A Model for Studying Display Methods of Statistical Graphics

William S. Cleveland

October 18, 1993

Abstract

A method of statistical graphics consists of two parts: a selection of statistical information to be displayed and a selection of a visual display method to encode the information. Some display methods lead to efficient, accurate visual decoding of encoded information, and others lead to inefficient, inaccurate decoding. It is only through rigorous studies of visual decoding that informed judgments can be made about how to chose display methods. A model has been developed to provide a framework for the study of visual decoding. The model consists of three parts: (1) a two-way classification of information on displays — quantitative-scale, quantitative-physical, categorical-scale, and categorical-physical; (2) a division of the visual processing of graphical displays into pattern perception and table look-up; (3) a specification of visual operations that are employed to carry out pattern perception and table look-up. Display methods are assessed by studying the visual operations to which they lead. Studies use the theory and experimental technique of various areas of vision research including psychophysics, cognitive psychology, and computational vision. This process is illustrated by studies of three display methods: visual reference grids for graphs with juxtaposed panels and common scales, encoding a categorical variable on a scatterplot by the type of plotting symbol, and choosing the aspect ratio of a factor-response graph.

1 Introduction

1.1 Scientific Underpinnings

A graphical method for statistical analysis has two components . The first is a statistical component, a selection of information to be shown by the method. The second is a visual component, a selection of a method of display. For example, for a normal probability plot (Wilk and Gnanadesikan 1968), the statistical information is the order statistics of the data together with approximations of their expected values under sampling from a specific theoretical distribution, and the display method is a Cartesian graph with plotting symbols. For a boxplot (Tukey 1977), the statistical information is a five-number summary of the empirical distribution together with end values, and the display method is the familiar box with a line inside, with two appendages, and with plotting symbols to show end values. For a scatterplot matrix (Chambers et al. 1983), the statistical information is simply measurements of three or more quantitative variables, that is, the raw data itself, and the display method is a square array of scatterplots.

Display methods are critical to the success of graphical methods; it is not true that the selection of statistical information is the only critical part, and that virtually any method of display suffices. A display method that leads to inefficient visual decoding can prevent important aspects of data from being detected or can lead to distortions in the perception of information. This is illustrated in Figure 1, which twice graphs the yearly sunspot numbers from 1749 to 1924, the time series used in the landmark paper by G. Udny Yule on autoregressive processes (Yule 1927). The two panels have different aspect ratios. The aspect ratio is critical to visual decoding from factor-response graphs because it affects the judgment of orientations of line segments that encode information about the rate of change of the response as a function of the factor. In Figure 1, the dominant frequency component of variation is the cycles whose

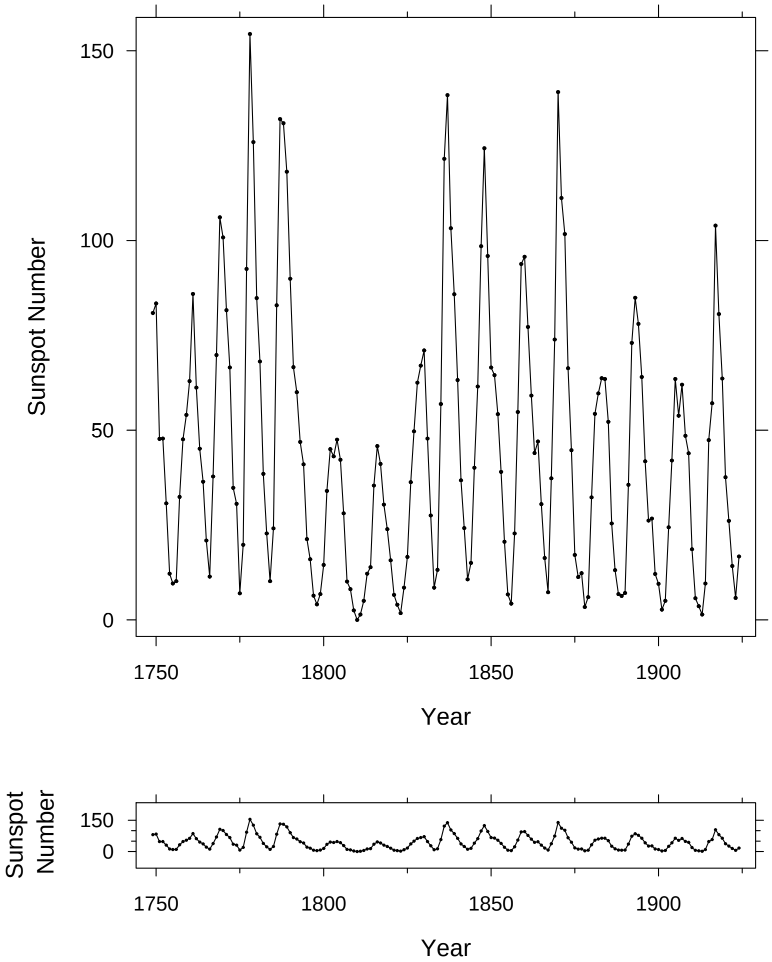

Figure 1: Yearly sunspot numbers with two different aspect ratios.

average period is about 11 years. These cycles are evident in both panels of the figure, but the top panel fails to reveal an important property — the cycles rise more rapidly than they fall.

The theory and methods of statistical science guide the choice of the statistical information contained in a graphical method. The theory and methods of vision science must guide the choice of a display method. Only rigorous study of visual decoding can form a scientific basis for display methods.

There is much to draw on to carry out rigorous study. First, we can exploit the massive body of knowledge that resides in several scientific disciplines that study vision generally — psychophysics (Baird and Noma 1978), cognitive psychology (Treisman 1986), and computational vision (Marr 1982). Theory and experimental results from these disciplines are readily applicable to issues of visual decoding. Second, we can carry out controlled experiments, systematically varying aspects of display methods, and study the effects on the decoding process. We can observe effects in two ways. One is just simple observation of visual displays (Bertin 1967 and 1983). For example, we can readily observe that changing the aspect ratio affects our visual decoding of the rate of change of sunspots. This *simple demonstration* mode of experimentation is sufficient in cases where mechanisms are relatively simple. In more complicated situations we must use systematically varied displays as stimuli in more formal experiments with subjects and response measures (Cleveland and McGill 1984 and 1985). We can borrow from the experimental technique of vision research to run these *measured-response experiments*.

1.2 A Model for Graphical Perception

While the methods outlined above provide ample means for analyzing display methods, more is needed — a basic framework within which to formulate issues of visual decoding. Section 2 of this paper presents a model that provides such a framework.

In the model, information in the display region of a graph is partitioned by a two-way classification scheme. There is *quantitative* information and *categorical* information. Each of these is represented in two ways on a graph — as *scale* information, quantities in units of the data and category names, or as *physical* information, quantities in physical units and categories of graphical elements. Thus the four classes of the two-way classification of information are categorical-physical, quantitative-physical, categorical-scale, and quantitative-scale.

The visual processing of graphical displays is divided into *pattern perception*, a visual decoding of physical information, and *table look-up*, a visual decoding of scale information.

Three visual operations of pattern perception are defined — *detection*, *assembly*, and *estimation*. Three visual operations of table look-up are defined — *projection*, *interpolation*, and *matching*. The central thesis is that the effectiveness of display methods can be investigated by studying the speed and accuracy of these visual operations. This is illustrated in Sections 3 to 5. In Section 3, simple demonstration and Weber's Law, a basic law of psychophysics, are used to investigate visual reference grids on displays where different panels with the same scales must be compared. Section 4 draws on theory and experiment from texture perception to investigate a commonly used display method: encoding a categorical variable on a scatterplot by using different plotting symbols. In Section 5, the aspect ratio of factor-response graphs is investigated by invoking previous measured-response experiments, results from the psychophysical literature about the perception of orientation, and a new derivation.

1.3 Other Work in Graphical Perception

For most of the 20th century there has been a vigorous literature on display methods of statistical graphics. Examples are Brinton (1914), Bertin (1967 and 1983), and Schmid and Schmid (1979)). But until the 1980s, there was very little rigorous study of visual decoding for statistical graphs. The few investigations that explicitly studied issues of visual perception before the 1980s tended not to be guided by existing knowledge from the more general fields of vision research (Wainer and Thissen 1981). With few facts about visual perception to provide guidance, there was little consensus on many display issues. For example, recommendations in the display-method literature for the aspect ratio of a graph abounded and were contradictory (Cleveland, McGill, and McGill 1988).

Bertin (1967 and 1983) partially countered this trend. He formulated issues of visual decoding and proposed techniques for enhancing it. In fact, several of the distinctions made in the model presented here are also made by Bertin. But Bertin, relying solely on simple demonstrations, eschewed principles of vision theory, measured-response experiments, and deductive derivations from basic hypotheses. The result, as we will see later in the paper, is mistaken conclusions about visual decoding.

In the 1980s, research in the visual decoding of information on statistical graphs substantially increased. Section 6 of this paper cites some of this work, in particular, work that serves as the foundation for the model of Section 2.

2 The Model

2.1 A Two-Way Classification of Information

Information in the data region of a graph can be partitioned into *quantitative* and *categorical*. The first consists of numerical values and the second consists of categories. This is illustrated in Figure 2, a multiway dot plot (Cleveland 1993) that shows data from a full factorial experiment (Immer *et al.* 1932). The data have been corrected for an apparent error. The response is barley yield and

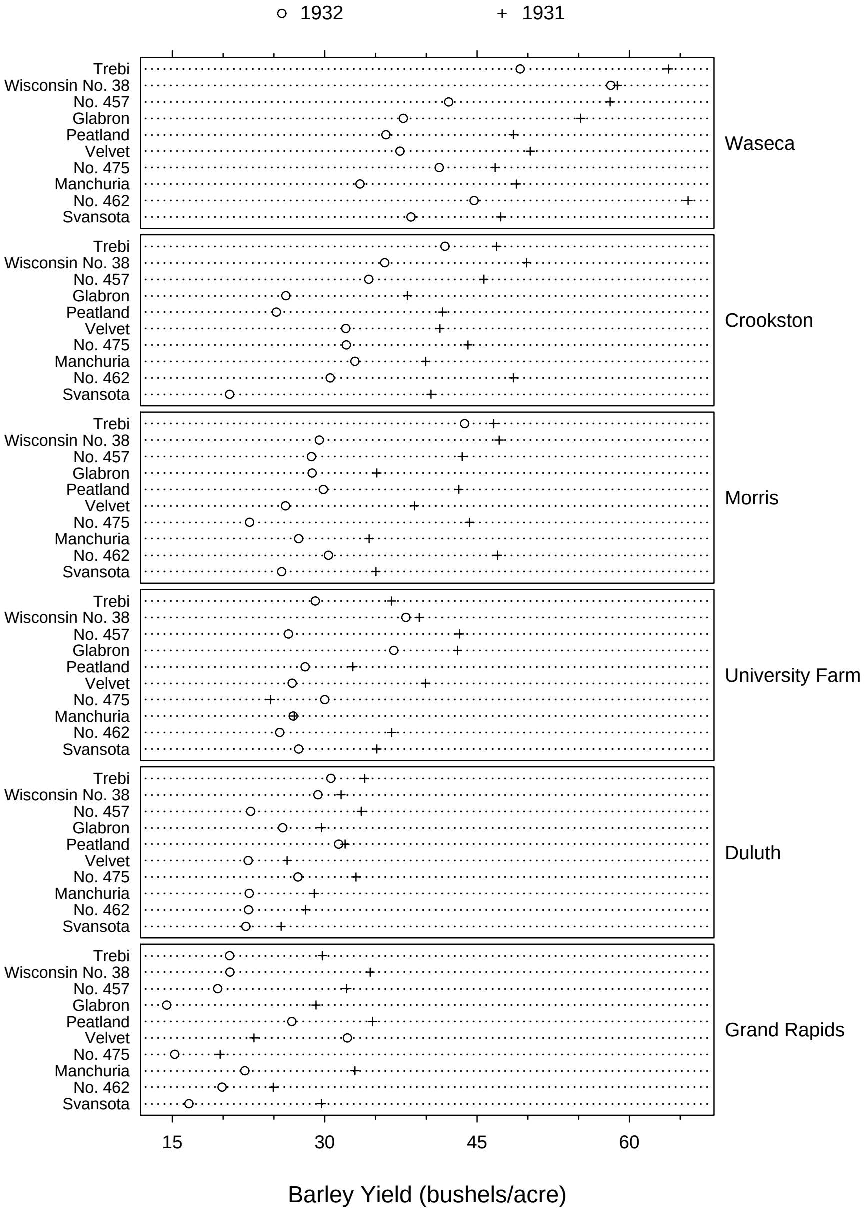

Figure 2: Multi-way dot plot of barley yields.

there are three factors: site (10 levels), variety (6 levels), and year (2 levels). There is one quantitative variable, yield, and there are three categorical variables, the three factors. Thus the information encoded by each plotting symbol consists of four values, one quantitative and three categorical.

Categorical information and quantitative information each can be described in two ways: *scale* and *physical*. For a quantitative variable, scale information is the graphed values in the units of the data and for a categorical variable it is the names of the levels of the graphed values. Consider the data for the largest yield in Figure 2. The scale information is the following: (1) yield = 65 bushels/acre; (2) site = "Waseca"; (3) variety = "No. 462"; and year = "1931". The physical information can be thought of as a description of the quantitative and categorical information on the graph after the removal of the tick labels and the names of levels of categorical variables. We retain the geometry of graphical elements and a breakup of the data into different subsets according to the levels of categorical variables. For the observation with the largest yield, the physical information for the four variables is the following: (1) yield = 7.3 cm (along the horizontal scale); (2) site = "top panel"; (3) variety = "panel-level two from the bottom"; (4) year = "+".

2.2 Pattern Perception and Table Look-Up

The division of information into scale and physical partitions the visual processing of graphical displays into two sets of processes. The first is *table look-up*, a decoding of scale information. For example, we use operations of table look-up to take away the scale information about the observation with the largest yield — 65 bushels/acre, "No. 462", "Waseca", and "1931". The second set of processes is *pattern perception*, a decoding of physical information that leads to the perception of geometric patterns. For example, in Figure 2, we can perceive that on each panel, the "+" symbols as a group are shifted to the right with respect to the "o" and that the overall magnitude of the shift varies with site.

It is crucial to distinguish between pattern perception and table look-up. The output is different, and the visual processing that is necessary to produce the output is different. Table look-up tends to be focused on individual values of quantitative and categorical variables, one after another, with no visual connection of successive look-ups. In pattern perception, we detect and visually group a collection of geometric objects, each carrying physical information, to see patterns. The visual processes can vary from slower sequential operations where we compare physical magnitudes by focused attention, to exceedingly fast processes that operate in parallel, or nearly so, and produce objects, or gestalts, on graphical displays, much as trees, dogs, and cars coalesce when we view everyday scenes (Treisman 1986).

2.3 Three Operations of Pattern Perception

The model specifies three operations of pattern perception — *detection*, *assembly*, and *estimation*. Each has associated with it a notion of efficiency that is invoked in experimental or theoretical studies of the operations.

Detection is the visual recognition of a geometric aspect that encodes a physical value. For example, in Figure 2, each plotting symbol encodes a single value of each of the four variables on the graph — yield, variety, site, and year. In this case, detection is the recognition of the symbol. The line segment connecting the "+" to the "o" on each level of each panel encodes the yield difference between years for one combination of variety and site. In this case, detection is the recognition of the segment. The efficiency of a detection operation depends on its *latency*, the time it takes to carry out the operation. If a physical characteristic fails to trigger our sensory processing — for example, if one symbol is obscured by the overplotting of other symbols — then we fail to detect at all and the latency is infinite. In Figure 2, the use of "o" and "+", as we will see in a later section, enhances this efficiency.

Assembly is the visual grouping of detected graphical elements. For example, on each panel of Figure 2, we can cluster the "o" symbols as a whole, mentally filtering out the other elements. As with detection, an assembly operation has a latency associated with it that determines the perceptual salience with which the pattern forms. In

Figure 2, the use of "o" and "+", as we will see in a later section, enhances this efficiency.

Estimation is the visual assessment of the relative magnitudes of two or more quantitative physical values, a and b. There are three progressive levels of this operation: *discrimination*, *ranking*, and *ratioing*. Discrimination refers to simply determining if $a = b$ or if $a \neq b$. The next level is ranking, deciding among $a > b$, $a < b$, and $a = b$. Ratioing is the judgment of a/b. In Figure 1, estimation of the orientation of line segments gives us information about the rate of change of the sunspots through time. In Figure 2, estimation of the lengths of line segments connecting the two symbols on each dotted line gives us information about the yield differences between years. The efficiency of the estimation operation is the accuracy of the estimates. It is not necessary that a ratioing operation provide several significant digits of accuracy for pattern perception to be informative. As Tukey (1993) points out, the result of an efficient estimation operation may simply yield one of the following statements: "is way below", "is below", "is a little below", "is almost on", "is a little above", "is above", and "is way above". Still, in running measured-response experiments to probe the efficiency of estimation, it can make sense to ask subjects to provide a quantitative value for a judged ratio, simply as an experimental expedient, even though pattern perception does not entail such a cognitive result (Cleveland and McGill 1984 and 1985). In Figure 2, aligning the panels vertically enhances the accuracy of estimation of differences of yields — for example, comparing the overall locations of 1931 yields at two sites — because all data are graphed along a common scale (Cleveland and McGill 1984 and 1985). For example, the accuracy would decrease were we to arrange the panels in a three by two rectangular array.

2.4 Three Operations of Table Look-up

Three operations of table look-up are *scanning*, *interpolation*, and *matching*. They are all invoked in decoding the scale information of the observation with the largest yield in Figure 2.

To get the scale value of the largest yield we scan and interpolate: (1) scan up or down to the horizontal scale line to fix a point along the axis; (2) interpolate by estimating the distance from the point to the tick mark to the left as a fraction of distance between tick marks. Then another process takes over; we read the tick mark labels and use the physical interpolation to convert to an interpolation in data units. For most people, the decoded scale value would be about 65 bushels/acre. By scanning to the left we get "No. 462" for the variety. The dotted lines of the dot plot increase the speed and accuracy of this table look-up operation. Note that visual guides, say in the form of classical grid lines, are not used for the table look-up of yield since without them there is sufficient accuracy to give a rough sense of the scale of the data. But for the variety variable, we cannot allow a rough decoding; for example, decoding the nearby label "Svansota" in place of "No. 462" would not be acceptable. Finally, by scanning to the right we get "Waseca" for the site.

Matching provides the table look-up for the year variable. By matching the "+" encoding the value with the "+" in the key, and then reading the label, we get "1931" for the year.

2.5 What the Model Achieves and What it Does Not

The model posits a collection of visual operations that are to be considered in the study of a display method. It does not prescribe which of its visual operations are salient in any particular case, nor does it provide information about the efficiency of the salient visual operations once they have been determined. Determinations of salience and efficiency must come from basic reasoning, the theory of visual perception, simple demonstration, and formal experimentation. But the model does play a vital role in specifying a sufficiently rich collection of operations that progress can be made in understanding the efficacy of many display methods by considering them. The examples in the next sections demonstrate this.

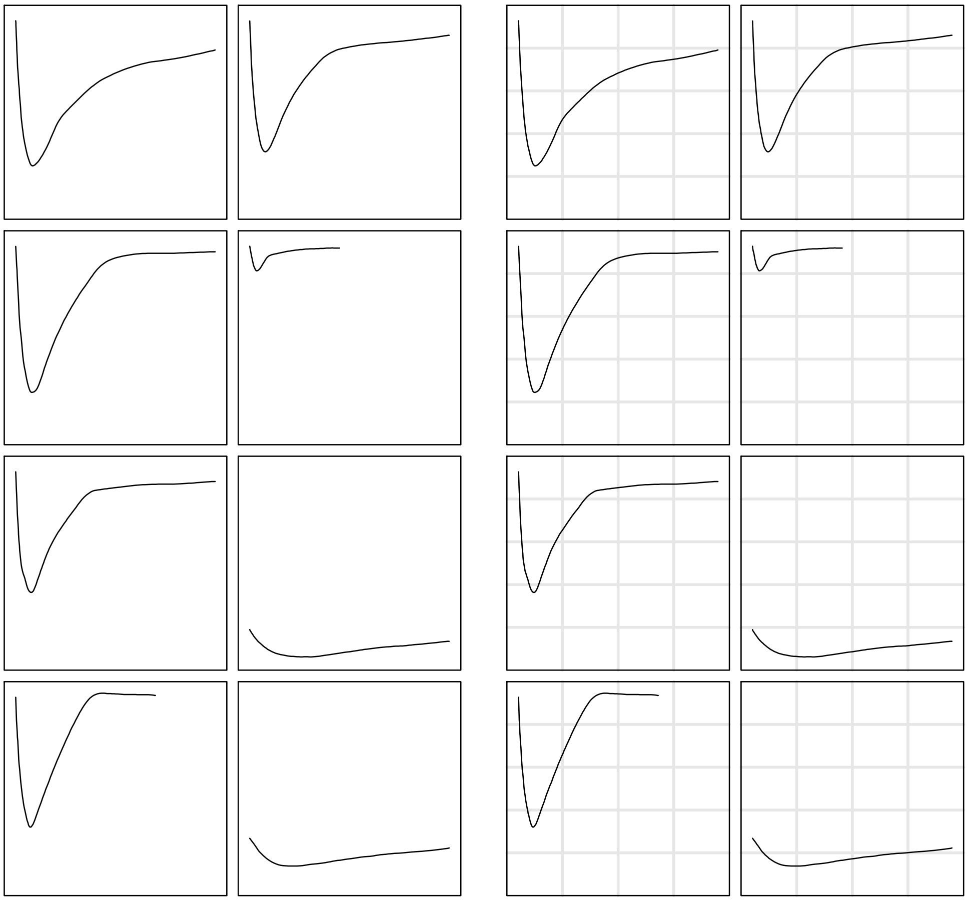

Figure 3: A visual reference grid for juxtaposed panels.

3 Visual Reference Grids

3.1 Operations of Pattern Perception

The left two columns of Figure 3 graph eight curves by juxtaposition: each data set is on a separate panel. We can study the properties of an individual curve from its geometric aspects. For example, as with the sunspots, we detect, assemble, and estimate local orientations to study the relative rate of change. For the curve in the top row and the left column, this visual process conveys the following: from the starting value to the minimum, the curve is convex; from the minimum to the final value, it is concave; and the concave section, overall, has greater curvature than the convex section.

We can also compare corresponding aspects of different curves. For example, we can compare the minima. There are a variety of visual processes for doing this. One is to detect and cluster the vertical line segments that extend from the bottom lines of the boxes to the points on the curves where the minima occur, and then carry out estimation of the lengths of these segments. The assembly of these segments must be carried out by a highly attentive sequential search process with rapid shifts of our eyes from one panel to the next. Contrast this with the decoding of the rate of change of the curve in the upper panel. The detection and assembly of the local line segments along the portion of the curve from the beginning to the minimum is carried out considerably faster. In fact, in doing this, our visual system produces a gestalt, a curve object, that we see as a single perceptual unit. A widely accepted model of vision proposes that such an object forms in the following way: initial parallel operations simultaneously extract local features such as the orientations of local segments; different local features are saved in different files; by focused attention, the visual system assembles the features to form the object (Treisman 1986). The much slower sequential operations that we must use to compare minima do not result in the formation of an object. We cannot readily hold all segments in short term memory, and even our cognitive processing, our conscious conclusion about relative values, must be built up from a sequence of comparisons of subsets of values. Nevertheless, the comparison of minima and other commensurate aspects of the curves is an informative visual process.

3.2 Grids and Table Look-Up

Grid lines are an old-fashioned display method that has largely been abandoned in the display of scientific data because of the ubiquitous use of computers to produce graphs and communicate data. In the era when graphs were made by hand, they were also data transmission devices, and grid lines were an important display method for increasing the accuracy of table look-up. Today, such accuracy is not needed, and the use of grid lines has declined dramatically.

3.3 Grids and Pattern Perception

Grid lines, however, still have a role to play — the enhancement of pattern perception on graphs with juxtaposed panels. In the right two columns of Figure 3, grid lines have been added to graphs of the curves from the left two columns. The entire *visual reference grid* consists of these new lines together with the solid box. Since the purpose of the grid is no longer to enhance drawing or data transmission, the grid lines need not be drawn at simple numerical values. Rather, the important property is that they form equal intervals in the horizontal direction and in the vertical direction.

Simple demonstration shows that visual reference grids enhance the estimation of commensurate values from different panels. Figure 3 is one example. Consider just the five curves that have deep troughs. Among other things, the visual reference grid enables us to more readily carry out the estimation operation for the minima. We can do this by detection, assembly, and estimation of the five line segments extending from the points of the minima on the curves to the horizontal grid lines just above them, or by estimation of the five lines extending down to the grid lines below. The estimation operation involves ranking and crude ratioing. Weber's Law (Baird and Noma 1978, Chapter 3) implies that the accuracy of estimation for each of these two collections of segments is greater than the accuracy of estimation of the collection of segments from the minima to the bottom box lines. Each of the former has lengths whose pairwise percentage differences are greater than those of the latter. Weber's Law states that the accuracy of discrimination of two lengths l and $l + \delta l$ increases with $\delta l / l$, and is independent of l except for very small l.

It might be argued that a better solution than visual reference grids is to superpose the curves in the same data region. There are two reasons why this is not a good general solution. First, the overlap of curves can slow down the detection and assembly of values of each individual curve, or even prevent it altogether. Second, vertical differences between curves are typically poorly detected. Instead, our system detects minimum rather than vertical distances, with a result of frequent distortion of patterns (Cleveland 1985, pp. 271-274; Warren 1989).

4 Plotting Symbols On Scatterplots

4.1 Operations of Pattern Perception

Figure 4 has scatterplots in which symbol type encodes a categorical variable. The plotting symbol set affects our ability to carry out the visual operations of detection and clustering for each group of symbols. For example, assembly is more efficient in the top panel of Figure 4 than in the bottom. In the bottom panel we must use slower sequential search to locate the symbols of each group, and this detracts from our ability to assemble, but in the top panel, the group gestalts form with strong perceptual salience. The top panel uses a symbol set whose origin will be described shortly. The bottom panel uses letters that are visually confusable (Lewandowsky and Spence 1989).

If color is available, simple demonstration shows convincingly that it provides highly efficient assembly that typically performs noticeably better than different symbol types in black and white. Experimental results support the results of theses demonstrations (Lewandowsky and Spence 1989). So too do theories of vision which stipulate that color is extracted locally and in parallel by fast initial visual operations (Treisman 1986).

4.2 Detection

Suppose the medium is black and white and that symbols might overlap. The second requirement restricts us to symbols whose abstract representations, like regular polygons and the letters of the alphabet, consist of *corecue symbols*. Such a symbol is a connected set consisting of a finite number of rectifiable curves. Using these symbols enhances the visual operation of detection when the symbols overlap. Symbols such as a filled circle whose abstract mathematical representations have Lebesque measure greater than zero, tend to form uninterpretable blobs in the presence of overlap, so detection is degraded. We do not consider unconnected symbols because in the presence of overlap, different pieces of an unconnected symbol can too easily be rendered visually separated, so again, detection is degraded.

4.3 Assembly and Texture Perception

Picture a room with a plaid rug and a polka-dot wall. In the language of vision theory, the wall and the rug have distinct *textures*, micropatterns with much local variation that nevertheless take on a uniform appearance to our visual system. Where the wall and rug meet, the visual system perceives a clearly defined boundary between the texture patterns. One current of thinking in vision theory proposes that boundaries are formed through parallel processing of the local features of the micropatterns (Julesz 1981). But some micropatterns with very different local features nevertheless appear much like one another. Consider the seven panels of Figure 5. In each, there are two micropatterns left and right with different local features, but there is substantial variation in the perceptual salience of the boundaries between the patterns. For example, there is a strong boundary for "+" and "▫", a weaker one for "T" and "L", and none for "R" and "mirror-R".

What properties of the local features result in strong boundaries and what properties are ignored? This question has been intensively stud-

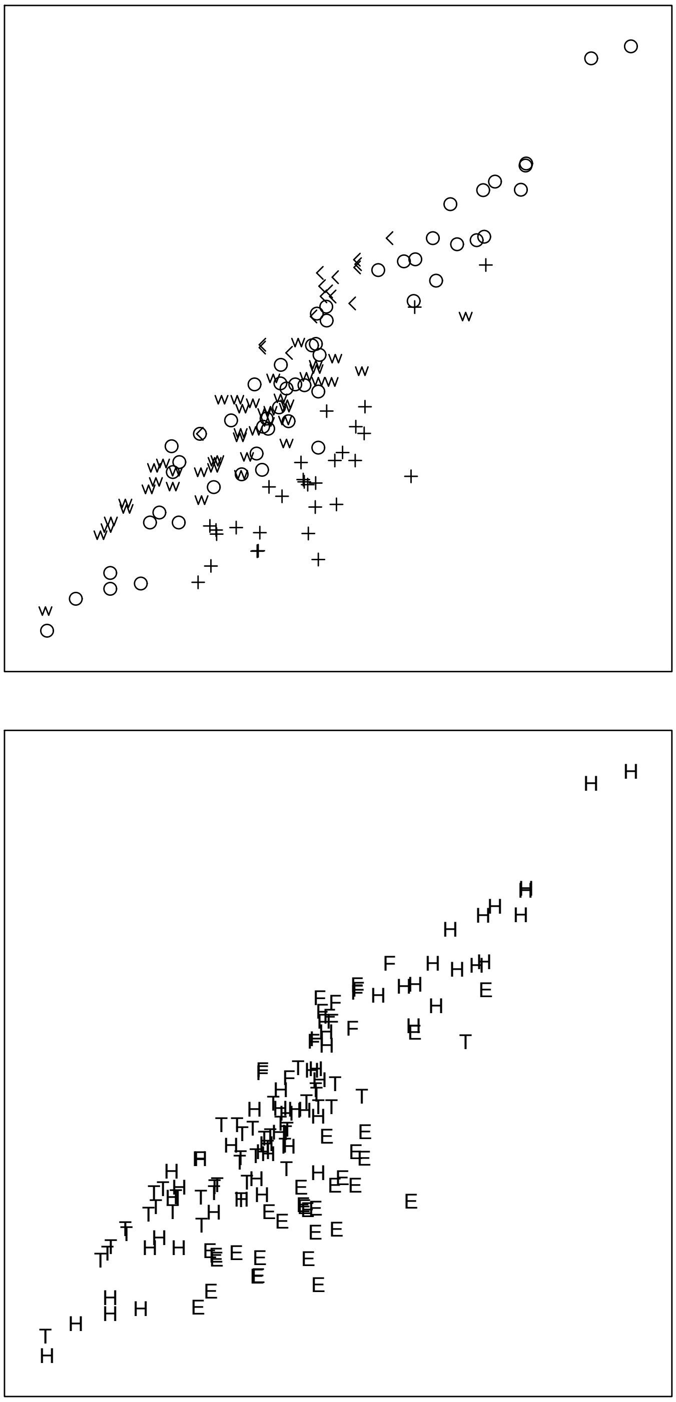

Figure 4: Encoding of categorical variable by symbols from vision experiments (top panel) and by confusable letters (bottom panel).

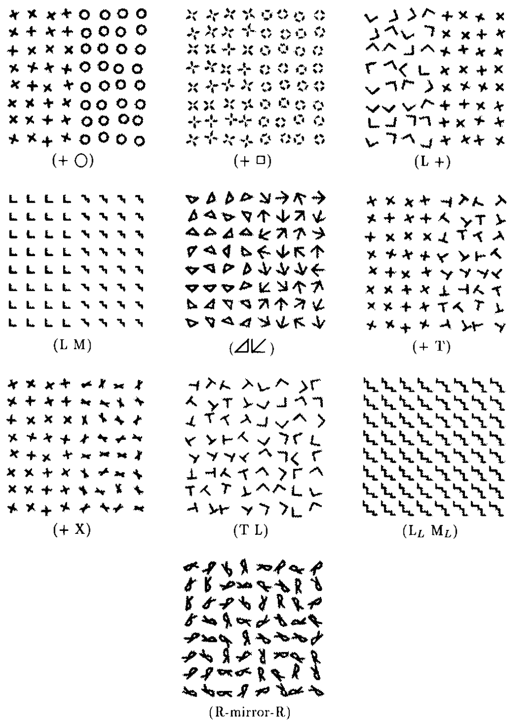

Figure 5: Texture patterns. Reproduced with permission from Malik and Perona (1990).

Texture Pair	Experiment	Algorithm
$+$ o	100	407
$+$ □	88.1	225
L $+$	68.6	203
L M		165
Δ $\downarrow$	52.3	159
$+$ T	37.6	120
$+$ X	30.3	104
T L	30.6	90
L_L M_L		85
R mirror-R		50

Table 1: Texture Discrimination.

ied by vision researchers because texture perception is one of the important initial processes of our visual algorithms. In many cases, researchers have used texture patterns made up of corecue symbols (Julesz 1981; Marr 1982; Treisman 1986; Kröse 1987; Voorhees and Poggio 1988; Fogel and Sagi 1989; Malik and Perona 1990). An example is Figure 5, which is from a study of computational vision (Malik and Perona 1990).

One method for studying texture perception is measured-response experiments. For example, Kröse (1986 and 1987) describes an experiment in which subjects had to determine the presence or absence of a single occurrence of a corecue symbol of a certain type drawn among a texture pattern of corecue symbols of another type. The target was drawn just once in a random position in the texture for 60% of the stimuli, and was omitted in the others. Each stimulus was shown for 80 ms, and subjects indicated whether or not they saw the target symbol. The percentage of correct answers, corrected for guessing, measures the degree of texture discrimination. Column 2 of Table 1 shows percentages from the Kröse experiments for seven of the symbol pairs in Figure 5.

The second method of study of texture perception is image processing of texture patterns using algorithms of computational vision that attempt to reproduce the algorithms of the human visual system (Marr 1982; Voorhees and Poggio 1988; Fogel and Sagi 1989; Malik and Perona 1990). The algorithms employ banks of linear and nonlinear filters. Malik and Perona (1990) applied such a texture algorithm to each panel of Figure 6 to produce a texture image. As a measure of the degree of texture discriminability, they computed an average of the image gradient over the vertical boundary between the two regions of symbols. This measure is shown in the third column of Table 1. Clearly the experimental and algorithmic measures agree quite well.

Two corecue symbols that provide strong boundaries also provide efficient assembly on scatterplots. In studies of texture perception, different patterns almost always do not cover regions that overlap, as would the point clouds of many scatterplots, but Tremmel (1992) gives compelling arguments that the results of the studies are relevant to overlapping patterns as well.

4.4 A Set of Four Corecue Symbols That Enhance Assembly

Bertin (1967 and 1983, p. 67) demonstrates the poor performance of corecue symbols consisting of simple geometric figures such as circles and regular polygons. But Bertin ends his discussion with this limited experimentation, not considering any other corecue symbol, and not invoking the large literature on texture perception available even in 1967, and declares *shape (forme)* not to be useful for assembly, even though he correctly recognizes that orientation is an important element of pattern perception.

Here, we allow a much wider range of corecue symbols, and use the results of studies of texture perception — in particular, those reported in Table 1 — to choose a symbol set. Table 1 suggests that the four symbols labeled "o", "L", "+", and "M" are good candidates. The discrimination measures for those pairs of these symbols that appear in the table are high. Simple demonstrations — for example, the top panel of Figure 4 — show that this selection does quite well in practice.

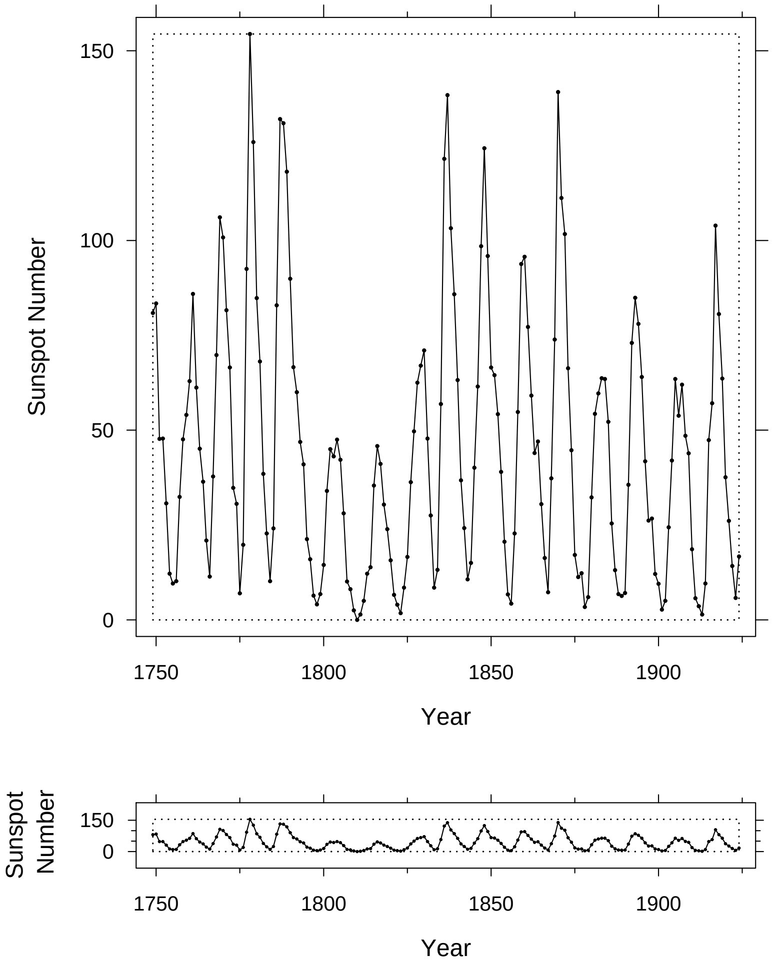

Figure 6: Yearly sunspot numbers with data rectangles.

5 Aspect Ratio

5.1 Definitions

Figure 6 graphs the sunspots again, but this time with a *data rectangle* on each panel, a box that just encloses the data. We saw earlier that local properties of the sunspots were better revealed in the bottom panel. A derivation in this section leads to a display method for choosing the aspect ratio that maximizes the accuracy of orientation estimation, which leads to an enhanced pattern perception of curvature.

Let v be the length of a vertical side of the data rectangle in physical units, and let h be the length of a horizontal side. Then the aspect ratio is $\alpha(h, v) = v/h$. In the top panel of Figure 6, $\alpha(h, v) = 1$ vcm/hcm, where "vcm" means vertical cm and "hcm" means horizontal cm, and in the bottom panel, $\alpha(h, v) = 0.055$ vcm/hcm. Let $\ddot{v}$ be the vertical length of the data rectangle in scale units. In Figure 6, $\ddot{v} = 154.4$ sn, where sn stands for sunspot number. Let $\ddot{h}$ be the horizontal length of the data rectangle in scale units. In Figure 6, $\ddot{h} = 174$ yr. Now $v/\ddot{v}$ is the conversion factor that takes the length in the units of the data and changes it to a length in physical units on the graph. Similarly, $h/\ddot{h}$ is the conversion factor for the horizontal scale.

A line segment on a graph has a *scale slope* — a slope measured in data units (sn/yr in Figure 6) — and a *physical slope* — a slope measured in physical units (vcm/hcm). Consider a collection of n line segments on a graph. Let $\ddot{v}_i$ be the change in data units of the ith line segment along the vertical scale, and let $v_i(v)$ be the change in physical units when the vertical length of the data rectangle is v. Define $\ddot{h}_i$ and $h_i(h)$ similarly. Then $v_i(v)/h_i(h)$ are the physical slopes and $\ddot{v}_i/\ddot{h}_i$ are the scale slopes. Now,

$$\frac{v_i(v)}{h_i(h)} = \alpha(h, v)\frac{v_i(1)}{h_i(1)}$$

so if $\alpha(h, v)$ is multiplied by a constant, all of the physical slopes are multiplied by the constant. For example, in going from the top to the bottom panel of Figure 6, all physical slopes are multiplied by 0.055 because in the top panel $\alpha(h, v) = 1$ vcm/hcm, and in the bottom panel, $\alpha(h, v) = 0.055$ vcm/hcm.

5.2 The Operation of Orientation Estimation

Consider a collection of line segments on a graph for which detection, assembly, and estimation provide us with information about rate of change. What estimation task is the basis for the pattern perception that gives us information about rate of change? Simple arguments suggest that it is not estimation of the physical slopes of the segments but rather estimation of their orientations,

$$\theta_i(h, v) = \arctan\left(\frac{v_i(v)}{h_i(v)}\right).$$

First, Marr (1982) and Stevens (1983) argue that the estimation of slant and tilt in real 3-D scenes is based on orientation and not slope. Their studies apply equally well to graphical displays. Second, suppose we have two line segments with orientations θ and $\theta + \delta$. Suppose the physical slopes of the segments are s and $s(1+\tau)$. Our ability to carry out the estimation operation of discrimination for the slants of the two segments is nearly independent of θ (Rochlin 1955) and depends on δ. As δ increases, we can more readily discriminate. Similar statements do not hold for s and τ. As s gets either large or small for fixed τ, we lose all ability to discriminate because the value of δ goes to zero. In other words, δ, not τ, is the determining factor of discrimination.

5.3 Optimization of the Discrimination of Orientation

Suppose the line segments to be judged are the successive segments that form a curve on a graph. One example is Figure 6. This is a special case, but the result of the ensuing derivation holds generally. Let the orientations be $\theta_i(h, v)$, where i increases as we go from left to right on the graph. In view of the discussion in Section 5.2, we can enhance our perception of the local rate of change of the function by making the values of $|\theta_{i+1}(h, v) - \theta_i(h, v)|$ as large as possible through the choice of $\alpha(h, v)$; this makes the discrimination of the orientations of successive local segments as accurate as possible, and thus enhances estimation generally.

The ensuing derivation will be simpler if we consider a curve that is twice differentiable over an interval of interest. Let the physical coordinates of the curve be $(x, \alpha f(x))$ when the aspect ratio is α. (For this derivation we will not show the dependence of α and h on v.) The goal of having orientations of successive line segments be as different as possible in the discrete case becomes the goal of having

$$g(x) = \arctan(\alpha f'(x))$$

change as much as possible at each x, which means having the derivative

$$g'(x) = \frac{\alpha f''(x)}{1 + \alpha^2 f'^2(x)}$$

be as large as possible in absolute value. Thus, for fixed x in $[a, b]$, we seek the α that maximizes

$$h(\alpha) = \left|\frac{\alpha f''(x)}{1 + \alpha^2 f'^2(x)}\right|.$$

Suppose $f''(x) \neq 0$; otherwise, α has no effect on the variation of g at x. Suppose also that $f'(x) \neq 0$; otherwise, there is in some sense no solution since no matter what the value of α, there is an interval about x in which $g(x)$ is arbitrarily close to a constant. Since $h(\alpha) \to 0$ as $\alpha \to 0$ or $\alpha \to \infty$, we know that a maximum exists. Setting $h'(\alpha)$ equal to zero yields the solution

$$\hat{\alpha}(x) = |f'(x)|^{-1}.$$

In other words, the α that makes $g(x)$ change as much as possible is the α that makes the orientation of the tangent at x equal to $\pi/4$ radians, or $45°$, if $f'(x) > 0$ and $-\pi/4$ radians, or $-45°$, if $f'(x) < 0$.

Thus the derivation shows that the perception of rate of change in a neighborhood of x is optimized by this $45°$ principle. A simple heuristic supports the principle. If $y = f(x)$ is monotone in the neighborhood, then symmetry suggests that optimum pattern perception of curvature should not depend on whether we think of y as a function of x or x as a function of y. A similar derivation of the optimum angle, taking x as a function of y leads to a result of $45°$ with the vertical, which implies $45°$ with the horizontal, the result when y is taken as a function of x. $45°$ is the only angle that can result in the same answer for the two derivations.

5.4 A Compromise Method

Putting this $45°$ principle to use in practice requires a compromise; different values of x typically have different solutions, $\hat{\alpha}(x)$, so we cannot optimize simultaneously for all x with a single α. The following describes one compromise method.

Let us return to the general case now of n line segments whose orientations are to be judged. Let

$$\bar{v}_i = \ddot{v}_i / \ddot{v}$$

and

$$\bar{h}_i = \ddot{h}_i / \ddot{h} .$$

The orientation of the ith segment is

$$\theta_i(h, v) = \arctan\left(\frac{v_i(v)}{h_i(v)}\right) = \arctan\left(\alpha(h, v)\bar{v}_i/\bar{h}_i\right) .$$

The physical length of the ith segment is

$$\ell_i(h, v) = \sqrt{h_i^2(h) + v_i^2(v)} = h\sqrt{\bar{h}_i^2 + \alpha^2(h, v)\bar{v}_i^2} .$$

The mean absolute orientation weighted by the line segment lengths is

$$\frac{\sum_{i=1}^{n} \theta_i(h, v)\ell_i(h, v)}{\sum_{i=1}^{n} \ell_i(h, v)} = \frac{\sum_{i=1}^{n} \arctan\left(\alpha(h, v)\bar{v}_i/\bar{h}_i\right) \sqrt{\bar{h}_i^2 + \alpha^2(h, v)\bar{v}_i^2}}{\sum_{i=1}^{n} \sqrt{\bar{h}_i^2 + \alpha^2(h, v)\bar{v}_i^2}} .$$

One compromise method is to choose $\alpha(h, v)$ so that this weighted mean is $45°$. There is no closed form solution, but $\alpha(h, v)$ can be determined very rapidly by approximation since the weighted mean is monotone in $\alpha(h, v)$. For the sunspot data, the segments are the local segments connecting successive yearly values. The resulting aspect ratio is 0.055, the value used in the bottom panel of Figure 6, which, as we have seen, leads to successful pattern perception of local curvature.

The $45°$ compromise method is general and applies to the estimation of orientations of any collection of line segments. For example, it was applied to all of the local segments of the eight curves in Figure 3 to choose the aspect ratio of the common scales. Of course, any compromise method, in attempting to optimize overall, can optimize nowhere, but in general the mean-orientation method does remarkably well. In part, its good performance is due to an invariance property. Suppose a curve is well-approximated by two different collections of line segments. Then the two compromise values of $\alpha(h, v)$ will be nearly the same.

For a single set of data on a graph there may be two or more sets of line segments that convey quite different blocks of information that are all of interest. This can result in the necessity to draw the graph more than once with different aspect ratios. The above application to the sunspot numbers targeted local line segments to enable a study of local properties. If the compromise method is applied to the segments connecting the cycle peaks, which enhances the perception of the variation of peak values through time, the resulting aspect ratio of the graph is 0.30.

One reaction might be that if studying rate of change is so vital, then we should measure rate of change and graph it directly. This is undoubtedly sometimes true. But data analysis is not always that obliging. Graphs of a response against a factor often lead to the discovery of a relationship between rate of change and level, and a graph of rate of change by itself would not, of course, reveal this. Such a relationship exists for the sunspots. The faster rise than fall is most pronounced for the cycles with high peaks, is less pronounced for those with medium peaks, and disappears for those cycles with the very lowest peaks. Thus, we cannot avoid judgments of dy/dx from a graph of y against x. Furthermore, as soon as we graph dy/dx against x, we find ourselves judging d^2y/dx^2.

5.5 Other Work

The 45° principle arose first from measured-response experiments of Cleveland, McGill, and McGill (1988) who show that estimation of rate of change is most accurate when orientations are centered on 45°. The work of these authors, however, rests on a premise of slope estimation as a basis of pattern perception. The theoretical reasoning presented here suggests this is not the case. But the empirical results of the experiment remain valid and are complemented by the arguments given here. In addition, Cleveland, McGill, and McGill suggest a compromise method in which the median absolute orientation is set to 45°. This method does not have the invariance property of the weighted-mean method, and does not do as well in practice.

Bertin (1967 and 1983, p. 178) also addresses the effect of the aspect ratio on judging rate of change. He correctly identifies the role of orientation and the effect of aspect ratio on orientation. Bertin identifies two cases. The first is the *elementary level (niveau élémentaire)*. By this, he means the local properties conveyed by orientations of local line segments. His solution for the elementary level is to choose the aspect ratio to center orientations on 70°, a number given without any supporting argument. The above arguments and empirical results demonstrate that this value is incorrect. Bertin's second case is the *ensemble level (niveau de l'énsemble)*. By this he means the information contained in a low-frequency trend curve on a time series graph or a regression curve on a more general factor-response graph. His solution, again, given without supporting arguments, is an aspect ratio of one. This is also incorrect. The 45° principle applies equally well to to a trend curve, and the resulting aspect ratio depends on the properties of the curve and often deviates substantially from one.

6 Discussion

While the composition of the model is new, each of its components has been selected from among many visual operations discussed elsewhere, either implicitly or explicitly. Bertin (1967, p. 13, and 1983, p. 11) defines a visual operation called *selection (sélection)* that corresponds, for practical purposes, to assembly. This assembly operation is also defined and investigated by Cleveland (1985, p. 191) using simple demonstration, by Lewandowsky and Spence (1989) and Tremmel (1992) using measured-response experiments, and by Kosslyn (1989) and Pinker (1990) using the theory of cognitive psychology. Cleveland (1985, p. 291) defines and investigates the detection operation, which is similar to the operation of *absolute discrimination* of Kosslyn (1989). Bertin defines an operation of *identification (identification)* (1967 and 1983, p. 19), but it is much broader and refers to a cognitive process whose end result is an understanding of the information that is encoded by a set of graphical elements. Cleveland and McGill (1984 and 1985) present a comprehensive study of the ratioing operation of estimation, drawing on both measured-response experiments and vision theory. But they do not consider discrimination and ranking. This and other work on ratioing is reviewed by Spence and Lewandowsky (to appear). Bertin (1967 and 1983, p. 96) defines *quantitative perception (perception quantitative)*, an operation that refers to estimation of *size (taille)*. He does not allow for estimation of orientation beyond discrimination, and does not differentiate different size encodings by accuracy.

The salient aspect of the model is its composition — the judgment of which components of the vastly complex process of vision to include. The choice of components does not have the goal of providing a complete description of the entire process that goes from sensory input on the retinas to conclusions from the data (Kosslyn 1989; Pinker 1990). Rather, the goal is an engineering one: devise a framework that can guide investigations of display methods of statistical graphics in order to improve the methods. Thus the development of the model requires judgments of the likely benefit of many different potential aspects that could be included. What is left out is as important as what has been included. For example, the model contains no delineation of graph schemata, which are templates stored in memory that are matched to input and aid in our recognition of objects (Kosslyn 1989; Pinker 1990). On the other extreme, the model includes more than an earlier model of Cleveland and McGill (1984 and 1985), which attempted to assess display methods solely on the basis of the accuracy of estimation.

The selection of the model components, the resources outlined for investigating them, and the examples used to illustrate their application to studying display methods, all combine to make the framework one that focuses on perceptual processing. Other currents of work have had different emphases.

Haber and Wilkinson (1982), Kosslyn (1985 and 1989), Simkin and Hastie (1987), Tversky and Schiano (1989), Pinker (1990), and Lohse (1991), all within a with a more general framework of visual information processing, have investigated sensory, short-term memory, and long-term memory processes. Spence and Lewandowsky (to appear) provide a review of this current of research. Much in these treatments draws heavily on the more general theory of mental imagery (Hastie 1981; Kosslyn 1980; Spoehr and Lehmkuhle 1982; Ullman 1984). The Pinker work, widely disseminated in preprint form for a number of years, consists of a model with a sequence of stages that describe the movement of information from sensory processes to conscious interpretation of the data. A critical element is the delineation of schemata that control the organization of sensed visual information and that provide for interpretation. Simkin and Hastie make hypotheses about operations of visual processing, and probe these hypotheses with experiments. Kosslyn presents a comprehensive analytic scheme whose goal is to provide a method of assessing the effectiveness of graphs. The physical layout of a graph is broken up into four pieces (framework, specifier, labels, and background), and there are three levels of analysis (syntactic, semantic, and pragmatic). Tversky and Schiano study long-term memory processes, demonstrating distortions of graphical patterns held in memory. Finally, Lohse (1991), in an interesting paper, describes a computer program that models perceptual and cognitive processes.

Another current of research takes a more behavioristic approach. For example, Dickson, DeSanctis, and McBride (1986) and Benbasat, Dexter, and Todd (1986) describe experiments that probe the decisions people make from studying graphical displays. Maceachren and Ganter (1990) discuss behavior, but focus on epistemology, and present a model of how insight is provided to geographers and earth scientists by cartographic visualization.

The justification for limiting the model presented here to perceptual processing is that it provides satisfactory analysis of many display methods of statistical graphics. Thus, while limited, the model has useful applications. This should come as no surprise. Studies of the early stages of visual processing have shown that perceptual processes, described in the form of parallel local operations on images, can provide representations of visual scenes (Julesz 1981; Marr 1982; Treisman 1986). While these representations are crude, it must be remembered that the visual world of most 2-D graphical data display is itself crude, lacking the richness of natural images. The visual world of 3-D graphical data display on computer screens is another matter (Livingstone and Hubel, 1988; Wanger, Ferwerda, and Greenberg, 1992).

7 Rejoinder

The comments of the discussants reflect a thorough knowledge of visual perception and add substance to the issues of the paper. This speaks well of the ability of a new journal to attract researchers in vision, bringing greater depth to the important area of graphical perception.

Since the journal sits at the interface between computing and statistics, I will use the e-mail style of comment and response, but showing the comments in italics rather than the usual ">" at the beginning of each line.

7.1 Comments of Tremmel

Obviously, Cleveland's 'pattern perception' is related to the 'preattentive vision', and his 'table look-up' to the *'attentive' component. However, there are differences:* This comment focuses on an issue of great importance for graphical perception. But I did not intend to equate table look-up with attentive vision and pattern perception to preattentive vision. In fact, I think it important to define the two as I did in terms of scale and physical information, and then determine the visual operations of each. However, as Tremmel suggests, table look-up tends toward the attentive and pattern perception toward the preattentive.

The author also states that table look-up is the first process, and pattern perception the second one.

I did not mean to imply this order and agree with the statement of the discussant. For the most part, operations of table look-up have as their starting point features recognized by the operations of pattern perception.

Unfortunately, these concepts ('unconnected rectifiable curves', 'planar measure', 'Connectedness' and 'Closure' - both of which is not the same) are not well defined or explained, and no references are given.

A corecue symbol is an abstract mathematical object with no planar area. The symbol is rendered by drawn curves, which have a planar area. In most cases we want to minimize this area, using drawn curves that are as thin as possible without losing our ability to detect them. By "curve" I mean a continuous function from R_n into the plane. By "rectifiable" I mean that the curve can be arbitrarily well approximated by an inscribed polyline. (A good thing, since otherwise no finite resolution device can show the properties of the curve.) By "connected" I mean that the symbol cannot be covered by two open sets whose intersections with the symbol form two disjoint sets.

But still I felt compelled to comment on the phrase that "our ability to carry out discrimination of the slants of the two segments is nearly independent of theta" (the orientation) because it is not true.

The important word here is "nearly". As Tremmel later points out, this was an engineering decision: ignoring an effect judged not to be important for the application. The judgment was based on the quantification of the effect by Rochlin (1955), who carried out a very thorough analysis. However, it would be possible to carry out a derivation similar to that in the paper that took account of the small perturbation in discriminability.

7.2 Comments of Holmes

Table lookup and pattern recognition characterize two separate brain functions that I will label, for historical reasons, left-brain and right-brain functions. I like this description, not because it is literally true that brain functioning is so neatly divided, but rather because it nicely delineates the tasks we perform into cognitive and perceptual. Table look-up has a cognitive component. So do other tasks of visualizing data. For example, drawing conclusions about the data by interpreting the patterns that form as a result of the visual tasks of pattern perception, is a cognitive task.

Training is necessary for efficient graph reading because access to the left brain interferes with efficiency of the right brain, if one has to 'work out' the meaning of the graph, that is, if it has an unfamiliar or unintuitive key, then it isn't effective. Indeed, training is necessary for the interpretation of perceived effects, but not to perceive them in the first place. A five-year-old child who looks at my Figure 3 will see the same curve gestalts as a trained scientist. But the scientist knows how to translate the perceived patterns into information about the data. This is not just an esoteric aside. It has important implications for the choice of subjects in many experiments probing pattern perception. Subjects do not need to know how to translate. In a number of experiments in which I have taken part, subjects needed only a sufficiently rich vocabulary to understand instructions and interpret very basic geometric properties such as the relative lengths of lines segments. In the experiments of Kröse described in my Section 4.3, subjects only needed the ability to determine if a corecue symbol was present or not; a five-year-old with enough maturity to sit still for a while could be the subject of such an experiment.

Unfortunately the sophistication of some of today's statistical analyses have resigned the researcher to entrust his data to black boxes. This could actually be avoided through step by step interactive graphical decomposition ... An excellent recent example is the several decades of work on kernel methods (including adaptive kernel methods) for estimating densities and regression curves. These methods yield to asymptotic derivations. But they have the disadvantage that they do not work. Anybody who chose to move from the black box of computer implementations of kernel methods to interactive graphical decomposition would quickly discover this. (But it takes new theory to derail old theory and it has been only been recently that theoretical work has revealed the problems of kernel methods (Fan, 1993; Hastie and Loader, 1993).)

7.3 Comments of Wilkinson

As Cleveland has noted, however, the slopes procedure does not specify which frequency to highlight in the series. The upper cat plot is scaled to reveal a low frequency (approximately 2.5 second) wave in the firings.

This point was treated only briefly in the paper and is vital to the usage of the method, which I call "banking to 45°" (Cleveland, 1993). It can be applied to any set of line segments on a graph. Time series have different frequency components of variation. The method can be applied any of these components. For the bottom panel of Wilkinson's Figure 1, it is applied to the high-frequency variation by banking the local segments to 45°. In the top panel, it is applied by banking a respiration component, presumably in an approximate way just by visual assessment. A fit that describes the component is not drawn but we can visualize it, a slow wave with about two complete cycles over the record. It cannot be visualized in the bottom panel because the local orientations of the wave are too close to the horizontal. In the top panel, the location of these absolute orientations is closer to 45°, so it can be perceived, although my guess is that the component is not quite banked to 45° and needs a larger aspect ratio.

There are likely configurational effects when more than two of these symbols are used in a plot.

While it is true that configurational effects exist, they are not typically salient except in the situation where there is so much data that overlap causes detection problems. As discussed earlier, for the engineering goals of developing display methods for data, we have the right to ignore an effect that is not salient.

The symbols Cleveland offers are discriminable mainly because they stimulate different sets of feature detectors in the visual cortex, namely verticals, horizontals, and diagonals. Why not use a set based on the elements which stimulate these feature detectors more exclusively?

One comforting aspect of purely scientific enquiry is that we can conjecture, experiment, and report. No such luxury exists in an engineering discipline, such as building display methods for data, because we must decide. (And no one understands this point better than the discussant.) Going from the experimental information to an actual choice puts our judgment fully on the line. Despite the objectivity provided by experimentation, we must balance different effects uncovered by experiment. Often there is more than one solution. For plotting symbols we must balance detection and assembly. The two symbols "o" and + have very high scores in Table 1. Thus I included them in the set, which precluded "—" and "|". But if a common experiment shows that the second pair beats the first, we should indeed be prepared to switch.

Cleveland wishes to stay grounded in perceptual psychology, but the topics he discusses also involve areas of higher cognition. Cognitive psychologists such as Pinker (1990), Simken and Hastie (1987), and Kosslyn (1989) have proposed process models for graphical perceptual processing.

True, I have stayed grounded in perceptual psychology, but not through personal preference. Again, we must remember the goal — engineering better visual displays. The classical work in perceptual psychology (e.g., Baird and Noma, 1978) has proven enormously useful. So too has some of the modern work in cognitive psychology that focuses on specific processes of elementary vision (e.g., Treisman, 1986). But I have not yet succeeded in finding a convincing application of the work in cognitive psychology that conjectures a full prescription of all of the processes that lead from the retinal image to conclusions about visual scenes. Nor has anyone else succeeded. I suspect no will until the theory is better conceived and is more firmly grounded in empirical observation. But the onus is not on me to prove the ideas are useless. The onus is on those who advocate the broader context to demonstrate that it is useful. I am all eyes.

Acknowledgements

Comments from R. A. Becker, N. I. Fisher, T. E. Graedel, A. M. MacEachren, C. L. Mallows, L. Tremmel, J. Tukey, and four referees on the original manuscript led to a substantial improvement in both the content and exposition of the paper. Also, Tukey, a creative inventor of terminology, is responsible for the term "assembly", which does a better job of connoting the operation than did the author's initial term, "clustering".

References

[1] J. C. Baird and E. Noma. *Fundamentals of Scaling and Psychophysics*. John Wiley & Sons, New York, 1978.

[2] I. Benbasat, A. S. Dexter, and P. Todd. An experimental program investigating color-enhanced and graphical information presentation: an integration of the findings. *Communications of the ACM*, 29(11):1094–1105, 1986.

[3] J. Bertin. *Semiologie Graphique: Les Diagrammes, Les Réseaux, Les Cartes*. Mouton & Co, The Netherlands, 1967.

[4] J. Bertin. *Semiology of Graphics: Diagrams, Networks, Maps*. The University of Wisconsin Press, Madison, WI, 1983. Translation of the French edition.

[5] W. C. Brinton. *Graphical Methods for Presenting Facts*. Engineering Magazine Co., New York, 1914.

[6] J. M. Chambers, W. S. Cleveland, B. Kleiner, and P. A. Tukey. *Graphical Methods for Data Analysis*. Chapman & Hall, New York, 1983.

[7] W. S. Cleveland. *The Elements of Graphing Data*. Chapman & Hall, New York, 1985.

[8] W. S. Cleveland. *Visualizing Data*. Hobart Press, Summit, New Jersey, 1993.

[9] W. S. Cleveland, M. E. McGill, and R. McGill. The shape parameter of a two-variable graph. *Journal of the American Statistical Association*, 83(402):289–300, 1988.

[10] W. S. Cleveland and R. McGill. Graphical perception: theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984.

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

[12] G. W. Dickson, G. De Sanctis, and D. J. McBride. Understanding the effectiveness of computer graphics for decision support: a cumulative experimental approach. *Communications of the ACM*, 29:40–47, 1986.

[13] J. Fan. Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, 87:998–1004, 1992.

[14] I. Fogel and D. Sagi. Gabor filters as texture discriminator. *Biological Cybernetics*, 61:103–113, 1989.

[15] N. H. Haber and L. Wilkinson. Perceptual components of computer displays. *IEEE Computer Graphics & Applications*, 2:23–35, 1982.

[16] R. Hastie. Schematic principles in human memory. In E. T. Higgens, C. P. Herman, and M. P. Zaana, editors, *Social Cognition: The Ontario Symposium*, pages 39–88, Hillsdale, NJ, 1981. Lawrence Erlbaum.

[17] T. Hastie and C. Loader. Local regression: automatic kernel carpentry (with discussion). *Statistical Science*, 8:120–143, 1993.

[18] F. R. Immer, H. K. Hayes, and Le Roy Powers. Statistical determination of barley varietal adaption. *Journal of the American Society of Agronomy*, 26:403–419, 1934.

[19] B. Julesz. Textons, the elements of texture perception, and their interactions. *Nature*, 290:91–97, 1981.

[20] S. M. Kosslyn. *Image and Mind*. Harvard University Press, Cambridge, MA, 1980.

[21] S. M. Kosslyn. Graphics and human information processing. *Journal of the American Statistical Association*, 80:499–512, 1985.

[22] S. M. Kosslyn. Understanding charts and graphs. *Applied Cognitive Psychology*, 3:185–226, 1989.

[23] B. Kröse. *A description of visual structure*. PhD thesis, Delft University of Technology, Department of Industrial Design Engineering, 1986.

[24] B. Kröse. Local structure analyzers as determinants of preattentive pattern discrimination. *Biological Cybernetics*, 55:289–298, 1987.

[25] S. Lewandowsky and I. Spence. Discriminating strata in scatterplots. *Journal of the American Statistical Association*, 84:682–688, 1989.

[26] M. Livingstone and D. Hubel. Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science*, 240:740–749, 1988.

[27] J. Lohse. A cognitive model for the perception and understanding of graphs. In S. P. Robertson, G. M. Olson, and J. S. Olson, editors, *CHI '91 Conference Proceedings*, pages 137–144, New Orleans, Louisiana, 1991.

[28] A. M. MacEachren and J. H. Ganter. A pattern identification approach to cartographic visualization. *Cartographica*, 27(2):64–81, 1990.

[29] J. Malik and P. Perona. Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America, A*, 7:923–932, 1990.

[30] D. Marr. *Vision*. W. H. Freeman, San Francisco, 1982.

[31] S. Pinker. A theory of graph comprehension. In R. Friedle, editor, *Artificial Intelligence and the Future of Testing*, pages 73–126, Norwood, NJ, 1990. Ablex.

[32] A. M. Rochlin. The effect of tilt on the visual perception of parallelness. *American Journal of Psychology*, 68:233–236, 1955.

[33] C. F. Schmid and S. E. Schmid. *Handbook of Graphic Presentation*. Wiley, New York, 1979.

[34] D. Simkin and R. Hastie. An information-processing analysis of graph perception. *Journal of the American Statistical Association*, 82:454–465, 1987.

[35] I. Spence and S. Lewandowsky. Graphical perception. In J. Fox and Long S., editors, *Modern Methods of Data Analysis*, Beverly Hills, CA, to appear. Sage Publications.

[36] K. T. Spoehr and S. W. Lehmkuhle. *Visual Information Processing*. W. H. Freeman, New York, 1982.

[37] K. A. Stevens. Slant-tilt: the visual encoding of surface orientation. *Biological Cybernetics*, 46:183–195, 1983.

[38] A. Treisman. Features and objects in visual processing. *Scientific American*, 255(1):114B–125, 1986.

[39] L. Tremmel. *Untersuchungen zu optimalen Symbolen in graphischen Darstellungen*. Verlag Dr. Kovaĉ, Hamburg, 1992.

[40] J. W. Tukey. *Exploratory Data Analysis*. Addison-Wesley Publishing Co., Reading, Massachusetts, 1977.

[41] J. W. Tukey. Graphic comparisons of several linked aspects: alternatives and suggested principles (with discussion). *Journal of Computational and Statistical Graphics*, 2:1–49, 1993.

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

[43] S. Ullman. Visual routines. *Cognition*, 18:97–159, 1984.

[44] H. Voorhees and T. Poggio. Computing texture boundaries from images. *Nature*, 333:364–367, 1988.

[45] H. Wainer and D. Thissen. Graphical data analysis. *Ann. Rev. Psychol.,*, 32:191–241, 1981.

[46] L. R. Wanger, J. A. Ferwerda, and D. P. Greenberg. Perceiving spatial relationships in computer-generated images. *IEEE Computer Graphics & Applications*, 12(3):44–58, 1992.

[47] R. E. Warren. Judging differences in data displays. Technical report, AT&T Bell Laboratories Technical Memorandum, 1989.

[48] M. B. Wilk and R. Gnanadesikan. Probability plotting methods for the analysis of data. *Biometrika*, 55:1–17, 1968.

[49] G. U. Yule. On a method of investigating periodicities in disturbed series, with special reference to wolfer's sunspot numbers. *Philosophical Transactions, A,*, 226:267–298, 1927.