Graphics and Human Information Processing

A Review of Five Books

STEPHEN M. KOSSLYN*

Five books on charts and graphs are reviewed from the standpoint of cognitive psychology. The review focuses on properties of the human visual information-processing system, and it considers how well the recommendations offered by each book work within the strengths and weaknesses of that system. The books reviewed are *Semiology of Graphs,* by J. Bertin (translated by W. J. Berg) (1983); *Graphical Methods for Data Analysis,* by J. M. Chambers, W. S. Cleveland, B. Kleiner, and P. A. Tukey (1983); *Mapping Information,* by H. T. Fisher (1982); *Statistical Graphics,* by C. F. Schmid (1983); and *The Visual Display of Quantitative Information,* by E. R. Tufte (1983).

KEY WORDS: Human factors; Cognitive engineering; Statistical graphics.

CONTENTS

1. General Overview of the Books (p. 500)

The five books have very different characters, ranging from a basic "how to" book to an extended theoretical treatise.

2. Effective Graphics (p. 501)

Graphs are intended to be read by human beings, who have information-processing systems with specific properties. Thus effective graph design requires an understanding of human visual information processing.

Human Visual Information Processing: A Brief Overview (p. 501)

Human visual information processing can be divided into three general phases, each of which has properties that place constraints on display design. The central question pursued in this review is: How well do the recommendations offered by the books respect the strengths and weaknesses of the user? The three phases provide a structure for the next three major parts of the review.

3. Getting Information Into the System (p. 502)

The first phase of visual information processing involves detecting edges and isolating regions of a stimulus. A number of factors affect how well this can be done; we consider how well the books provide guidelines for avoiding problems relating to the following (3.1–3.4).

3.1 Adequate Discriminability (p. 502)

Marks must be discriminable on the page and from each other.

3.2 Visual Properties (p. 503)

Variations in different kinds of marks can be used to convey information, some better than others.

3.3 Processing Priorities (p. 503)

Some marks are noticed and attended to more than others.

3.4 Perceptual Distortion (p. 504)

The visual system sometimes results in our having distorted impressions of visual dimensions.

4. Short-Term Memory Constraints (p. 505)

Short-term memory is a limited-capacity "working memory." We consider how well the books help one avoid problems due to the following (4.1 and 4.2).

4.1 Perceptual Grouping (p. 505)

Principles of perceptual organization delineate what we see as a "perceptual unit." The capacity limit of visual short-term memory is defined in terms of such units. Furthermore, if we organize a display into the wrong units, it will be uninterpretable. Four of the more important principles are reviewed.

4.2 Memory-Capacity Limitations (p. 506)

Only about four perceptual units can be held in short-term memory at once. Thus displays must not require the reader to consider more than this number of perceptual units.

5. Long-Term Memory Processing (p. 507)

Long-term memory is the repository of all one knows. Thus input must enter long-term memory before it can be recognized or interpreted (which requires matching against stored information). We consider how well the books help one avoid problems due to the following (5.1 and 5.2).

5.1 Ambiguity in Labels and Design (p. 507)

A pattern is ambiguous when it accesses more than one stored memory. Inappropriate titles, labels, and the juxtaposition of parts of the display all can cause this problem.

5.2 Inferences (p. 508)

A display can be drawn so that one will access the wrong associations, resulting in a display that "lies."

6. Formats and Purposes (p. 509)

Different graph types are appropriate for different purposes, given the way we interpret patterns of marks (e.g., line graphs are better for depicting trends and interactions, and bar graphs

* Stephen M. Kosslyn is Professor of Psychology, Harvard University, 1236 William James Hall, 33 Kirkland Street, Cambridge, MA 02138. Preparation of this review was supported by Office of Naval Research Contract N00014-83-K-0095 and National Institute of Mental Health Grant MH 39478. The ideas developed in this article have profited from many discussions with Steven Pinker. The author also wishes to thank E. E. Smith for useful comments on an earlier draft.

Journal of the American Statistical Association
September 1985, Vol. 80, No. 391, Invited Essay in Review

are better for depicting individual point values). We consider the guidance provided regarding the following (6.1 and 6.2).

6.1 Communication, Analysis, and Storage (p. 509)

Graphs can be used for three basic purposes, and a given type may be best suited for a given purpose.

6.2 Type of Question (p. 510)

Different types of graphs make different kinds of information readily accessible (e.g., interactions versus point values). Thus a given type will be more or less appropriate depending on the kind of question the reader is supposed to be able to answer about the data.

7. Data and Formats (p. 511)

Different graph types are appropriate for different types of data (e.g., as is obviously true for pie charts and proportional or percentage data).

8. Conclusions (p. 512)

Table 1 summarizes the evaluation of each book on each of the factors noted, in addition to some more general considerations. Research on graphics is mutually beneficial for statisticians and psychologists; we conclude by considering one way that important, specific empirical issues can be addressed systematically.

INTRODUCTION

It has often been said that graphs capitalize on many of the virtues of pictures, being worth the numerical equivalent of a thousand words. This is without question true of some graphs, but unfortunately it is not true of many—perhaps most—graphs (as even a casual reader of the national news magazines will readily acknowledge). Indeed, all five books discussed in this review begin with the premise that there is much to be learned about effective graphic communication. This review focuses on how well the books succeed in providing guidelines for the effective construction and use of charts and graphs.

1. GENERAL OVERVIEW OF THE BOOKS

This is a collection of unusually good books. I learned something from each of them. Before turning to a detailed review, it will be helpful to present a brief overview of the books.

In his foreword to Bertin's (1983) book, Howard Wainer (the technical editor) says that the "book has become a classic in its own strange and wonderful way" (p. vii). And indeed, it is strange and wonderful. Bertin loves taxonomies and leaves virtually no graphical stone unturned. This book has more wisdom and astute observations packed into every section than is present in most other books on the topic. I am convinced that I could read this book three times and still miss interesting insights and analyses; it is an understatement to say that this book is "meaty." I especially recommend this book to anyone considering doing research on graphics; the book has numerous ideas and opinions that are begging to be studied and developed further. I could not think of anything that was not discussed at some point, and usually at multiple points. And this brings me to my only general complaint: The book was difficult to read, and it probably would be difficult to use if you came to it with data in hand and wanted to know the best way to depict them. There are too many cross-references to material on other pages. It would have helped if there were figure captions. Nevertheless, this book is *must* reading for any serious student of graphics. It is a superb treatment of the area, and it is unlikely to be surpassed in the foreseeable future by a deeper consideration of these topics.

Chambers et al. (1983) have written a very creative, timely book. It fills a glaring gap in the literature by providing a substantive discussion of uses of graphs in data analysis. I strongly advise anyone who works with data to buy this book. Part of its value is simply as an inspiration, showing one what can be done using graphs. But a larger part of its value is as a handbook of methods that can be used practically, to help one in the day-to-day task of analyzing data. This book is well written and well organized; it could easily be used as a supplementary text in statistics courses. Indeed, the authors include exercises at the end of each chapter, and difficult sections and exercises are flagged as such.

Fisher (1982) has written an exceedingly thoughtful book on maps. He has provided a detailed discussion of ways of aggregating data, of the problems and procedures of interpolation, and the uses of specific types of symbols. As promised on the cover flap, the book is a mix of manual, text, and encyclopedia. This book is very accessible and provides much information that could be used immediately by an inexperienced mapmaker. The terminology is sensible, and the organization is straightforward and practical. It is especially good at a kind of "consciousness raising," leading the reader to understand the importance of considering different ways to map data.

If you are interested in having naive students get a good start at drawing graphs, Schmid's (1983) book is not a bad place to start. It has a very clear description of numerous graphical forms and some sound "nuts and bolts" advice about creating them. The book is very easy to read, and the figures have useful captions. This book makes one aware of the relevant variables and sensitizes one to many common errors. Indeed, even if some of the elements of some of his critiques are debatable, the book is good enough to make you aware of this.

Tufte's (1983) book is a joy to read. This book was by far the most entertaining of the set. Following the spirit of Tufte we can create an "Enjoyment Index," which is the number of smiles per page. On this criterion, this book is a winner. We can also create a "No-Pain Index," which is the number of facts times the number of smiles; it is a winner again. Tufte writes with a wry sense of humor and has a good eye for detail. The book was not just lavishly illustrated; the illustrations were selected with such care and are so well set off by the text that the ensemble is a work of art. It is a fascinating collection of illustrations and interesting facts about graphics. It also incorporates numerous novel ideas that are clearly presented and easily understood. This book is another source of inspiration for the aspiring researcher in the field, and it is "must" reading for anyone interested in the area. Unlike Bertin's book, which I would recommend only to the serious student, I recommend Tufte's book to all. Reading this book is an experience. No well-equipped library (and some might say living room) should be without this book.

2. EFFECTIVE GRAPHICS

But how good is the detailed advice given in each of the books? How well do they lead one to design "effective" charts and graphs? Any such evaluation can only be made relative to the nature of the human information-processing system. There are uncountable ways of representing information, ranging from drawings of circles divided into wedges to patterns of cracks on tortoise shells to differences in the size of bubbles in a fluid. Not all of the possible ways of representing information are useful for human beings. A display will be effective when it takes advantage of the capabilities of our perceptual, memory, and conceptual abilities; an optimal graphical display for a human might not be optimal for robots or creatures from another planet. Indeed, graphics are effective *because* they exploit properties of human information-processing abilities. This fact has been recognized at least since the time of William Playfair, whose 1786 *The Commercial and Political Atlas* is widely credited with introducing most of the contemporary forms of charts and graphs. Playfair wrote:

> Information, that is imperfectly acquired, is generally as imperfectly retained; and a man who has carefully investigated a printed table, finds, when done, that he has only a very faint and partial idea of what he has read; and that like a figure imprinted on sand, is soon totally erased and defaced. The amount of mercantile transactions in money, and of profit or loss, are capable of being as easily represented in drawing, as any part of space, or as the face of a country; though, till now, it has not been attempted. Upon that principle these Charts were made; and, while they give a simple and distinct idea, they are as near perfect accuracy as is any way useful. On inspecting any one of these Charts attentively, a sufficiently distinct impression will be made, to remain unimpaired for a considerable time, and the idea which does remain will be simple and complete, at once including the duration and the amount. (pp. 3–4)

Playfair's psychology was apparently influenced by Plato, as suggested by the following quote from the *Theatetus:*

> I would have you imagine, then, that there exists in the mind of man a block of wax. . . . When we wish to remember anything which we have seen, or heard, or thought in our own minds, we hold the wax to the perception and thoughts, and in that material receive the impression of them as from the seal of a ring; and . . . we remember and know what is imprinted as long as the image lasts; but when the image is effaced or cannot be taken, then we forget and do not know.

Experimental psychology, however, has made some headway since the time of Playfair, and we can use the results of psychological experimentation to guide us in designing and using statistical graphics. There are two ways one can discuss the psychology of graphics. Perhaps the most straightforward is to consider the literature that directly addresses the topic (for reviews see Kruskal 1982, Macdonald-Ross 1977, and Wainer and Thissen 1981). However, this research tends to be highly focused on particular issues (e.g., the best way to scale circle sizes, the relative ease of using graphs compared to using tables, etc.), and it is too narrow to use in evaluating general recommendations for graph design. An alternative approach is to consider the psychological literature on basic visual information processing and to use this literature to evaluate the perspicacity of the proffered guidelines for good display design and use. That is, over a hundred years' worth of research has provided a rich body of information about human perceptual, memory, and conceptual abilities. Some of these research results have direct bearing on the design and use of statistical (and other) graphics. In this review we will briefly consider the most important of the relevant psychological factors and discuss how well the books deal with these factors when specifying acceptable graphic design and use. We will ask, How well do the recommendations offered respect the strengths and limitations of the user?

Human Visual Information Processing: A Brief Overview

Figure 1 presents an overview of the contemporary "canonical theory" of human visual information processing (see Marr 1982 for a more detailed version). This scheme divides processing into three general phases. At each phase information can be difficult to process if the limitations of the system are not respected. In this review I will consider major constraints on display design that are dictated by the nature of each phase of processing. Let us begin with a brief overview, considering each box of Figure 1; details will be added later as they become relevant.

When you first look at a display, the light reflected from it [or emitted by it, if it is on a cathode-ray tube (CRT)] strikes the retinas and is immediately converted into a pattern of neural impulses. These neural impulses go from the eyes back into the brain. When reaching the brain, they first are processed to detect patterns of lines, blobs, colors, and textures (as well as movement; see Van Essen and Maunsell 1983). These visual properties are not "semantically interpreted" initially; they are simply seen as visual patterns. This sort of "pre-semantic" processing appears to take place in the occipital lobes, posterior temporal lobes, and posterior parietal lobes of the brain (see Cowey, in press). This processing leads to the formation of what we can call a "visual sketch."

This first phase of processing is sensitive to the discriminability of lines and marks (if they are too small or dim, or differences are too subtle, they will not be seen), and to orientation and various visual relations that should be exploited in conveying information (see Section 3.2). In addition, large differences along a visual dimension are noticed before more subtle differences (which can emphasize the importance of the right information or result in interference from extraneous information, such as grid lines in a graph), and physical variations in marks are not always registered accurately; indeed, our perceptions are sometimes systematically distorted (see Section 3.4).

The output from the first phase of processing is organized into "perceptual units," which are operated upon in the second phase of processing. For example, three lines that meet to form an enclosed figure are not seen as three lines, but as a triangle. Many of the rules that dictate how lines and marks will be grouped to form perceptual units have been clearly described (see Section 4.1), and these rules must be respected if the elements of a display are to be perceived in the proper juxtaposition.

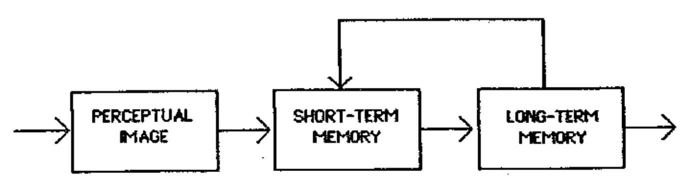

Figure 1. Three Levels of Visual Information Processing.

Table 1. Summary of Evaluations

Section	Criterion	Bertin	Chambers	Fisher	Schmid	Tufte
	Readability	4	9	9	8	10
	Originality	10	10	7	3	10
	Generality	10	4	4	8	7
3.1	Discriminability	8	2	7	3	6
3.2	Visual Properties	9	7	7	5	7
3.3	Processing Priorities	7	6	3	5	5
3.4	Perceptual Distortion	7	9	5	5	7
4.1	Perceptual Grouping	8	7	5	5	2
4.2	Memory Limits	9	3	5	4	7
5.1	Ambiguity	9	5	8	7	6
5.2	Inferences	7	8	7	6	8
6.1	Purposes	9	4	4	5	4
6.2	Questions	8	8	6	5	3
7	Data Formats	10	9	8	7	7

NOTE: 1 indicates very poor; 10 indicates excellent.

The perceptual units are held in a "short-term memory," which is also sometimes called "working memory" (for further details, see Anderson and Bower 1973 and Lindsay and Norman 1977). Information in short-term memory can be reorganized and interpreted in various ways. In addition, short-term memory has a notoriously limited capacity, which is one reason why the reader would have difficulty if I required him or her to hold more detail about short-term memory in mind until it becomes relevant; thus I will defer presenting more detail about it until this information will be used (Section 4).

Finally, the input ultimately must access the relevant information stored in "long-term memory." This information constitutes one's knowledge about how charts and graphs serve to convey information. Without accessing such stored knowledge, one would not know how to relate points on two axes, how to compare areas of two wedges, and so on. Indeed, "comprehension" depends on accessing the proper stored information. Thus, lines and regions must be drawn to be unambiguous; they should access only the appropriate set of stored information. In addition, part of this comprehension process involves drawing inferences about the input, some of which can result in a display presenting "lies" about the data (see Section 5.2).

In the following sections of the review we will consider in more detail the constraints on visual display design posed by the nature of our information-processing system. Table 1 summarizes the evaluations of the books on each of the criteria we will develop; space limitations preclude a detailed treatment of each book on each criterion. The ratings in Table 1 vary from 1 to 10, with 1 being "extremely poor" and 10 being "excellent."

3. GETTING INFORMATION INTO THE SYSTEM

The first phase of visual information processing involves detecting edges and isolating regions of the stimulus. A number of factors affect how well this can be done, and these factors must be respected in display design if the display will be readily apprehended. The most relevant factors are adequate discriminability, visual properties, processing priorities, and perceptual distortion.

3.1 Adequate Discriminability

Variations in marks, either on a page or on a CRT screen, must be great enough to be easily noticed. If they are not, we will never consciously even notice that they are there. There are two aspects to this observation: First, a mark must be a certain minimal size or it will not be seen. This "absolute threshold," as it is called, has been computed for many marks, particularly letters of the alphabet (see Smith 1979). Photoreduction of displays often renders parts of the display unreadable because this factor has not been taken into account. Second, to be discriminated from each other, two or more marks must differ by a minimal proportion. This is especially important when marks are superimposed on each other (e.g., when labels are placed directly on bars) or are in close proximity (e.g., bars are divided into two segments, each corresponding to a different independent variable). In addition, if colors or different symbols (e.g., different sorts of dashed lines) are used to discriminate among bars, lines, or wedges, they must be clearly discriminable from each other. Some of the books have useful advice on how to ensure that displays and parts of displays will be discriminable.

Bertin provides the most useful discussion of these aspects of graphic design (primarily in Part I, Section IIIE, although supplementary information is provided as relevant elsewhere, e.g., p. 300), with a concise summary on page 190. This discussion includes very specific guidelines about how different two marks must be in order to be discriminable, and for the number and magnitude of intermediate variations that are discriminable between two values of a visual dimension. The detail given here is almost overwhelming. Although most of these recommendations seem reasonable, it is not clear how they were derived and some are debatable. For example, he recommends 2 mm as the minimal size for a "separable shape"; this seems not unreasonable advice, but the eye can do far better than that, and other recommendations have been made elsewhere (see the comments in this section on Tufte).

Chambers et al. confront the problem of points that are overprinted or so close as to be indiscriminable by suggesting the use of "jitter," a vertical staggering of points. In addition, Chambers et al. worry about the discriminability of symbols used on plots, and they offer useful suggestions about what should and should not be used (pp. 179, 181).

Fisher spends considerable time discussing how to ensure adequate discriminability by correctly selecting the levels of darkness or contrast used when different numbers of symbols are employed. His first suggestion is that values should be assigned to discrete classes whenever possible, ideally with no more than five or six classes; it is much easier to ensure that five or six symbols are discriminable than that all those representing a wide range of distinct values are discriminable. Fisher considers extensively how one should classify values, and he produces some very specific suggestions. In addition, he developed his own set of symbols (called Fisher Symbolism, patent number 4,148,507), fields of which should be discriminable from each other. The new symbols are an attempt to deal with his observation that

> if differentiability depends on (1) darkness alone or with change of size and arbitrary shape, it will be inadequate; (2) darkness and change of size alone,

> it will be better but still inadequate; (3) darkness, change of size, and change of direction, it will be better, but an arbitrary and meaningless variable has been introduced. Darkness with change of size and meaningful shapes aided by count, will be most successful. When the designer is not limited to black and white, the addition of progressive change in color is optimum. (pp. 289–290)

It was not clear how this advice was derived, and at least some of it is debatable (as will be discussed in Section 3.2).

Tufte has a nice example in which he shows how our eyes can make 25,281 distinctions within a square inch (as seen from a normal viewing distance). He cites a cartographer as saying that "the resolving power of the eye enables it to differentiate to 0.1 mm where provoked to do so" (p. 162). This is true (e.g., see Frisby 1980, figure 92, p. 96), but this does not imply that it is a good idea to assume that such discriminations are easily made or desirable in normal viewing conditions. Indeed, on Tufte's page 169 we see an example of this very problem; at the reduced size several of the labels are difficult to make out.

3.2 Visual Properties

The eye and brain are sensitive to a number of visual properties (e.g., size, orientation, darkness, hue, intensity, texture, etc.), variations of which can be used to convey information. In many cases, variations of one property can be combined with variations of others, allowing a single mark to convey a large amount of information. How should visual properties be used most effectively in display design? One answer to this question is clear: Differences in quantities should not be represented by differences in color. Differences in color are more like differences on a nominal scale; shifting from red to green does not result in "more of something" in the same way as shifting from a small dot to a large one does. Indeed, the best depiction of the psychological similarity-space for colors is not a single line, but a circle (see Kaufman 1974, chap. 6). Thus, it is difficult to use progressive differences in color to stand for progressive increases or decreases in some quantity. A nominal scale and a quantitative one do not naturally line up. You have probably seen many maps that use color differences to stand for different populations or the like. And you may have also noticed that it is hard to tell which areas have more without repeatedly returning to the key. How good is the advice offered in these books for use of visual properties?

Bertin refers to variations of a mark that co-occur at a single location as "retinal variables," and he offers an encyclopedic and useful taxonomy and treatment of such variations (Part I, IIC). A useful table of visual variations is presented on page 96. This table is particularly useful because the variations are broken down in terms of whether they are ordered (e.g., levels of darkness) or not (e.g., different colors, which are not seen as arranged along a quantitative dimension). Visual variations that themselves are not seen as being ordered are not useful for depicting ordered information. Bertin also presents a table of a "natural" series of graduated sizes of circles, which may be useful for some applications (pp. 368–369). In addition, Bertin discusses which sorts of variations are effective in different circumstances, but I found this discussion difficult to use.

Chambers et al. conjecture that the eye is able to perceive a variety of relations very easily: Locations along an axis are presumed to be more easily seen than other graphical aspects (e.g., size); straight lines are presumed to be seen more clearly than curves, and so on. Some of these conjectures seem more plausible than others, but none should be taken too seriously (as the authors realize) until they are investigated empirically. In addition, there is a useful discussion of coding schemes for plotting symbols (pp. 179–181). Although this discussion is not exhaustive, the advice given here about symbol use is very clear and concrete.

Fisher presents a detailed discussion of different ways of visually representing information. A useful table of "map symbolism" appears on pages 66 and 67, which considers ways in which extent, darkness, and countability can be used with spots, bands, and fields. Fisher provides a way of determining spot symbol sizes that, unfortunately, assumes that there is a linear relation between area and amount (p. 119). (This assumption is unfortunate because it is incorrect, as will be discussed in Section 3.4.) The discussion of appropriate darkness levels (p. 283) was considerably more useful, being based in part on empirical research (see Williams 1956). This research produced a function (relating number of classes to percentage of ink) that differed from Fisher's theoretical curve. Although Fisher recommends use of his theoretical curve anyway, the presentation of both curves leaves it up to the reader to decide which function to adopt. The book contains plenty of useful advice about how to depict in black and white. Fisher has an interesting chapter on color that contains much sound advice; unfortunately, his recommendation that one use color to depict an ordered series is misguided (p. 294). Color is not psychologically ordered along a continuum.

Schmid's presentation is organized around different graphic formats, and visual properties are discussed almost entirely in this context. However, he does devote chapters to two- and three-dimensional "graduated point" symbols, to three-dimensional projections, and to pictorial charts. Numerous specific recommendations are offered about three-dimensional symbols (e.g., he prefers a specific kind of projection), but they often are not well justified and are debatable. This tendency is particularly evident in discussions of the use of area to represent quantities, as noted in Section 3.4.

Tufte provides a useful discussion of color and its drawbacks, and he suggests the use of varying shades of gray to represent ordered quantities (p. 153 ff). He also presents valuable observations about the undesirability of "Moire effects," which are a kind of vibratory, oscillating, "unintentional optical art" that arises from dense fields of stripes, hatching, checks, or other regular patterns. He also offers sound advice to use words as labels, rather than keys that depend on filling in the graph with such patterns. Tufte amply justifies his concern by presenting a survey of how often such patterns appear in textbooks and handbooks of statistical graphics and in manuals for computer graphics programs (p. 112). Tufte takes issue with Bertin's advice that one ought to make use of this effect, pointing out that there are no good examples of such effective use.

3.3 Processing Priorities

The visual system detects differences in line weight, orientation and length, shading, colors, and other visual properties.

And larger differences are more easily detected than smaller ones (up to an asymptote). This characteristic leads us to detect heavier lines before lighter ones, given the difference between the line and the background. Similarly, brighter colors are detected before dimmer ones, and larger bars before more slender ones. Graphs should be constructed so that one notices the more important things first. If there is a difference in the heaviness of lines, it should reflect a difference in their importance. Inner grid lines should never be darker than the content lines. And background patterns should never be as noticeable as the components of the graph itself. As another example, sometimes a wedge is exploded out from the pie in a pie chart; this difference leads the eye to focus on this wedge, so it should be the most important one. Some computer graphics programs automatically explode the wedge that happens to be largest, whether or not that is the most important one for the user. What guidance do we get on controlling for, and using, our processing priorities?

Bertin has a nice discussion of the role of such processing priorities in labeling (Part I, IA) and in depicting information (Part I, IIIE). This issue is discussed in various places throughout the book, usually in the context of discussions about specific graphic formats.

Chambers et al. point out that "a good display is one in which the visual impact of its components is matched to their importance in the context of the analysis" (p. 323). Some of their suggested formats explicitly take this factor into account; for example, in the box and whisker plot the "whiskers" are less important—and prominent—than the central 50% of the data depicted by the box. They do not, however, offer any specific principles or guiding rules on such matters (they cite Tufte's admonitions).

Schmid repeatedly brings up the "principle of contrast" in the context of pointing out that the "visual importance and distinction assigned to the various elements of a chart should be commensurate with the intellectual significance of the ideas being presented" (p. 15). The most complete discussion of this factor occurs when he considers problems in displaying information on maps (p. 118 ff). He notes that visually, "certain units may seem very important merely because of the relatively large area they comprise, but actually, in terms of the number of cases or size of population each areal unit contains, they may rank very low." He then proposes five solutions to this problem. Some of these solutions, however, are as problematical as the original, but for different reasons. For example, one proposal is to place small pie charts over each area, which also purportedly allows more variables to be displayed simultaneously. This solution results in information overload (see p. 120). Alternatively, he suggests that three-dimensional forms be placed over the locations. Unfortunately, even in the best of circumstances, humans are not good at estimating volumes of objects and are even worse with pictures of objects (see Teghtsoonian 1965). Some of the other solutions, such as "value-by-area" maps (where the map is stretched and bent so that area is proportional to amount), are reasonable if the reader is supposed to see only ordinal relations. Reading this book was sometimes frustrating because Schmid frequently made recommendations without justifying them, and I found that good advice (from a psychological point of view) was inte mixed with advice of arguable worth.

Tufte offers part of a solution to the problem of incorre processing priorities: He suggests that the designer try to max imize the share of "data ink" in the display; data ink is th "non-erasable core of a graphic, the non-redundant ink arrange in response to variation in the numbers represented" (p. 93) Tufte suggests, literally, that one erase as much nondata ink a possible (within reason). This technique has the desirable con sequence of eliminating "chart junk," meaningless, useless and distracting decorations. The notion that the designer shoul strive to maximize data ink is a fascinating idea, which lead to the introduction of new formats (e.g., bar graphs in whicl one side of each of the bars is erased). A problem, however is that it is not clear how to select data ink. The framewor and tick marks are treated as nondata ink, even though the are essential if one wants to know a specific value or the mag nitude of a difference; both the point on a function *and* th labeled axes are needed to convey the precise data. This tech nique does serve to eliminate the clutter in a display, but i may be too extreme as here specified. In addition, at one poin Tufte suggests that a superior grid can be created in a bar grap not by drawing horizontal grid lines, but by erasing regula horizontal stripes across the bars. This idea is somewhat of target: The eye fills in "virtual lines" connecting the corresponding erased parts of the bars, which seems just as distracting as real lines.

3.4 Perceptual Distortion

The visual impressions we have are not always a linear function of the physical parameters of the input; we sometimes systematically err in how we perceive patterns. In addition to illusions (see Dodwell 1975 and Frisby 1980), we also have nonveridical perceptions of specific visual dimensions. A classic example of this sort of distortion occurs in the perception of area, which is systematically underestimated. Thus, to get the same increase in perceived area from one form to another, one must consistently over-represent the area. [The perceived area is usually proportional to the actual area raised to an exponent of about .8; see Stevens 1974 and Teghtsoonian 1965.] This problem is especially vexing when one wants to portray equal step sizes visually. How should one cope with the human tendencies to distort input?

Bertin has a good discussion of perceptual factors underlying the perception of equidistant steps of value. He points out that the eye sees ratios, and he presents examples of common progressions with appropriate light/dark ratios (p. 75). I could not find a discussion of Weber's law (which states that equally discriminable steps along a perceptual dimension are produced by adding an amount to a previous value to produce a constant ratio of the added amount over the previous value), which turns out to be only a rough approximation to the truth, nor of Stevens's power law (which indicates that the perceived value is proportional to the physical value raised to some power). Bertin's advice for compensating for distortions is based on a formula he presents but then modifies (intuitively, it appears) to produce an "adjusted" function; I wonder how well this technique approximates the psychologically correct function.

Chambers et al. have something better than a good discussion of visual distortions: In some of their examples they use data from an experiment on visual distortions. They are clearly aware of the distorting effects of area, but they point out that it is still unclear whether to use the diameter, area, or some other measure of size in order to scale symbols to convey detailed quantitative information. There is no discussion of equidistant scaling of most visual dimensions. Some of the authors of this book have conducted empirical investigations of the perception of graphs, and the book provides us with interesting aspects of the results. For example, when asked to judge the linear association in a scatterplot, people tend to see r squared, not r. The authors mention that "construction factors, such as the size of the point cloud relative to the size of the frame around it, can affect our perception" (p. 320), but they do not tell us exactly how.

Fisher is concerned that readers be able to discriminate among close values. Thus, at times, he recommends using roots or powers of the values—which results in the symbol values in the display (level of darkness, size, etc.) presenting a distorted perceptual impression of the actual distribution. In these cases, the key must be studied carefully and the reader must try to overcome the perceptual impression. Fisher realizes that these manipulations can cause difficulties, and he recommends this technique only for specific applications, such as when one part of the range is of special interest (p. 147). Even in these cases, however, it may be better to present a separate display of that portion of the range than a distorted version of the entire range.

Schmid warns against shading and cross-hatching patterns that produce optical illusions, which sometimes result in "tilting" of the bar. His solution to this problem is to superimpose preprinted, self-adhesive tapes that are available commercially (p. 43). Having seen some of these, I did not find this advice entirely adequate. In addition, Schmid decries the use of diameters of circles to represent amount and then goes on to point out that area is in itself not well perceived (he mentions Weber's law and Stevens's power law but does not make much of it). This leads Schmid to advise that "Rather than include several pie charts of varying size in a single illustration, it would be more expedient and effective to use a bar or column chart" (p. 66). This is sound advice, and I could not understand why Schmid went on to violate his own advice later in this book (e.g., his use of pie charts, noted in Section 3.3). He suggests that the diameters of circles be made proportional to the square roots of the values they represent (p. 140). When discussing problems in perceiving area he recommends including a key with reference sizes, which can serve as "anchor stimuli" (p. 143), but no evidence is presented that this technique is effective (such matching becomes quite difficult when the sizes are relatively similar). One interesting tidbit was mention of an empirical study on how people read pie charts (p. 66). The results indicated that in estimating the sizes of sectors, 51% of the subjects used arc length; 25% used area; 23% used central angles; and 1% used chord length. Such variability does not bode well for effective communication of quantitative information using pie charts (Eells 1926).

Tufte has a lovely section on perceptual distortions (in Chapter 2). He presents a clear summary of the power law. However, Tufte's suggestion for dealing with the problem is "to use a table to show the numbers. Tables usually outperform graphics in reporting on small data sets of 20 numbers or less. The special power of graphics comes in the display of large data sets" (p. 56). This assertion is highly debatable, especially if the reader is supposed to see and compare trends, in which case graphs are useful even if there are only two sets of two numbers each. Tufte then goes on to advise that "the representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented" (p. 56). This advice simply ignores the problem altogether and invites readers to continue to produce distorted displays. There is no general discussion of the problem of perceptually equivalent equal step sizes, and most of the advice is at a fairly general level.

4. SHORT-TERM MEMORY CONSTRAINTS

Cognitive psychologists make a distinction between "short-term memory" and "long-term memory." Short-term memory is a "working memory" in which information can be reorganized and reinterpreted. The material in short-term memory is organized into "perceptual units," which can be reorganized only with effort. For example, when we view a Star of David we typically will visually organize it (according to grouping principles discussed in Section 4.1) into two overlapping triangles; it requires effort to mentally reorganize it into a hexagon surrounded by six small triangles. A hallmark of information's being in short-term memory is that we are aware of perceiving it (Norman 1978; Posner 1978).

Information in short-term memory is transient (it can be stored for a matter of seconds, such as by saying a telephone number over and over silently to oneself or holding a mental image), and only a very limited amount of information can be stored; only about four groups of items can be held in short-term memory at once (see Ericsson et al. 1980). In contrast, long-term memory is the "permanent," very large capacity (its limitations are as yet unmeasured) memory in which all knowledge is stored. Thus, short-term memory is an important bottleneck in information processing, and its limitations must be respected by display designers.

4.1 Perceptual Grouping

Visual information is stored as perceptual units in short-term memory. The nature of these units was studied in depth by the German "Gestalt psychology school," which was founded in 1912 and had its heyday in the 1920s. These experimental psychologists discovered numerous laws that describe how we organize marks into forms. They began with the observation that we do not see each and every little dot and smudge before us; rather, we see the patterns they form. The Gestalt Laws of Organization must be respected if parts of a framework are to group together, if individual function lines are to be discernable, if labels are to be correctly associated with parts, and so on (see Kaufman 1974, chap. 12).

In recent research in my laboratory we have found that when a visual display is difficult to interpret, violations of these grouping laws are often the root of the problem. Thus I was surprised to discover that none of the books had a detailed treatment of these organizational principles. It seems worthwhile to summarize the more relevant laws here.

Proximity. Marks that are close together will tend to be grouped together. For example, "xxx xxx" is seen as two groups, whereas "xx xx xx" is seen as three. This principle is often relied upon to associate labels with scales. The label should be closest to the part it is labeling.

Good Continuation. Marks that suggest a continuous line will tend to be grouped together. For example, "----------" is seen as a single perceptual unit, not 10 separate ones. In line graphs, content lines that cross can sometimes be confused if the segment from one flows naturally from a segment of another. In addition, when lines are labeled at the ends, the label should be a continuation of the line itself.

Similarity. Marks that have similar shapes, orientations, colors, and so on will tend to be grouped together. For example, "ooo|||" is seen as two groups. This principle is often used to associate a label in a key with a content element (line or bar or wedge); that is, there will be a short symbol or color next to a label in one corner of the display, and the reader uses this to figure out which content element is being labeled. In addition, when multiple panel displays are used, this principle is used to pair up the corresponding sections (usually by their having corresponding positions in the graphs or charts). A common violation of this principle occurs when two pie charts are used, each having the same cases but rearranged in different positions in the two pies.

Common Fate. Marks or lines that seem to be going the same way will be grouped together. Thus, content lines that are all going in the same direction will tend to be grouped together. In light of the capacity limits of the system (discussed in Section 4.2), this is particularly important because it allows the designer to pack many lines into an easily understood display if they happen to be parallel.

Although Bertin never summarizes these principles in one place, he makes effective use of them throughout the book. For example, similarity of shape, orientation, and darkness of symbols is used to group locations on a map according to the relevant criteria (Part II, IIID). The grouping principles are discussed largely in the context of factors that produce a single "image" [in Bertin's jargon, a "meaningful visual form, perceptible in the minimum instant of vision" (p. 142); see Part I, IIIB]. In his simplification procedures (e.g., pp. 36, 168, and 271), Bertin also makes effective (and creative) use of the Gestalt organizational principles. The reader is provided with ample examples of these principles at work, but some effort is required to integrate the information presented in different places and contexts throughout the book.

Chambers et al. introduce the idea of "jitter" as a way of breaking up spurious perceptual groupings in a scatterplot. They are concerned with ways of plotting data to allow meaningful visual patterns to emerge, but for the most part they adopt an exploratory try-it-and-see attitude. They have some interesting suggestions about how to arrange panels of multi-paneled displays so that they group in interesting ways (p. 163), but these suggestions are not precise enough to provide a clear-cut method for grouping.

Fisher implicitly bows to these grouping principles when he discusses discriminability, with the goal of ensuring that different regions on maps do not group to form a single perceptual unit. In addition, he suggests that to make displays easier to comprehend, bars should sometimes be grouped into clusters (p. 273). There is no general discussion of the implications of these principles for display design.

Schmid urges that if there is more than one curve on a chart, the lines should be differentiated by distinct patterns or colors (along with appropriate labels). He also notes the problem of superimposing curves, and he urges an ordering of bars from smallest to largest, whenever possible. Although he does not explicitly develop the grouping principles, he uses them in some of his recommendations (e.g., placement of labels relative to axes and tick marks). Unfortunately, Schmid's recommendation of the preferred three-dimensional rendition of a pie chart fails to take account of the operation of such principles: He recommends drawing a circle, then adding a "front lip" to it to give the impression of viewing the pie from the top. This front lip in fact results in the display appearing elliptical, because the bottom of the lip and the back part of the circle group together (via "good continuation"), resulting in a distorted impression of the front area.

Tufte never discusses these grouping principles, and in fact offers suggestions that explicitly violate them. That is, Tufte is interested in removing all redundant ink, including parts of the framework of graphs (producing lines only when they are under data points). This practice disrupts the "good figure" of displays; instead of seeing the framework as a single unit, to which the lines are anchored, the lines float in space. Tufte acknowledges this consequence of following his advice, but cites one study showing that people only study half of a symmetrical face as evidence that the other half is redundant and not necessary to include (p. 97). This result does not show that processing is equally effective in the two cases, or that memory would be equally effective; indeed, there is evidence to the contrary (e.g., see chapters 2 and 3 of Glass et al. 1979). Tufte usually hedges his recommendations, but I worry that there is not enough emphasis on the purpose-specific uses of graphic displays. For example, a gridwork is useful if one wants to compare specific values in addition to seeing trends.

4.2 Memory-Capacity Limitations

The limited capacity of short-term memory is a bottleneck in processing. If a display is to be easily understood, it should not require the reader to hold more than the capacity limit of short-term memory in mind at once. The capacity limit is roughly four units (see Ericsson et al. 1980); recall that a perceptual unit is not always a single line—the principle of Common Fate (Section 4.1) can result in numerous lines forming a single unit and hence being readily comprehensible. The constraints on information processing imposed by our limited short-term memory capacities often make it difficult to read a visual display, for two reasons: First, too much material is placed in the display. Second, too much material can be placed in a key, forcing the reader to engage in an arduous memorization task. What are we to do with complicated data displays? Some of the books provide useful advice about how to circumvent problems stemming from our limited short-term memory capacity.

Bertin has a fairly lengthy discussion of problems that have more than two variables, producing complex displays (e.g., Part II, IB and IC). Basically, he claims that the eye can only see three variables at once (the position of a point on X and Y coordinates and some "retinal" value of the point), and the difficulty of using a graph is determined by how many fixations are required. Both assertions are debatable, however: First, psychologists (e.g., Clement 1978 and Garner 1974) distinguish between "separable" dimensions, such as orientation and size or height and hue, and "integral" dimensions, such as hue and saturation or height and width. Separable dimensions are processed individually; integral dimensions are processed together. Thus the x and y coordinates are integral dimensions, as are many combinations of "retinal" variables. Depending on the symbols, more or less information can be taken in at one time (see Clement 1978 and Garner 1974). Second, the total processing of a display depends on a host of psychological processes, of which taking in the information is only one. Bertin points out that the reader must remember and integrate the separate "images" (sets of three variables taken in at a single glance), but this will be more or less difficult depending on how easily the separate images can be integrated, which is, in part, a consequence of the reader's previous experience and knowledge. Nevertheless, Bertin offers a very large number of useful examples of different ways of displaying the data, and he discusses ways of constructing displays to produce an "overall image" (integration) and ways of simplifying different types of displays. He suggests that it is often necessary for the designer to choose a "preferred question," and then to construct the graph so that it is biased to allow readers to easily answer this question. Bertin suggests that the graph designer consider this procedure even if such biasing requires sacrificing some of the information. In addition, he introduces a kind of "external memory" card file system to help one comprehend complex data (see p. 218 ff; however, I fear that this system has been superseded by the computer).

Chambers et al. do not have explicit suggestions for dealing with memory overload. This book focuses on graphics for the working scientist, as opposed to graphics for publication or communication. Thus they are less concerned with possible difficulties encountered by naive readers, assuming—correctly—that sophisticated readers can learn to organize complex patterns into fewer perceptual units.

Fisher's emphasis on forming classes from a set of raw data is in part a response to limitations of human information-processing abilities. His elaborate procedures for forming classes are, then, specific suggestions for dealing with these limitations. Some of his suggestions, however, result in perceptual distortions (as noted in Section 3.4), and others rely on unequal class sizes—which is a dubious procedure for many purposes. Nevertheless, his classification procedures would be useful for some purposes, and he has some suggestions as to when such procedures would make sense. In addition, in particular examples there is advice on selecting symbols and constructing keys that addresses the problem of memory limitations. For example, "The main disadvantage of bars in multi-subject mapping is that they lack personality. Remembering what each bar stands for is difficult except when there is a logical relationship . . ." (p. 273). Similarly, a concern with memory overload produces the suggestion that when there are 8 or 10 different variables, a set of separate maps might be better than a single one. I usually found these suggestions to be reasonable, but they are not systematically discussed or developed.

Schmid offers a number of specific suggestions designed to deal with potential memory overload. For example, he suggests putting curve labels contiguous to curves instead of using a legend or key, and limiting pie charts to five or six segments. He explicitly states, however, that "It is not possible to prescribe definitive rules or criteria that might serve as guides for determining the optimal number of curves or other symbols for any particular graphic form, since so many factors are involved in the design process" (p. 33).

Tufte's concern with the relative amount of data ink is driven in part by a realization of human memory limitations. In his efforts to reduce the nondata ink in the display, Tufte makes numerous interesting proposals for formats that may also be more easily processed. He has a very stimulating chapter entitled "Multifunctioning Graphic Elements," in which he advocates: "Mobilize every graphical element, perhaps several times over, to show the data" (p. 139). The chapter reviews numerous examples of formats that fulfill this aim, but it is neither exhaustive nor generative.

5. LONG-TERM MEMORY PROCESSING

Long-term memory is the repository of all one knows. Thus one must access long-term memory to locate stored information about a stimulus (see Spoehr and Lehmkuhle 1982). Locating such information is necessary to "recognize" the stimulus; only by making contact with the appropriate stored information can one apply previously acquired knowledge to a stimulus. Recognition of the graph type, graphic elements, and their interrelations is a critical step in the comprehension process. (If one has never seen a display type before, it is a problem to be solved—not a display to be read.) Another issue in display design, then, is to devise ways of drawing displays so that each part will easily make contact with the previously stored relevant information. That is, displays should not be ambiguous (i.e., subject to more than one interpretation) and should not lead one to access inappropriate information (i.e., such as occurs when one is led to draw incorrect inferences). How well do these books guide the designer to produce unambiguous, straightforward displays?

5.1 Ambiguity in Labels and Design

Parts of displays can be ambiguous for two reasons: First, labels can be missing or unclear. This is not uncommon and can be a real impediment to comprehension. Second, a pattern can be amenable to more than a single description. It can be ambiguous because it is not clear how parts fit together or because the parts themselves are not clear. As is evident in Table 1, the books have different amounts to say about how to avoid ambiguity in display design.

Bertin has an excellent section on the selection of labels, clearly the best of any of these books. The basic idea is that all displays have an "invariant" and "components." The invariant is what is common to all material presented in the display ["the complete and invariable notion common to all the data" (p. 16)], which is what the display is about (e.g., the price of tea in China). The components are the variables ("variational concepts," such as year and region); each component is represented by a separate visual dimension (e.g., height, width, size, darkness, and so on), and values of the component are represented by variations along the dimension. Bertin provides concrete advice about how to describe the invariant and how to label the display effectively. He distinguishes between "external identification," in which the headings permit the reader to identify the invariant and components involved, and the "internal identification," which allows one to recognize which visual variables represent which components. Both topics are discussed systematically in detail. Bertin does not devote a detailed discussion to graphical ambiguity per se, but the topic is mentioned occasionally in the context of specific examples.

Chambers et al. are less concerned with clearly interpretable displays than with those that will help one to analyze and explore data. Indeed, they acknowledge that "the interpretability of nearly every graphical method in this book hinges to some extent on the user's familiarity with the technique. Perhaps the correct question to ask is not how inherently interpretable a display is, but how much background and training—of a general or specific nature—is required to make interpretation easy" (p. 319). There is something to this point of view (especially when displays are used in exploratory data analysis), as they convincingly demonstrate in their book. They also offer some sensible, general advice about labeling, but nothing very detailed (p. 328).

Fisher does not deal with problems of ambiguity directly, but he does include a nice section on how to select titles for maps. He provides some heuristics for deriving a good title, which are similar to those offered by Bertin. Fisher's advice leads to very long titles, however. In addition, he recommends that we "attach a note to most value keys, unless they are *absolutely* clear" (p. 112). Furthermore, "The key must be rational, sensible, and memorable, so that it does not have to be referred to constantly by the user" (p. 112). Unfortunately, little guidance is provided in helping one decide what is rational, sensible, and memorable.

Schmid provides a simple heuristic for formulating a title: Try to answer three questions—What? Where? and When? The What certainly seems always appropriate, but the Where and When would definitely be out of place in most graphs published in the journals I read. Schmid provides sensible, concrete advice about the placement of labels along axes and about how to avoid ambiguous titles of scales. Schmid points out the usefulness of "blow up" and "total" inserts (i.e., subgraphs with additional information) for purposes of clarification, elaboration, or emphasis (p. 50). This advice was sketchy, however, and no specific recommendations were given (e.g., of positioning of inserts, labeling, etc.). The book also includes a reasonable discussion of possible ambiguities in three-dimensional displays, and it is rather sanguine about the possibility of overcoming such problems. I was not convinced by the author's optimism. Indeed, I found that purportedly good examples of three-dimensional displays (e.g., figures 8–16) were somewhat ambiguous, requiring a moment's thought to discover which face of the bar (front or back) portrayed the amounts. The discussion of ambiguities in pictographs, and ways of overcoming them, was convincing, however; this discussion included useful details about the construction of "pictorial unit charts" (essentially bar graphs in which the bars are formed by series of small pictures).

Tufte urges the reader to use clear, detailed, and thorough labeling, and to label important events in the data. There is no detailed procedure that could be used to determine what is important or what should be included in a label. He also points out some of the kinds of ambiguities that can arise when people perceive a surface and convert the area into a number. This point is especially well developed in his discussion of how graphs can be misleading (see Section 5.2). In some of his slimmed-down displays, other forms of ambiguity seem likely to be a potential problem.

5.2 Inferences

Comprehension is more than simply recognizing parts and relationships among them in a display. We often draw inferences after recognizing the key elements and relations. Indeed, patterns of lines can invite the reader to draw conclusions that are sometimes unwarranted. For example, if the appearance of the line itself suggests a rapid or shallow rise, the reader may interpret the content as being consistent with the appearance—even if the numbers tell a different story. That is, we get a visual impression on the basis of the appearance of the display, which can lead us to draw inferences about the actual information content. Thus, graphs can "lie" (e.g., with a vertical axis truncation producing a sharper rise) and distort information as well as make a point especially clearly. How well do these books guide us to avoid creating displays that lie?

Bertin's book is very "question-oriented." He says, "One does not 'read' a graphic; one asks three questions of it" (p. x). Thus he does not consider at length the "bottom-up," "data-driven" processing of the display (i.e., processing of input independent of one's goals and expectations; see Lindsay and Norman 1977), and it is this sort of "taking in what is there" processing that often can lead to different sorts of inferences being drawn by different formats. Bertin does not devote a separate section to a discussion of how graphics can lie. He points out in individual discussions (e.g., of the use of log scales) the sorts of misimpressions that can result (also see p. 235 for scale variations on multiple-scale displays), but his treatment of inferences likely to be drawn from given displays is not easily accessed.

Chambers et al. advise looking at multiple formats for the same data, which is a useful strategy for avoiding unwarranted inferences. They also stress our ability to detect linearity and deviations from it, and to draw the appropriate inferences. In fact, in many of the formats they suggest, they try to arrange for "reference situations" to correspond to straight lines, and they suggest two general ways to "link reference situations to straight lines." These suggestions are: "(i) to plot some set of observed quantities against a set of fitted or standard values, that is, values that they can be anticipated to have on average under the reference assumptions and (ii) to plot residuals from some fitted model against fitted values, or against a variety of other things" (p. 322).

Fisher says that there is a genuine need for a small book entitled *How to Lie With Maps.* He says that the same factors that make "a totally honest display difficult make a misrepresentation easy" (p. 10). He also says, however, that it may sometimes be legitimate to "stress a particular viewpoint" (p. 10), but that it usually is unnecessary and undesirable to slant the display. His concern focuses on the procedure for dividing the data into classes, which can conceal extreme values, the number of locations being sampled, and so on. He makes no hard-and-fast specific recommendations. In addition, as noted in Section 3.4, some of his recommendations for transformations and grouping values into classes may result in displays that lead readers to draw the wrong inferences.

Schmid does not devote a separate section to this topic, but does discuss it in a number of places. For example, he has a nice illustration of the dangers of plotting data using unequal class size (p. 74), and he points out some of the errors that can be introduced in three-dimensional projections (e.g., p. 157) and pictographs (e.g., p. 178). No specific, general principles are offered about ways to encourage readers to draw only proper inferences.

Tufte presents a very cogent and thoughtful discussion of inferences that can be drawn from visual patterns. His discussion of how graphs can lie is interesting and informative, as are his suggestions for avoiding such problems. The main points are presented in the context of well-chosen examples. Again, however, the advice is at a fairly general level (e.g., "do not quote out of context"). Tufte introduces what he calls the "Lie Factor," which is the size of the effect shown in the graph divided by the size of the effect in the data. A problem with this idea is that it is not always clear in three-dimensional displays how big the size of the effect is in the graph. That is, the eye and mind introduce "size constancy," which is the tendency of the system to compensate for distance, so that objects do not seem to be of reduced size when they are farther away. Size constancy results in our seeing small, far away patterns as representing larger objects; simply measuring lines on the display does not convey their psychological magnitude.

6. FORMATS AND PURPOSES

To be read, the stimulus first must be entered into long-term memory and identified as a graph. Following this, a novice graph reader will access and use a set of rules to interpret the display. Pinker (in press) points out that these rules center on a simple principle that is general to most charts and graphs: Greater quantities are indicated by "more" of a mark (higher lines or bars, larger areas, etc.). Scales can be included as a kind of "ruler," allowing one to mark off precise quantities. With practice, however, one can learn to read graphs without actually having to use this rule: One becomes able to recognize that specific *patterns* of marks have specific meanings. For example, an experienced graph reader will see an "X" pattern of lines as indicating a specific kind of statistical interaction and will not need to evaluate each point separately (which is to say that an experienced graph reader will be able to recognize interactions, simply matching the perceived pattern to its stored representation in long-term memory). A bar graph of the same data will not display such an easily seen pattern and hence is less useful for communicating information about interactions. Thus, if you want experienced graph readers to notice interactions at first glance, use a line graph (see Pinker, in press).

Numerous formats usually can be used to display any given set of data. As should be clear in the example just discussed, however, not all of the possible formats are equally easily used for specific purposes. That is, visual displays are used for communicating, analyzing, and storing data. Furthermore, different sorts of information can be communicated, sought, or stored (e.g., point values, point value differences, trends). The books are of more or less use in helping one select the right display type for a given purpose.

6.1 Communication, Analysis, and Storage

It is common to see display types divided into two general classes, those used for communication and those used for analysis; either type of display can also be used to store information. Depending on the intended use of the display, one would take care to make a pretty display or a crude one, would place a premium on ease of immediate comprehension or on completeness, and so on.

Bertin explicitly discusses the three functions of graphic representations (Part I, IIIC) and uses this distinction to discuss alternative formats throughout Part II of the book. Part of his recommendations stem from a consideration of the trade-off between memory capacity and comprehensiveness. Bertin claims that the memory constraints are particularly important for displays intended to communicate information (it is not clear, however, why they are also not important for someone trying to "take in" patterns when using graphics for analysis). This leads him to the problem of how displays can be simplified, and he then proceeds to discuss ways of making displays more easily processed (also see Part I, IIIB). For displays having numerous components, the purpose of the display is considered before suggestions are made about how to simplify it. There are, as usual, some very interesting suggestions here, but they often are embedded in a discursive treatment of examples. This book is unusual in that it offers detailed treatments of plots for use in analysis and in communication. It would be difficult, however, simply to pick up this book and use it as a guide for plotting a set of data to, say, help one compare two distributions (although this problem is in fact treated). Nevertheless, once one has mastered the taxonomies and jargon, and has become familiar with the organization and contents, this book would be a valuable guide for helping one to select an appropriate way to graph a set of data.

Chambers et al. focus almost entirely on graphics as a discovery tool. They are interested in how to use graphics to analyze data, and they offer a rich collection of suggestions about various ways of plotting data for various purposes. One of the more interesting aspects of the discussion is that they consider alternative plots for the same data, suggesting that one begin by plotting the raw data and then move on to using plots to fit plausible models. Depending on the pattern of anomalous points, one can then go on to consider other plots examining other possible models. Although the book has a distinct heuristic flavor, the suggestions are very systematic and useful. The authors do not discuss in any detail how one would construct graphs for communication purposes.

Fisher's book focuses almost exclusively on the communication aspects of cartography. The discussion of different data transforms, however, would be useful in analysis (perhaps more so than in communication).

Schmid declares at the outset that "in this book the orientation and emphasis are frankly focused on presentation" (p. 3). The best discussion of the importance of considering the different purposes is in the chapters on three-dimensional projection and pictorial charts, where he defends the use of such displays in certain communicative contexts.

Tufte emphasizes that "graphics should be reserved for the richer, more complex, more difficult statistical material" (p. 30). He asks, "Why waste the power of data graphics on simple linear changes, which can usually be better summarized in one or two numbers?" But I found myself asking, what if the point is to communicate convincingly that data *are* linear? For most people, seeing a line makes a much stronger impact than reading a correlation coefficient or the like. Tufte takes graphics at their best to be instruments for enhancing human reasoning, and he evaluates their appropriateness primarily in this context. Even so, we do not get much in the way of guidelines about which types of graphs will help us to reason about particular types of data.

6.2 Type of Question

The pattern of lines in a line graph is often organized perceptually into a single unit, which must be broken up in order to see a single value for one case of the independent variable. In contrast, the units of this sort of display make it ideal for seeing patterns and interactions in the data. The situation is exactly reversed for bar graphs, which have one or more bars over each case on the independent variable scale. The corresponding bars often will not form perceptual groups, and it is easy to pick up individual values but more difficult to see trends. These sorts of differences are pervasive, with different formats being more or less well suited for helping a reader to answer specific questions. How well do the books guide us to select formats that will be best for answering particular kinds of questions about the data?

Bertin's discussions of graphical efficiency are in part grounded on the observation that displays are more or less useful for answering particular types of questions. He distinguishes between different "levels of reading," ranging from the "elementary" (at the level of individual values) to the "overall" (at the level of general trends or patterns), and he notes that different visual properties are more effective in the context of different questions. Indeed, in much of his discussion of formats he considers the usefulness of the format for answering different types of questions. In some cases there are very useful tables indicating appropriate uses and question types for different formats (e.g., see pp. 200 and 201). Bertin's recommendations are based largely on his intuitions, which should be empirically investigated. Bertin often makes judgments of the efficiency of a display; he defines "efficient" displays as those that require the shortest period of perception to comprehend well enough to answer a given question. Bertin proposes a set of "rules of construction" to allow one to select the variables to construct the most efficient representation. The "rules" were somewhat disappointing to me, however; they are really general goals and are not algorithmic. This is not a straightforward "how to" book; it requires time and patience to dig out information pertaining to a specific set of data to be used for a specific purpose.

Chambers et al. discuss plotting techniques in the context of different kinds of questions one would be asking of the data. The questions here, however, are closely tied to data analysis; they consider the usefulness of various plots to answer questions addressed by different statistical techniques. Much of the book is concerned with graphical techniques to aid one in comparing data distributions or in using regression analyses. The advice hinges almost entirely on trying to make explicit properties of the data that are usefully considered in data analysis and, as they acknowledge, much of the advice would be inappropriate for other purposes of graphs. If you want to know about ways of plotting data to see symmetry, covariation among values, local densities, variability, and so on, this is your book.

Fisher's book is somewhat slanted toward professional cartographers; he spends time discussing the role of the "sponsor," the one commissioning a map, and provides a sample dialogue between a designer and a sponsor (Chapter 2). This orientation leads him to be sensitive to the composition of the "audience" and their goals. However, the most explicit and interesting advice here focuses on ways of classing data and mapping the classes. Fisher is concerned about ways in which the range and distribution can result in an obscuring of particular differences, and he discusses ways of bringing out distinctions in the map. As noted in Section 3.4, however, this approach does not simply vary the ease of extracting specific information; it actually distorts the visual impression. I was not sure whether there was enough stress on the importance of using these transformations only when the designer is sure that the display will be used for a specific purpose.

Schmid also notes the importance of different kinds of questions that a reader will be trying to answer, resulting in a kind of tension between his offering strong, rigid advice and his being cautious in presenting "hard and fast rules." It is not clear to me how he decided when it was possible to have a rule and when it was not. For example, on page 30 there is a quotation from a publication of the American Society of Mechanical Engineers: "There should be as many horizontal rulings as needed to show the reader the amount value of the plotted points *to the degree of accuracy required.*" The line following the quote is: "As a supplementary rule, there should be a coordinate or grid line for each labeled division of the ordinal axis." However, if the point of the display is to convey the shape of a trend, it is not clear why *any* grid lines should be required.

Tufte advises the reader to cram as much information as possible into a graph. Indeed, he introduces a measure of "data density" (number of entries in the data matrix divided by the area of the display) and argues that this index should be maximized. This advice seems debatable, for two reasons: First, a maximum-information display will not be best for all purposes; extraneous information only serves to distract. Tufte does not make use of the fact that different information may be appropriate depending on the purpose of the display. For example, he decries the use of smoothed curves, even though such general trends are often more appropriate for pedagogical purposes than the actual data (with all of its kinks and random perturbations). For instance, a simple inverted U pattern is sufficient (and possibly better than the actual data) for teaching the Yerkes-Dodson law (relating performance to amount of arousal). Second, the perceiver may well have difficulty processing displays above a given density (this is an empirical question).

7. DATA AND FORMATS

A graphics designer also needs to know which formats are appropriate for which sorts of data. Given the way displays work, then, some formats are more appropriate for specific types of data than others. For example, pie or divided bar charts are only appropriate for proportions and percentages, whereas bar graphs can be used for these sorts of data plus interval or ratio data. The pie and divided bars are "wholes" whose parts must always sum to unity, whereas the height of a bar reflects "how much" and virtually any scale can be used. How well do the books provide guidelines for using different display types for different sorts of data?

Bertin offers a taxonomy (the book is permeated with taxonomies) of levels of organization of data types (essentially nominal, ordinal, and interval-ratio; see Part I, ID), and develops this taxonomy in Part II, I of the book. The taxonomy is used in conjunction with other distinctions (e.g., types of uses of displays) to develop recommendations for acceptable formats for data. Part II is organized hierarchically, with sections on diagrams, networks, and maps; each section is broken down into large subsections according to the number of variables ("components," in his jargon), and these subsections are in turn broken down into sub-subsections devoted to combinations of different types of data (e.g., see p. 199 for graphs for nominal entities paired with quantities, pp. 212–216 for simple time series, and p. 270 for networks). This discussion is extremely thoughtful. It is sometimes cryptic, however, and often compressed. There is so much information here that a well-developed treatment of all points probably would have resulted in a book twice as long as this one. Examples abound, but sometimes they are not developed very well and, occasionally, the underlying principle is difficult to infer. At times I had the feeling that the underlying conceptual system was obscured by the many special cases; sometimes the forest seemed lost for the trees. Much of this required two readings.

Chambers et al. present a very strong discussion of how data can best be graphed to discover underlying patterns. Indeed, this is the core of the book. Unlike Bertin's book, I suspect that this one would be relatively easy to use if you approached it with data in hand. They present numerous well-developed suggestions about how to plot specific sets of data. I found this book to be the most useful for purposes of analysis and reasoning about data.

Fisher offers much useful advice on how to select the right map for a given sort of data. He works through a wide range of different types of symbols that are appropriate for different types of data, and he evaluates the pros and cons of the different formats (see especially Chapters 9 and 10); there is considerable overlap with Bertin here, but Fisher's treatment is easier to read. In fact, for maps illustrating values of only a single variable, he provides a flow-chart-like, step-by-step, procedure for producing the map (pp. 171–174). This procedure is in fact three procedures, one each for nominal, ordinal, and interval-ratio data; numbered footnotes serve as "subroutines" that are called when relevant. This procedure is by no means an algorithm, but it is definitely a step in that direction. In the following chapter the procedure is generalized (less successfully, however) to maps depicting multiple variables. There is a lengthy discussion of the appropriateness of different symbols, and there are some useful observations about circumstances in which different forms of symbolism would be best used or avoided (e.g., a wedge may be better than a bar if it is to refer to a small location on a map; in situations where the bottom of a bar would not fit, the tip of the wedge can point to the location of the quantity). Some of this advice seems highly debatable, however. For example, Fisher advocates use of multiple pie charts superimposed on a map (p. 273), a practice that is condemned (with justification) by Bertin and Tufte.

Schmid's "how to" book sometimes includes fairly detailed directives about when to use particular formats (although the detail pales in comparison to Bertin). For example, on pages 39–42 we are introduced to eight different types of bar charts, with the distinctions usually being described in part in terms of the type of relevant data. Of those reviewed here, this book probably requires the least background knowledge to use and provides a good overview of ways of graphing different sorts of data. Indeed, the overview seems most intent on simply covering the ground, and dubious procedures are included along with the rest (e.g., see Section 3.4). In some cases we get a catalog of ways of displaying data, with little advice about how to select among them (e.g., the discussion of error bars and alternatives on p. 201 ff). This book has a good introduction to semilogarithmic graphs, with illustrations of specific types of data that may be most appropriately plotted using this format (e.g., see p. 92). Advice is often based on unjustified opinion, however, which sometimes is confusing. For example, I had difficulty with the section on "Chernoff faces"; these cartoon faces are used to represent multidimensional data by varying the length of the mouth, width of the nose, shape of the face, and so on, to depict values of different variables, creating an overall expression. H. Wainer is quoted, "We found that the use of Chernoff *faces* for this task yields an evocative and easy to understand display" (p. 187). Schmid then writes, "In light of this experiment, it seems that the utilization of the Chernoff cartoons as a technique of visual communication leaves much to be desired" (p. 188). I could not locate the source of the Wainer quote (it was from a newsletter), and I was frustrated by the apparent non sequitur.

Tufte makes a case for his suggestion that the number of information-carrying dimensions should not exceed the number of dimensions in the data, which would eliminate most graphs depicting three-dimensional objects and many graphs depicting two-dimensional objects (e.g., dollar bills of different sizes, each representing purchasing power in a different year). This is useful advice. Tufte has a section presenting advice about when to use sentences, text-tables, tables, semigraphics, and graphics (p. 178 ff), in which the most useful information concerns how simple tables should be composed, but there is very little guidance about which kind of graph to use for specific data.

8. CONCLUSIONS

These books provide many valuable insights into the design and use of graphs for communicating and understanding data. In most cases, the advice is well-founded. In other cases, however, the advice is less clearly well-founded and could have profited from a consideration of empirical facts about how human readers process visual information. In reading the books, I was repeatedly struck by the potential value of continued research on graphics for both statisticians and psychologists. Both groups are interested in ways of making explicit patterns in data, and both have recognized the value of graphics in data display. Graphics obtain their power because they tap into the very powerful (but at times eccentric) human visual information-processing system. Thus a further understanding of this system, in addition to continued creative exploration of display design, can only be to the good of both fields. To quote Kruskal (1982):

> Statistical graphics has, of course, a double connection with psychology. First, there is the application of psychological ideas, findings, and theories to statistical graphics, for example, paradigms of care in carrying out and reporting empirical work. Second, there is traffic in the other direction: the use of statistical graphics, and other parts of statistics too, by psychologists in their properly psychological planning and analysis of experiments. Given this double connection, I am somewhat surprised that there has not been more joint research on graphical methods by psychologists together with statisticians. (p. 290)

Given the desirability of continuing to develop new types of graphs (which would make explicit different sorts of patterns in data) and of improving the use of graphs as a communication tool, the question becomes how to devise specific, important research questions. Although these books offer many insights and intuitions that should inspire empirical research, it is desirable to devise a systematic research program aimed at investigating basic issues about graphic communication. One way to systematically develop a program of empirical research is to consider how one would program a computer to emulate an expert human graph designer (compare Kosslyn 1980, 1983). Formulating alternative ways to program a computer would help one to define the research questions (see Kosslyn 1980 for a detailed illustration of this approach). Indeed, the first thing one would discover is that much research would be required before one could even begin to write such a program. For example, one would need to know which visual dimensions "naturally" (for humans) best convey different types of relationships in data, and what sorts of questions are best answered by making specific regularities in the data explicit. In addition, one would want to know the specifics of the parameters to be used in producing optimally readable displays (e.g., minimal sizes at given light/dark contrasts, etc.). Developing the design for an "artificial expert graphmaker" would force one to consider the issues systematically and in depth, and actually writing such a computer program is a sure-fire way of coming face-to-face with the fundamental problems. Without question, there is much room for further research on the topic!

It is clear from reading the books that much of display design is still an art, and it will probably be some time before a machine can emulate the intuitions and wisdom that leaven all of these books. As Table 1 makes clear, the books have different strengths and weaknesses; depending on one's purpose, any one of the books may be the one to read. These are good books. There is much to be learned within these pages.

[*Received September 1984. Revised March 1985.*]

REFERENCES

Anderson, J. R., and Bower, G. H. (1973), *Human Associative Memory*, Hillsdale, NJ: Lawrence Erlbaum.

Bertin, J. (1983), *Semiology of Graphs* (translated by W. J. Berg), Madison, WI: University of Wisconsin Press.

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

Clement, D. E. (1978), "Perceptual Structure and Selection," in *Handbook of Perception* (Vol. 9), eds. E. C. Carterette and M. P. Friedman, New York: Academic Press, pp. 49–84.

Cowey, A. (in press), "Aspects of Cortical Organization Related to Selective Attention and Selective Impairments of Visual Perception: A Tutorial Review," in *Attention and Performance XI*, eds. M. I. Posner and O. S. Martin, Hillsdale, NJ: Lawrence Erlbaum.

Dodwell, P. C. (1975), "Pattern and Object Perception," in *Handbook of Perception* (Vol. 5), eds. E. C. Carterette and M. P. Friedman, New York: Academic Press, pp. 267–300.

Eells, W. C. (1926), "The Relative Merits of Circles and Bars for Representing Component Parts," *Journal of the American Statistical Association*, 21, 119–132.

Ericsson, K. A., Chase, W. G., and Faloon, S. (1980), "Acquisition of a Memory Skill," *Science*, 208, 1181–1182.

Fisher, H. T. (1982), *Mapping Information*, Cambridge, MA: Abt Books.

Frisby, J. P. (1980), *Seeing: Illusion, Brain and Mind*, New York: Oxford University Press.

Garner, W. R. (1974), *The Processing of Information and Structure*, Hillsdale, NJ: Lawrence Erlbaum.

Glass, A. L., Holyoak, K. J., and Santa, J. L. (1979), *Cognition*, Reading, MA: Addison-Wesley.

Kaufman, L. D. (1974), *Sight and Mind: An Introduction to Visual Perception*, New York: Oxford University Press.

Kosslyn, S. M. (1980), *Image and Mind*, Cambridge, MA: Harvard University Press.

——— (1983), *Ghosts in the Mind's Machine*, New York: W. W. Norton.

Kruskal, W. H. (1982), "Criteria for Judging Statistical Graphics," *Utilitas Mathematica*, 21B, 283–310.

Lindsay, P. H., and Norman, D. A. (1977), *Human Information Processing*, New York: Academic Press.

Macdonald-Ross, M. (1977), "How Numbers Are Shown: A Review of Research on the Presentation of Quantitative Data in Text," *Audiovisual Communications Review*, 25, 359–409.

Marr, D. (1982), *Vision*, San Francisco: W. H. Freeman.

Norman, D. A. (1978), *Attention and Memory*, San Francisco: W. H. Freeman.

Pinker, S. (in press), "Understanding Graphs," *Cognitive Science*.

Playfair, William (1786), *The Commercial and Political Atlas*, London: Corry.

Posner, M. I. (1978), *Chronometric Explorations of Mind*, Hillsdale, NJ: Lawrence Erlbaum.

Schmid, C. F. (1983), *Statistical Graphics*, New York: John Wiley.

Smith, S. L. (1979), "Letter Size and Legibility," *Human Factors*, 21(b), 661–670.

Spoehr, K. T., and Lehmkuhle, S. W. (1982), *Visual Information Processing*, San Francisco: W. H. Freeman.

Stevens, S. S. (1974), "Perceptual Magnitude and Its Measurement," in *Handbook of Perception* (Vol. 2), eds. E. C. Carterette and M. P. Friedman, New York: Academic Press, 361–389.

Teghtsoonian, J. (1965), "The Judgment of Size," *American Journal of Psychology*, 78, 392–402.

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

Van Essen, D. C., and Maunsell, J. H. R. (1983), "Hierarchical Organization and Functional Streams in the Visual Cortex," *Trends in NeuroSciences*, September, 370–375.

Wainer, H., and Thissen, D. (1981), "Graphical Data Analysis," *Annual Review of Psychology*, 32, 191–241.

Williams, R. L. (1956), *Statistical Symbols for Maps: Their Design and Relative Values*, New Haven, CT: Yale University, Map Laboratory.