Perception & Psychophysics
1989, 46 (4), 365-374

Efficiency of graphical perception

GORDON E. LEGGE, YUANCHAO GU, and ANDREW LUEBKER
University of Minnesota, Minneapolis, Minnesota

The term *graphical perception* refers to the part played by visual perception in analyzing graphs. Computer graphics have stimulated interest in the perceptual pros and cons of different formats for displaying data. One way of evaluating the effectiveness of a display is to measure the *efficiency* (as defined by signal-detection theory) with which an observer extracts information from the graph. We measured observers' efficiencies in detecting differences in the means or variances of pairs of data sets sampled from Gaussian distributions. Sample size ranged from 1 to 20 for viewing times of 0.3 or 1 sec. The samples were displayed in three formats: numerical tables, scatterplots, and luminance-coded displays. Efficiency was highest for the scatterplots ($\cong$60% for both means and variances) and was only weakly dependent on sample size and exposure time. The pattern of results suggests parallel perceptual computation in which a constant proportion of the available information is used. Efficiency was lowest for the numerical tables and depended more strongly on sample size and viewing time. The results suggest serial processing in which a fixed amount of the available information is processed in a given time.

Computer graphics have stimulated lively interest in the design of new formats for displaying data. Numerical data can be displayed in countless formats, but only some of them are well suited to the information-processing capacities of human vision. The phrase *graphical perception* has been coined to refer to the role of visual perception in analyzing graphs (Cleveland 1985, chap. 4; Cleveland & McGill, 1985). These authors have studied several elementary visual tasks relevant to graphical perception, such as discrimination of slopes or lengths of lines.

Cleveland and McGill attribute the great advantage of graphical displays over numerical tables to the capacity of human vision to process pattern information globally at a glance. Julesz (1981) has used the term *preattentive* to refer to perceptual operations that can be carried out in parallel across the visual field.

While it is clear that we can make relative statements concerning the perceptual superiority of one type of data display over another, can we also make absolute statements about the perceptual effectiveness of a given type of display? *Efficiency*, as defined by signal-detection theory, provides a measure on an absolute scale of how effectively information is used. Efficiency ranges from zero to one and represents the performance of a measuring instrument or real observer relative to the performance of an *ideal observer*. The ideal observer makes optimal use of all available information. H. B. Barlow and colleagues have measured human efficiency for several visual information-processing tasks, including detection of mirror symmetry (Barlow & Reeves, 1979), discrimination of dot density (Barlow, 1978), detection of modulation of dot density (van Meeteren & Barlow, 1981), and discrimination of the number of dots in displays (Burgess & Barlow, 1983).

Our purpose was to use the concept of efficiency to study graphical perception. Two common graphical tasks are the estimation of means and variances in sets of noisy data. We have measured observers' efficiencies for estimating these two statistical parameters from samples of data drawn from Gaussian distributions.

We compared efficiencies for data displayed in three different formats: numerical displays, scatterplots, and luminance-coded displays. Numerical tables are the traditional means for displaying data. How well can observers estimate the means and variances of columns of numbers? Scatterplots are representative of pictorial methods for displaying data. They take advantage of the substantial capacities of spatial vision. Luminance-coded displays are ones in which salient differences are conveyed by luminance contrast. A large body of recent research points to the importance of contrast coding in vision.

Burgess and Barlow (1983) described two generic ways in which human performance can be suboptimal. Observers might simply fail to use some of the information available to them, yet optimally process the remaining information. On the other hand, observers might use all the information, but contribute imprecision (intrinsic noise) due to errors of internal representation. Burgess and Barlow (1983) showed how these two factors—incomplete sampling and internal noise—can be teased apart by measuring discrimination thresholds as a function of the level of externally added visual noise. (See also Barlow, 1977; Burgess, Wagner, Jennings, & Barlow, 1981; Legge, Kersten, & Burgess, 1987; Pelli, 1981.) We used this

Preliminary reports of this work were given at the 1987 annual meeting of the Association for Research in Vision and Ophthalmology, and the Spatial Displays and Spatial Instruments Conference sponsored by the NASA Ames Research Center and UC Berkeley, Asilomar, CA, August 31-September 3, 1987. We thank Traci Toomey for many hours of dedicated service as an observer. The research was supported by US Public Health Service Grants EY02857 and EY02934 and AFOSR Grant 86-0280. Correspondence may be addressed to Gordon E. Legge, Department of Psychology, Elliott Hall, 75 East River Road, Minneapolis, MN 55455.

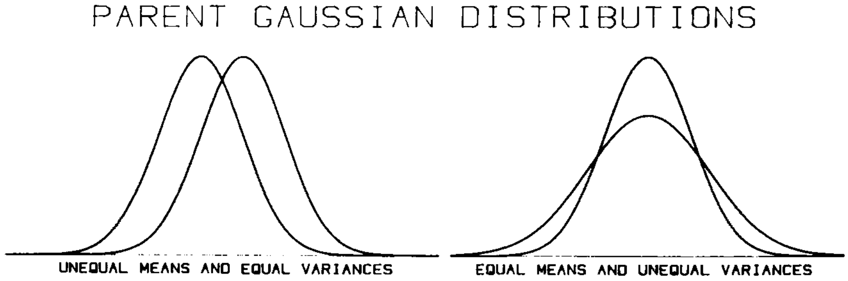

Figure 1. Stimuli were based on values sampled from two Gaussian distributions. In one series of experiments, these distributions had equal variances but means that differed by one standard deviation (right). In a second series, the distributions had equal means but their variances differed by a factor of two (left).

technique to evaluate the roles of the two factors in limiting perceptual estimation of means and variances.

METHOD

The stimuli were derived from numbers drawn at random from two parent Gaussian distributions (Figure 1). In one series of experiments, the two Gaussian distributions had identical variances, but their means differed by one standard deviation. In a second series, the two Gaussian distributions had identical means, but their variances differed by a factor of two.

In an experimental trial, an observer was shown N random samples from each distribution for a time T. The observer was instructed to choose which set of samples was drawn from the distribution with the higher mean (or larger variance). An observer's performance was measured as percent correct in blocks of 300 trials. At least two blocks, collected in separate sessions, were averaged to measure performance levels.

In order to compute efficiency, percent correct was transformed to accuracy, D, using formulae from signal-detection theory. For discrimination of means, given equal variance, D is equivalent to detectability d'. An observer's accuracy D_{obs} is given by

$$D_{obs} = \sqrt{2}Z, \qquad (1)$$

where Z is the standardized normal deviate corresponding to proportion correct in the two-alternative forced-choice procedure (Green & Swets, 1974). D_{ideal} is the accuracy obtained if an optimal strategy is used. In the case of means, the optimal strategy is straightforward: compute the means of the two sets of samples and choose the one with the higher value. When there are N samples,

$$D_{ideal} = \sqrt{N}(M_1 - M_2)/SD, \qquad (2)$$

where M_1 and M_2 are the means of the two parent distributions and SD is the common standard deviation of the parent distributions. In our experiments, the difference of the means was equal to the standard deviation, so

$$D_{ideal} = \sqrt{N}. \qquad (3)$$

In the case of equal means but unequal variances, the optimal strategy is to compute the sum of squared deviations of sample values from the parent mean for each set of samples, and choose the sum with the higher value. When there are N samples, D_{ideal} is given by

$$D_{ideal} = \sqrt{N}(SD_1/SD_2 - SD_2/SD_1), \qquad (4)$$

where SD_1 and SD_2 are the standard deviations of the parent distributions (Egan, 1975, p. 136).[1] The accuracy of the real observer is obtained from proportion correct as follows. With formulae given by Egan (1975, pp. 239–240), receiver-operating characteristics (ROC curves) can be computed for different values of SD_1 and SD_2. Equation 4 associates a value of D with each ROC curve. The area under the ROC curve is proportion correct in forced choice (Green & Swets, 1974). Therefore, the ROC forges the link between the observed proportion correct and a value of accuracy attributable to the real observer.

An observer's efficiency, E, is defined to be (Tanner & Birdsall, 1958)

$$E = (D_{obs}/D_{ideal})^2. \qquad (5)$$

Efficiency ranges from zero to one and provides an absolute scale for judging the effectiveness of an observer's performance.

As illustrated in Figure 2, the stimuli were displayed in three formats: numerical displays, scatterplots, and luminance-coded displays. Figure 2 shows examples in which the sample size N was 10. In each case, the observer was asked to choose which member of the stimulus pair was drawn from the parent distribution with higher mean or larger variance. In the case of scatterplots, the observer judged the mean or variance of the vertical positions of dots on the screen. For luminance coding, the judgment referred to luminance levels of the bars.

All stimuli were displayed on a Conrac SNA 17/Y monochrome video monitor with P4 phosphor, at a viewing distance of 64 cm. For numbers and scatterplots, the stimuli had a luminance of 300 cd/m² on a black background. Stimuli were generated with an LSI-11/23 computer and Grinnell GMR274 frame buffer with a display resolution of 512×480 pixels. Stimuli were presented for either 300 or 1,000 msec. A masking pattern of Xs was presented immediately upon termination of each display so that afterimages could not be used by the observers. The following three paragraphs contain details of the three types of display formats.

Numerical Displays

The parent Gaussian distributions had means of 45 and 55, and standard deviations of 10. Sampled values were rounded off to the nearest integer before being displayed. Accordingly, there was a quantization error corresponding to about 5% of the standard deviation. (We did not take quantization error into account in computing the accuracy of the ideal observer. The issue of quantization noise is dealt with in detail by Burgess [1985]. In general, the effects of this type of noise are negligible if the quantization error

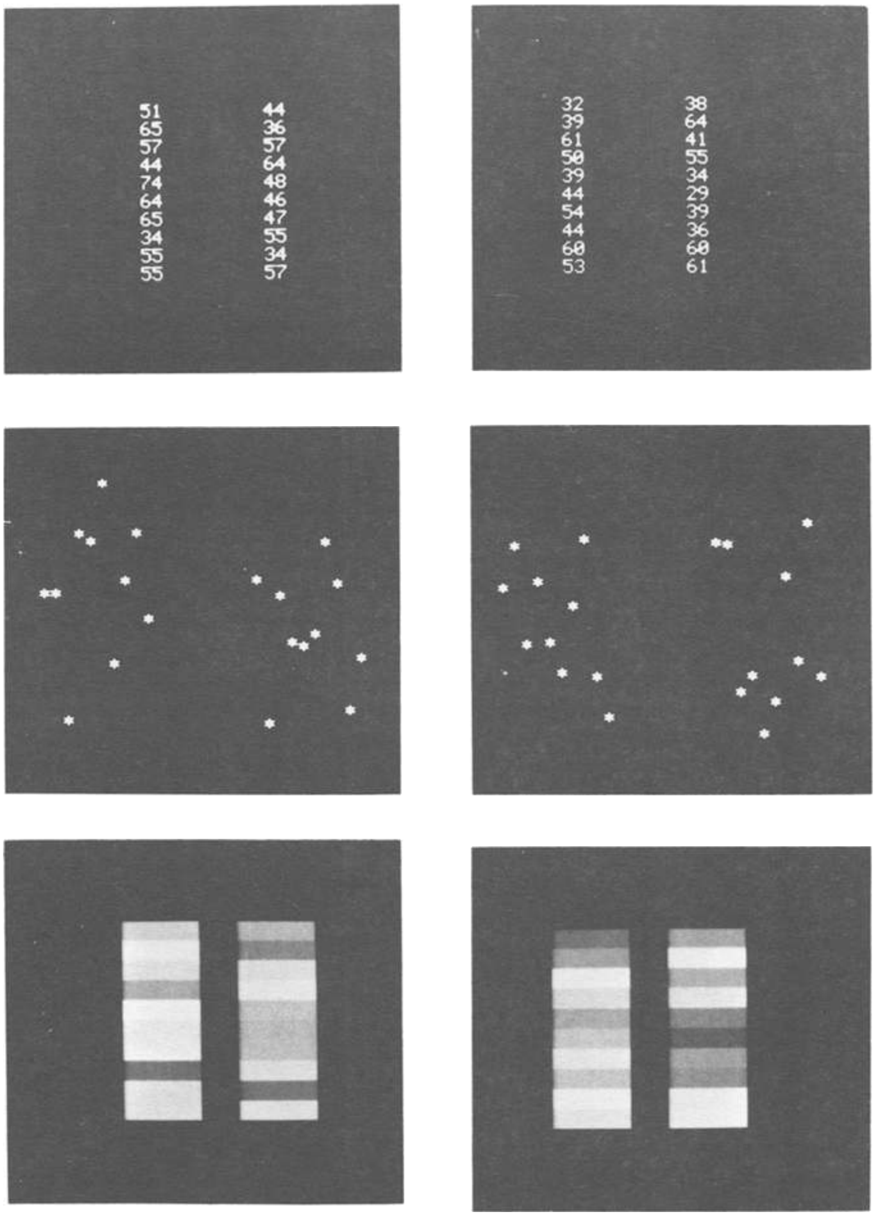

Figure 2. The three display formats. The column of three panels on the left shows the same trial displayed as numbers, scatterplots, and luminance-coded bars. Values plotted top to bottom in the numerical and luminance displays are plotted right to left in the scatterplots. The subject's task was to decide which set of samples was drawn from the Gaussian distribution with the higher mean. The three panels on the right depict a trial in which the subject's task was to decide which set of samples came from the distribution with larger variance.

is small compared with the standard deviation of the noise process.) The center-to-center spacing of the digits was .51° and the empty space separating the two columns subtended 4.5°.

Scatterplots

Numerical values drawn from the Gaussian distributions were transformed linearly to vertical position on the screen. A + symbol centered on the screen specified the vertical position midway between the two means. One standard deviation of the Gaussian distribution subtended 2.6°. Each sample was displayed as a * symbol subtending .51°, placed on the screen with an accuracy of 2% of one standard deviation. The horizontal spacing between samples was .51° and the horizontal separation between the two sets of samples was 4.5°.

Luminance-Coded Displays

The numbers drawn from the Gaussian distributions were mapped to luminance and displayed as bars on the screen. The digital-to-analog converter of the frame buffer quantized values to 256 levels. A look-up table was used to correct for the video monitor's nonlinear transformation from voltage to luminance. The two distributions had mean luminances of 150 and 185 cd/m², and the standard deviation was 35 cd/m². In terms of luminance, the quantization error was 2% of one standard deviation near the mean luminance levels of the two distributions. Each bar subtended 3.5° × 0.86°. The columns of bars were separated by 1.8°.

In a separate experiment, we measured *thresholds* for discriminating differences in mean values. A threshold was found by reducing the difference between the means of the parent Gaussian distributions until a criterion accuracy was achieved. The criterion was 75% correct, corresponding to D_{obs} = .95. The QUEST procedure was used to find the threshold difference in means (Watson & Pelli, 1983). As discussed below, the threshold data were used to distinguish between internal noise and inappropriate sampling as reasons for an observer's deviation from ideal performance.

Two highly practiced observers participated in the experiments.

Monte Carlo computer simulations were used to evaluate the efficiencies of some putative perceptual strategies. The simulations were programmed in Pascal and run on a Sun 3/160 with a floating point accelerator board. Gaussian random numbers were obtained from a uniform random distribution (the RANDOM function under Sun UNIX), then transformed to a Gaussian random variable based on the cumulative normal distribution (Abramowitz & Stegun, 1972, p. 933). Each value of efficiency was computed from 100,000 simulated trials.

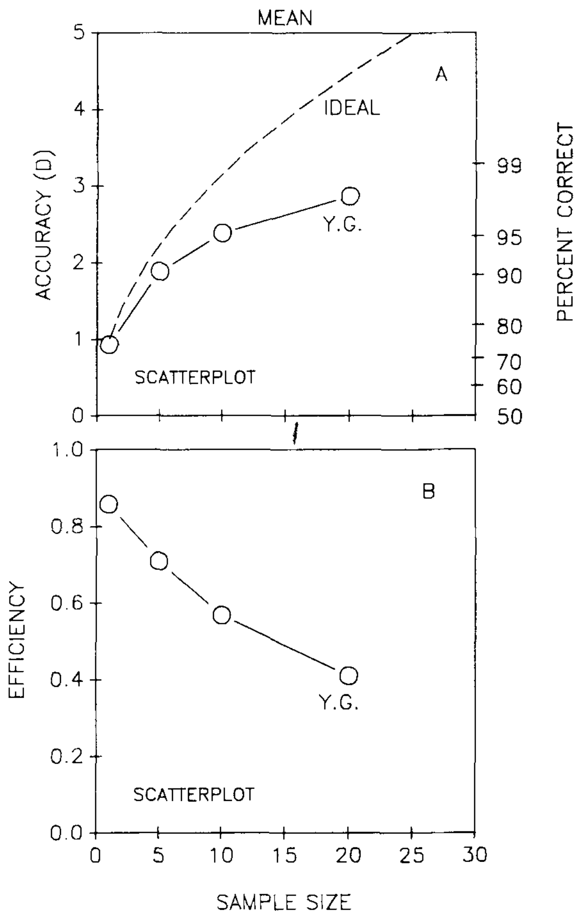

Figure 3. Illustration of the use of *efficiency* as a performance measure. In Panel A, Y.G.'s performance for discriminating scatterplot means is plotted as a function of sample size. Accuracy, *D*, is plotted on the left vertical scale and corresponding values of percent correct on the right. The performance of the ideal observer, Equation 3, is also shown. Panel B shows corresponding values of Y.G.'s efficiency, computed using Equation 5.

RESULTS AND DISCUSSION

Effect of Sample Size

Figure 3A illustrates Subject Y.G.'s performance with scatterplot means for sample sizes ranging from 1 to 20. Accuracy D_{obs} is plotted on the left vertical scale and corresponding values of percent correct on the right. The dashed line shows the performance of an ideal observer. Notice that when the sample size is one, both Y.G. and the ideal observer have accuracies near 1.0 (percent correct near 76). When only one sample is drawn from each of the overlapping parent Gaussian distributions, there will be occasions when the sample from the distribution with lower mean is greater than the sample from the distribution with higher mean. The ideal observer will select the higher value and be counted wrong. In this way, the accuracy of the ideal observer is limited by the variability (noise) inherent in the parent distributions. As the sample size increases, the standard deviations of the sampling distributions decrease in proportion to $\sqrt{N}$, and the performance of the ideal observer steadily improves. Figure 3A illustrates that the performance of the real observer also improves, but not as fast as that of the ideal. Y.G.'s efficiency can be computed from D_{obs} and D_{ideal}, using Equation 5. Figure 3B shows the data of Figure 3A transformed to efficiency on the vertical scale. Even though Y.G.'s accuracy increased with sample size (Figure 3A), his efficiency declined slowly for the same range of sample size (Figure 3B). This decline reflects the growing vertical separation between the curves for Y.G. and the ideal observer in Figure 3A as sample size increases. In subsequent figures, we plot efficiency as the dependent variable.

In Figure 4, efficiency is shown for the discrimination of means. The viewing duration was fixed at 0.3 sec. Data are shown for the three display formats. Best-fitting straight lines have been fit through the sets of data. Since both scales are logarithmic, a slope of −1.0 would represent inverse proportionality between efficiency and sample size. Values of the slopes are shown on the figure and are summarized in Table 1.

Efficiencies are highest for scatterplots, >50%. These values are a little higher than the 25% efficiencies obtained by Barlow and Reeves (1979) for the detection of mirror symmetry, and equivalent to the 50% efficiency observed by Barlow (1978) for the discrimination of dot density. The scatterplot data show a weak dependence on sample size (slopes of −.23 and −.36). Had the slopes been zero (constant efficiency), subjects would have been processing new samples with equal effectiveness. Instead, both observers made substantial, but incomplete, use of additional information available from additional samples.

Efficiencies are lowest for numerical displays and show a more pronounced decline with increasing sample size. The decline means that subjects are less able to use the additional information provided by the extra samples.

The results for luminance lie intermediately between those for scatterplots and numerical displays.

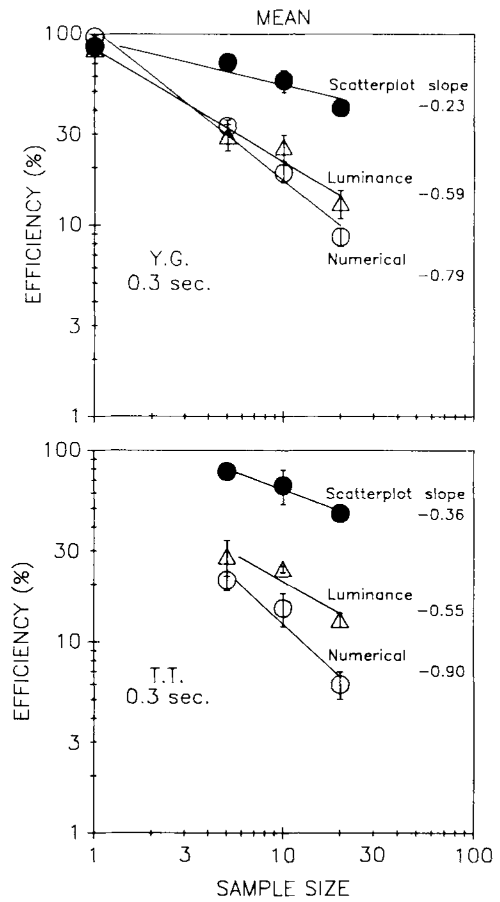

Figure 4. Efficiency for discriminating means is plotted as a function of sample size on log-log coordinates. Data are shown for the three display formats. Slopes are given for the best-fitting lines through the sets of data. These slopes are listed in Table 1. The two panels are for Observers Y.G. and T.T.

Notice that Y.G.'s efficiencies for sample sizes of one were close to 100%. The $N = 1$ condition is really a control to verify that performance is not limited by visual resolution or other forms of sensory discrimination. Small deviations from 100% efficiency in the $N = 1$ condition can probably be attributed to occasional nonperceptual errors (e.g., "finger errors"). The high efficiencies in the $N = 1$ condition lead us to believe that the substantial departures from 100% efficiency in other conditions are not sensory in origin but result from more central processes. Brief experiments with two low-vision observers—one with central-field loss due to optic atrophy and the other with severe corneal vascularization (Snellen acuity = 20/1,000)—confirm this point. Both had about the same efficiencies for scatterplots as normal observers.

In these experiments, the means of the parent Gaussian distributions (with standard deviations equal to the difference of the means) were mapped into arbitrary values along a stimulus dimension. In the case of scatterplots, the means were separated by 2.6° on the face of the video monitor. In the case of luminance, the means were 150 and 185 cd/m^2, the separation being 35 cd/m^2. These values were dictated in large part by constraints of the video monitor. It is possible that our estimates of efficiency are highly dependent on the particular mappings. To take an extreme example, suppose we had mapped the parent Gaussian distributions into luminance distributions with means of 150 and 151 cd/m^2 and standard deviations of 1 cd/m^2. We can be sure that limitations on sensory discrimination would reduce efficiency substantially.

We conducted two ancillary experiments to see whether our estimates of efficiency were influenced by the particular stimulus mapping. In the case of scatterplots, we compared efficiencies for conditions in which the means were separated by 2.6° (original mapping), 1.73°, and 0.86°. (Wider separations would have resulted in loss of points off the top and bottom of the screen, and narrower separations would have resulted in unacceptable spatial quantization errors.) T.T. was the subject, and the sample size was 10. Average values of efficiency changed only slightly, from 52.6% (0.86° condition) to 64% (2.6° condition). In the case of luminance displays, we compared efficiencies for three mappings. Again, the subject was T.T. and the sample size was 10. The pairs of mean luminances and corresponding average efficiencies were: 90 and 105 cd/m^2 with efficiency of 12.8%; 90 and 120 cd/m^2 with efficiency of 38%; 90 and 135 cd/m^2 with efficiency of 40.7%. Clearly, T.T. was less efficient when the distributions of luminance values were close (i.e., means of 90 and 105 cd/m^2), but her performance leveled out for the more widely separated mappings. We conclude that estimates of efficiency are likely to be unaffected by the particular mapping, as long as the mean stimulus values are well separated (i.e., highly discriminable) from one another.

Table 1
Linear Regression Slopes

	Scatterplots		Luminance		Numerical	
Observer	Mean	Var	Mean	Var	Mean	Var
Log (Efficiency) versus Log (Sample Size), t = .3 sec						
Y.G.	−.23	.26	−.59	−.37	−.79	−1.00
T.T.	−.36	.03	−.55	−.22	−.90	−1.38
Log (Efficiency) versus Log (Viewing Time), Sample Size = 10						
T.T.	.08	.08	.67	.55	.70	1.13

Figure 5 shows data for discrimination of variance. On the whole, the pattern of results is very similar to those for means. Observers are as efficient at estimating variances as means. This may be surprising because, as every neophyte statistics student laments, numerical calculation of variance is much harder than that of means. Unlike the data for means, the scatterplot curves for variance in Figure 5 are nonmonotonic with a peak at a sample size of 10. We will return to this point below, when we discuss algorithms that subjects may use in estimating variances.

Effect of Viewing Time

Figure 6 shows the effect of viewing time on efficiency with a fixed sample size of 10. Data are shown for Subject T.T. Once again, the results are quite similar for means and variances.

There is only a slight increase in efficiency for scatterplots as exposure time increases from .3 to 1.0 sec. Little increase is to be expected, because efficiency is already high for the shortest exposure. In comparison, efficiencies for numerical displays rise more rapidly, with log-log slopes much nearer one. For viewing times even

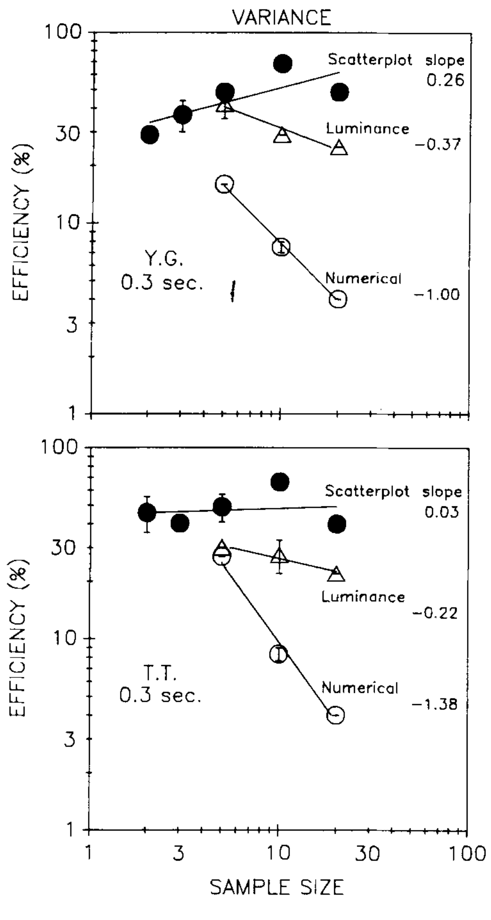

Figure 5. Efficiency for discriminating variances. Other details as in Figure 4.

longer than 1 sec, we might expect efficiencies for numerical displays to keep rising. Eventually, they might catch up with the scatterplots. In the extreme, a subject might consciously compute statistics on the columns of numbers and possibly achieve very high efficiencies (limited only by skill at mental arithmetic).

In visual search tasks, perceptual processing is said to be parallel if performance is independent of exposure time and the number of elements. For our discrimination task, parallel processing would mean constant efficiency. It is not quite the case that scatterplot perception is a purely parallel process. Efficiency rises slowly with viewing time and drops slowly with sample size. Relatively speaking, however, scatterplots are processed in a more parallel manner than are numerical displays. This difference between the two types of displays quantifies and supports Cleveland and McGill's (1985) view that the preattentive (and hence parallel) processing of graphs is the most fundamental reason for their superiority to tabular displays.

In visual search, processing is said to be serial when performance increases linearly with time and declines in inverse proportionality to the number of elements. We can see that these relations apply to the efficiency of a serial processor as follows. Suppose that an observer processes samples serially at a rate of N_0 elements in time T_0. Suppose that this observer makes optimal use of the available information and is limited only by the rate at which samples can be processed. In time T, such an observer can process $(N_0/T_0)T$ samples. Substituting this for N in Equation 3, we obtain the accuracy of the serial processor for discriminating means:

$$D_{\text{serial}} = \sqrt{(N_0/T_0)T}\,.$$

Substituting this expression in Equation 5 for efficiency, we find that

$$E = (D_{\text{serial}}/D_{\text{ideal}})^2$$
$$= (N_0/T_0)\ (T/N) = k(T/N),$$

where k is a rate constant. (E can never exceed 1.0.) Such an observer's efficiency is linearly related to viewing time T and inversely proportional to sample size N. This is roughly the pattern of results we found for numerical displays.

The results for luminance-coded data lie between those for scatterplots and numerical displays, both in values of efficiency and in slopes. In the experiments with means, the task amounted to discrimination between the average luminances of two patches of static noise. We are not aware of any experiments in which this capacity has been studied. The variance experiment amounts to discrimination of the r.m.s. contrasts (contrast power) of two displays of one-dimensional static visual noise. The efficiencies we found—typically 30%—are in good agreement with values measured by Kersten (1987). He measured efficiencies for the detection of static visual noise on uniform fields or on fields of dynamic visual noise.

There is evidence for a compressive power-function relationship between brightness (a perceptual dimension)

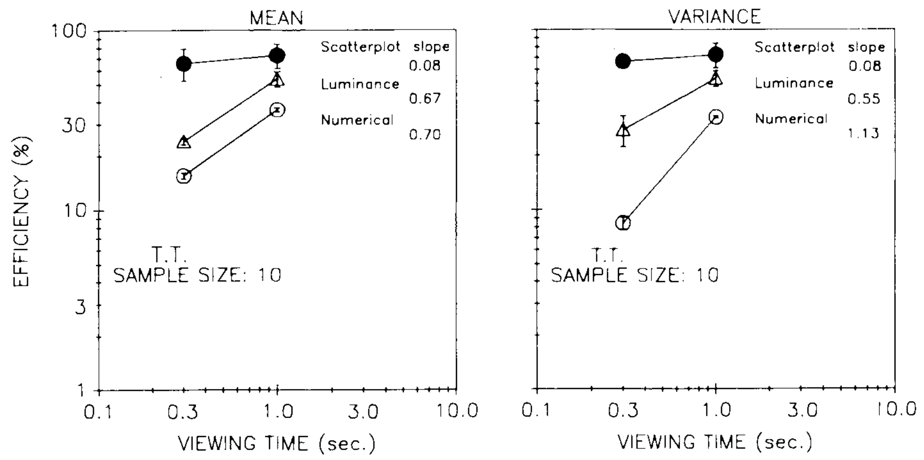

Figure 6. Efficiency is plotted as a function of viewing time for Observer T.T. The two panels show results for discrimination of means and variances.

and luminance (a physical dimension). This transformation might be a source of inefficiency in luminance-coded displays. While the exponent of the brightness power function depends weakly on a number of stimulus variables (Mansfield, 1973; Stevens & Stevens, 1963), a value of 0.3 is representative. Suppose that in our luminance experiments, real observers are ideal, except that they apply a power-function transformation with exponent 0.3 to each luminance sample prior to computation of decision variables. In our experiments, the luminance distributions had means of 150 and 185 cd/m². Brightness values corresponding to these means would differ by only about 6.5% (i.e., they would differ by a factor of $[185/150]^{0.3} = 1.065$). Therefore, all the luminance values used in the experiment would map into a very small range of brightness values. Over this small range of values, the brightness function would be approximately linear, and the luminance-to-brightness transformation should have little effect on efficiency. To check this, we used Monte Carlo computer simulations (see the Method section) to compute efficiency for an observer who is ideal except for the brightness transformation. For sample sizes of $N = 5$ and $N = 10$, the efficiencies were 0.99 and 0.98. Clearly, the reduced efficiency due to the brightness transformation does not account for the efficiencies of 20% to 30% measured with luminance-coded displays.

Sorted and Unsorted Samples

Our data indicate that perceptual analysis is serial for numerical displays but more parallel in character for scatterplots. A serial processor can handle only a few samples in a brief exposure. Its performance might be enhanced, however, if those few samples were well chosen. For example, if only one of N samples can be processed in a brief exposure, the median, maximum, or minimum value would be more useful than a value selected at random. We can make it easier for a real observer to use such values by presenting displays in which the samples are sorted in ascending or descending order. Then, in a brief exposure, an observer will know where to look in the set of samples for the extrema or median values. Since a parallel processor has simultaneous access to all values, sorting should be of less advantage. Sorting has no effect on the performance of the ideal observer that uses all information in the set of samples, regardless of presentation order.

These considerations led us to predict that sorting should benefit perceptual analysis of numerical displays more than scatterplots. We measured efficiencies for sorted and unsorted sets of 10 samples for Observer T.T. The results are shown in Table 2. In all four comparisons with numerical displays, efficiency was higher for sorted samples, but the advantage was significant only for discrimination of variances with 1,000-msec exposures ($p < .01$). Sorting had no systematic effect on efficiency for scatterplots. As predicted, sorting was more beneficial for numerical displays than for scatterplots, but the effect was relatively weak. For luminance-coded displays, sorting did not produce significant changes in efficiency for discriminating means, but sorting actually hampered performance on variance. Perhaps the r.m.s. contrast (corresponding to variance in the luminance-coded displays) was harder to estimate in the sorted displays because local-contrast steps were much smaller.

Table 2
Comparison of Efficiencies for Sorted and Unsorted Samples

	Mean			Variance	
	Sorted	Unsorted	Sorted		Unsorted
Luminance (300 msec)	.23	.19	.25	*	.39
Scatterplot (300 msec)	.54	.60	.66		.60
Numerical (300 msec)	.18	.16	.10		.08*
Numerical (1,000 msec)	.38	.34	.40	*	.21

Note—Average efficiencies are shown for Observer T.T. The sample size was 10. *Significant difference ($p < .01$).

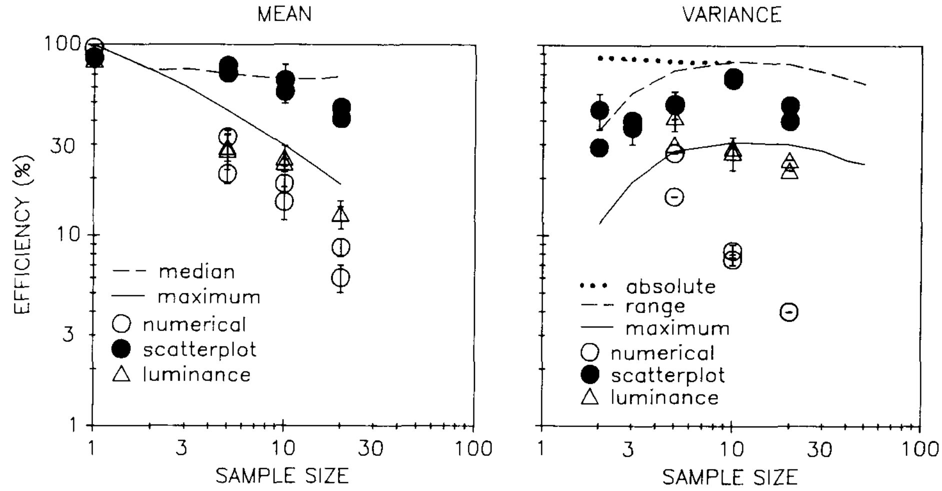

Figure 7. The five curves, two for means and three for variance, show how efficiency depends on sample size for five strategies (Table 3). The curves were derived from results of computer simulations. Data for Observer Y.G. have been replotted from Figures 4 and 5 for comparison.

Perceptual Strategies

How efficient would we expect subjects to be if they based their decisions on median values or extrema? There is a variety of strategies if one uses these values, and the efficiency can be computed for each of them. If we find that an observer's efficiency exceeds that of the strategy, we can be sure that the observer is not using the strategy. On the other hand, an observer whose efficiency is less than that of the strategy may be attempting to use the strategy, but failing to execute it perfectly.

The curves in Figure 7 show efficiency as a function of sample size, using two strategies for discriminating means and three strategies for discriminating variances. The strategies are described in Table 3. The curves are based on computer simulations (see the Method section) run with sample sizes ranging from 1 to 50. Data points have been replotted from Figures 4 and 5 for comparison with the simulation results.

The median strategy for discriminating means is very efficient. In the limit of large sample size, its efficiency drops to $2/\pi$, which is close to 64% (Freund, 1962, p. 219). Less efficient is the maximum strategy, especially as sample size grows large. Neither strategy provides a good fit to any of the data sets, including the sorted-list data.

We considered three strategies for discriminating variances. The absolute-value strategy is very efficient and outperforms real observers even for scatterplots. The range strategy (Table 3), however, shows the same sort of nonmonotonic dependence on sample size as the scatterplot data of real observers. This strategy has a peak efficiency of 85% at a sample size of 11, which is a little better than the real observers. The rough correspondence, however, suggests that real observers may rely heavily on extreme values in their estimates of dispersion of data in scatterplots. As might be anticipated, the curve for the maximum strategy roughly parallels the curve for the range strategy, but with overall lower efficiency.

As statisticians have long known, order statistics such as maxima, minima, or medians are particularly informative. The curves in Figure 7 indicate that means and variances can be efficiently discriminated with simple strategies based on these values. Only in the case of the range strategy for variance discrimination, however, is there evidence that subjects actually adopt one of the strategies we propose.

Thresholds for Discriminating Means

The two panels of Figure 8 show threshold data rather than efficiency. For a given value of variance (the same for the two parent Gaussian distributions), the difference

Table 3
Some Possible Perceptual Strategies

Strategy	Description of Statistic
	Discrimination of Means
Median	Median value
Maximum	Maximum value
	Discrimination of Variances
Absolute Value	Sum of the absolute values of the maximum positive and negative deviations from the mean
Range	Difference between maximum and minimum values
Maximum	Maximum

Note—In a trial, the subject is presented with two sets of N samples, each set drawn from a parent Gaussian distribution. The subject attempts to identify the set of samples that came from the parent distribution with higher mean (or larger variance). The subject who adopts one of these strategies picks the set of samples having the higher value of the indicated statistic.

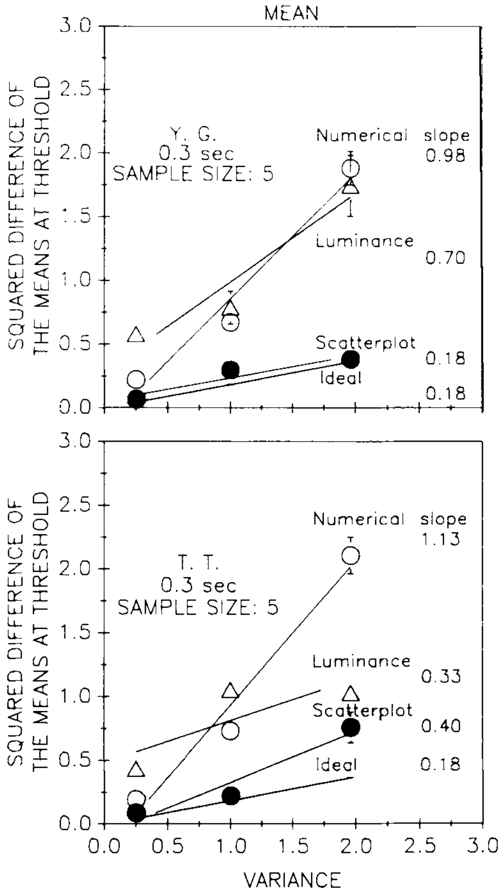

Figure 8. The results of a threshold experiment are shown, in which the difference between the means of the parent Gaussian distributions was reduced until the observers were 75% correct in the forced-choice procedure. The squared difference of the means at threshold (see text) is plotted as a function of the common variance of the Gaussian distributions. The intercepts and slopes of straight lines fit to the data are used to partition losses of efficiency into two sources. Results are shown for the three display formats and for the ideal observer. The two panels give results for Observers Y.G. and T.T.

between the means was adjusted by a forced-choice staircase procedure to find the threshold difference. The squared difference between the means at threshold (see next paragraph) is plotted as a function of variance. Thresholds are shown for a sample size of 5 and a viewing time of 0.3 sec. Each point represents the mean of 2 to 6 separate threshold estimates. Best-fitting straight lines have been fit to the data.

These threshold plots are analogous to plots of signal energy versus noise spectral density (see Burgess et al., 1981; Legge et al., 1987). Slopes and intercepts of lines through the data can be used to distinguish between two generic forms of inefficiency: nonoptimal sampling and internal noise. The slope of a straight line through such data is related to the efficiency with which the observer processes stimulus samples. The greater the slope, the lower the observer's efficiency. Burgess et al. (1981) have defined *sampling efficiency* as the ideal observer's slope divided by the real observer's slope. Values less than 1.0 reveal nonoptimal sampling. The magnitude of the internal noise is proportional to the x-axis intercept's distance to the left of the origin. The farther to the left of the origin the intercept, the greater the internal noise. The ideal observer's straight line passes through the origin (no internal noise) and has a slope of 0.18.[2] These concepts are described in more detail by Legge et al. (1987).

For both subjects, the scatterplot lines have slopes and intercepts close to the ideal. This is not surprising, because efficiencies for scatterplots are very high.

Both subjects had slopes that were much higher than the ideal, but intercepts near zero, for numerical displays. This reveals that the source of inefficiency is almost entirely due to incomplete sampling, confirming the serial-processing interpretation. In the brief exposure interval, the subjects could sample only a small fraction of the available information, but this they processed quite flawlessly.

For both subjects, the luminance displays had intercepts substantially to the left of the origin, suggesting the existence of internal noise. This noise might have a sensory origin. It might be related to the precision with which contrast is coded in the nervous system (Legge et al., 1987).

SUMMARY

Efficiency provides an absolute measure of perceptual performance. We have used efficiency to study graphical perception. Our findings extend those of Cleveland (1985) and Cleveland and McGill (1985) by quantifying the superiority of graphs over numerical tables. Specifically, our results tell us how well real observers can estimate statistical parameters—means and variances—compared with an ideal observer (or a statistical test), who uses all information optimally.

Perceptual efficiencies are very high for scatterplots, often 60% or more. Efficiencies are much lower for numerical tables (< 10% for a moderate number of samples and short exposures). Efficiency for the luminance-coded displays lies intermediate between those for scatterplots and numerical tables.

Performance with scatterplots has the earmarks of a parallel process: weak dependence on sample size and viewing time. This confirms the view that the major contributor to the superiority of graphical displays is spatial parallel processing. Our simulations of possible perceptual strategies, however, indicate that relatively high efficiencies can be achieved without analyzing all the samples. Instead, observers may use a parallel spatial process to quickly identify samples (e.g., minimum and maximum values) that are particularly informative. Their decisions may depend only on these values.

Real observers appear to process tables of numbers in a much more serial fashion. Their efficiencies drop roughly linearly with increasing sample size and increase in rough proportion to time. A plausible interpretation is that entries in tables are processed sequentially at a fixed rate. Given enough time, efficiencies might become quite high even for numerical displays.

REFERENCES

Abramowitz, M., & Stegun, I. A. (1972). *Handbook of mathematical functions*. Washington, DC: U.S. Government Printing Office.

Barlow, H. B. (1977). Retinal and central factors in human vision limited by noise. In H. B. Barlow & P. Fatt (Eds.), *Vertebrate photoreception* (pp. 337-358). New York: Academic Press.

Barlow, H. B. (1978). The efficiency of detecting changes of density in random dot patterns. *Vision Research*, **18**, 637-650.

Barlow, H. B., & Reeves, B. C. (1979). The versatility and absolute efficiency of detecting mirror symmetry in random-dot displays. *Vision Research*, **19**, 783-793.

Burgess, A. E. (1985). Effect of quantization noise on visual signal detection in noisy images. *Journal of the Optical Society of America*, **A2**, 1424-1428.

Burgess, A. E., & Barlow, H. B. (1983). The precision of numerosity discrimination in arrays of random dots. *Vision Research*, **23**, 811-820.

Burgess, A. E., Wagner, R. F., Jennings, R. J., & Barlow, H. B. (1981). Efficiency of human visual signal discrimination. *Science*, **214**, 93-94.

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

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

Egan, J. P. (1975). *Signal detection theory and ROC analysis*. New York: Academic Press.

Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh: Oliver & Boyd.

Freund, J. E. (1962). *Mathematical statistics*. Englewood Cliffs, NJ: Prentice-Hall.

Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics*. Huntington, NY: Robert E. Krieger.

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

Kersten, D. (1987). Statistical efficiency for the detection of visual noise. *Vision Research*, **27**, 1029-1040.

Legge, G. E., Kersten, D., & Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of the Optical Society of America*, **A4**, 391-404.

Mansfield, R. J. W. (1973). Brightness function: Effect of area and duration. *Journal of the Optical Society of America*, **63**, 913-920.

Pelli, D. G. (1981). *The effects of visual noise*. Doctoral dissertation, Physiology Department, Cambridge University.

Sakitt, B. (1973). Indices of discriminability. *Nature*, **241**, 133-134.

Stevens J. C., & Stevens, S. S. (1963). Brightness function: Effects of adaptation. *Journal of the Optical Society of America*, **53**, 375-385.

Tanner, Jr., W. P., & Birdsall, T. G. (1958). Definitions of d′ and η as psychophysical measures. *Journal of the Acoustical Society of America*, **30**, 922-928.

van Meeteren, A., & Barlow, H. B. (1981). The statistical efficiency for detecting sinusoidal modulation of average dot density in random figures. *Vision Research*, **21**, 765-777.

Watson, A. B., & Pelli, D. G. (1983). Quest: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, **33**, 113-120.

NOTES

1. The definitions of D_{ideal} are different for means and variances, because the underlying distributions are different. The distributions are chi-square for variances, but Gaussian for means. While the definition of accuracy for Gaussian distributions with equal variance, d', is well known, there is no commonly accepted definition of accuracy for chi-square. We adopted the definition proposed by Sakitt (1973) for non-Gaussian distributions, where

$$D = [\text{mean}(s) - \text{mean}(n)]/\sqrt{[SD(s) \cdot SD(n)]}.$$

For chi-square distributions, this reduces to Equation 4. Notice that for both means and variances, D_{ideal} is proportional to $\sqrt{N}$. Therefore, our definitions of efficiency (see Equation 5) for both means and variances are consistent with Fisher's (1925) original definition. He defined efficiency as the ratio of the number of samples required by an ideal observer to the number required by a real observer to achieve the same performance level.

2. The slope value of 0.18 can be derived from Equation 2. Squaring both sides and denoting the threshold difference of means by Δm, we have

$$(\Delta m)^2 = (D_{ideal}{}^2/N)SD^2.$$

A threshold criterion of 75% correct corresponds to a value of D_{ideal} of .95. For $N = 5$, the slope of the relation between squared difference of means and variance is $(.95)^2/5 = .18$.

(Manuscript received April 13, 1988;
revision accepted for publication March 30, 1989.)