure in Cleveland and McGill's scheme (rightly so because they are investigating preattentive vision) is the degree of textual encoding. When a user approaches a graph does he or she read it, view it or do both? I personally find line slopes, tick marks and even scale indicators only partially meaningful until I read whatever verbal description the author has chosen to provide. Such textual elements comprise heirarchies, i.e., inside or around the display (names of axes and lines, etc.), outside the display (legends, figure captions etc.) and the main body of the text that explains and accompanies the display. How much of this verbal encoding corrects and structures purely visual encoding? Such a question and related ones must of course be addressed by other paradigms.

Mr Vivian A. Daniels (Hull University): The authors assert that the shape parameter "has a substantial effect on our ability visually to decode the slopes". Justification for this assertion is less than clear.

Would our slope judgements on 1 cycle of a sine wave, shape 0.5, differ substantially from judgements on a graph representing 2 cycles, shape 0.25, or on a graph showing n cycles, shape $1/2n$?

On regarding a symmetrical parabolic graph, constructed within a data rectangle measuring 1×1, is our ability correctly to read the slopes in the left half of the graph altered by simply blanking out the right half, thereby altering the shape parameter from 1.0 to 2.0?

The "failure" of the graph in Fig. 3 is attributed to the choice of shape parameter, but is this really the cause?

Would the slope characteristics of Fig. 2 be different were we to digitise the curve into discontinuous graph segments, each with shape parameter 1.0 and each "graph" separated from the next by a hiatus of 1 millimetre? It would seem that the shape parameter may be more a red herring than the villain of the piece.

If we accept that slope resolution is maximised when the median slope of the graph is 1, then in seeking to maximise slope resolution, what is required is a function specific to the data under consideration, but independent of any shape parameter.

Such a function would be expressed, not in pure numeric terms, but in units of the actual data concerned and would describe the "scale ratio". In the case of a sine wave, since equal increments in x and y are always given equal length increments on the axes, a scale ratio of 1.0 is ensured. This ratio, being independent of shape parameter, remains unchanged however many cycles are drawn.

If, in the sunspot data for Figs 2 and 3, analysis were to yield a median slope of, say, 23 sunspots per 5 years, then the "scale ratio" would be expressed as (23 sunspots)/(5 years). A choice of scales giving the scale ratio a value of 1 would then maximise slope resolution in the graph. On this basis, is not the scale ratio of Fig. 2 *closer* to a value of 1 than the scale ratio in Fig. 3?

The use of scale ratio avoids all the problems set out above, and should surely be the graph constructor's guide, rather than H/W.

Dr Geert De Soete (University of Ghent, Belgium): Cleveland and McGill's work convincingly demonstrates the need for empirical research on graphical perception. Their approach differs from much of the recent experimental work in the area in that they provide a theoretical framework. Whereas too many studies are limited to evaluating specific (usually existing) graphical methods as to accuracy and efficiency (e.g., De Soete and De Corte, 1985), Cleveland and McGill formulate a paradigm which can guide empirical research and which can help integrate empirical findings. Although empirical studies that evaluate a particular graphical method are interesting by themselves, their usefulness is limited because little theoretical explanation can be offered for the results and the results can often not easily be transferred to other (possibly new) graphical displays. By constructing a theory of graphical perception, and by systematically testing predictions derived from the theory, basic principles and processes involved in graphical perception can be identified and characterized. Moreover, such results will not be limited to the particular graphical methods included in the study, but may be generalized to other graphical procedures.

However, as important as a theory of graphical perception may be, it should be realized that perception is only one (though important) aspect of understanding graphs. In addition to decoding the information accurately, one has to make certain inferences, comparisons, etc., in order to fully understand graphical display in statistics. Therefore, besides studying graphical perception, attention should be paid to the basic cognitive processes that play a role in understanding graphs. This ultimately should lead to a theory of graphical cognition. For the same reasons that statisticians should be involved in the study of graphical perception, statisticians should take part in the study of the cognitive processes involved in understandings graphs.

Dr William F. Eddy (Carnegie-Mellon University, USA): The authors have provided us with an important beginning to the study of how people interpret graphical information.

I am not happy with the authors' use of the word perception. Perception is the process of becoming *directly* aware of something. Weber's law, which was created and named by the psycho-physicist (and proto-statistician) Gustav Fechner, is an example of a law of perception. Essentially, it is an attempt to characterize the effect of a sensory transducer, for example, the eyes. I am not a psychologist but I suspect that Weber's law applies, in varying degrees of accuracy, to every animal for which the relevant data can be collected. On the other hand, the elementary codes given in Table 1 of the paper implicitly assume that there is an information processing capability available. The ability to make comparative statements about lengths or angles presumes the ability to identify the concept of length or angle. I suppose that not every sighted animal can identify these concepts and thus is not able to glean the same information as a human can from them. My point is simply that there is some interpretation of the information that is not performed in the eyeball but rather in the brain. The authors' use of the word *judgement* in their description of the experiment in Section 5.4 seems to suggest that they agree that the brain is involved. While this is all really just a minor quibble it suggests to me the need for close collaboration between statisticians concerned with graph construction and psychologists interested in visual perception and cognition.

I agree quite strongly with the authors that we must study graphical *interpretation* experimentally if we are going to put graph construction on a firm foundation. We must be careful however to replicate experiments and not fall in the trap that other experimental sciences have: Once a single experiment shows a particular result is "statistically significant" the result is taken to be true. Surely (we think) the statistical profession can rise above this. I hasten to note that if I repeat the experiment of Section 5.4, I doubt I could publish it, no matter what the result.

Dr P. J. Green (University of Durham): It is pleasing to see statistical and psychological science brought to bear on the subject of perception in statistical graphics.

In particular, it is good to see an actual experiment reported in the proceedings of the Society. However, like previous discussants, I have some doubts about the way in which this one was designed. Our students learn that an experiment should be *relevant* to the real phenomenon being studied, and that in assessing the effects of some treatment, other sources of variation should be *controlled.* Now even if we are prepared to say that slope judgements can be reduced to the simple matter of comparing two straight line segments, there remain 8 degrees of freedom in specifying these segments. The explanatory variables considered here are functions of two of these, the slopes. It is convenient to parameterise the other six as follows. We would presumably agree intuitively that to within reasonable limits the two components of location and one of overall scale are irrelevent. But surely the remaining three: the two relative lengths of the segments and their degree of 'overlap', might be expected to influence perception substantially? In the experiment reported in Section 5.4, these seem to have been ignored or confounded with other variables. Have these issues been studied elsewhere? Further, in most of the applications mentioned in the paper, the line segments of interest are actually contiguous, quite a different configuration from that in the experiment.

Dr Perry D. Haaland (Becton Dickinson Research Center, NC, USA): Cleveland and McGill's work has made it clear that there are quantifiable differences between good graphical displays and bad graphical displays. They have also shown that it is possible to conduct formal research into these differences. However, it is still not necessarily easy to create *good* graphical displays, and our understanding of how and why individuals create good graphical displays is not clear.

For example, Cleveland and McGill's work has clearly demonstrated that the shape parameter of an x-y plot strongly affects the viewer's ability to make accurate judgements about slopes. However, not many creators of statistical graphics have the hardward and software which really make it *easy* to control this important parameter. By "easy", we mean the ability to dynamically control the parameters and see the effects in real time.

When these dynamic tools become available, will we actually use them to make better graphical displays? Cleveland and McGill have provided a paradigm for how we decode the information in statistical graphics. They have extended this paradigm to cover the construction of graphs under the assumption that we should construct graphs which can be most accurately decoded. However, to what extent does their paradigm capture the dynamic aspects of graphical construction?

The following experiment might be used to address this question: suppose that under properly controlled experimental conditions, subjects were allowed to freely vary the shape parameter of a given display. (This is actually possible with at least one commercially available software package – DataDesk (1) for the Macintosh.) It would probably be important to provide the subject with a hypothesis to best illustrate by choice of the shape parameter. Will subjects actually choose the optimal shape parameter that Cleveland and McGill's current work has suggested? Given the proper tools, is the construction of good graphical displays natural or is this a learned response?

Cleveland and McGill have provided a soiid foundation for investigating these and other new questions in the construction area of statistical graphics. In addition, their results so far have made a significant impact on our understanding of what makes a good graphical display. Finally, their paradigm for the encoding of quantitative information provides an interesting starting point for investigating the potential impact of the powerful new tools of dynamic graphics.

Drs G. J. Hahn and F. W. Faltin (G E, Schenectady): Incisive graphical displays provide a key tool for conveying information. One well chosen plot is, indeed, often worth a thousand statistics! This is true, moreover, both for presentation and for exploratory graphics. We fully agree that how to effectively impart quantitative information by graphical displays should be an object of scientific study, and applaud the authors' contributions in this regard. Our comments deal primarily with some related topics to which this philosophy is equally applicable.

First, we wish to observe that the current paper deals mainly with HOW to do good plots. We feel that the question of WHICH plots to do, would benefit from a similar approach. The issue is a particularly sticky one, since which plots are most revealing may well depend on the statistical expertise of the observer.

A common example of deciding on the right plot is provided by multiple regression, especially of highly non-orthogonal data. Often, practitioners wish to obtain an understanding of the relationship between individual predictor variables and the response, with other predictor variables "held constant." Numerical outputs, such as regression coefficients, tell only part of the story—just as single-variable regressions are incomplete without the associated scatter plots. Adjusted variable plots (see Chambers *et al.*, 1983) and partial residual plots (see Larsen and McLeary, 1972) have been suggested for this purpose. But how well do these represent the true relationship? Can we do better? What should we do when the predictor variable under evaluation interacts with other predictor variables? There appears to have been only limited research on this topic (e.g., an unpublished manuscript by Feder).

Another natural example is that of statistical process control, where graphical displays have traditionally been used extensively. Shewhart control charts, introduced about 60 years ago when calculational simplicity was a major issue, are the predominant tool—even though other methods, such as CUSUM charts, can be shown to be more effective in detecting certain types of "out of control" situations. What practitioners find most informative—in our experience—is not the statistical signals, but the graphical display of process performance over time. (Concentrating on such displays also sidesteps the sometimes thorny practical question of deciding what really represents chance variability.) Shewhart-style plots appear to be the favourite of the statistically uninitiated, whereas CUSUM plots seem more enlightening to those with more background in the subject. But is this informal observation objectively defensible? What then is the most incisive way of plotting process performance over time under different circumstances? How do we best display process data when we are data-rich (e.g., automated continuous measurements), rather than data-poor—the situation for which control charts have been traditionally used? How can we guide graphical perception in displaying multivariate data (e.g., correlated process variables) or multisource data (e.g., "in-parallel" data from different machines, operators, etc.)?

As is evident from the authors' analyses, a major challenge is how to go about such evaluations. Unlike the study of, say, the properties of a new test statistic, the assessments will often be experimental, rather than analytical. Thus, we will need to develop criteria for selecting one display over another. Our statistical upbringing will be handy. For example, we will favour graphical displays that not only convey the most information on the average, but that do so as consistently as possible from one viewer to the next. Moreover, we will have to beware of "interactions" between the media under consideration, and the characteristics of the observer. As the authors indicate, we can use all the help we can get in such assessments—and that surely includes that of researchers in visual perception.

Finally, we wish to comment briefly on the importance of taking fuller advantage of the exciting opportunities posed by computer graphics. We share the authors' enthusiasm about what we—and our clients—will soon be able to do routinely using interactive personal workstations. Most past work has been directed toward what one might call static, or "hardcopy" graphics. A vast additional array of possibilities is posed by animation, the use of perspective, point tracing of graphs, multiple windows, etc. Much software research has been devoted to such areas, but the human interface is one which would benefit from study via the fusion of the theoretical and empirical methods advocated by Cleveland and McGill.

In conclusion, we agree that recent developments in computing have not merely relieved the tedium of statistical calculations, but, more fundamentally, have opened whole new vistas of data acquisition, computational techniques, and information representation (among others). The problem is no longer that our resources permit us too few alternatives, but that they offer us so many. If we are to master the information at our disposal, rather than be mastered by it, the tools we employ must be wisely chosen.

Professor M. J. R. Healy (London School of Hygiene and Tropical Medicine): After following with great interest Dr Cleveland's published work on graphical methods, it is a pleasure to have him in person at this meeting. His main topic, that of choosing the height/width ratio for a graph, is not new in the statistical literature—it is a familiar topic in discussions of cusum charts, where the 45° recommendation is common—but it is good to have some experimental evidence.

As usual the epithet "good" invites the question "good for what?". Fig. 10 is poor for slope comparisons, but it does convey well the near-parallelism of the lines and the rough equality of the distances between them.

Fig. 11 introduces the issue of comparing separate graphs. Here a general principle worth testing is that comparisons are easier when distances are smaller. This would suggest a 2×2 layout for the component tables rather than the 4×1 layout used at the meeting. Another issue, which is illustrated by comparing Fig. 11 with Fig. 4 parts I and II, is that vertical comparisons of horizontal lengths are easier than horizontal comparisons.

Dr David C. Hoaglin (Harvard University): This thought-provoking account leads me to ask for more, primarily in two areas. First, the ranking of the elementary codes in Table 1 provides a valuable start, but can one go beyond that ranking to measure relative accuracy (or inaccuracy) on a suitable scale? For example, it might be possible to say that so-and-so many judgments of position along a common scale "equal" one judgment of length. In an experiment where subjects see displays for a fixed, brief time, perhaps the accuracy of most elementary codes has some useful inverse relationship to their difficulty. Even rough information in this area would be valuable in judging reductions in complexity associated with redesign of displays.

Second, exploratory data analysis also emphasizes calculating and displaying residuals, and plotting the residuals from a fitted line is a leading example. From their work with slope judgments, what can the authors say about the relative effectiveness of (i) a plot of the data that has a mid-angle around 45° and (ii) a plot of the residuals from a well-chosen line? Of course, one can no longer use the ratio of two slopes when one of those slopes is zero, but most of us do expect plots of residuals to reveal smaller differences in slope. For example, it would have been routine to make a version of Fig. 12 in which the vertical coordinate was the residual (speed) from Dewey's model. The resulting display would have one slope to study per distance.

Incidentally, in applying the median-absolute-slope procedure to a plot of residuals, what slopes should one use?

May we please have more motivation for the model that summarizes the dependence of m_{ij} on a_{ij} and p_i? Some of its qualitative features seem to accord less well with Fig. 6 than I would expect.

Finally, despite the nice advances in software (Section 2.1), I would like to urge even more effort to make flexible graphics software widely available, as an effective way of encouraging people to experiment, both informally and formally, in graphical perception. Much remains to be done.

Ms Carolyn Backer Cave and Professor Stephen M. Kosslyn (Harvard University):

Only a Piece of the Pie

The work of Cleveland and McGill is highly laudable, particularly regarding their use of psychological methodologies to study rigorously issues in statistical graphics. We enthusiastically agree that research on graph perception would greatly benefit from collaborations between statisticians and psychologists. However, it seems to us (perhaps egocentrically) that the number of psychological issues has been underestimated. We would like to point out two such areas of weakness in Cleveland and McGill's conception of "graphical perception."

1. Cleveland and McGill rightly point out the relevance of perceptual processes in graphical perception, but seem to minimize the relevance of cognitive processes. The purely sensory aspects of visual perception occur within a very brief period after a stimulus is presented (typically under 1 second), and probably are not consuming the bulk of the multisecond period required to read a graph. In particular, the information conveyed in the elementary codes, which is indeed relevant for statistical interpretation of graphs, does not seem to be the product of purely perceptual processes (as Cleveland and McGill seem to imply). Cognitive processes must also be at work. For instance, "short term" memory (*STM*) is clearly a relevant factor in graph interpretation; indeed, if one is comparing two or more stimuli sequentially, *STM* is recruited almost by necessity. A consideration of the well-known capacity limits of *STM* seems critical for understanding graphical perception. In addition, other cognitive factors, such

as ease of access to "long term" memory (*LTM*), are important. Variables such as familiarity of the graphic format affect *LTM* access.

2. The graph designer must have a sense of what needs to be conveyed to the reader and what the reader will want to know in a given context. Therefore, it seems that the ordering of the elementary codes should not be fixed, but instead vary with the intent of the graph. For example, line length would probably be better than slope for representing a set of amounts if the reader wanted to read the absolute values.

Without a doubt, Cleveland and McGill are contributing valuable insights and discoveries. However, a careful analysis of the information processing steps involved in graphic perception would be a valuable addition to their work.

Drs John LaPrelle and Forrest W. Young (University of North Carolina, Chapel Hill, NC, USA): Cleveland and McGill make an important contribution by suggesting a theoretical foundation for the way we decode graphical information from visual displays. They also provide experimental evidence supporting their position. However, we believe that researchers interested in understanding the principles of graphical perception should consider the advantages of dynamic, high-interaction graphics.

In particular, aspect ratio can be investigated more directly and simply in experiments in which the judge uses a dynamic computer graphics system which provides high-interaction adjustment of aspect ratio (Buja, Hunley and McDonald, 1986). With such a system the judge could continuously adjust the aspect ratio until a subjectively "optimal" shape is obtained. The experimenter would then be able to directly record the optimal aspect ratio for each judge in each situation, avoiding the complex and indirect "mean absolute error" minimization procedure.

When compared to static graphics, dynamic graphics are much more like our everyday visual environment. People interact with and interpret a dynamic world, not a static one. Moreover, the familiar world is three-dimensional (3*D*), not two-dimensional (2*D*). Thus, compared to static 2*D* graphics, it seems more likely that dynamic 3*D* graphics (Donoho, Donoho and Gasko, 1985; Young, Kent and Kuhfeld, 1986) will exercise cognitive and perceptual decoding mechanisms that are closer to those of "real-world" visual situations. Dynamic 3*D* graphics should provide a powerful vehicle for understanding graphical perception and should yield principles more harmonious with those discovered in other areas of vision research.

The point just made brings to mind a more general (and more crucial) point that seems to be under-represented in the Cleveland and McGill discussion: Namely, that the type of graph about which perceptual or cognitive judgements are made can have a drastic effect on the ease or difficulty with which a judge can decode a particular display. Just as we would expect to find different cognitive and perceptual effects in a 3*D* world than in a 2*D* world, we would expect to find other differences specific to a pie chart "world" or a bar chart "world". At the very least, if we ask subjects to judge area of bars in a bar chart *versus* area of slices in a pie chart, we should expect different results since area is confounded with length in a bar chart and with angle and length of arc in a pie chart. Such considerations lead us to doubt that the hierarchy of Elementary Codes presented in Table 1 can be reasonably taken as context free.

Dr A. D. Lovie and Dr P. Lovie (University of Liverpool and University of Keele): We are grateful for the opportunity to comment on this interesting paper by Drs Cleveland and McGill. May we say at the outset that we wholeheartedly support their call for more collaborative work between statisticians and cognitive scientists. It is in no-one's interest for either group to lose touch with the expertise that the other can bring.

Our comments have two linked strands: firstly, it is impossible to separate out the parts of a graph from their field or setting, and secondly, statistical graphics are influenced by illusory effects which work against the essential message of the display. For example, in Fig. 1 of the paper, the shape parameter of the graph is estimated by the ratio of the height to the width of the data rectangle. Any direct reading of this parameter from the display would be inflated by the "vertical-horizontal" illusion in which a vertical line is judged to be longer than a horizontal one of the same length.

Context effects can also apply to the interactions between the components of the graph within its frame. We can illustrate this latter point from some of our own work on visual factors in the perception of box plots. Subjects' judgments about the relative lengths of the boxes (midspreads) of pairs of box plots were influenced by "irrelevant" features such as the width of the box and the length of the whiskers. For instance, subjects tended to overestimate the lengths of boxes which were narrower or had shorter whiskers than their partners and *vice versa*. Even the simple boxplot appears to hide a multitude of perceptual problems for the analyst.

We would not wish to argue that illusory context effects account for an overwhelming portion of a judgement. Nevertheless, in marginal cases where one might require a graphical display to approach the discrimination power of a more conventional numerical data analysis, such visual effects could have serious consequences.

Our final point is a cautionary one. Psychologists have long argued that perception is not a passive, mechanical process. Data analysts bring a *conceptual* context to their interpretation of a graphical display; in other words, they attempt to induce meaning into the material. Such an operation selects, sharpens and supresses evidence. How we should design displays to let the data tell *alternative* stories is a challenge for the future.

Professor R. M. Loynes (University of Sheffield): I imagine that most of us have, at least occasionally, felt that there is something unsatisfactory in the way we choose our ways to draw graphs, but equally most of us have done nothing to help change this; it is particularly pleasing therefore to have a study which goes beyond the usual rather general prescriptions, but I would like to make a few comments.

1. Fig. 5 illustrates that we often have rather more information than merely the slopes themselves: here, for instance, we can compare the slopes of direct interest with the axes and I think if one does this, which one would probably do quite unconsciously, one can see that the slopes differ in panels *B* and *C*, just as one can in *A*.
2. The suggestion that we would like the orientation resolution of two segments to be as large as possible is, I think, sidestepping the question, since this presupposes that the eye is unaffected by the orientation of a configuration: I would have guessed however that the eye does not perform equally well in the horizontal and vertical directions. There is perhaps some support for this in the experimental results reported towards the end of Section 5.4, showing that the apparent optimum orientation is significantly different from 45 degrees.
3. The question under consideration changes from time to time: for example, at first we are considering whether two line segments have equal slope, but later the experiment itself is concerned not with that but with judging a quantitative characteristic of a configuration, the ratio of the slopes of two line segments.
4. I seem to disagree with the authors when they say that certain things cannot be seen, since I think that one can in Fig. 8 notice that the graph is not linear. This may be connected with the fact that there are axes quite close by (cf 1 above), but it may also be connected with another aspect of graphs, which is the relative position of the various features: in this particular example the slopes are lined up end-on to one another whereas in Fig. 5, for example, they are at quite different positions, and I think this may well have an effect on our ability to judge things. I am not altogether sure whether one should not distinguish carefully between exploratory data analysis and data display. For data display of some particular feature we could indeed use something like the median absolute slope procedure, but often it is not clear which particular feature we are interested in and in that case it will be a little difficult to decide which scale to use. The fact that one could use more than one scale for different purposes is of course only a partial answer to this.

Mr M. S. Ridout (Institute of Horticultural Research, East Malling): It is a pleasure to join other discussants in congratulating the authors on an interesting and well presented paper.

Many of the authors' examples are time series. My own interest is in biological response data which often exhibit one of a small number of patterns (linear, sigmoidal, increasing or decreasing to an asymptote, etc.). I wonder if guidelines for the choice of shape parameter can be developed which depend only on the broad pattern of response: an example might be "for a sigmoidal response choose the shape parameter. so that the central 'linear' part of the curve has a physical slope of one".

Consider, for example, the authors' Fig. 7 showing a non-linear trend. Despite the legend I can also detect this in Fig. 8—I just have to look more carefully. Suppose I regard this trend as approximately quadratic. For a smooth curve I can consider a median-absolute-derivative procedure and when, as here, the absolute derivative is monotonic, the procedure simply requires me to choose the shape parameter so that the absolute slope at the mid-point on the x-axis is one. For a segment of quadratic curve which does not include a turning point, the absolute slope at this mid-point is the same as the absolute slope of the line joining the end-points of the segment. In other words the shape parameter should be one—roughly that obtained by the authors.

Lastly, I wonder if the authors have considered the influence of the shape parameter on the assessment of correlation from a scatter diagram. My intuition is that the correlation will be judged to be greatest if the physical slope of the (undrawn) principal axis is one.

Mr D. K. Simkin (Northwestern University): Dr Cleveland and Dr McGill's collection of elementary codes of graphs moves the focus of investigations to a more molecular level than the type of graph used. This is an important step in the establishment of a scientific foundation for graphical perception and many suggestions for the improvement of data display will continue to come from investigations at this level. As the authors anticipated, a case can be made for an even more molecular level. Improvement of data display can also be achieved as an indirect product of investigations into the cognitive processes which underlie these judgements. What processes allow for the more accurate decoding of position than length? The graph reader must be studied as well as the graph.

The point can be made by analogy to the study of reading, or more technically, text comprehension. As in graphical perception, information is encoded in the text and must be decoded by the reader. Stretching the analogy now, the collection of elementary codes might be considered the surface structure of graphs. We have learned that studying text comprehension at the level of surface structure only, leads to an incomplete picture. Focus is on the flow of information in only one direction, starting with a very crude analysis and working through successively more complicated analyses. In cognitive psychology this is referred to as *bottom-up* processing. However, information is also flowing in a *top-down* direction. Information at higher levels is organizing incoming information, filling in missing information, and setting up expectations for as yet unperceived aspects. The difficulty in discovering errors while proofreading or the ease with which we connect pronouns with their referents while reading are testament to the powerful and automatic expectancies set up by top-down processing.

Various graphs as well as patterns of data must certainly set up expectancies as to what information is to be extracted and how it will be extracted. Such considerations have prompted me to investigate whether the position-length-angle ordering which holds for comparative estimation also holds for other tasks. Pie charts are routinely used to convey information about the proportion a segment represents of the whole. When subjects are asked to make this proportion-of-the-whole estimate, judgements based on angles (the code of a pie chart) are at least as accurate as judgements based on position. Results such as this task dependent ordering of codes contribute to an expanding data base which is required as we increase our magnification and define a finer grained collection of elementary particles of graph perception.

In conclusion, I find myself echoing the final thesis of the paper: that investigation into graphical perception is an interdisciplinary endeavour. I humbly submit that cognitive psychologists can be added to the list which already includes statisticians and researchers in visual perception.

Dr A. R. Unwin (Trinity College Dublin): There are many patterns that could be present in a data set (rising trends, periodic outliers, slower declines than rises and all manner of others). Most data sets will exhibit several features and several different graphs will be needed to display the evidence effectively. Rather than talking of one graph we should be thinking of best sets of graphs. Saying that "One solution is to make more than one graph" (Section 5.7) is not putting it strongly enough.

The authors do not make a clear distinction between exploratory graphics and presentation graphics. Given a data set of a particular kind (e.g. a univariate time series like Wölfer's sunspots) we need a statistical strategy to seek out features of interest. The power of to-day's micros and the good graphics they provide make it feasible to do this by searching graphically as well as analytically. The graphic exploring of data can be done using spreadsheets like Lotus on the IBM or Excel on the Macintosh. Excel has instantaneous resizing of graphs, allowing fast adjustment of the scale parameter.

For presentation graphics there are different aims. Having found what we think is in the data, we have, usually, only one graph with which to convince others of our views. This means that one feature may be emphasised at the expense of others. In future, as a direct result of Cleveland and McGill's work, writers may be required to state what criteria they have used in designing the graphs in their papers.

The examples in the paper are interesting for in some ways they contradict the principles the authors espouse. If you want to show that the sunspot numbers rise more rapidly than they decline, should you not plot the slopes? If they were plotted as yearly differences you could compare positions among a common scale and thus employ an elementary code that is "as high in the ranking as possible" (Section 1.2). It would be still better to plot the times between successive turning points. Are the authors following an unstated deeper principle that the raw data should always be shown? This would explain why figure

7 is scaled with respect to line segments which are not actually drawn. If the trends in maxima and minima are interesting, why not just show the turning points?

This paper concentrates on the shape of a graph. Other factors such as size, clutter and quality of reproduction are relevant too. With these factors in mind, how many of the figures in the paper would Cleveland and McGill like to redraw?

The authors deserve every encouragement in their efforts to provide statisticians with a theory of graphics. Their emphases on experimentation and an interdisciplinary approach are especially welcome.

Mr Leland Wilkinson (University of Illinois at Chicago and Systat, Inc): As a psychologist who has worked in the area of graphical perception, I am once again impressed by the psychological insight of statisticians Cleveland andMcGill. What makes their research on graphics so useful in practice is that they go beyond *ad-hoc* comparisons of specific graphs. Human factors experiments comparing specific graphical types are relatively easy to design. We need theories of graphical construction to improve what we already have. I hope psychologists will give as much attention to their work as they have given to psychologists.

As a statistical graphics programmer, my first thought after reading this article was to design the median absolute slope procedure into the microcomputer graphics system I am developing. Wouldn't it be nice for the computer to choose an optimal scaling for scatterplots (using *LOWESS* or robust regression for identifying slopes), *ANOVA* interaction plots, and time series plots? I am worried about several consequences, however, particularly since so many users trust the computer and fail to explore nuances.

Cleveland and McGill point out some of these problems in their article. First, we must know in advance whether we want to enhance local or global slope contrasts. Second, enhancing local contrasts in a single frame may be impossible. We need to know when to subset our data. Finally, even if we have a plot as straightforward as Fig. 5, I am not yet convinced the median absolute slope procedure will give us the most discernable plot because there are several visual illusions in these graphs that the procedure cannot correct.

We know, for example, that small (acute) angles tend to be overestimated and large (obtuse) angles underestimated (Coren and Girgus, 1978). Furthermore, frames (axes, CRT screens, pages of a book) influence the perception of angular orientation. If the Maximum Resolution Theorem describes human perception, then we should expect that the orientation mid-angle of 45 degrees is optimum for all slope ratios. The data indicate otherwise. Fig. 6 shows a dip at around 30 degrees for slope percentages above 70. If we use weighted least squares on this subset of the data (with standard errors taken from the graphs), Cleveland and McGill's function yields an optimal angle of about 39 degrees. The estimate is even lower for the higher slope ratios. This trend makes sense for larger slope ratios because as angles approach 90 degrees, veridical judgements can be improved by vertical and horizontal frame comparisons. This conjecture could be tested by including higher slope ratios and larger mid-angles in a replication of the study.

Nevertheless, this study represents an important advance in identifying a optimal frame ratio for most data. The median absolute slope procedure surely will be a basic tool in future adaptive graphics systems. Such systems will operate on graphs as objects, automatically manipulating features to reduce the effects of visual illusions and perceptual error.

Drs Edward J. Wegman and Donald T. Gantz (George Mason University, Virginia, USA): We would, first of all, like to convey our thanks to Dr Cleveland and Dr McGill not only for this particular paper, but also for the larger body of their highly innovative work. There is no doubt that Cleveland and McGill are prominent among the most important contributors to this area and deserve the thanks of the entire statistics community.

The Maximum Resolution Theorem is a geometric result which suggests that in comparing two lines we adjust the construction of the statistical graphic so that the mid-angle is 45°. Interestingly enough, the experimental evidence suggests that the optimal mid-angle is 41.1°. While we agree with Cleveland and McGill who conclude that there is little practical difference for purposes of statistical graphics, there is an annoying matter that 45° is "just outside a probable range". The question then becomes is this a small real effect or just bad luck?

The formula $\arctan(\gamma^* \times s) = a(\gamma^*) = 45°$ suggests that a pair of lines whose inverse geometric average slope is 1 should have an optimal shape parameter of 1, i.e. $\gamma^* = \tan[a(\gamma^*)] = \tan[45°]$. If the optimal angle is legitimately 41.1°, then $\gamma = \tan[41.1°] = .8723$. This suggests better discrimination when the statistical graphic is slightly elongated in the horizontal direction. There seems to be a human preference for horizontally elongated displays possibly connected with the fact that the eyes are located horizontally rather than vertically. We note that a shape parameter of .8 would correspond to an optimal mid-angle of 38.7°. [38.7° = arctan(.8)]. And that 41.1° is actually more compatible with the optimal mid-angle being 38.7° than with 45°. The Maximal Resolution Theorem is based on a geometric symmetry between horizontal and vertical directions. However, there is an asymmetry in the human visual system and consequently, it is not the least surprising that the empirical optimal angle is less than 45°.

Turning now to the exploitation of these ideas in data analysis and in presentation graphics, it is clear these ideas will have differing impacts on these two roles. The data analyst needs to perceive patterns in the data. Formalizing the visual analytic process is certainly appropriate and perhaps even necessary. The impact of this work on the presentation of data is somewhat different because the whole environment that surrounds the presentation of statistical findings introduces requirements and constraints. Take, for example, the statistical expert's presentation of data analysis to a judge or jury. The task of the expert is to convince the jurist that the analysis discriminates a real difference of sufficient magnitude to compel a decision. The realness of a difference is a statistical "black box" call from the point of view of a jurist. However, the decision-driving magnitude of a difference is a call based on the jurist's own judgement. This aspect of data analysis is often called practical significance (cf. Meier *et al.*, 1986). Our approach has been to present the jurist with a striking graphic from which the jurist will reduce a compelling magnitude of difference. The present work not only suggests a best shape parameter, but provides rebuttal to suggestions that the selected scales distort the data. However, complex and nonstandard summaries that provide substantial information to the data analyst often raise suspicion in the layman. Indeed, an extreme shape parameter, say less than .1 or greater than 10 may very well leave the jury suspicious that they are being manipulated.

Our last comment involves the implementation of the Cleveland-McGill ideas. It is clear that more than one shape parameter may be appropriate for a given data set. This is very suggestive that a high-interaction implementation of a variable shape parameter may be fruitful. We intend to implement such a tool in our work.

The authors replied later, in writing, as follows.

We would like to congratulate the discussants for squarely facing the intellectual issues, providing many excellent suggestions, and raising what are mostly reasonable caveats about our work. We are quite pleased to see the scientific process at work. In the past, the construction domain of statistical graphics has too often suffered from a reign of pure opinion. The discussion here demonstrates that scholarly debate of intellectual issues is possible. We hope this demonstration will break down tolerance of pure speculation and unsupported statements about data display.

Several of the comments made us realize that a deficiency in the paper is a precise statement about what we are *not* doing. We will correct that in the first two sections below and then respond to other comments. The issues relating to the investigation of shape are discussed in much more detail by Cleveland, McGill and McGill (1986).

Perception, Cognition, Experimentation, and Training

When we study a graphical display, there are various types of visual processing that we invoke to extract the quantitative information. One is a highly perceptual processing consisting of very rapid judgements of geometric, textural, and colour aspects of the graph. This processing is what we are studying; it involves more than the rapid 5-50 ms processing that Julesz (1981) calls *pre-attentive vision*, because we do employ eye movement and attentive search, but it nevertheless is carried out quickly, perhaps on the order of a second or two. Another processsing is scale reading; it begins with the perceptual process of scanning horizontally or vertically to project visually a point onto a scale line, but then involves a more highly cognitive process of noting numbers associated with positions on the scale line.

The information extracted from the perceptual processing and the scale reading is combined with concepts and knowledge about the data in a highly cognitive way to derive quantitative statements such as "over most of the range of the data, solar radiation is attenuated in sea water by a constant factor of about one order of magnitude per 40 metres." This is the contemplation to which Seheult refers. We are not attempting to study such cognitive processing. Cave and Kosslyn, Costigan-Eaves, De Soete, Eddy, Gower, and Simkin have argued for its importance. We concur. In Section 3.3 we cite some good early work in this area.

Suppose one asks a very specific question about some small amount of quantitative information on a graph. We can always read off information from the scales and do some mental arithmetic to get an answer. But this process is not what we use in general to extract the large amount of information that is on many graphs; most of the information is obtained by the rapid perceptual processing described above. (An exception is a graph with a very small amount of information that we use more like a table.) When we run experiments to probe the perceptual processing, specific questions are asked about a small amount of information. We must limit the time a subject has to see a stimulus to prevent him or her from resorting to more highly cognitive processing.

Gower has raised the issue of training. The rapid visual processing we study is something the human visual system does using its extraordinary algorithms for seeing in general. Thus, a subject's level of technical training is not a factor, or to be more specific, we have never been able to detect any dependence on it from our experiments.

Shape and Two-Variable Graphs

Our results on the choice of shape do not address all two-variable graphs but rather just those made to see how one variable depends on the other. Slope judgements are relevant for this case. We do not address graphs made to probe the bivariate structure of measurements of two variables; on such graphs one looks for outliers, clusters, and sets of points that lie along one-dimensional manifolds. Judging the orientation of line segments is a relevant task but typically not judging slope. In particular, our results do not apply to the plots referred to by Gower—those used to show the results of correspondence analyses—or any other graph where one is probing the position of points with respect to some metric in x and y. In such a case the scales must be chosen to show faithfully the metric; another example, a much more common one than correspondence analysis, is a map.

Discrimination and Comparative Estimation

Loynes has alertly pointed out that in one place we raise the issue of whether two slopes are equal but then carry out work on judging the ratio of two slopes. Using the terminology of Follettie (1986), the first task is discrimination and the second is comparative estimation. Comparative estimation is the visual workhorse of graphical perception while discrimination plays a minor role. It is almost never the case that we care only whether two things are exactly equal or not; in any application of which we can think, it would matter not at all to us whether the ratio of two slopes is exactly 1 or $1+10^{-10}$. Hand's comments appear to indicate that he agrees with this statement. However, thinking about detection can sometimes shed light on comparative estimation. Suppose $s_1 < s_2$ are two positive slopes and our goal is to judge s_1/s_2. If we cannot detect a difference between the two, we know that the absolute error in judging the ratio is $1-s_1/s_2$. Here, for example, is one application. We know that as the shape goes to 0 or ∞, the orientation resolution goes to $0°$; thus the absolute error of comparative estimation goes to $1-s_1/s_2$.

Also, we confess that to some extent we discuss detection because it provides vivid examples, as illustrated in Figs 2, 3, and 5. We apologize for not keeping the two issues clearly separated in the paper.

Context, Other Factors, and Experimental Design

LaPrelle and Young, Lovie and Lovie, and Loynes correctly point out that there are factors other than elementary code that affect our visual decoding. Indeed we have probed some of these factors in our work (Cleveland and McGill, 1986). The settings of the other factors affect the amount by which the accuracies of the codes differ, but the ordering tends to remain intact. In fact, we have experimental evidence about the specific context issue raised by LaPrelle and Young. Their rather definite statement about pie charts and bar charts does not agree with our experimental information (Cleveland and McGill, 1984 and 1986). (Again, we must emphasize the importance of basing statements, at least rather definite ones, on experimentation or solid theory.) Judgements of the wedges in pie charts behave very much like angle judgements and not like area judgements, and judgements of the bars of bar charts behave very much like length judgements and not like area judgements. Our experiments show that bar charts convey information about relative magnitudes more accurately than pie charts, as the ordering of the elementary tasks predicts.

Green asks about other factors that might affect slope judgements and Jennison questions the range of the mid-angles in the experiment on slope reported in Section 5.4. It is very tempting to try and learn everything from a single experiment particularly, as in this case, when it is a first one in some area; experience has taught us that good experimentation necessitates resisting such temptation. We have found that in graphical perception, as in many other fields, it is far better to learn with substantial precision about a narrow domain. It should be remembered that our slope experiment was not merely an accumulation of evidence unguided by theory; we were probing a highly plausible hypothesis suggested by the Maximum Resolution Theorem. We allocated the finite resources of the experiment to make our prior probability high that we could get reliable results. We considered the factors mentioned by Green and held fixed what we could beyond the variation in mid-angle and true percent. We have demonstrated in other experiments (Cleveland and McGill, 1986) that distance, as Healy has suggested, is an important factor. Thus distance was held fixed; line segments were always at the northwest and southeast corners of the screen and there was no overlap. We also included no mid-angle above 55°. The reason is that we have a reasonable theoretical argument that the error for a mid-angle of $45° + \theta$ is about the same as that for $45° - \theta$. Because of this theory, we decided that our finite resources would be better spent on replication than on expanding the range of the mid-angles. A more detailed discussion of these issues is given by Cleveland, McGill, and McGill (1986).

Plotting Slopes Directly

Unwin has called attention to an important point that we discuss in Cleveland and McGill (1985) and Cleveland, Cleveland, and McGill (1986). We can greatly improve our visual decoding of any particular set of slopes by computing them and graphing them directly so that the elementary code is position along a common scale. This, however, is not a practical general solution partly because we often make judgements of many different sets of slopes. But it is an important alternative to consider in certain cases.

The Median-Absolute-Slope Criterion

The comment of Seheult on the robustness of the median-absolute-slope criterion is perceptive and well expressed. Ridout makes another important observation: if the criterion is applied to a curve that is nearly quadratic, the shape parameter is one. The comments illustrate one advantage of an explicit criterion—it provides something concrete that can be analyzed.

Ridout and Ross have raised an important topic. Suppose we make a scatterplot for the purpose of assessing the broad pattern of the dependence of y on x. To which set of slopes should we apply the criterion? A method for doing this, discussed and illustrated by Cleveland, McGill, and McGill (1986), is to smooth y as a function x using a procedure such as robust locally-weighted regression, plot the smooth curve on the scatterplot, and use the slopes of the curve as the input to the criterion. This might serve as a solution to the question of Hoaglin about plotting residuals against dependent variables or fitted values, but more experience is needed to see if this would lead to plots with a shape parameter that is typically too large.

Dynamic Graphics

Buja, Haaland, LaPrelle and Young, Lewis, and Wegman and Gantz have raised several issues relating to dynamic graphics. Dynamic control of the shape parameter provides a good solution to the problem of needing different shapes for the same graph. This is discussed by Cleveland, McGill, and McGill (1986). It also provides one solution to the user interface problem raised by Wilkinson and Haaland. A particularly clever implementation of this idea is given by Buja *et al.* (to appear).

Haaland and LaPrelle and Young have suggested an interesting experiment. Suppose we give a subject dynamic control of shape, specify a set of slopes, and ask the subject to choose the shape and then study the slopes. How close would the subject come to the shape arising from the median-absolute-slope criterion? We hope someone will run such an experiment.

Time Series

Chatfield has pointed out an important consequence of the paper for plotting a time series—the choice of the shape parameter depends on the frequency of the component under study. Buja makes a similar point and gives an example.

Lawrance makes an important observation about the sunspot numbers. There is no small amount of irony here. The canonical example of a linear and gaussian autoregression is nothing of the sort. In fact, we can take this statement even one step further. The irreversibility shown in Fig. 2 makes it clear that the series is not adequately described by any model that is both stationary and gaussian. Since stationarity of the sunspots seems like a reasonable assumption, the gaussian property must fall. We were first made aware of the nongaussian behaviour from an estimate of the bispectum presented by Brillinger and Rosenblatt (1967); the estimate has an argument that is significantly different from 0 or

π, which indicates the series is not reversible. But as Lawrance points out, the very simple plot with the right shape parameter shows this convincingly and easily. It turns out that similar statements hold for another canonical time series: the Canadian lynx data (Elton and Nicholson, 1942). This is discussed by Cleveland, McGill, and McGill (1986).

Graphic Designers

Hand has suggested that there may well be expertise in the world of graphic design that bears on the issues discussed here. We think not. Graphic design, as a profession, is true to artistic principles, is true to aesthetics, is true to the surrounding medium in which a graph appears, and, particularly in the mass media, is true to commercial interests. Unfortunately, graphic design is not overly concerned with being true to the data. Tufte (1983) puts it aptly: "Illustrators too often see their work as an exclusively artistic enterprise .... Those who get ahead are those who beautify data, never mind statistical integrity."

Miscellaneous Issues

Two discussants made comments about the quality of the graphs in the manuscript that was distributed. The problems arose because our originals were, unfortunately, redrawn. The final version of the manuscript will contain graphs drawn by us.

Chatfield (1977) showed an intriguing example of how the method of plotting can affect the perception of the quantitative information on a graph. Two factors varied on these graphs – the shape and the method of connecting successive values. To determine which of the two has the bigger effect it would be interesting to see four graphs that show all combinations of the two levels of the two factors.

Hahn and Faltin have asked how we can assess the efficacy of competing graphical methods. Principles of graphical perception can often provide answers to questions about the details of graph construction but they rarely can tell us about the relative merits of two competing methods that address the same data analytic task but rely on two very different sets of quantitative information. For example, the work here tells us little about whether a partial residual plot or an adjusted variable plot (Chambers *et al.*, 1983) is a better tool for regression diagnostics. To answer such a question one needs to carry out rigorous field testing. This is discussed in detail by Cleveland (to appear).

The information processing algorithms of the human visual system cause visual illusions. For example, as Lovie and Lovie and Wegman and Gantz have pointed out, vertical extents are judged to be somewhat longer than horizontal ones (Coren and Girgus, 1978). Also, there are consistent biases in judgements of orientations of line segments (Fisher, 1974), so Loynes is correct in raising rotation as an issue; an orientation distortion may account for the effect of rotation noted by Preece. These illusions are discussed in Cleveland, McGill, and McGill (1986), and our conclusion is that they play a minor role in the overall error of slope judgements.

The arrangement issue raised by Healy is a very important matter and is treated in substantial detail by Cleveland (1985) under the topic of juxtaposition and superposition of graphs. Healy's statement about vertical and horizontal arrangement expresses the ordering of the first two codes in Table 1.

In the second paragraph of her contribution, Costigan-Eaves has expressed, incisively and elegantly, what we have been trying to do in our work on graphical perception and how it contrasts with work in the past.

Healy points out some things that one can see in Fig. 10. This is indeed true, but we would add to this that it is an inefficient way to convey the information, which could be conveyed more easily by a simple numerical summary of the data.

Clifford has reminded us about the visual method of sighting along a piece of graph paper to help judge slopes. This is effective for judging a set of smoothly changing slopes of connected (actual or virtual) line segments, but does not provide a general solution to improving slope estimates.

Despite the objective of Eddy, we believe we are using the word *perception* correctly or at least as psychologists would. Perception is not limited to optics and the retina, and is mostly a study of a processing that is carried out in the brain. Eddy underestimates the information processing ability of animals, in particular, fish. The Müller-Lyer illusion, an incorrect judgement about the relative lengths of two line segments, is best explained by incorrect performance of the brain's algorithms (Rock, 1984). It catches fish as well as humans (Coren and Girgus, 1978).

Unfortunately, Daniels may well have missed the point of the paper entirely. The shape parameter *per se* does not have any effect on a graph. It causes an effect only in so far as changing it will typically change elements on the graph, such as the slopes of line segments. Clearly, if we change the shape and change what we graph in such a way that the configuration of certain graphical elements remains exactly the same, then the change in shape has little or no effect on our perception of these elements.

Stewart commented at the meeting, but not in writing, that we have concealed the 10% or so increase in the carbon dioxide concentrations in our figures. The proof that we did not is that he determined the magnitude of the increase. The scales are on the graph for anybody willing to take the time to look at them. A lot of silliness has been put forward about the zero point of a scale. This is discussed in great detail by Cleveland (1985).

REFERENCES IN THE DISCUSSION

Brillinger, D. R. and Rosenblatt, M. (1967) Computation and interpretation of *k*th order spectra. In *Spectral Analysis of Time Series* (B. Harris, ed.) New York: Wiley.

Buja, A., Asimov, D., Hurley, J. A. and McDonald, J. (to appear). Elements of a viewing pipeline for data analysis. *Dynamic Graphics for Statistics*, W. S. Cleveland and M. E. McGill, editors. Monterey, California: Wadsworth.

Buja, A., Hurley, C. and McDonald, J. (1986) A data viewer for multivariate data. In *Proc. 18th Symp. on Interface of Statistics and Computer Science*.

Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A. (1983) *Graphical Methods for Data Analysis*. Monterey, California: Wadsworth.

Chapman, L. (1986) *Plain Figures*. London: HMSO.

Chatfield, C. (1977) Some recent developments in time-series analysis. *Journal of the Royal Statistical Society A*, **140**, 492–510.

Cleveland, W. S. (1985) *The Elements of Graphing Data*. Monterey, California: Wadsworth.

Cleveland, W. S. (to appear) Research in statistical graphics. *Journal of the American Statistical Association*.

Cleveland, W. S., Diaconis, P. and McGill, R. (1982) Variables on scatterplots look more highly correlated when the scales are increased. *Science*, **216**, 1138–1141.

Cleveland, W. S., McGill, M. E. and McGill, R. (1986) The shape parameter of a two-variable graph. *Technical Memorandum*. Murray Hill, New Jersey: AT&T Bell Laboratories.

Cleveland, W. S. and McGill, R. (1984) Graphical perception: theory, experimentation, and application to the development of graphical methods. *J. Amer. Statist. Assoc.*, **79**, 531–554.

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

Cleveland, W. S. and McGill, R. (1986). An experiment is graphical perception. *International Journal of Man-Machine Studies*, **25**, 491–500.

Coren, S. and Girgus, J. S. (1978) *Seeing is Deceiving*. Hillsdale, New Jersey: Lawrence Erlbaum.

Donoh, A. W., Donoho, D. L. and Gasko, M. (1985) *MACSPIN-Graphical Data Analysis Software*. Austin, TX: D^2 Software.

Ehrenberg, A. S. C. (1977) The rudiments of numeracy (with Discussion). *J.R. Statist. Soc.* A, **140**, 277–297.

Ehrenson, S. (1982) Correlations perceived and measured. *Science*, **218**, 918–919.

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.

Julesz, B. (1981) Textons, the elements of texture perception, and their interactions. *Nature*, **290**, 91–97.

Larsen, W. A. and McClearly, S. J. (1972) The use of partial residual plots in regression analysis. *Technometrics*, **14**, 781–790.

Mahon, B. H. (1977) Statistics and decisions: the importance of communication and the power of graphical presentation (with Discussion). *J.R. Statist. Soc.* A, **140**, 298–323.

Meier, P., Sacks, J. and Zabell, S. L. (1986) What happened in Hazelwood: statistics, employment discrimination, and the 80% rule. In *Statistics and the Law*, (DeGroot, M. H., Fienberg, S. E. and Kadane, J. B., eds.), New York: John Wiley and Sons, Inc., pp. 1–48.

Mosteller, F., Siegel, A. F., Trapido, E. and Youtz, C. (1985) Fitting straight lines by eye. In Exploring Data Tables, Trends and Shapes (D. C. Hoaglin *et al.*, eds) New York: Wiley.

Rock, I. (1984) *Perception*. New York: W. H. Freeman.

Tufte, E. R. (1983) *The Visual Display of Quantitative Information*. Cheshire, Connecticut: Graphics Press.

Velleman, P. F. and Velleman, A. Y. (1986) *Data Desk Handbook*. Ithaca, N.Y.: Data Description Inc.

Weiss, G. (1975) Time reversibility of linear stochastic processes. *J. Appl. Prob.*, **12**, 831–836.

Young, F. W., Kent, D. P. and Kuhfeld, W. S. (1986) *VISUALS: Software* for dynamic hyper-dimensional graphics. In *Proc. Amer. Statist. Ass.* Annual Meeting, Section on Statistical Graphics, pp. 69–74.