APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. **15**: 413–431 (2001)
DOI: 10.1002/acp.714

The Discrimination of Graphical Elements

J. G. HOLLANDS[1]* and IAN SPENCE[2]

[1]*Defence and Civil Institute of Environmental Medicine, Canada*
[2]*University of Toronto, Canada*

SUMMARY

A model is proposed to account for how people discriminate quantities shown in pie charts and divided bar graphs (i.e. which proportion is larger, A or B?). The *incremental estimation model* assumes that an observer sequentially samples from the available perceptual features in a graph. The relative effectiveness of sampled perceptual features is represented by the spread of probability distributions, in the manner of signal detection theory. The model's predictions were tested in two experiments. Participants took longer with pies than divided bars and longer with non-aligned than aligned proportions in Experiment 1. In Experiment 2, participants took longer with divided bars than pies when graphs were of unequal size. Generally, graphical formats producing longer response times incurred a greater time penalty when the difference between proportions was reduced. These results were in accordance with the model's predictions. Implications for graphical display design are discussed. Copyright © 2001 John Wiley & Sons, Ltd.

INTRODUCTION

Whose market share is larger – Ford or GM? Who sold more albums in CD format in 1999 – Shania Twain or Celine Dion? We often ask ourselves such questions as we scan graphs shown in newspapers, in magazines, or on the Internet. Discriminating quantities shown in graphs – determining which of two quantities is larger – is a fundamental graph-reading activity (Kosslyn, 1993; Simkin and Hastie, 1987; Spence and Lewandowksy, 1991), and perhaps the most common task people perform with graphs. Discrimination may also be involved in more complex graph reading tasks (see Carswell, 1992; Hollands and Spence, 1992, for task descriptions). It is important, therefore, to understand the psychological processing that underlies how quantities are discriminated in graphs. In this paper, we propose a processing model that predicts when discrimination of graphical elements should be more difficult, test those predictions in two experiments, and discuss the implications of the results for the design of graphical displays.

Discrimination of graphical elements is not as straightforward as might appear at first glance. Examine the eight pairs of graphs (pies and divided bars) shown at the top of

*Correspondence to: Justin G. Hollands, DCIEM, Human–Computer Interaction Group, 1133 Sheppard Ave West, P.O. Box 2000, Toronto, Ontario, Canada M3M 3B9.

Contract grant sponsor: Natural Sciences and Engineering Research Council of Canada.
Contract grant number: A8351.

Copyright © 2001 John Wiley & Sons, Ltd.

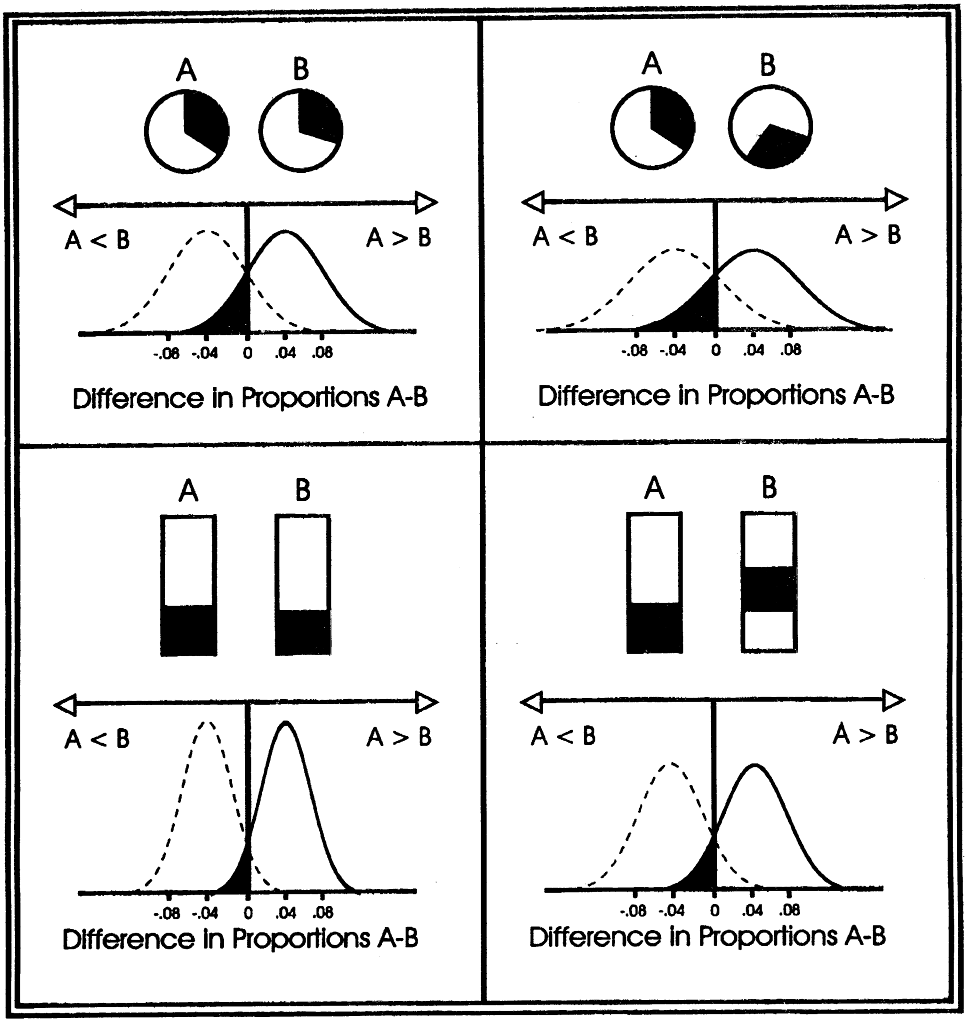

Figure 1. An illustration of the decision model as applied to aligned and non-aligned segments used in Experiment 1. See text for details

each panel in Figures 1 and 2. For each pair, ask yourself: Which shaded region depicts the larger proportion, A or B? You will probably find that some of the pairs are more difficult to discriminate than others. We note five important properties of these situations.

First, the observer has a choice of perceptual cues or *features* to make the decision: angle, area or slope with pie charts; bar top position, area, or height with divided bars. Second, as the above example illustrates, the available perceptual features vary across graph types. Third, perceptual features are likely to vary in terms of their effectiveness: lengths being more effective than angle, for example. Fourth, available perceptual features change when quantities are depicted in different positions (Figure 1) or with different scaling (Figure 2), and the task subjectively appears more difficult. The former

Copyright © 2001 John Wiley & Sons, Ltd.

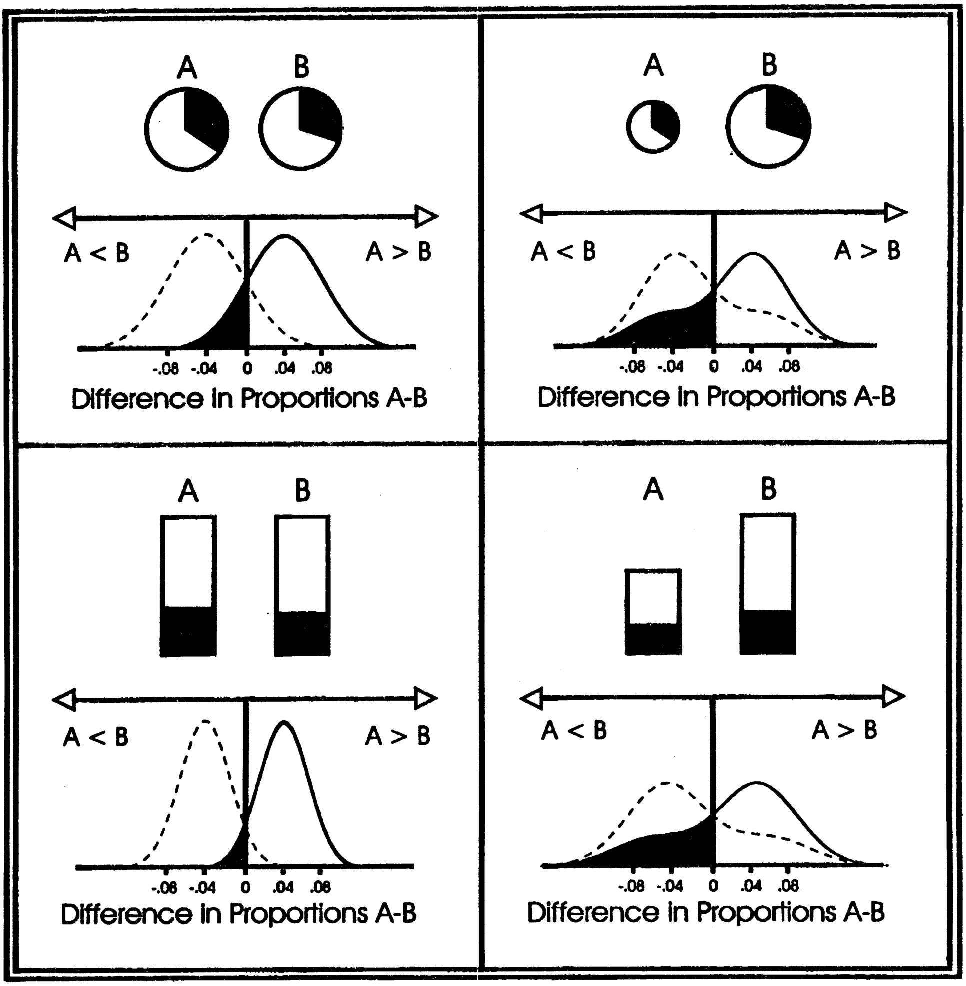

Figure 2. An illustration of the decision model as applied to equal-size and unequal-size pies used in Experiment 2. See text for details

situation occurs when the position of a graphical element corresponding to one variable changes from graph to graph, usually because another variable also changes (e.g. because Ford's market share increased from one year to the next, the position of the elements representing GM's market share is shifted up in a divided bar graph, or clockwise in a pie chart). The latter situation (Figure 2) occurs because the size of a single divided bar or pie can be used to show changes in total amount (e.g. to indicate that, in total, more CDs were sold in 1999 than in 1998). Finally, there are situations where different cues provide contradictory information, and this seems to make the decision particularly difficult. Note, for example, that for the pie charts on the right of Figure 2, angles indicate that A > B, whereas areas indicate that A < B. In the next section we shall describe a model that takes these five properties of the situation into account.

Copyright © 2001 John Wiley & Sons, Ltd.

THE INCREMENTAL ESTIMATION MODEL

There is a large literature examining perceptual discrimination dating back to the pioneering work of Fechner, and there are good theoretical models of how the discrimination process works, most taking a signal detection approach (e.g. Petrusic, 1992; Vickers, 1980). Nonetheless, these models cannot properly account for the discrimination of data shown in graphs because there are several, sometimes redundant, perceptual features that can be judged with most graph types, as noted above. In addition, data shown in graphs are often expressed as proportions or percentages (e.g. market share) and thus a common task for the graph reader involves the discrimination of *proportions* rather than absolute amounts. Simple perceptual discrimination cannot lead to accurate performance when the overall size of one graph is changed relative to the other (as in Figure 2). It is unclear how existing models of perceptual discrimination could deal with this situation.

The *incremental estimation model* to be described here accounts for the discrimination of proportions by proposing that the five properties affect both the internal representation of the difference and how that representation changes over time as perceptual features are repeatedly sampled. The modelled internal representation can then be used to derive specific predictions that can be tested experimentally.

Consider the two pie charts A and B shown at top-left of Figure 1. We have depicted proportion A as being larger than proportion B, without loss of generality. The participant's task is to respond with a key press to the question: 'Which proportion is larger, A or B?' The decision process may be described by a two-alternative forced-choice (2AFC) signal detection model (Green and Swets, 1966; Macmillan and Creelman, 1991, Chapter 5). This approach to modelling the discrimination process has its origins in the work of Thurstone (1927) and since then similar models have often been used to account for performance in discrimination tasks (e.g. Ashby, 1983; Link and Heath, 1975; Petrusic, 1992; Pike, 1973; Smith and Vickers, 1988; Treisman and Watts, 1966; Vickers, 1980). Although the theory of signal detection has seen limited application in the graphical perception literature (Legge *et al.*, 1989; McEwan, 1994), it has not been applied to the discrimination of *proportions* as examined here.

Perceptual features

When comparing A with B using the pies on the upper-left of Figure 1, an observer may choose to attend to a feature such as *angle* and estimate the difference in the two angles. Many sources of error will impose a probability distribution on the observer's estimate of the difference in angle (A–B). We assume this *difference distribution* to be Gaussian. In the top-left panel of Figure 1, the true difference in proportions is 4% and the standard deviation of the distribution is 4%, both values arbitrarily chosen for illustration. The shaded portion to the left of the criterion represents the probability of a decision error. On a given trial, the observer samples from the difference distribution (A–B) and depending on the magnitude of the sample and the location of the criterion the decision will be '$A < B$' or '$A > B$'. The 'correct' distribution is shown using a solid line and an 'incorrect' distribution [corresponding to the case when E $(A–B) = -4\%$] by a dashed line. The criterion is set at $(A–B) = 0$, a neutral criterion, which is typical in the 2AFC paradigm (Macmillan and Creelman, 1991).

Copyright © 2001 John Wiley & Sons, Ltd.

Table 1. A hierarchy of perceptual features for discriminating proportions in graphs (see Cleveland, 1985)

Accuracy	Perceptual feature
Most accurate	Position along a common scale
	Position along identical, non-aligned scales
	Length
	Slope
	Angle
	Area
	Volume
Least accurate	Color hue–saturation–density

With a larger difference between the true proportions, the observer's estimate of difference (A–B) is farther from 0 on average. In other words, the centre of the difference distribution is farther from the criterion, and the probability of a decision error decreases.

In addition to *angle*, the observer could base the judgment of difference on *area* or – if the segments are aligned – the *slopes* of the non-aligned radii. The variances of distributions corresponding to different features would differ if some perceptual features were more effective at conveying quantity than others. Cleveland (1985) has ordered several perceptual features used in graphs (e.g. length, angle, area) in terms of their accuracy. Cleveland ranked angle and slope as being about equally effective. However, recent studies have shown that angle discrimination is less accurate than slope discrimination (e.g. Regan *et al.*, 1996; Snippe and Koenderink, 1994) and we have modified Cleveland's ranking to reflect this, as shown in Table 1.

Cleveland's ranking is based on a task in which participants estimated the proportion one graphical element was of another (e.g. what percentage is A of B?). As such, the task involved estimating relative magnitude. It is an empirical question whether Cleveland's ranking is effective for other graph-reading tasks, such as discrimination. In this paper we test Cleveland's ranking (Table 1) for the discrimination of proportions: Are higher-ranking perceptual features associated with smaller estimation errors?

We propose that an observer discriminating two proportions will sample from the set of available features – on one occasion concentrating on angle, on another area, and so forth. The accuracy with which the observer computes (A–B) is apt to vary if different features are attended to and thus the overall difference distribution will be a mixture of distributions, all of which have the same positive expectation but possibly differing spreads. For simplicity we assume this mixture of Gaussians with equal expectations to be approximately Gaussian.[1]

In the two next sections we detail the situations where graphical elements are aligned and not aligned, respectively, for both pies and divided bars. Following that, we examine the situation where the overall size of one graph is different from the other.

[1] The mixture of Gaussians with different variances will be more platykurtic than a comparable Gaussian. This fattening of the tails does not alter the predictions of our model.

Copyright © 2001 John Wiley & Sons, Ltd.

Aligned segments

Pies

In the top-left panel of Figure 1, all features that indicate the magnitudes of the two proportions are in accord. If proportion $A > B$, then slope $A > B$, angle $A > B$, and area $A > B$. In the absence of errors of estimation, the observer will come to the same correct conclusion and respond '$A > B$', independent of the feature(s) used to make the judgement. In the presence of error, the observer samples from the difference distribution (A–B) and makes the decision '$A < B$' or '$A > B$'. The accuracy of different features will vary from most accurate to least accurate in the following order: slope, angle, area (see Table 1). Although the optimal feature for the observer to attend to is slope, we assume that the observer *samples* from the set of available perceptual features. We test this assumption later.

Divided bars

Consider two divided bars A and B, with shaded areas representing two proportions as shown at bottom-left in Figure 1. As with pies, the observer samples from the difference distribution and decides '$A < B$' or '$A > B$'. If the segments are aligned, and proportion $A > B$, as shown, then (bar top) *position* $A > B$, *length* $A > B$, and *area* $A > B$. An observer can judge position on a common scale with divided bars but cannot with pies. On average, perceptual features for pies rank lower in the Table 1 hierarchy than those for divided bars and therefore the variance of the mixture of distributions of difference should be greater with pies than divided bars (as illustrated in Figure 1). This should result in greater error for aligned pies than aligned divided bars.

Non-aligned segments

Pies

Consider the two pies at upper right in Figure 1. If the two segments are not aligned the observer can no longer discriminate on the basis of slope. Hence the variance of the probability distribution for non-aligned pie charts will be greater than for aligned pie charts, since the former lacks the high-ranking slope feature. Thus decision errors will be more probable with non-aligned than with aligned segments. Also, the error rate for non-aligned proportions will decrease more rapidly as (A–B) is increased, since the probability of a decision error decreases more rapidly as a Gaussian distribution with larger spread is shifted to the right.[2]

Divided bars

Position along a common scale (as occurs with bars whose bases are aligned on a common axis) ranks highest in the Table 1 hierarchy. However, if the two segments are not aligned, as in the bottom-right panel of Figure 1, the position feature is no longer available; the observer must judge length or area. The variance of the mixture difference distribution will be larger and the probability of a decision error correspondingly greater. As with pies, and for the same reason, differences in decision error when judging aligned and non-aligned segments will be smaller with larger differences in the proportions A and B.

[2] This statement is correct when error rates are small. In our experiments participants were instructed to be as accurate as possible; hence, error rates were small. Consequently the error rate should decrease more rapidly as (A–B) is increased when the difference distributions have greater variability.

Copyright © 2001 John Wiley & Sons, Ltd.

Pies versus divided bars

The variability of the difference distribution should be greater for non-aligned pies than for non-aligned divided bars, since the available features (angle, area) rank lower on average than those for non-aligned divided bars (length, area). Thus, the probability of a decision error should be greater with non-aligned divided bars than non-aligned pies. Differences in decision error between non-aligned pies and divided bars should decrease as the difference in proportions increases, given the greater variability in the difference distribution for non-aligned pies.

In the next section we consider the discrimination process when pairs of pies and divided bars are not equal in size.

Overall size

Pies

If the pies are not equal in size all features may not yield the same sign for (A–B). Consider the two pie charts shown in the-top right panel of Figure 2. In the absence of error, comparing slopes or angles will produce the correct response '$A > B$', but comparing areas will produce an incorrect response of '$A < B$'. Note, however, that this will not always be the case. If proportion B were reduced to 11%, for example, A would now be greater than B for areas as well. Thus, even though unequal sizes will increase the likelihood of error when areas are compared, sometimes a comparison of areas will yield the correct response.

In the top right panel of Figure 2, we show the distributions that would result if the observer attends to angle 75% of the time and area 25% of the time, resulting in a mixture of Gaussians. One Gaussian has a positive and the other a negative expectation, producing an asymmetric mixture distribution. The shape of the mixture would change according to the probabilities of attending to different features. For example, the shape of the difference distribution for an observer who attended to slope 60%, angle 30%, and area 10% of the time would be somewhat different. However, regardless of such variation, the important point is that if the observer attends to a contradictory feature some of the time the difference distribution will no longer be symmetrical because some of its probability mass will be shifted to the 'wrong' side of the distribution.

Thus the difference in the probability of error between equal and unequal-size cases can be considerable. If the difference (A–B) is increased, the expectation of the difference distribution moves away from the criterion, just as in the equal-size case. Unlike in the equal-size case, the variance also changes since there is a component of the mixture on the 'wrong' side of the criterion. Although we have shown this component with location at about −4%, the positioning of such a component will depend on the perceived difference in proportion *associated with the attended-to feature.* As the true difference in proportions increases, the component on the 'wrong' side will not necessarily move in mirror fashion to the component(s) on the 'correct' side. The consequential effect is that the overall variance of the difference distribution is increased when contradictory features are attended to.

However, when there is a sufficiently large difference between the proportions A and B in the unequal-size case, the problem of contradictory features goes away. Consider the top-right panel of Figure 2 (33% in A and 29% in B) and assume that the observer uses either angle or area to make the decision. Angle yields the correct decision whereas area misleads. Imagine again that the proportion in B is reduced to 11% thus increasing

Copyright © 2001 John Wiley & Sons, Ltd.

the difference in proportions to 22% from 4%. Attending to *either* angle or area will yield the correct decision since the shaded area in B will now be smaller than in A. Hence, the probability of decision error is small with unequal-size displays showing large differences, as it is with equal-size displays, because both components of the mixture distribution have the same (large) positive expectation.

Divided bars

With pies the observer has the possibility of avoiding the contradictory features – always attend to angle or slope. With divided bars there is no way out. If the bars are not equal in size, as at bottom right in Figure 2, it makes no difference whether the observer attends to position, length, or area. When one of these features produces a contradictory result, so do the others.

Hence, it is not possible to discriminate the proportions shown in unequal-size divided bars accurately if the decision is based solely on perceptual features. Cognitive operations are needed to perform the task accurately. We propose that the observer performs a *ratio estimation* operation (Hollands and Spence, 1992, 1998) that takes as input the length (or area) of the segment and the length (or area) of the whole bar. The *mental ratio* is the observer's estimate of the ratio of the part to the whole (e.g. 33%).

Mental ratios for the two divided bars, A and B, are incorporated into the model by treating them like stimulus features. In the absence of error, a comparison of mental ratios (A–B) yields the correct response, '$A > B$' while attending to perceptual features such as length or area produces a negative (A–B) and a decision error. In the presence of error an asymmetric difference distribution results (similar to that obtained with unequal-size pies) for observers who use perceptual features some of the time, either voluntarily or involuntarily. As with pies, the likelihood of decision error will be larger and the effect of varying the difference in proportions should be more pronounced with unequal sizes.

Mental ratios are likely to be less accurate than perceptual features. Each mental processing step (estimating magnitudes of part and whole, and estimating their ratio) introduces a potential source of error with the resulting ratio estimate likely to be inaccurate. In judgements of proportion with similar bar graph stimuli, for example, Hollands and Spence (1998) obtained average errors of $\pm$3–5%. Therefore we shall assume that the difference distribution for mental ratios has higher variability than for perceptual features. Observers will be more likely to err with unequal-size divided bars than with unequal-size pies, where relatively accurate perceptual features (such as angle or slope) are always available. In addition, the probability of a decision error should decrease with difference in proportions more quickly with unequal-size divided bars than pies, because of the greater variability in the difference distributions for unequal-size divided bars.

Incremental estimation and response time

In many real-world contexts (e.g. a banker studying a graph in a financial report; a physician examining a graph of epidemiological data in a medical journal), the reader is not under strict time pressure and there will generally be sufficient time to discriminate the proportions shown accurately. Although the basic elements of the decision process may be modelled as described above, we do not think that the observer makes only one estimate of the difference using only one perceptual feature. Instead, we assume that the observer

Copyright © 2001 John Wiley & Sons, Ltd.

accumulates incremental estimates – indexed by ν – and computes a running average as ν increases. This kind of accumulation process has been used in the past to model RTs in discrimination tasks (see Petrusic, 1992 for a review). The most important effect of the repeated sampling is to reduce the variance of the averaged difference, as Figure 3 illustrates using pie charts as stimuli. If the observer averages the incremental estimates, the variance of the final average will be smaller than that of a single estimate. Indeed, the variance will fall as ν increases – we have shown the standard deviation dropping as the square root of ν, as it would with independent estimates and the computation of a running mean.

The repeated sampling will also symmetrize a skewed distribution, as shown in the right panels of Figure 3. The sampling distribution tends towards symmetry fairly quickly, in the same fashion that the sampling distribution of a mean tends towards the Gaussian by the Central Limit Theorem. This has the effect of pulling in the tail of the distribution and reducing the error rate. Our illustration shows only a small degree of skewness when $\nu = 2$ and approximate Gaussian form when $\nu = 4$. The repeated sampling has an especially important consequence when there are components on the 'wrong' side of the difference distribution – asymmetries should quickly be reduced, although the variance of the distribution will still be larger than otherwise.

When the graphs are of unequal size the model predicts a higher error rate than for equal-size graphs, for any value of ν. However, if the observer tries to match some internal error criterion, the number of incremental estimates in the unequal-size case must be greater and hence the RTs should be longer. In other words, the number of incremental estimates ν should be greater in the unequal-size case than in the equal-size case to make the error rates roughly equivalent. For example, in the illustration in Figure 3, four incremental estimates are needed in the unequal-size case to achieve the same level of accuracy obtained with two incremental estimates in the equal-size case.

Similarly, if the difference between the proportions is made smaller, and the observer is trying to match the same internal error criterion level, RTs should increase. Also, whenever the difference distribution exhibits greater variability, for whatever reason, we expect the observer who is attempting to minimize error to take longer. That is, the same predictions we have made for error should be seen in RTs in an experiment where accuracy is emphasized.

Evaluating the incremental estimation model

Two experiments were conducted to evaluate specific predictions of the model with divided bar graphs and pie charts. Experiment 1 was designed to investigate the effects of alignment and difference in proportion on RTs. In Experiment 2 we manipulated overall size and difference in proportion in a similar manner. In both our experiments, accuracy was emphasized to the participants. To keep accuracy levels high over multiple trials, we gave participants feedback after each judgement.

No previous experimental work has examined the effects of alignment and size on an observer's ability to discriminate proportions shown in graphs. In addition, the incremental estimation model makes several predictions that other models in the graphical perception literature would not make without modification. Some predictions made in Experiment 1 are consistent with Cleveland's (1985) ranking of perceptual features but since the ranking has not previously been applied to discrimination judgements it remains untested for that task. Moreover, some predictions made by the incremental

Copyright © 2001 John Wiley & Sons, Ltd.

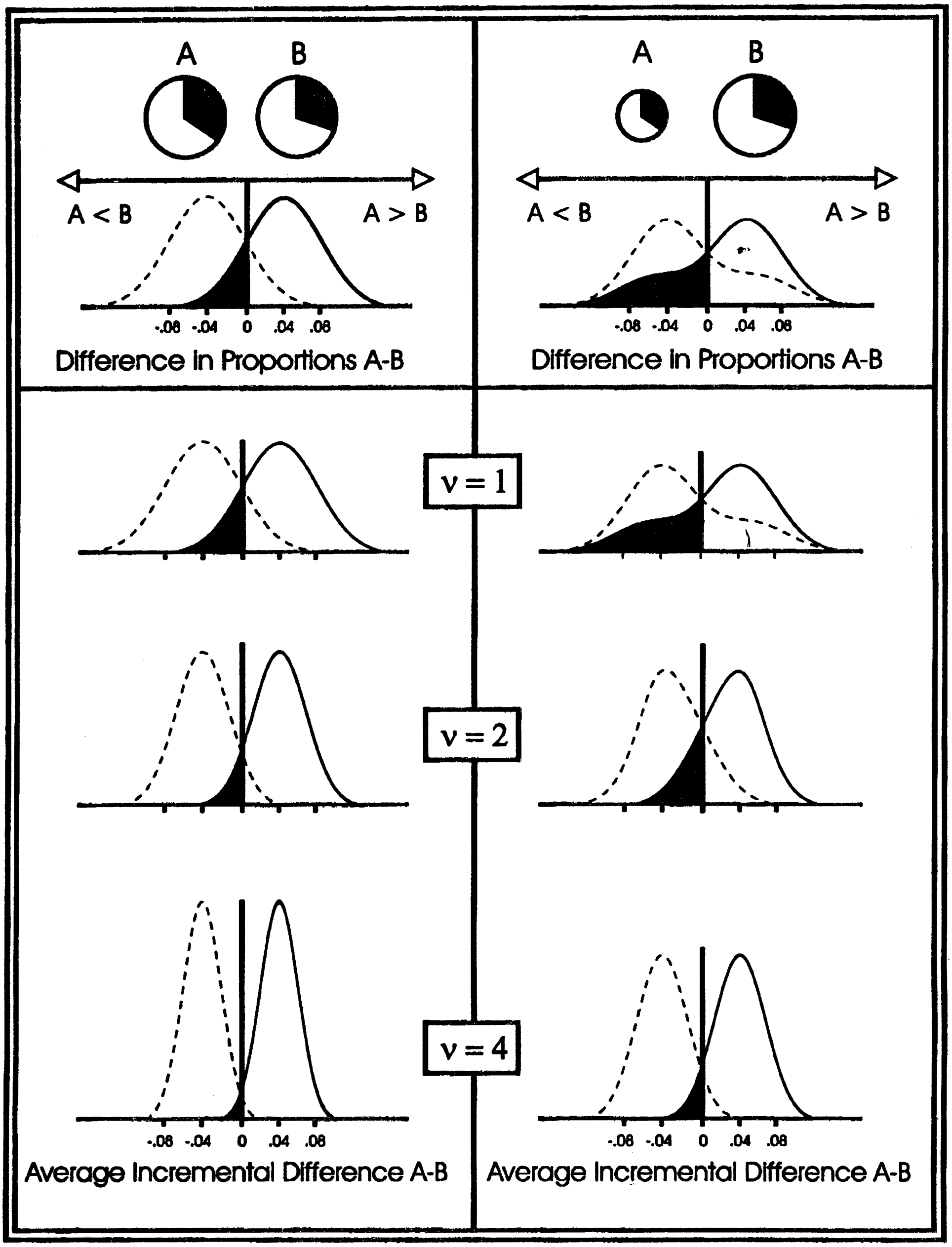

Figure 3. An illustration of the effect of repeated sampling in the incremental estimation model with equal sizes (Experiment 1) shown on the left and unequal sizes (Experiment 2) shown on the right. Note that error decreases with repeated sampling and that the distributions become more symmetric with repeated sampling in the unequal-size case

estimation model in Experiment 2 are inconsistent with a simple application of Cleveland's ranking.

EXPERIMENT 1

The participant's task in Experiment 1 was to discriminate between two proportions shown in graphs. Graph type, alignment, and difference in proportion were manipulated. RT and error measures were obtained. Participants were instructed to be as accurate as possible, and received feedback after each response. The incremental estimation model makes the following predictions:

(1) RTs will decrease monotonically as the difference in proportions increases.
(2) RTs will be shorter with aligned than non-aligned segments.
(3) RTs will be shorter with divided bars than pies.
(4) The monotonic decrease in RT with increased difference in proportions will be greater for non-aligned than aligned segments.
(5) The monotonic decrease in RT with increased difference in proportions will be greater for pies than divided bars.
(6) Similar patterns are predicted for error.

Predictions (2) and (3) follow from Cleveland's ranking, although this ordering has not been previously applied to discrimination tasks. In addition, by itself, Cleveland's ranking does not specify the mechanism by which errors and response times should increase or decrease.

Method

Participants
Thirty-two students at the University of Toronto participated in the experiment. They received course credit for their participation.

Materials and apparatus
Twenty integer proportions were randomly generated in the range from 0.05 to 0.50. No proportion was within 2% of any other in the set.

Each of the twenty proportions was paired with another proportion determined by adding 0.01, 0.02, 0.04, or 0.08 to the first. On a given trial, a pair of proportions was displayed using two divided bar graphs or two pie charts. Figure 1 shows examples of each graph type for aligned and non-aligned conditions. The left-right position of the larger proportion was randomized over trials. On half the trials (the non-aligned condition) one of the segments (selected at random) was displaced upwards (or rotated clockwise) corresponding to a proportion of 0.30. Both segments were displaced from the base (divided bars) or 12 o'clock position (pies) by a random value.

The height of each divided bar was 80 mm, and its width was 15 mm. The radius of each pie chart was 28 mm. For both pies and divided bars, the two members of the pair were separated by a 100 mm horizontal distance (centre to centre). The viewing distance was approximately 50 cm. All graphs were white on a black background. The two segments representing the proportions were coloured red. The graphs were shown

Copyright © 2001 John Wiley & Sons, Ltd.

on a 30 cm display (measured diagonally) at 640×480 (horizontal by vertical) resolution, controlled by a program running on an IBM personal computer, which also measured RTs.

Design and procedure

A $2 \times 2 \times 4$ (Graph Type by Alignment by Difference in Proportions) within-subjects design was used: The two graph types were divided bar graphs and pie charts, the segments were drawn aligned or non-aligned, and the differences in proportion were 0.01, 0.02, 0.04, and 0.08. The order of trials was randomized within each of the 16 conditions, and the order of conditions was counterbalanced across participants using Latin squares. There were 20 trials per condition for a total of 320 trials per participant.

Participants were told that they would be shown pairs of graphs, and that their task would be to indicate which proportion was larger. Participants pressed the left key (the '1' key on the numeric keypad) when the left proportion was larger, and the right key (the '2' key) when the right proportion was larger.

Participants were provided with feedback after each trial; the computer produced a high tone for a correct response, and a low tone for an incorrect response. They were instructed to be as accurate as possible, and to take as much time as they needed. Participants were debriefed at the end of the experimental session. The experiment required approximately 40 minutes to complete. To minimize learning effects during the experiment, participants completed 64 practice trials beforehand.

Results

Data from three new participants replaced the data for three participants who produced more than 40% errors in one or more conditions. For each participant, the median RT for correct trials in each condition was computed, and the means of these medians (averaged across participants) are shown in Figure 4.[3] A $2 \times 2 \times 4$ (Graph Type by Alignment by Difference in Proportions) within-subjects analysis of variance (ANOVA) was performed. We also performed a trend analysis to determine if the changes in RT with difference in proportion were monotonic. RTs decreased linearly as the difference in proportion increased, $F(1,31) = 73.13$, $MS_e = 2.16$, $p < 0.0001$.[4] There was no evidence for a quadratic or cubic trend (both $ps > 0.05$). RTs were shorter when segments were aligned, $F(1,31) = 28.11$, $MS_e = 0.57$, $p < 0.0001$. RTs were shorter with divided bars than pies, $F(1,31) = 25.60$, $MS_e = 1.35$, $p < 0.0001$. Tests for differences in linear, quadratic, and cubic trends as determined by graph type and alignment were carried out (see Keppel, 1982, pp. 232–234). The linear decrease in RT with difference in proportion was greater for non-aligned than aligned segments, $F(1,31) = 6.89$, $MS_e = 0.33$, $p < 0.05$. The linear decrease in RT with difference in proportion was also greater for pies than divided bars, $F(1,31) = 5.37$, $MS_e = 0.83$, $p < 0.05$. There was no evidence for interactions having quadratic or cubic components ($ps > 0.50$).

[3]If RTs for error trials were included, the same general pattern was obtained. This was true for both experiments.
[4]To correct for violations of sphericity and homogeneity of variance, a Geisser–Greenhouse adjustment was used for both experiments. After adjustment, all reported effects from analyses of variance are associated with p-values less than 0.05. However, unadjusted degrees of freedom and p-values are reported in the text to simplify exposition.

Copyright © 2001 John Wiley & Sons, Ltd.

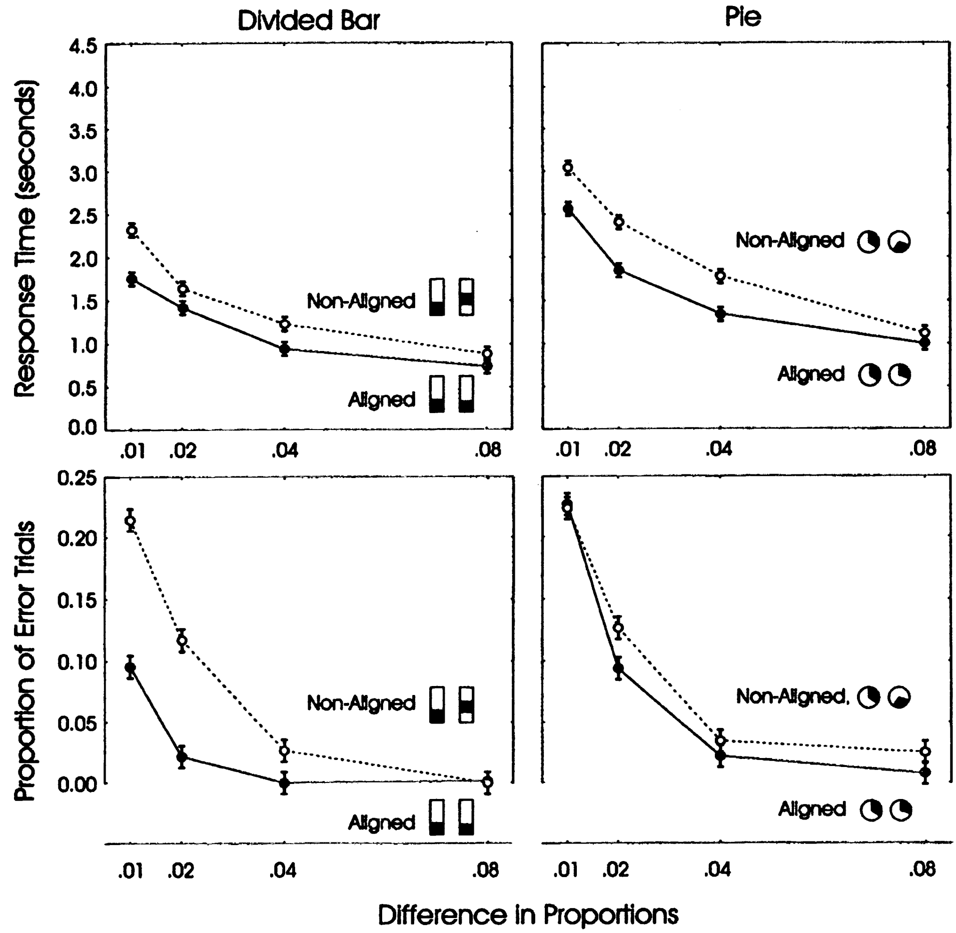

Figure 4. Experiment 1 results. Top: mean RTs for correct trials, as a function of graph type, alignment, and difference in proportions. Bottom: proportion of trials in which participants made an error. Error bars indicate the standard error of the mean

Errors occurred on 8% of the trials on average and they tended to occur in those conditions in which participants produced the longest RTs. The lower part of Figure 4 shows the mean error data.

Discussion

The predictions of the model were supported. Participants took less time with larger differences in proportion (because the expected value of the difference distribution was located farther from the criterion), and the decrease in RT was monotonic. Participants took longer with pie charts than divided bars (because the perceptual features available with divided bars ranked higher in the Table 1 hierarchy than those available with pies, leading to smaller variability in the difference distribution). Participants took longer with non-aligned segments (because a higher-ranking perceptual feature was lost when segments were not aligned, leading to greater variability in the difference distribution).

Copyright © 2001 John Wiley & Sons, Ltd.

RTs for non-aligned proportions decreased more rapidly as the difference in proportion was increased (because of lower variability in the difference distribution for aligned segments). Finally, RTs decreased more rapidly for pies than divided bars as the difference in proportions was increased (because of lower variability in the difference distribution for divided bars).

The error data were in approximate correspondence with the RTs, consistent with the model's prediction that RTs and errors should be correlated.

EXPERIMENT 2

In Experiment 2, graph type, size, and difference in proportion were manipulated and RT and error measures were obtained. Participants were instructed to be as accurate as possible and received feedback after each response. The incremental estimation model makes the following predictions:

(1) RTs will decrease monotonically as the difference in proportions increases.
(2) RTs will be shorter with equal than unequal sizes.
(3) With unequal sizes, divided bars will produce longer RTs than pies, but with equal sizes, the reverse will occur.
(4) The monotonic decrease in RT with increased difference in proportions will be greater for the unequal than equal-size case.
(5) For the unequal-size case, the monotonic decrease in RT with increased difference in proportions will be greater for divided bars than pies.
(6) For the equal-size case, the monotonic decrease in RT with increased difference in proportions will be greater for pies than divided bars.
(7) Similar results are predicted for error.

A simple application of Cleveland's ranking predicts that divided bars should generally produce better performance than pies, since position on a common scale can be judged with divided bars, but observers can only use slope or angle with pies. Thus, Cleveland's ranking is in accord with the second part of prediction (3) (that divided bars should produce better performance than pies for the equal-size case), but contrary to the incremental estimation model, would predict the same result for the unequal-size case (the first part of prediction (3)).

Method

Thirty-two University of Toronto students served as participants. They received course credit for their participation. Experiment 2 had a 2 (graph type) $\times$ 2 (size: equal-size or unequal-size) $\times$ 4 (difference in proportions) within-subjects design. The stimuli and procedure were the same as in Experiment 1 with the following differences. On half the trials (the unequal-size condition) the height of one divided bar (or radius of one pie) was 60% of the other. Hence, a small divided bar was 48 mm in height; a small pie chart had a 17 mm radius. For the other half, the sizes were equal. Each of the twenty proportions in a condition was paired with another proportion determined by adding 0.04, 0.08, 0.16, or 0.32 to the first. The difference in proportions was increased to keep error trials few and RTs in the same range as they were in Experiment 1. Participants were instructed to judge the relative proportions, rather than absolute sizes, in the unequal-size condition.

Copyright © 2001 John Wiley & Sons, Ltd.

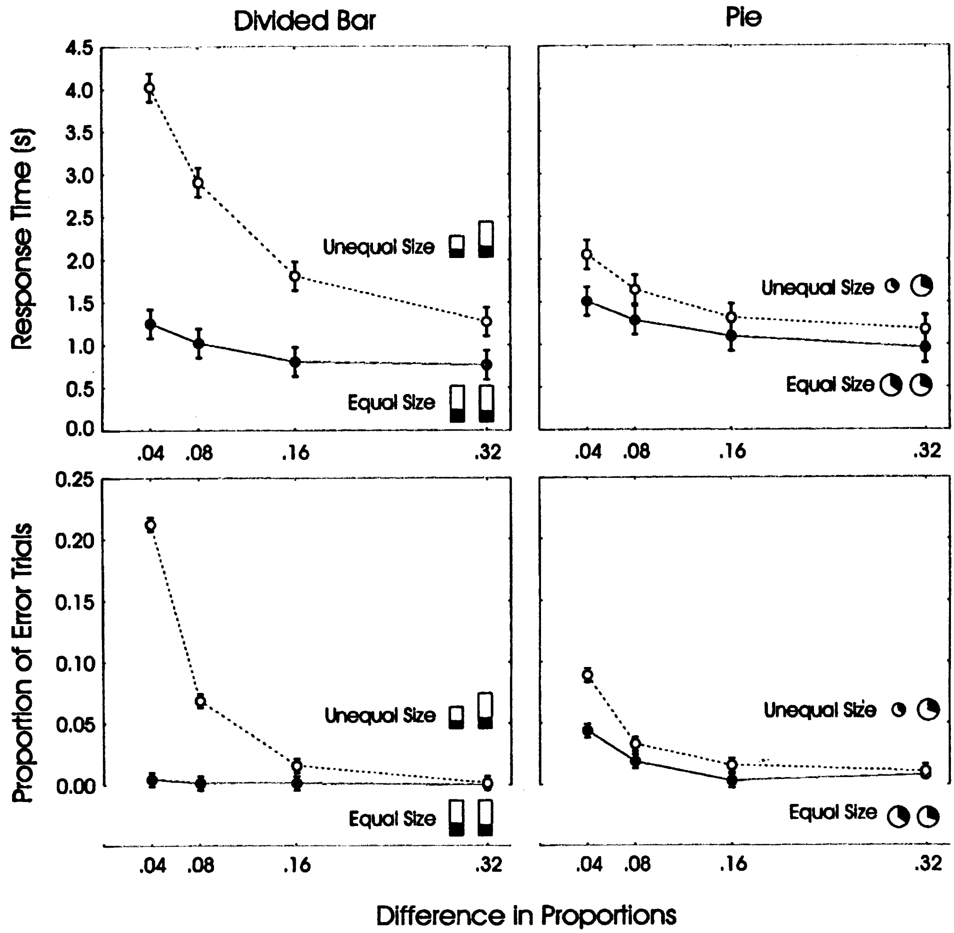

Figure 5. Experiment 2 results. Top: mean RTs for correct trials, as a function of graph type, size, and difference in proportions. Bottom: proportion of trials in which participants made an error. Error bars indicate the standard error of the mean

Results

Data from two new participants replaced the data for two participants who produced more than 40% errors in one or more conditions. For each participant, the median RT for correct trials in each condition was computed, and the means of these medians (averaged across participants) are shown in Figure 5. A $2 \times 2 \times 4$ (Graph Type by Size by Difference in Proportions) within-subjects analysis of variance (ANOVA) was performed on the medians. A trend analysis was also performed, as in Experiment 1. RTs decreased monotonically (with significant linear and quadratic components) as the difference in proportion increased ($F(1,31) = 40.3$, $MS_e = 2.50$, $p < 0.0001$ for linear; $F(1,31) = 4.91$, $MS_e = 0.52$, $p < 0.05$ for quadratic). There was no evidence for a cubic component ($p > 0.15$). RTs were shorter with equal sizes than unequal sizes, $F(1,31) = 29.36$, $MS_e = 3.87$, $p < 0.0001$. Divided bars produced longer RTs than pies in the unequal-size case, but in the equal-size case, the reverse occurred, $F(1,31) = 13.7$, $MS_e = 3.40$,

Copyright © 2001 John Wiley & Sons, Ltd.

$p < 0.001$. The linear decrease in RT with increased difference in proportions was greater for the unequal than the equal-size case, $F(1,31) = 14.60$, $MS_e = 2.11$, $p < 0.001$. (Quadratic and cubic components were not significant, $p > 0.25$.) Thus, predictions (1) to (4) were supported.

As the difference in proportions was increased, the rates of linear decrease for equal and unequal sizes were different for the graph types, with the rates converging more quickly for divided bars than pies, $F(1,31) = 8.06$, $MS_e = 2.12$, $p < 0.01$. An interaction comparison (Keppel, 1982, pp. 232–234) showed that the linear decrease in RT with an increased difference in proportion occurred at a faster rate with unequal-size divided bars than unequal-size pies, $F(1,31) = 7.87$, $MS_e = 4.15$, $p < 0.01$, supporting prediction (5). The same analysis applied to equal-size graphs showed that the linear decrease in RT with increased difference in proportion did not differ between divided bars and pies, $F(1,31) = 0.36$, $MS_e = 0.06$, $p > 0.60$. Hence, the effect of graph type and difference in proportions appeared to be additive with same-size graphs, inconsistent with prediction (6). There was no evidence for quadratic or cubic components to the interaction between size and difference in proportions for either graph type (all ps > 0.15).

Errors occurred on 3% of the trials, on average. Conditions in which participants produced the longest RTs produced the most errors. Inspection of Figure 5, which shows the mean error data for each condition, shows a similar pattern for the error data in accordance with prediction (7). In addition, it is apparent that the error for equal-size divided bars stayed relatively constant as the difference in proportions increased, but error decreased gradually for equal-size pies.

Discussion

The model predicted that RTs should decrease monotonically when the difference in proportions was larger (because the difference distribution should be farther from the criterion). The model predicted longer RTs for unequal sizes than for equal sizes (because the asymmetric distributions require more incremental estimates to minimize decision error). With unequal-size divided bars, where direct use of perceptual features is unreliable, observers must compute mental ratios, increasing RTs relative to unequal-size pies where perceptual features of angle or slope yield the correct response. The results were consistent with these predictions.

The model predicted that equal-size divided bars would produce shorter RTs than equal-size pies. This is because the perceptual features that can be used with divided bars (position along a common scale, length, and area) rank higher on average than those that can be used with pies (slope, angle, and area). Thus variances in the difference distributions in our model should be smaller for divided bars than for pies. The results were consistent with this prediction.

The model predicted that the monotonic decrease in RT with increased difference in proportions should be greater for the unequal than equal-size case. The probability of an error is larger with an asymmetric distribution and should decrease more rapidly as difference in proportion is increased. This should lead to non-additivity in size and difference in proportion for the RTs. This prediction was supported.

It was predicted that the effects of graph type and difference in proportions should be non-additive in the unequal-size case, since mental ratios should produce greater variability in the difference distributions than slope or angle features, and the probability of an error decreases more quickly with greater variability in the difference distributions.

Copyright © 2001 John Wiley & Sons, Ltd.

The results supported this prediction. A similar prediction for the equal-size case was not supported by the RT data although the error data were consistent with the prediction.

The results support our assumption that observers sample from the set of available perceptual features rather than choosing the most effective perceptual feature and using it consistently. If an observer only paid attention to slope we should expect to see no difference in performance between unequal- and equal-size pies. The data do not support this. Instead performance was worse with unequal-size pies than with equal-size pies suggesting that observers sample from the set of available features (slope, angle, area). With unequal-size pies this strategy will lead to an asymmetric difference distribution with a correspondingly larger variance, increasing decision error and thereby RT. This result supports a core assumption of the incremental estimation model.

GENERAL DISCUSSION

A common graph reading task involves the discrimination of quantities shown in graphs. Real-world data are often depicted in proportion or percentage form – most typically in pie charts and divided bar graphs. The psychological processes that people use to discriminate proportions, and how the efficiency of those processes is affected by factors such as alignment and size, have not been previously examined. In this paper, we have proposed an incremental estimation model to account for how people discriminate proportions shown in graphs and successfully tested the model's predictions in two experiments.

According to the model, the observer samples from the difference distribution (A–B) and produces a response of 'A < B' or 'A > B'. The model proposes that the position of the distribution is determined by the difference between proportions, and that its spread is determined by the relative effectiveness of the perceptual features available. Finally, the model proposes that the shape of the difference distribution (symmetric or asymmetric) will depend on whether the graphs are the same size or not. With unequal-size graphs, when more of the distribution is on the 'wrong' side of the criterion, errors are more probable. An observer may minimize error by increasing the number of incremental samples, which should increase RTs in situations where accuracy is stressed.

The model's predictions were tested in two experiments. In Experiment 1, participants were shown pairs of equal-size pies and divided bars. The segments were aligned or non-aligned and the difference in proportions shown was varied. RTs and errors were recorded. Accuracy was stressed and participants received feedback after each trial. The model made six specific predictions, all of which were supported.

In Experiment 2, equal- and unequal-size graphs were used. Graph type and difference in proportion was varied. Six of the seven predictions made were supported by the RT data. Although there was little support for the prediction that the RT difference between same-size pies and divided bars would be greater with smaller differences in proportion, the pattern of error data was consistent with the hypothesis.

Conclusions

The results from Experiment 2 suggest that a simple application of Cleveland's (1985) hierarchy of perceptual features is not always appropriate to explain performance. For example, Cleveland would rank divided bars above pies because an observer can judge

Copyright © 2001 John Wiley & Sons, Ltd.

position along a common scale with divided bars but must judge slope or angle with pies. Our results show that there is not always an advantage to the divided bar. The incremental estimation model provides a better account because it presumes that several different perceptual features may be used – an observer will sample from the available set. Our model accounts for differences in accuracy by assuming that the spread of the difference distribution in proportions varies with the perceptual features sampled and also by assuming that observers make multiple incremental estimates before coming to a judgement. The model assumes that observers sometimes attend to features that yield incorrect judgements, thus reducing accuracy and increasing time to respond. This assumption was validated in Experiment 2. In conditions where perceptual features cannot provide the necessary information, the model proposes that mental ratios (derived quantities) are computed. The use of mental ratios incurs a considerable cost in terms of processing time. Features that make equal-size divided bars effective, such as length or position on a common scale, are not useful in the unequal-size case. Participants performed more poorly with unequal-size divided bars than they did with pies in Experiment 2.

The discrimination of proportions in graphs cannot be considered a simple process in which an observer uses a single stimulus feature. One might expect that an observer would choose the most effective perceptual feature in a graph and use that feature consistently. However, our results suggest that observers do not do this. Although perceptual features such as area, length, or position cannot be the basis for accurate discrimination with unequal sizes, they appear to be used nonetheless. It is difficult for an observer to ignore stimulus features, such as area, that afford efficient discrimination in some cases (e.g. equal-size pie charts) but impair processing in other situations (e.g. unequal-size divided bars).

Practical implications

The discrimination of quantities shown in graphs is one of the most common tasks graph readers perform (Gillan and Lewis, 1994). Therefore, it is important that graphs be designed to make it easy for people to do this.

Usually, when a series of divided bars or pies shows proportions over time, corresponding segments are not aligned because other segment sizes have changed. Our results show that when segments are not aligned, the discrimination of proportions is slowed, especially with small differences in proportion. Divided bars are faster than pies, regardless of alignment, and this advantage is largest with small differences in proportion.

Proportions are often shown with different-size graphs to show variation in total amount and this may lead to difficulties for the reader. For example, Shania Twain may sell fewer total CDs than Celine Dion but Twain may have a larger proportion of sales in her respective market (country) than does Dion in hers (adult contemporary). When such data are graphed, our study shows that it is more time-consuming to discriminate two proportions with respect to different wholes (total country sales vs. total adult contemporary sales), especially if the two proportions are of similar magnitude. The disadvantage is less pronounced with pies than divided bars. For same-size graphs, however, the advantage appears to belong to the divided bar format.

Thus if a series of graphs depicting proportions or percentages consists of equal-size wholes, divided bars are preferred. If, however, the wholes are of unequal size, a series of pies is the better choice.

Copyright © 2001 John Wiley & Sons, Ltd.

ACKNOWLEDGEMENTS

This research was supported by grant A8351 from the Natural Sciences and Engineering Research Council of Canada to the second author. Preliminary results were presented by the first author at the 36th Annual Meeting of the Human Factors Society in Atlanta, Georgia, October 1992.

We thank Joe Baranski, Brian Dyre, Lisa Fournier, Michael Kahana, Jocelyn Keillor, Allan Stanley, and Megan Thompson for helpful comments, and Curt Braun and Faith Lai for assistance with the data analysis.

REFERENCES

Ashby FG. 1983. A biased random walk model of two choice reaction times. *Journal of Mathematical Psychology* **27**: 277–297.

Carswell CM. 1992. Choosing specifiers: An evaluation of the basic tasks model of graphical perception. *Human Factors* **34**: 535–554.

Cleveland WS. 1985. *The Elements of Graphing Data.* Wadsworth: Pacific Grove, CA.

Gillan DJ, Lewis R. 1994. A componential model of human interaction with graphs: I. Linear regression modeling. *Human Factors* **36**: 419–440.

Green DM, Swets JA. 1966. *Signal Detection Theory and Psychophysics.* Wiley: New York.

Hollands JG, Spence I. 1992. Judgments of change and proportion in graphical perception. *Human Factors* **34**: 313–334.

Hollands JG, Spence I. 1998. Mental operations in graphical perception: The summation model. *Applied Cognitive Psychology* **12**: 173–190.

Keppel G. 1982. *Design and Analysis: A Researcher's Handbook* (2nd edn). Prentice Hall: Englewood Cliffs, NJ.

Kosslyn SM. 1993. *Elements of Graph Design.* Freeman: San Francisco.

Legge GE, Gu Y, Luebker A. 1989. Efficiency of graphical perception. *Perception & Psychophysics* **46**: 365–374.

Link SW, Heath RA. 1975. A sequential theory of psychological discrimination. *Psychometrika* **40**: 77–105.

Macmillan NA, Creelman CD. 1991. *Detection Theory: A User's Guide.* Cambridge University Press: Cambridge.

McEwan JSA. 1994. *Optimising Line Graph Aspect Ratio.* PhD thesis, University of Waikato, New Zealand.

Petrusic WM. 1992. Semantic congruity effects and theories of the comparison process. *Journal of Experimental Psychology: Human Perception and Performance* **18**: 962–986.

Pike AR. 1973. Response latency models for signal detection. *Psychological Review* **80**: 53–68.

Regan D, Gray R, Hamstra SJ. 1996. Evidence for a neural mechanism that encodes angles. *Vision Research* **36**: 323–330.

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

Smith PL, Vickers D. 1988. The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology* **32**: 135–168.

Snippe HP, Koenderink JJ. 1994. Discrimination of geometric angle in the fronto-parallel plane. *Spatial Vision* **8**: 309–328.

Spence I, Lewandowsky S. 1991. Displaying proportions and percentages. *Applied Cognitive Psychology* **5**: 61–77.

Thurstone LL. 1927. Psychophysical analysis. *American Journal of Psychology* **38**: 368–389.

Treisman M, Watts TR. 1966. Relation between signal detectability theory and the traditional procedures for measuring sensory thresholds: Estimating d' from results given by the method of constant stimuli. *Psychological Bulletin* **66**: 438–454.

Vickers D. 1980. Discrimination. In *Reaction Times*, Welford AT (ed.). Academic Press: London; 25–72.

Copyright © 2001 John Wiley & Sons, Ltd.