Visual Psychophysics of Simple Graphical Elements

Article *in* Journal of Experimental Psychology Human Perception & Performance · December 1990

DOI: 10.1037//0096-1523.16.4.683 · Source: PubMed

1 author:

Ian Spence
University of Toronto

Journal of Experimental Psychology:
Human Perception and Performance
1990, Vol. 16, No. 4, 683-692

0096-1523/90/$00.75

Visual Psychophysics of Simple Graphical Elements

Ian Spence
University of Toronto, Ontario, Canada

The accuracy with which graphical elements are judged was assessed in a psychophysical task that parallels the real-life use of graphs. The task is a variant of the Metfessel-Comrey constant-sum method, and an associated model based on Stevens's law is proposed. The stimuli were horizontal and vertical lines, bars, pie and disk slices, cylinders, boxes, and table entries (numbers). Stevens's law exponents were near unity for numbers and 1-dimensional elements but were also close to 1 for elements possessing 2 or 3 apparent dimensions—subjects accommodate extraneous dimensions that do not carry variation, changing the effective dimensionality of the stimulus. Judgment errors were small, with numbers yielding the best performance; elements such as bars and pie slices were judged almost as accurately; disk elements were judged least accurately, but the magnitude of the errors was not large.

The usefulness of statistical graphs for the analysis and communication of data has increased as the costs and difficulties of producing charts and graphs have declined in recent years. Perhaps as a result of this growth in use, psychologists and statisticians have become interested in how people perceive statistical graphs, for both practical and theoretical reasons. Recent examples of empirical work include Cleveland and McGill (1984, 1986), Lewandowsky and Spence (1989), Spence and Lewandowsky (in press), Wainer and Thissen (1979), and Broersma and Molenaar (1985). More theoretical discussion—much of which is based on results from traditional psychophysics—may be found in Bertin (1983), Cleveland (1985), Macdonald-Ross (1977), and Kosslyn (1985, 1989).

In statistical graphs, quantities such as frequencies, cumulative frequencies, percentages, and proportions are often represented by the lengths of lines, the areas of bars or pie slices, and sometimes by the apparent sizes of more complicated elements such as cylinders, boxes (rectangular parallelepipeds), and other volumes of various kinds drawn in perspective. Using different graphical elements and also a table, Figure 1 shows a variety of ways of displaying the same data. Clearly, the accuracy with which an observer infers numerical values depends critically on the perception of the size of the elements used to represent the numbers. Elements associated with substantial perceptual bias are undesirable. There is much advice in the literature regarding the suitability of graphical elements for the representation of numerical quantities. Tufte (1983), for instance, recommends avoiding all unnecessary elaboration such as extraneous dimensions that do not carry information about the data—thus the use of cylinders or boxes is anathema—but he offers no data in support of this recommendation. Macdonald-Ross (1977) advocates the use of bar charts in preference to pie charts, assuming that the judgment of bars involves only an assessment of length, whereas estimating the size of a pie slice involves a combination of area, angle, and arc length. He makes the point that because the exponents of psychophysical power functions tend to be smaller for judged volume and judged area than for judged length, graph makers should avoid elements with high apparent dimensionality. A similar position is taken by Cleveland and McGill (1984), who presented judgmental accuracy data for several graphical elements but did not estimate Stevens's exponents. Croxton and Stein (1932) advise using bars in preference to other graphical elements: They showed that bars were judged more accurately than squares or circles, which were in turn judged more accurately than cubes. Their results are consistent with progressively smaller exponents as the apparent dimensionality rises.

Although many authors have appealed to the psychophysical literature to support their recommendations, and also as a basis for theoretical analysis, there appear to be no experiments that attempt to estimate Stevens's exponents in a psychophysical task that reflects the behavior of observers of graphs. In this article, I propose a psychophysical task and an associated judgmental model as well as two statistical estimators of the Stevens's exponent. The method is applied to graphical elements of various apparent dimensionalities.

The Constant-Sum Procedure

The Task

When we examine a graph, we usually compare the sizes of individual graphical elements. How much greater was the rainfall in September than May? Is the price of oil in constant dollars increasing or decreasing from year to year? Do more people subscribe to *Time* than *Newsweek*? Did the ABC Corporation pay the largest dividends last year, or did XYZ?

This research was supported by Grant A8351 from the Natural Sciences and Engineering Research Council of Canada to Ian Spence and by Grant G1779 from the Natural Sciences and Engineering Research Council of Canada to R. Baecker, A. Fournier, P. Muter, D. Olson, and I. Spence.

I thank Yoshio Takane for helpful comments on the manuscript and Linda Tilley for testing the subjects and preparing the graphs.

Correspondence concerning this article should be addressed to Ian Spence, Department of Psychology, University of Toronto, Toronto, Ontario M5S 1A1, Canada.

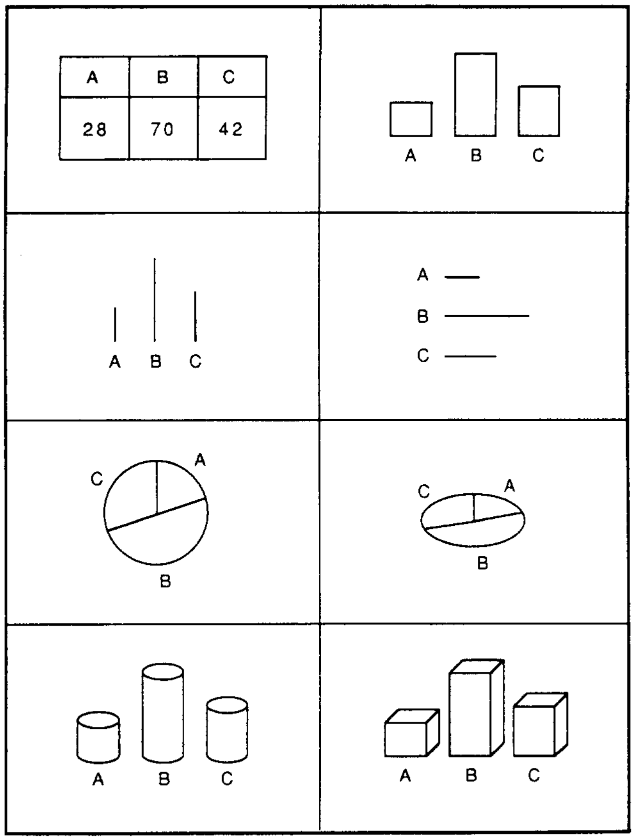

Figure 1. One table (top left) and seven graphs that display the same numerical data. The table elements are numbers. The graphical elements are bars, lines [vertical], lines [horizontal], pie slices, disk slices, cylinders, and boxes.

Indeed, the estimation of individual magnitudes, while not unimportant, is probably not as useful as the comparison of the relative sizes of quantities. It is the ability to perceive relations among the elements of graphs that makes graphs such powerful tools, and the use of particular graphical elements, rather than others, may make this process of comparison easier. Hence, the experimental task described here requires subjects to compare graphical elements, rather than make magnitude estimates of elements presented in isolation.

The method may be applied with any element, but the following illustration uses boxes: A subject is shown two boxes, generally of different sizes, and must estimate the apparent proportion of the whole that each represents. The instructions do not ask the subject to attend to "height" or "volume" but simply to "size." There are many ways of eliciting this judgment, but in the present experiment subjects divided a horizontal line, of fixed length, into two parts whose lengths were proportional to the perceived sizes of the boxes (see Figure 2). Some subjects saw a numerical scale below the line but others did not.

The constant-sum method is not restricted to studies involving graphical elements but has wider application. Its invention is usually credited to Metfessel (1947), but Comrey (1950) and Torgerson (1958) are mainly responsible for its development and popularization; see also Goude (1962) and Fagot (1981) for the closely related part-sum method. The present procedure differs slightly in that, at least within a particular experimental condition, the magnitudes of the two physical stimuli always sum to the same constant. The method may have the advantage (over conventional magnitude estimation tasks) of avoiding the problems of sequential effects (Cross, 1973; Morris & Rule, 1988; Ward, 1973) and other effects noted by Poulton (1968). Although the basic task is not new, its use to estimate Stevens's law exponents is novel.

Estimating the Exponent

On each trial, a subject judges the sizes of two quantities, Π and Ω, the sum of which is a constant value that may be assumed to equal unity, without loss of generality. Thus, $\Pi + \Omega = 1$, or, alternatively, $\Pi + (1 - \Pi) = 1$. Assume that the subject's (unspoken) magnitude estimate of the size of either quantity follows Stevens's law; for example,

$$\phi = \alpha\Pi^{\beta},$$

where α is a scale factor and β is the exponent of the power function. If, for either quantity (Π or Ω), the subject is asked to provide an estimate (P or Q) of its proportion of the whole, the judged proportion may be modeled as

$$P = \frac{\alpha\Pi^{\beta}}{\alpha\Pi^{\beta} + \alpha(1 - \Pi)^{\beta}} = \frac{1}{1 + [(1 - \Pi)/\Pi]^{\beta}},$$

for P, and a similar expression holds for Q. This implies that the judged proportions (P and Q) for the two objects sum to unity.

A slightly different conceptualization also leads to the same psychophysical function. If one assumes that the subject bases the judgment on the ratio of the subjective size of one object to the subjective size of the other subject, the same model results, because

$$\frac{(1 - P)}{P} = \frac{Q}{P} = \frac{\alpha(1 - \Pi)^{\beta}}{\alpha\Pi^{\beta}} = [(1 - \Pi)/\Pi]^{\beta};$$

hence,

$$\frac{1}{P} - 1 = [(1 - \Pi)/\Pi]^{\beta},$$

which is equivalent to

$$P = \frac{1}{1 + [(1 - \Pi)/\Pi]^{\beta}}.$$

Consequently, one cannot distinguish the two possible strategies with this experimental method.

In the typical case where $\beta \neq 1$, the function relating P and Π is S-shaped (e.g., see Figure 3). This type of function has been observed empirically (e.g., Nakajima, 1987; and also the present experiment).

The model

$$P = \frac{1}{1 + [(1 - \Pi)/\Pi]^{\beta}}$$

may be restated as

$$[(1 - P)/P] = [(1 - \Pi)/\Pi]^{\beta}.$$

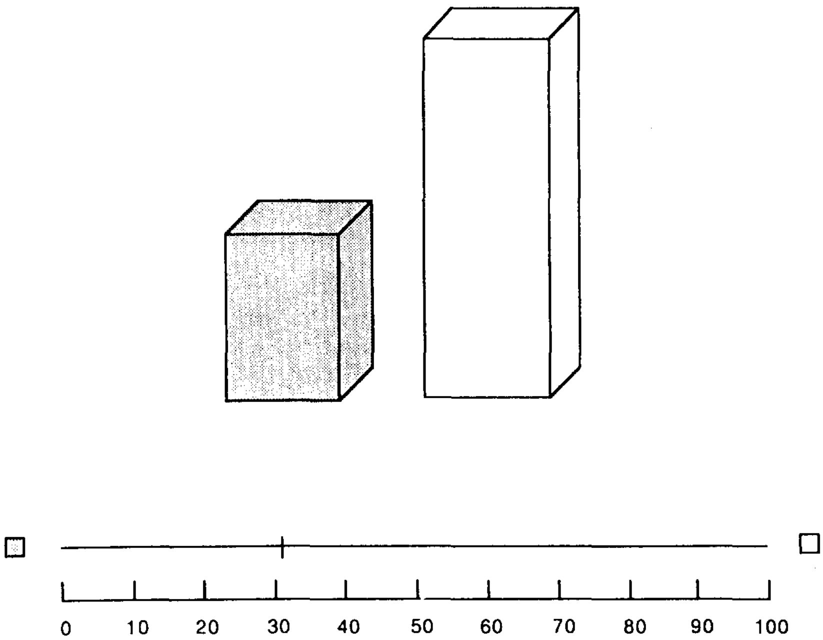

Figure 2. Task display. (Subject must position the cursor so that the horizontal line is divided in proportion to the apparent sizes of the elements, which are boxes in this example.)

Taking logarithms yields

$$\log[(1 - P)/P] = \beta \log[(1 - \Pi)/\Pi].$$

If, under replication, the subject produces judgments P_i corresponding to several values of Π_i, the exponent β may be estimated by least squares:

$$\hat{\beta} = \frac{\sum_i \log[(1 - P_i)/P_i]\log[(1 - \Pi_i)/\Pi_i]}{\sum_i \log^2[(1 - \Pi_i)/\Pi_i]}$$

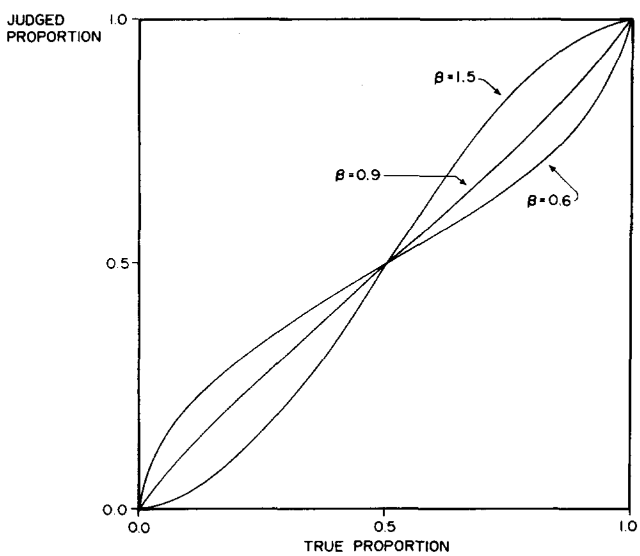

Figure 3. Three theoretical psychophysical functions for Stevens's exponents of 0.6, 0.9, and 1.5.

Optionally, a robust estimate may be used. One particularly simple method uses Siegel's (1982) repeated-median estimator, which, in this case, reduces to

$$\tilde{\beta} = \text{med}_i\{\log[(1 - P_i)/P_i]/\log[(1 - \Pi_i)/\Pi_i]\}.$$

In this article, robust estimates are preferred. Untrained subjects produce occasional aberrant responses that can exert considerable leverage on least squares estimates (Spence & Lewandowsky, 1989). These outliers are sometimes due to inattention or boredom but can also occur as the result of a reversal of the response scale. A reversal of the scale occurs, for example, if the subject divides the line in the ratio 30/70 when the true stimulus ratio is 70/30. It is clear that such reversals do not represent a failure to perceive the ratio accurately, but rather constitute a failure to execute the task properly. Occasional apparent reversals of scale were observed in the experiments reported here (see below).

Thus the Siegel repeated-median estimator was used with the Stevens's law exponent. Like the ordinary median, this estimator has an efficiency of about two thirds with Gaussian data, and although robust estimators with higher efficiencies are available, the median-based estimator was preferred because of its simplicity.

Experiment 1

Method

Subjects. Subjects were University of Toronto undergraduates who were recruited by means of sign-up sheets posted at various locations on campus. They were paid $5 for their participation.

Apparatus. Stimuli were presented by using an IBM PS/2 Model 80-071, with 80387 mathematics coprocessor and VGA graphics adaptor, on an IBM 8513 monitor with approximate screen dimensions of 22 × 17 cm. The distance between the subject's head and the monitor was approximately 1 m, with the screen subtending about 8° horizontally and 6° vertically. The room was illuminated by subdued incandescent lighting, and the monitor was positioned to minimize reflections from other objects in the room.

Stimuli. Each display was created and presented, virtually instantaneously, by using the graphics primitives available in Microsoft C, Version 5.0 (Microsoft, 1987). The element sizes on each trial were determined by sampling randomly from a discrete uniform distribution on the range 2–98, thus assigning this value to one element and its hundreds complement to the other. Seven different graph elements were used, as well as table elements (numbers). The element pairs were displayed in three different overall sizes: small, medium, and large, in the ratios 1:2:4. The largest element pair occupied roughly two thirds of the screen area.

The graphical elements were similar to those shown in Figure 1, and they had apparent dimensionalities of one (horizontal and vertical lines), two (bars and pie and disk slices), and three (boxes and cylinders). The different element types were presented to the subjects in a display like that shown in Figure 2. To help subjects properly associate each of the elements with the correct end of the to-be-divided line, the experimenter showed one element in outline form (white on black) and the other in a light-gray fill. Small squares, one in outline and the other shaded light gray, were positioned at the ends of the to-be-divided line.

The table elements (numbers) were positive integers. Because numbers are symbols and not geometrical objects like the seven graphical elements, they are assumed to have zero dimensionality. The table elements (numbers) were arbitrarily chosen to sum to 140 in the large condition, 70 in the medium condition, and 35 in the small condition.

Subjects moved the cursor using the left and right arrows on the cursor keypad to divide the line before pressing the Enter key. The initial cursor position was determined by moving 25 percentage points away from the optimal position and then superimposing a random normal deviate with standard deviation 3. If the resulting value fell outside the 0% and 100% boundaries, the cursor was placed at the closer boundary. The response latency, from display onset to pressing the Enter key, was obtained and recorded. The next trial started 500 ms after the preceding response.

Instructions. Subjects were given written instructions covering various aspects of the study before the practice and experimental trials. The most critical instructions concerned the task:

> You must move the cursor left or right to divide the line into two parts, with each part proportional to the sizes of the two parts in the figure above. Thus if the shaded part of the figure is about the same size as the unshaded part, you would position the cursor somewhere in the middle. If the shaded part is very small relative to the unshaded part, you would position the cursor near the left end of the line. If the shaded part is larger than the unshaded part, then the cursor should be nearer the right-hand end of the line.

Subjects were instructed to work quickly but also accurately.

Design and procedure. Each subject saw only one type of graphical element (e.g., boxes), and the overall size of the elements was varied among, but not within, three consecutive experimental sessions of 100 trials each, with two short rest breaks between sessions; small, medium, and large elements were used. The judgments of 12 subjects were obtained for each of the eight element types, requiring 96 subjects in total. Six of each subgroup of 12 subjects made their judgments with the assistance of a graduated scale, with divisions at 10% intervals labeled 0, 10, 20, . . .100. The remaining six used no scale. Thus, judgments were made with and without the presence of an explicit numerical scale.

All possible presentation orders of the element sizes were used with each group of 6 subjects, and thus the basic building block of the experiment was a changeover design (composed of two 3 × 3 Latin squares). A 10-trial practice period, with subjects using large elements, preceded the experimental sessions. The experimenter observed each practice trial; if subjects did not perform the task appropriately, they had to repeat the practice trials. The experiment lasted about 50 min for most subjects.

Results

Individual psychophysical functions. Figure 4 shows six psychophysical functions from the set of 288 (96 subjects × 3 stimulus sizes) generated in the experiment; each function was fit to the 100 judgments according to the robust procedure described earlier. The element type is shown at the top left of each panel, and in the two cases in which a scale was used, this is indicated. The estimated Stevens's exponent is shown at the bottom right of each panel. The examples shown are atypical inasmuch as five of the six have exponents that differ markedly from unity, illustrating the distinctive S-shape of the data in such cases. Nearly all of the 288 scatterplots exhibited variabilities similar to those shown in Figure 4: The variabilities were larger than that associated with the disk example (at lower left) in only a handful of cases, with the worst approximately 1.5 times as variable. Most of the 288 plots showed only small deviations from linearity, and when nonlinearity was present it displayed the characteristic S-shape present in five of the six panels of Figure 4. Visual inspection and various goodness-of-fit statistics (such as the mean- and median-squared errors) suggest that the model used to estimate the Stevens's law exponent generally fits very well. As discussed later, group-average estimated exponents did not deviate greatly from unity, but individual exponents as low as 0.70 and as high as 1.35 were observed, with the middle 50% lying between 0.84 and 1.01, and an overall median of 0.95.

Figure 4 contains data (in the two center panels) that suggest that these subjects reversed the scale on some trials (this occurred between 1% and 5% of the time with 14 of the 96 subjects.) As can be seen, the fitted regression has resisted the effect of such outliers. Visual inspection of all regressions in which scale reversals may have occurred confirm that this phenomenon did not adversely affect the estimation of the exponent. Another interesting feature is seen in the disk example, where the subject has divided the line only at points corresponding to the graduation marks on the accompanying scale.

Combined analysis. There was no effect involving presentation order for any of the three response variables (exponent, accuracy, and latency). The presence or absence of a numerical scale below the line was only significant with accuracy, $F(1, 80) = 5.24$, $MS_e = 2.58$, $p < .001$: Overall accuracy was slightly better with the scale (2.8%) than without (3.6%). There was no main effect of scale with either exponent or latency.

There were significant effects involving stimulus size, but only with exponent and not with accuracy or latency. The main effect of size, $F(2, 160) = 32.02$, $MS_e = 0.0043$, $p <$

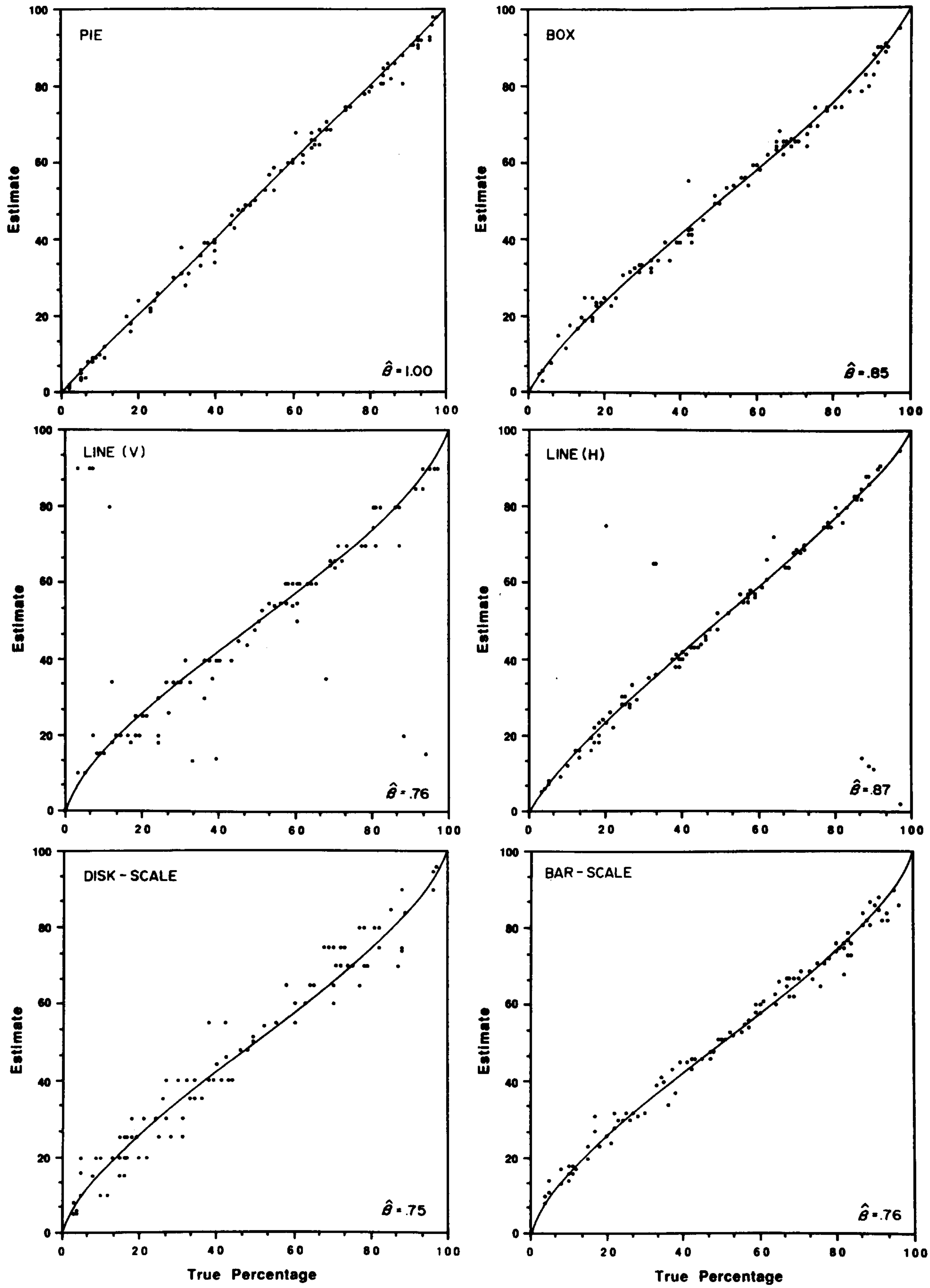

Figure 4. Six empirical psychophysical functions for a variety of element types (Some subjects used an accompanying numerical scale, but others did not. The estimated Stevens's exponents are shown.)

.001, the Size × Scale, $F(2, 160) = 4.24$, $MS_e = 0.0043$, $p < .05$, and the Size × Element, $F(14, 160) = 1.97$, $MS_e = 0.0043$, $p < .05$, interactions were significant. The main effect of stimulus size was not large: The average exponent was 0.99 for small stimuli, 0.96 for medium stimuli, and 0.91 for large stimuli.There was a small Size × Element interaction, however: The pattern was the same across elements—the smaller the stimulus, the larger the exponent—but differences for low dimensional objects were negligible (the maximum difference for horizontal and vertical lines, or numbers, was 0.05, and the associated 5% Fisher's least significant difference was 0.06).

Element was significant for all three response variables: exponent—$F(7, 80) = 4.16$, $MS_e = 0.033$, $p < .001$; accuracy—$F(7, 80) = 5.24$, $MS_e = 2.58$, $p < .001$; and latency—$F(7, 80) = 4.50$, $MS_e = 12.92$, $p < .001$. Apart from the interaction with size, no other interactions with element attained significance at the .05 level.

Figure 5 shows the estimated exponents for the element types. The figure is a dot chart (constructed after the fashion of Cleveland & McGill, 1984), where the dot position indicates the value of the response measure and the error bars are 95% confidence intervals for the means. For comparison, the bent arrows indicate typical values of exponents for one-, two-, and three-dimensional objects (Baird, 1970). The estimated exponents for the elements vary in the range 0.89–1.07, with mean 0.95, and thus resemble exponents typically obtained with one-dimensional objects. It is instructive to note that the theoretical function (in Figure 3) for an exponent of 0.9 deviates only slightly from the straight line associated with an exponent of 1.0. For an exponent of 0.9, the maximum percentage discrepancy associated with nonlinearity is about 2.4% and is typically much less. In practice, discrepancies will be larger because of other sources of error.

Figure 6 shows that the most accurate performers are table elements (numbers), pie elements, and bar elements, with accuracies better than 3%. Boxes, cylinders, and lines lie between 3% and 4%. The worst is the disk element, but even it has an average error of only 4.1%. There appears to be no association between apparent dimensionality and accuracy.

As shown in Figure 7, there are differences in the average time taken to respond, with element types of low apparent dimensionality requiring longer time. The two- and three-dimensional elements were judged almost 50% faster than the zero- or one-dimensional elements.

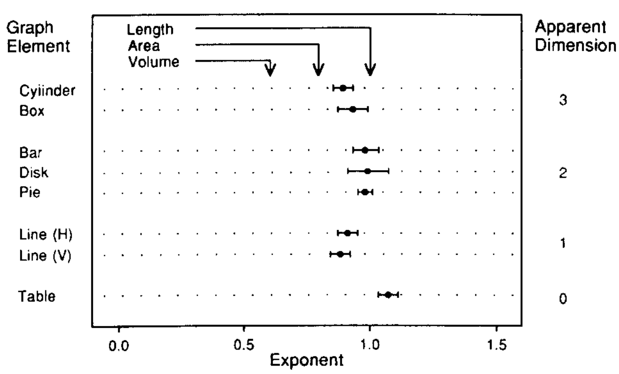

Figure 5. Dot chart with 95% confidence intervals showing the average exponent for each element.

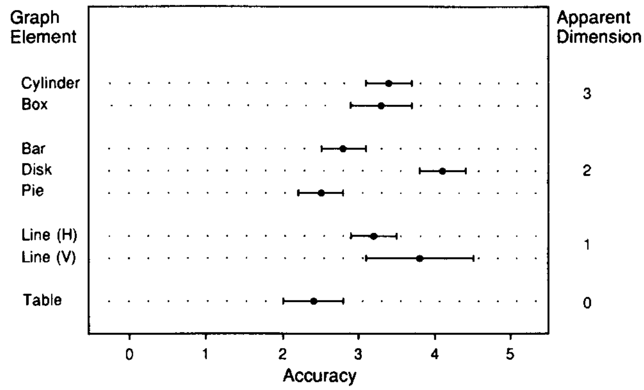

Figure 6. Dot chart with 95% confidence intervals showing the average accuracy (absolute discrepancy in percent) for the elements.

Discussion

Modeling the Metfessel-Comrey task. The data from 288 sessions were fit using the model proposed earlier. In virtually all cases it provided a very good description of the data, suggesting that the subjects' behavior was not inconsistent with the view that they either form the ratio of the apparent size of one object to the sum of the apparent sizes of both objects, or, equivalently, the ratio of the size of one object to the other. Also, although the model has not been used to fit data collected by others, it may be able to account for the shape of psychometric functions found by other investigators (e.g., Nakajima, 1987) using similar tasks.

Effect of dimensionality. Contrary to general opinion (e.g., Cleveland & McGill, 1984; Macdonald-Ross, 1977), an increase in the apparent dimensionality of a graphical element is not necessarily associated with a concomitant decrease in the Stevens's law exponent (see Figure 5). Likewise, the apparent dimensionality of a graphical element seems to bear no relationship to how accurately its magnitude is judged.

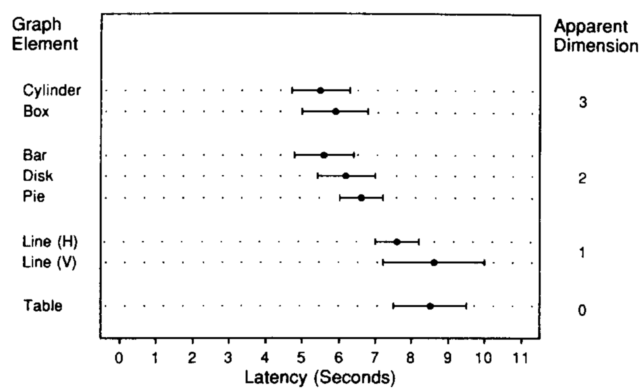

Figure 7. Dot chart with 95% confidence intervals showing the average latency (response time in seconds) for the elements.

Elements with apparent dimensionalities greater than one were judged as though the elements possessed only one dimension, but they were judged more quickly than lower dimensional elements. Of course, in general, Stevens's exponents for area and volume are considerably smaller than unity for most real and apparent objects (Baird, 1970), and higher dimensional objects are not usually judged as accurately (Cleveland & McGill, 1986; Croxton & Stein, 1932); however, care must be taken to distinguish situations in which the extra dimensions do not carry information and are merely decorative. When the additional dimensions of graphical objects do not show variability, the judgment of size becomes essentially one of length, and the present results are not surprising when considered in that light.

Apart from the pies and disks, only one dimension of the two- and three-dimensional objects was varied. This accords with normal practice in contemporary graphs, where it seems that the advice of authors like Karsten (1923), Mudgett (1930), Huff (1954), and Schmid (1983) has generally been heeded. Although area changes in a more complicated way in a pie or disk chart, it seems likely that most subjects do not use area as the primary basis for their judgments of size. Because the total area of the circles or disks did not vary, most subjects probably used the angle, arc length, or chord length rather than the slice area as a measure of size. Eells (1926) reported that only 25% of his subjects used area as a basis for their judgments of size, with the other 75% preferring to use the angle or arc/chord length. Several subjects in this study said they compared the segment angle with the angle formed by halves (180° = 50%) or quarters (90° = 25%). This suggests that judgments of pie and disk slices are effectively unidimensional and that some subjects create coarse (but linear) reference scales by using strategies such as the one just mentioned (Simkin & Hastie, 1987).

Effect of stimulus size. The effect of stimulus size on exponent was small but reliable, at least for the two- and three-dimensional elements, and may have been caused by the framing effect of the screen surround. Künnapas (1955) and Rock and Ebenholtz (1959), for example, have shown that a line enclosed by a small frame is judged to be longer than a line of equal length surrounded by a large frame. As previously noted, the effect is questionable with table elements (numbers) and lines but is convincing with the other elements. If framing is indeed responsible for these small effects, two- and three-dimensional elements may be more susceptible because they extend in both horizontal and vertical directions within the frame.

Tables. Tufte (1983) has suggested using tables instead of graphs when only a few numbers must be presented, and Ehrenberg (1975) has also championed the use of properly constructed tables. Although table elements (numbers) are judged accurately, subjects take more time than with almost any graphical element. Hence, tables are preferable only if the audience is able to devote sufficient time and energy to their interpretation. With casual readers, who are less likely to linger, graphs may be superior to tables. Spence and Lewandowsky (in press) compared the accuracy with which ordinal comparisons of size were made with complete bar graphs, pie charts, and tables to display proportions. They found that simple pairwise comparisons were made equally accurately with each of the display types. The results of the present study are consonant with that finding. However, Spence and Lewandowsky (in press) found that for comparisons involving three or more elements, the table was inferior to the graphs when processing time was restricted. The present study suggests that this may have been because subjects make judgments of table elements (numbers) more slowly.

Effect of stimulus variation. It may be argued that the task in Experiment 1 is somewhat unnatural: In the real world, numerical information is presented in a variety of formats, and one must switch rapidly among presentation styles within a single newspaper, book, magazine, or journal, and even within single articles. Readers forced to go back and forth among several formats may have to adopt different strategies with broader foci of attention. Experiment 2 was designed to explore the possibility that different results might be obtained when the subject is forced to switch rapidly from one format to another. Furthermore, in Experiment 1 comparisons among elements were between subjects, leaving open the possibility that observed differences among groups, despite random allocation, may have capitalized upon individual differences. In Experiment 2 all subjects judged all elements.

Experiment 2

Method

Subjects. Subjects were University of Toronto undergraduates who were recruited by means of sign-up sheets posted at various locations on campus. They were paid $20 for their participation in three separate experimental sessions of about 1 hr each.

Apparatus and stimuli. These were identical to those used in Experiment 1.

Instructions. These were the same as in Experiment 1.

Design and procedure. Each subject saw all eight graphical elements, as used in Experiment 1, but the size of the elements did not vary: All elements were of medium size. Subjects participated in four blocks of 80 trials, at approximately the same time on each of 3 consecutive days—making 12 blocks in total. Hence each subject made 960 judgments: 120 for each of the eight element types. The elements were presented in random order, with 10 presentations of each type in each block of 80 trials. There was a subject-paced break of a few minutes after each block of trials. As in Experiment 1, the numerical values of the percentages represented by the elements were chosen randomly for each trial.

All subjects made their judgments with the assistance of a graduated scale, with divisions at 10% intervals labeled 0, 10, 20, . . .100. A 16-trial practice period, with two exemplars of each of the eight elements placed in random positions, preceded the experimental sessions. The experiment required about 60 min on each of 3 consecutive days. Ten subjects were used.

Results

Robust estimates of Stevens's exponents were calculated for each element by amalgamating each subject's data from the 12 blocks. Thus each subject's exponent is based on 120 responses. There were significant element effects: exponent—$F(7, 63) = 4.50$, $MS_e = 0.002$, $p < .001$; latency—$F(7, 8640)$

= 39.07, MS_e = 32.33, $p < .001$; and accuracy—$F(7, 8640)$ = 10.68, MS_e = 28.59, $p < .01$. For both accuracy and latency, the Subject × Block and Subject × Element interactions were significant at the .05 level, but each accounted for a very small percentage of the variance in the experiment. Examination of the graphs of these interactions did not reveal systematic patterns.

Table 1 shows the element means for each of the response measures, alongside the means from Experiment 1. Both experiments show essentially the same patterns of results, with three exceptions. First, the accuracies for horizontal and vertical lines are reversed, with vertical lines judged slightly more accurately in Experiment 2. Second, the pie chart is judged much less accurately (3.8% error vs. 2.5% error) in Experiment 2. And third, the latencies for all elements are longer in Experiment 2 and no longer show a general reduction with increasing apparent dimensionality.

Learning effects over blocks were observed: latency—$F(11, 8640)$ = 57.22, MS_e = 32.33, $p < .001$, and accuracy—$F(11, 8640)$ = 5.24, MS_e = 28.59, $p < .01$. Subjects improved slightly in both speed and accuracy over the first three blocks of trials before settling down to relatively stable levels of performance.

It is interesting to note that the variability of the latency and accuracy responses (though not for the exponent response) has risen by a factor of 2 or 3 from Experiment 1 to Experiment 2. Closer inspection of the data reveals a lengthening and increase in weight of the tails of the distributions, probably reflecting the increased difficulty of the task. Whereas the medians and means for accuracy and latency differ little in Experiment 1, the differences are generally larger in Experiment 2, and in one case (pie chart: median = 2.3; mean = 3.8), the difference is very large indeed. This probably indicates that once subjects are used to the task, they are just as capable of making accurate judgments as in the first experiment, but the strain of having to switch quickly between eight different types of display induces inaccuracies or delayed responses in a significant proportion of cases.

Discussion

Many of the comments regarding the results of Experiment 1 apply equally well here. It seems that the primary consequence of forcing a subject to switch rapidly among different element types is to increase the variability of responding: There are larger differences between the means and medians of the accuracy and latency distributions, reflecting a lengthening and thickening of the tails. More gross errors are made by the subjects, and they also take longer to respond, sometimes considerably longer. These effects are presumably caused by the increased load on the subject. Indeed, subjects reported finding the task difficult and tiring, whereas no such complaints were voiced during the first experiment.

Because graphs are normally viewed in the context of other ongoing activities with varying demands that require immediate switching of attention, Experiment 2 may be a better indicator of real-world performance than Experiment 1. The first experiment required the subjects to repeat exactly the same task many times, and subjects probably became quite efficient at estimating the relative sizes of the elements partly as a result of not needing to switch rapidly and unpredictably among the several quite different strategies appropriate for each of the individual graph elements.

General Discussion

Related Studies

Mudgett (1930) proposed, but did not perform, a study similar to those reported here, and Croxton and Stein (1932) reported an experiment that is closely related to the present work. They presented pairs of figures, on cards, to subjects and asked them to say what fraction one was of the other. The figures were either horizontal bars, squares, circles, or perspective drawings of cubes. The bars varied in length, the circles in diameter, and the squares and boxes in length (or apparent length) of side. The mean judgmental accuracy for bars was found to be about 2.9%, for squares about 7.8%, for circles about 7.2%, and for cubes about 13.9%. Their value for bars, which is the only one that may legitimately be compared, accords remarkably well with our results. The accuracies for squares and circles are consistent with a Stevens's exponent of about 0.8, and about 0.6 for the cubes. But it must be remembered that Croxton and Stein simultaneously varied all dimensions present in their figures, except for bars in which only one dimension was varied.

Table 1
A Comparison of Experiments 1 and 2

Response measure	Apparent dimensions 0	1		2			3	
	Table	Line (V)	Line (H)	Pie	Disk	Bar	Box	Cylinder
Exponent								
Experiment 1	1.1	0.9	0.9	1.0	1.0	1.0	0.9	0.9
Experiment 2	1.0	1.0	0.9	1.0	1.0	1.0	1.0	0.9
Accuracy								
Experiment 1	2.4	3.8	3.2	2.5	4.1	2.8	3.2	3.3
Experiment 2	2.8	3.2	3.9	3.8	4.0	2.9	3.0	3.0
Latency								
Experiment 1	8.6	7.6	8.7	6.6	6.1	5.5	5.9	5.4
Experiment 2	12.5	10.1	10.3	12.2	11.4	10.1	10.1	10.1

Note. V denotes vertical and H denotes horizontal.

Cleveland and McGill (1986) used a similar task. They presented four stimuli and had subjects compare three of the stimuli to the fourth, estimating the percentage size of each relative to the standard. They did not calculate Stevens's exponents—although this could be done in a fashion similar to that proposed here—but they did report the accuracies. Where comparable, the accuracies were slightly poorer than those in the present study. This may be due to differences in the task and the experimental procedure: They required three judgments of four simultaneously presented stimuli, thus introducing dependencies, and the administration was by paper and pencil. They fitted quadratic functions to the accuracies (as a function of the true percentage) and found, surprisingly, that the maxima of these functions were generally located at true percentages greater than 50. Cleveland and McGill (1986) offered no explanation. If the judgment process is appropriately described by a Stevens's power function, the accuracy function should be

$$A(Q_1) = |Q_1/Q_2 - (Q_1/Q_2)^{\beta}|,$$

where Q_1 is the true size of the smaller stimulus and Q_2 is the size of the standard. The maximum may be found in the usual way, by equating the first derivative to zero, giving a true percentage of 25 for an exponent of 0.5. For exponents of 0.8, 1.2, and 2.0, the maxima fall at 33%, 40%, and 50%, respectively. Only for exponents greater than 2.0 do the maxima lie above 50%. For an exponent of unity, the accuracy function is flat. Some of Cleveland and McGill's functions appear to be fairly flat and are presumably consistent with a unit exponent.

Conclusions

Because one-, two-, and three-dimensional objects (such as lines, bars, and boxes) all yield exponents in the region of unity, we conclude that reduced exponents are not a simple function of the dimensionality of objects per se, but rather depend on the number of dimensions carrying variation. If rectangles with an invariant base length are presented to subjects, an exponent around one is obtained, whereas if rectangles with varying heights and bases were used, the estimated exponent would generally be lower (Baird, 1970). With boxes, the exponent may be expected to be around 0.6 if all three dimensions vary and around 0.8 if only two dimensions vary. This tuning of the exponent to the variability of dimensions deserves further investigation to determine how exponents for judged size depend on the amount of variation observed along each of the dimensions.

An alternative interpretation of the somewhat surprising finding that the exponents for two- and three-dimensional objects are in the region of unity is based on Teghtsoonian's (1965) finding that subjects can make proportional judgments of size, provided they are instructed appropriately. Teghtsoonian showed that judged area is linear with physical area when subjects are instructed to judge "real" rather than "apparent" area. In the present experiments, subjects were not instructed to judge the real size of the objects—indeed, the instructions were similar to those in Teghtsoonian's apparent size condition. Nonetheless, because graphical elements are typically seen in situations in which they are used to represent numerical quantities in a proportional fashion, subjects may have assumed that they should make judgments of real rather than apparent size. This possibility deserves further investigation.

From a practical viewpoint, and considering the nature of the tasks required in the present experiments, the best all-round candidates for representing numerical quantities are probably bars, boxes, and cylinders: All yield exponents close to unity, and all are judged accurately and quickly. Although the pie chart does almost as well, it may be more susceptible to gross errors when the viewer has to work quickly or is under some stress. Table elements are also compared accurately, but it should be remembered that they require more time. There seems little reason to prefer horizontal and vertical lines on the grounds of speed or accuracy to higher dimensional elements. For some applications, disks may not be a bad choice, because despite their relative inaccuracy, they exhibit no large systematic deviation from linearity.

Tufte (1983) has advocated using plain graphs that maximize the "data ink ratio." Thus, he prefers simple forms, such as lines, for representing numerical quantities over forms that possess irrelevant extra dimensions, such as cylinders. He explicitly recommends that "the number of information-carrying dimensions should not exceed the number of dimensions in the data" (Tufte, 1983, p. 71). The present results cast some doubt on the wisdom of this recommendation, and to the contrary, suggest that elements with high apparent dimensionality lack nothing in accuracy and may be processed faster under some circumstances. The nonessential dimensions must not carry information but be solely ornamental.

Attractive displays often result when high dimensional elements with irrelevant extra dimensions are used, and undoubtedly, attractiveness plays a role in drawing the attention of the reader. We may also speculate that attractive graphs containing elements resembling objects in the real world may be better remembered than those based on more abstract elements.

References

Baird, J. C. (1970). *Psychophysical analysis of visual space.* New York: Pergamon Press.

Bertin, J. (1983). *Semiology of graphics.* Madison, WI: University of Wisconsin Press.

Broersma, H. J., & Molenaar, I. W. (1985). Graphical perception of distributional aspects of data. *Computational Statistics Quarterly, 2,* 53–72.

Cleveland, W. S. (1985). *The elements of graphing data.* Monterey, CA: Wadsworth.

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association,79,* 531–553.

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

Comrey, A. L. (1950). A proposed method for absolute ratio scaling. *Psychometrika, 15,* 317–325.

Cross, D. V. (1973). Sequential dependencies and regression in psychophysical judgments. *Perception & Psychophysics, 14,* 547–552.

Croxton, F. E., & Stein, H. (1932). Graphic comparisons by bars, squares, circles, and cubes. *Journal of the American Statistical Association, 27,* 54–60.

Eells, W. C. (1926). The relative merits of circles and bars for representing component parts. *Journal of the American Statistical Association, 21,* 119–132.

Ehrenberg, (1975). *Data reduction: Analyzing and interpreting statistical data.* New York: Wiley.

Fagot, R. F. (1981). A theory of bidirectional judgments. *Perception & Psychophysics, 30,* 181–193.

Goude, G. (1962). *On fundamental measurement in psychology.* Stockholm, Sweden: Alqvist & Wiksell.

Huff, D. (1954). *How to lie with statistics.* New York: Norton.

Karsten, K. G. (1923). *Charts and graphs.* New York: Prentice-Hall.

Kosslyn, S. M. (1985). Graphics and human information processing. *Journal of the American Statistical Association, 80,* 499–512.

Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology, 3,* 185–225.

Künnapas, T. (1955). Influence of frame size on apparent length of line. *Journal of Experimental Psychology, 50,* 168–170.

Lewandowsky, S., & Spence, I. (1989). Discriminating strata in scatterplots. *Journal of the American Statistical Association, 84,* 682–688.

Macdonald-Ross, M. (1977). How numbers are shown: A review of research on the presentation of quantitative data in texts. *Audio-Visual Communication Review, 25,* 359–407.

Metfessel, M. (1947). A proposal for quantitative reporting of comparative judgments. *Journal of Psychology, 24,* 229–235.

Microsoft. (1987). *Microsoft C optimizing compiler, version 5.0.* [Computer program] Redmond, WA: Microsoft Corporation.

Morris, R. B., & Rule, S. J. (1988). Sequential judgment effects in magnitude estimation. *Canadian Journal of Psychology, 42,* 69–77.

Mudgett, B. D. (1930). *Statistical tables and graphs.* Cambridge, MA: Riverside Press.

Nakajima, Y. (1987). A model of empty duration perception. *Perception, 16,* 485–520.

Poulton, E. C. (1968). The new psychophysics: Six models for magnitude estimation. *Psychological Bulletin, 69,* 1–19.

Rock, I., & Ebenholtz, S. (1959). The relational determination of perceived size. *Psychological Review, 66,* 387–401.

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

Siegel, A. F. (1982). Robust regression using repeated medians. *Biometrika, 69,* 242–244.

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

Spence, I., & Lewandowsky, S. (1989). Robust multidimensional scaling. *Psychometrika, 54,* 501–513.

Spence, I., & Lewandowsky, S. (in press). Displaying proportions and percentages. *Applied Cognitive Psychology.*

Teghtsoonian, M. (1965). The judgment of size. *American Journal of Psychology, 78,* 392–402.

Torgerson, W. S. (1958). *Theory and methods of scaling.* New York: Wiley.

Tufte, E. R. (1983). *The visual display of quantitative information.* Cheshire, CT: Graphics Press.

Wainer, H., & Thissen, D. (1979). On the robustness of class of naive estimators. *Applied Psychological Measurement, 4,* 543–551.

Ward, L. M. (1973). Repeated magnitude estimations with a variable standard: Sequential effects and other properties. *Perception & Psychophysics, 13,* 193–200.

Received February 3, 1989
Revision received December 21, 1989
Accepted December 27, 1989 ■

1991 APA Convention "Call for Programs"

The "Call for Programs" for the 1991 APA annual convention was included in the October issue of the *APA Monitor.* The 1991 convention will be held in San Francisco, California, from August 16 through August 20. Deadline for submission of program and presentation proposals is December 14, 1990. This earlier deadline is required because many university and college campuses will close for the holidays in mid-December and because the convention is in mid-August. Additional copies of the "Call" are available from the APA Convention Office.