APPLIED COGNITIVE PSYCHOLOGY, VOL. 5, 61–77 (1991)

Displaying Proportions and Percentages

IAN SPENCE*
University of Toronto

STEPHAN LEWANDOWSKY
University of Western Australia

SUMMARY

Pie and bar charts are commonly used to display percentage or proportional data, but professional data analysts have frowned on the use of the pie chart on the grounds that judgements of area are less accurate than judgements of length. Thus the bar chart has been favoured. When the amount of data to be communicated is small, some authorities have advocated the use of properly constructed tables, as another option. The series of experiments reported here suggests that there is little to choose between the pie and the bar chart, with the former enjoying a slight advantage if the required judgement is a complicated one, but that both forms of chart are superior to the table. Thus our results do not support the commonly expressed opinion that pie charts are inferior. An analysis of the nature of the task and a review of the psychophysical literature suggest that the traditional prejudice against the pie chart is misguided.

The display of proportions and percentages is a common activity: opinion polls and market surveys appear almost daily in the popular press; scientific articles and technical reports frequently exhibit results expressed in percentage form; and proportional data are the very stuff of handbooks, almanacs, and compendia of every kind. Despite the ubiquity of these data, and their widespread presentation in tables and graphs, comparative evaluations of the relative merits of these displays have been largely based on intuition rather than experimental data.

Some modern commentators recommend the use of tables to display small data sets. Ehrenberg (1975), for example, prefers properly constructed tables to graphical displays, and Tufte (1983), in one of the most influential recent books on the visual display of data, suggests that tables are more appropriate than graphs for data sets containing fewer than 20 observations. However, contrasting views, advocating graphical displays over tables for most purposes (e.g. Mahon, 1977; Wainer and Thissen, 1988), are perhaps more common, and many authorities lean towards the bar chart as the most appropriate graph for the display of proportional data; evidence from early studies on graphical displays, and a large body of psychophysical work, has been interpreted to favour the bar chart over the pie chart (see MacDonald-Ross, 1977, for a review). In the face of this official disapproval, the continuing fondness for the pie chart on the part of practical graph-makers is intriguing.

*Correspondence to: Ian Spence, Department of Psychology, University of Toronto, Toronto, Ontario, Canada M5S 1A1.

0888–4080/91/010061–17$08.50

Received 24 May 1988
Revised 10 May 1989

EARLY EXPERIMENTS ON STATISTICAL GRAPHS

It has long been assumed that people are more accurate judges of line length than of area, angle, or arc length. Consequently, most authorities have championed the bar chart and disparaged the pie chart, even though the relevant evidence turns out to be ambiguous upon closer inspection.

Over 60 years ago a lively discussion concerning the relative merits of the pie and bar chart appeared in the *Journal of the American Statistical Association*. Walter Eells (1926) apparently had shown the superiority of the pie chart over the divided horizontal bar chart. In a simple but well-executed experiment he had subjects estimate the percentage associated with a single component (a slice of the pie or an individual bar) and found that the magnitude estimation was performed more accurately and more quickly when the data were in pie chart form. This attempt to confound the conventional wisdom did not go unchallenged, and within the year von Huhn (1927) and Croxton (1927) had taken up the cause of the bar chart, considering a variety of issues that Eells had not treated, but failing to refute Eells's basic finding that the pie chart is better than the divided bar chart for the rapid and accurate estimation of percentages.

The initial flurry of activity spawned a number of empirical studies of varying quality (e.g. Croxton and Stryker, 1927; Croxton and Stein, 1932; Peterson and Schramm, 1955; Culbertson and Powers, 1959). Each added something to the discussion, but none settled the question of superiority one way or the other. Croxton and Stein (1932) compared bars, circles, squares, and perspective drawings of cubes as the basis for representing magnitudes, and although they found bars to be superior to all other forms, their results have little bearing on whether a bar or a pie chart is the better display, since their subjects did not consider subdivisions or segments of circles, but rather compared circles of different sizes. Notwithstanding, in his authoritative and generally excellent review, Macdonald-Ross (1977) has leaned heavily on this study as a 'conclusive demonstration' that the bar chart is superior to the pie chart.

Most studies have required subjects to estimate the magnitude of a proportion directly from the graph. For this task at least, our interpretation of the body of published work is that the pie chart has not been demonstrated inferior to the bar chart. At first blush this may seem to be at variance with psychophysical work, which has shown people to be better judges of length than of area.

THE PSYCHOPHYSICS OF MAGNITUDE ESTIMATION

Psychophysical research has traditionally sought to determine the functions that relate physical variables such as length, area, or volume to their perceived magnitudes. Drawing on this work, Macdonald-Ross (1977), for example, assumes that the function

$$\text{Subjective area} = \text{Area}^{0.86}$$

provides an adequate description of the relationship between the perceived areas of circles and their physical areas, indicating that people underestimate area as the size of the circles increases. Macdonald-Ross cites an exponent of unity for length, implying that judged length differs from actual length by at most a scale factor. Thus, since length is more accurately judged than area, the bar chart is to be preferred to the pie chart. In our view this summary of the psychophysical literature is inadequate and, in any case, is not relevant to judging the sizes of segments in a pie chart, or the sizes of bars in a bar chart.

Macdonald-Ross says that Ekman and Junge (1961) found the exponent of the psychophysical function for judged area to be 0.86 (Macdonald-Ross, 1977, p. 373). However, the number 0.86 refers to the result of an experiment in a previous unpublished report, with no specification of the nature of the stimulus or task (Ekman and Junge, 1961, p. 1). Indeed, Ekman and Junge's motivation for the 1961 paper seems to have been to examine judgements of area more closely in the light of Stevens and Galanter's (1957) survey of several experiments that found exponents for both length and area to be close to unity. Ekman and Junge (1961, p. 7) concluded that 'results reported in previous studies were verified: subjective *length* and subjective *area* are not far from proportional to objective length and objective area'. Notwithstanding, Baird's (1970) extensive review shows that estimated exponents for judged area *can* range from about 0.7 to more than 1.0, and at least some of this variability may be ascribed to the instructions that subjects receive (Teghtsoonian, 1965). The nature of the stimuli can also affect the exponent: for example, Spence (1989) has shown that exponents for judged area and volume are in the region of unity if the stimulus set displays variation along only one of the dimensions (which is typically the case with bar and box charts).

Even if the exponent for judging areas of whole circles is less than one (and under most circumstances this is the case: Baird, 1970), we should be cautious in assuming this to be relevant. When estimating a proportion from a pie chart, the task is quite unlike those in traditional psychophysical studies: subjects judge sub-areas (the slices, segments, or components) of a larger circular area. In addition to the fact that the shape of the sub-area is idiosyncratic, subjects can use perceptual anchor points that are otherwise unavailable: with the pie chart it is easy to compare the size of a component with imaginary quarters, or halves, as well as the whole, before making either absolute or relative judgements. A crude linear scale is thus available to the observer, and subjects report making use of this scale (Eells, 1926; Simkin and Hastie, 1987; Spence, 1989). Indeed, Spence (1989) found that the estimated exponent for the judged size of pie chart segments is about one. Cleveland and McGill (1984) have observed that subjects probably estimate proportions by judging angles—if this is so, the results of psychophysical studies of how area is perceived can shed little light on how well proportions are estimated from the segments of a pie chart.

Thus, some of the evidence that has been cited to discredit the pie chart is, upon closer inspection, found to be non-committal: exponents for judged area can be similar to those for judged linear extent, and, in any case, it is probable that area is not the feature to which subjects attend when viewing a pie chart.

PERCEPTUAL TASKS

The estimation of magnitude may appear to be a sensible psychological task for experiments comparing different types of chart. However, it seems illogical to portray numerical data in graphical form if the principal reason for doing so is to communicate precise numerical magnitudes to the observer. Precise magnitudes are probably communicated most efficiently if the data are left in tabular form. If what is required is an assessment of relative magnitudes, involving groups of components, a graph may be better than a table. In real-life situations, questions of the following kind are often asked: 'Does IBM enjoy a larger market share than DEC?'; 'Do Olivetti and Bull together sell more computers than IBM?'; or 'Do BMW and Saab together make more cars than Fiat and Mercedes?' Empirical comparisons of charts and tables must involve judgements of the relative size of components, or combinations of components, in order to reflect how the displays are used in practice.

In summary, previous research has failed to identify the best way of presenting proportions and percentages for at least two reasons: first, most studies (e.g. Croxton and Stein, 1932), and virtually all psychophysical experiments have focused on comparing areas (for example circles or other regular shapes) when a comparison of *sub-areas* (for example slices) would have been more appropriate. Second, empirical studies have typically used magnitude estimation, which does not reflect how people use graphs in real life. The experiments reported here use tasks that simulate the everyday use of graphs, and the comparisons involve sub-areas of shapes, as opposed to the shapes themselves. We report four experiments that contrast pie and bar charts with each other and with tables, using tasks that range from simple comparisons to more complicated ones involving pairs of components.

GENERAL METHOD

Subjects and apparatus

Members of the University of Toronto community participated either for a remuneration of $4 per hour or in fulfilment of an introductory psychology course requirement. Subjects were statistically unsophisticated—none had completed a course in statistics—since our interest was in displays that are frequently used to convey numerical data to a lay audience. An IBM personal computer controlled stimulus presentation and scored and timed subjects' responses. Stimuli were presented on a 50×40 cm SONY colour monitor, at a viewing distance of approximately 1.2 m. The stimuli subtended a visual angle of approximately 15 degrees.

Stimuli

On a given trial the subject was shown either a pie chart, a horizontal divided bar chart, a vertical bar chart, or a table, as shown in Figure 1. Stimuli were displayed in light blue on a dark grey background, and contained between four and seven components, depending on the experimental condition. Components were labelled with the letters A, B, C, . . ., from left to right, or clockwise in the case of the pie chart. Letters were printed in white, and were centred below each component in the horizontal divided bar chart, below each bar in the vertical bar chart, and outside the perimeter of the circle for the pie chart. For the table, the letters were placed in a separate row above the numerical entry for each cell.

For each replication, percentages were sampled randomly from a discrete uniform distribution on the range 5 to 60, subject to the restriction that all possible experimen-

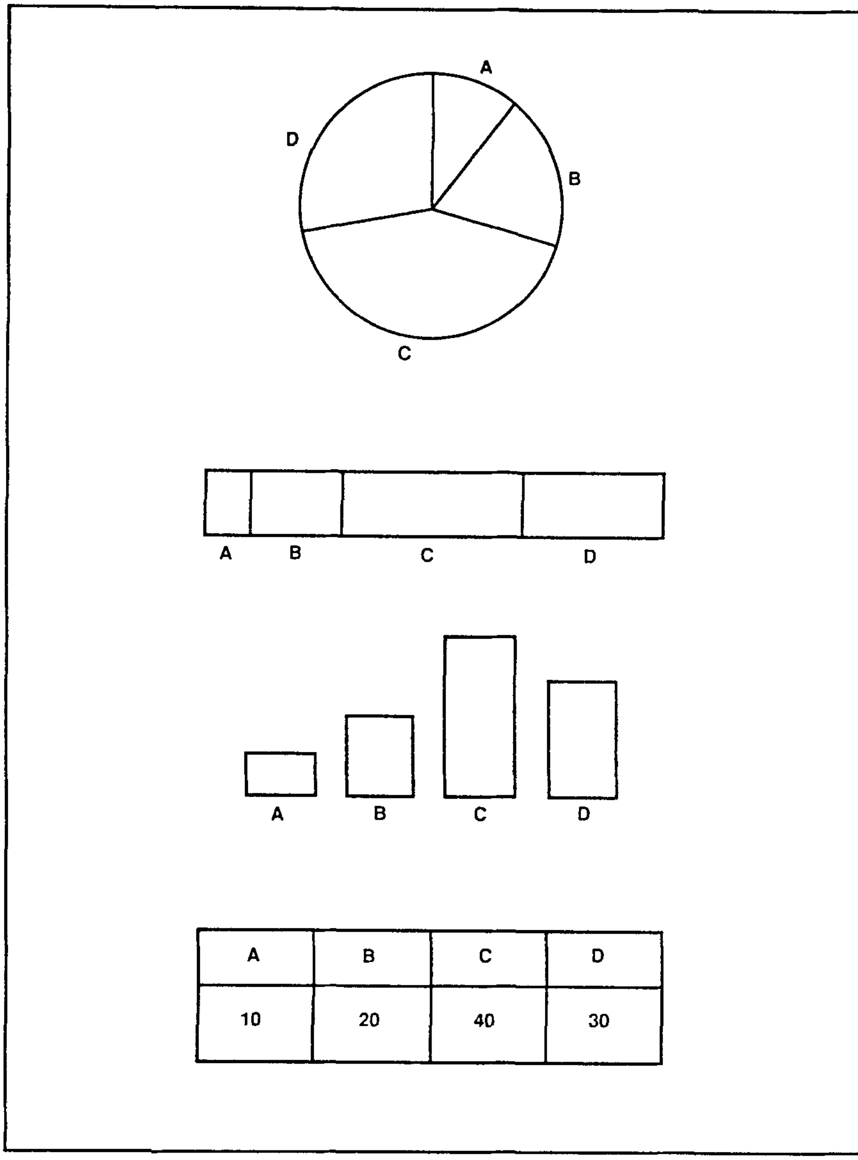

Figure 1. Examples of stimulus displays.

tal comparisons had to involve a difference of at least 1 per cent. Hence subjects were not required to make judgements of inequality with identical components, or combinations of components. The percentages for a given display always summed to 100, and each percentage was randomly assigned to a component.

Task

A two-alternative forced-choice task was used: subjects indicated which of two components, or combination of components, was the greater. The task might involve only single proportions with a question like 'Which is larger, A or B?', or combinations of proportions might be involved, leading to questions like 'Which is larger, A or B+C?', or 'Which is larger, A+B or C+D?' Letters indicating the components to be compared were shown beneath the stimulus, outside the frame of the graph, at the bottom left and right of the screen. Subjects responded by pressing either the '\' key (on the lower left of the keyboard) to indicate that the component, or combination, presented on the left was larger, or the '/' key (lower right) for the opposite judgement. The choice of components and their position—right or left—was determined randomly, subject to the constraints of the particular experimental condition. No position bias was associated with the correct response.

Procedure

Even statistically unsophisticated subjects can perform some perceptual tasks with almost perfect accuracy, provided that the time to respond is unrestricted (cf. Lewandowsky and Spence, 1989). In order to differentiate experimental conditions, a response signal method was used to make the task more difficult. Stimuli were presented for a predetermined period of time and the offset of the display was the response signal. Stimulus onset, as well as offset, was instantaneous and also involved a change in background: stimuli were presented on a dark grey background, but a blank, black screen was interposed between stimulus presentations. The change in background colour, from grey to black, when the display disappeared, served to mask the preceding stimulus.

Subjects were instructed to respond immediately upon disappearance of the stimulus, regardless of how certain they were of their decision, and responses with latencies greater than 1.5 seconds were scored as errors, although the results are not materially affected if slow responses are included in the analysis. The response-to-stimulus-onset interval was 1 second. The order of trials, corresponding to the cells of the experimental design, was randomized separately for each subject. The response variable was the percentage of correct responses.

Practice and instructions

Subjects were shown exemplars of each type of display, told that each component of the display represented a percentage, and that the percentages summed to 100. The sample charts included the actual percentages adjacent to the components of each graph. Before the experiment proper, subjects received four practice trials identical to the experimental trials except feedback was given after each response. All types of display were used during training. Subjects were prompted to respond faster if latencies exceeded 1.5 seconds during practice. At the discretion of the experimenter, practice was repeated if a subject responded too slowly or too inaccurately. Instructions were given in writing, and were recalled by subjects to ensure proper understanding.

EXPERIMENT 1

The firstexperiment explored some of the perceptual parameters of the task: the amount of time required to process the displays and the effect of the number of components. In addition, the experiment assessed the nature of the speed–accuracy trade-off (SAT) function for various graphs and types of task.

Design and procedure

The experimental design was a factorial combination of one between (task), and three within-subject (Display, Components, Deadline) factors. Ten subjects were randomly assigned to each of the three levels of the task variable: group 1 had to perform the simple comparison, involving only two components ('Which is greater, A or B?'), and groups 2 and 3 performed the A vs. B+C and A+B vs. C+D comparisons, respectively. Each group saw all three types of display; pie charts, bar charts, and horizontal divided bar charts. The graphs contained four, five, or seven components, and the display disappeared (according to the deadline) after either 1.5, 3.0, or 6.0 seconds. Seven replications were obtained in each of the cells formed by the three within-subject factors, yielding 189 trials. Sessions lasted under 1 hour, with two subject-paced break periods (after 60 and 120 experimental trials).

Results and discussion

The proportion of responses whose latency exceeded 1.5 seconds, and which were thus considered errors, was small, ranging from 1.9 per cent (group 1) to 3.2 per cent (group 2). These slow responses were distributed evenly across the levels of all within-subject

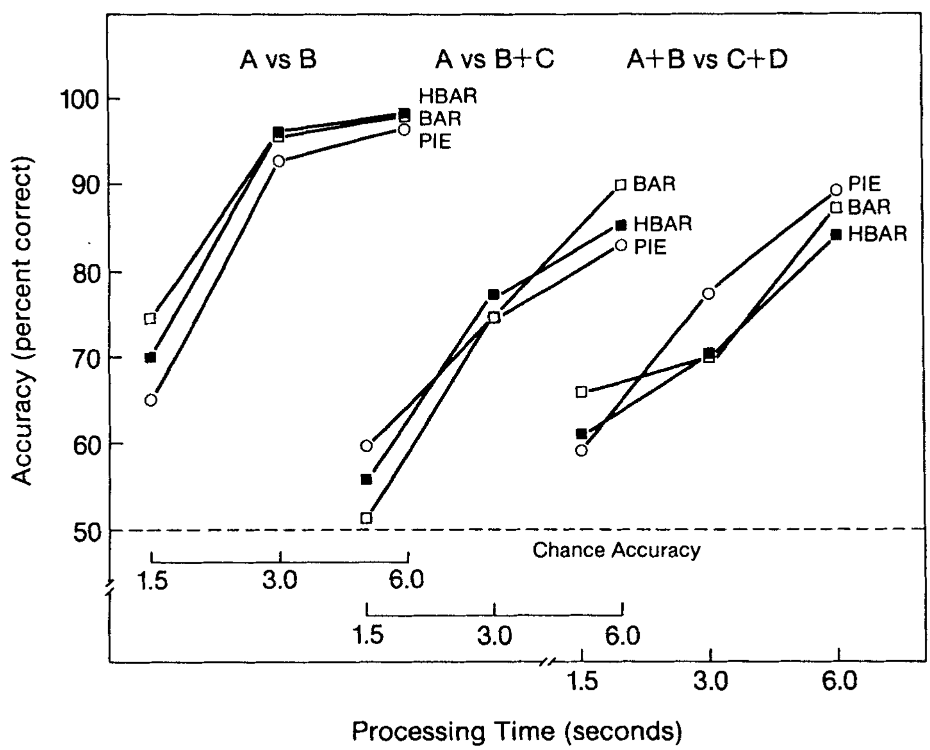

Figure 2. Speed–accuracy trade-off (SAT) functions for the three displays. Panels 1, 2, and 3 show SAT functions for the A vs. B, A vs. B+C, and A+B vs. C+D tasks, respectively.

variables, with the exception of deadline: more responses exceeded the cut-off with short presentation times than with longer ones.

Figure 2 shows the SAT functions obtained for the three different displays, with one panel devoted to each type of task. The effect of deadline is sufficiently obvious to warrant the omission of formal tests; accuracy increases with increasing processing times, up to a maximum value at the longest deadline. That maximum accuracy level is dependent upon the task: it is essentially 100 per cent for the A vs. B comparison (panel 1), and does not exceed 90 per cent, even at the longest deadline, if combinations of components are involved (panels 2 and 3).

Figure 2 shows that the shape of the SAT function is almost identical for all displays, with a divergence at shorter processing times: the bar chart was slightly better than the pie chart with the A vs. B task, whereas other differences between displays were negligible. Indeed, in separate analyses of variance for each of the three groups, the effect of display was significant only for group 1 [$F(2,18) = 4.54, p < .05, MSe=.01$]. A more detailed exploration of the effects of display is the focus of Experiments 3 and 4, described below.

The equivalent performance of the vertical bar chart and the horizontal divided bar chart is noteworthy, given some perceptually important differences between the two. Both charts represent a proportion by the length of a component, but in the vertical bar chart all components are referred to the same origin, whereas in the horizontal divided bar chart comparisons between segments must be made without a common point of reference. Cleveland and McGill (1984) have shown that magnitude estimations involving non-aligned scales are more prone to error—and hence are presuma-

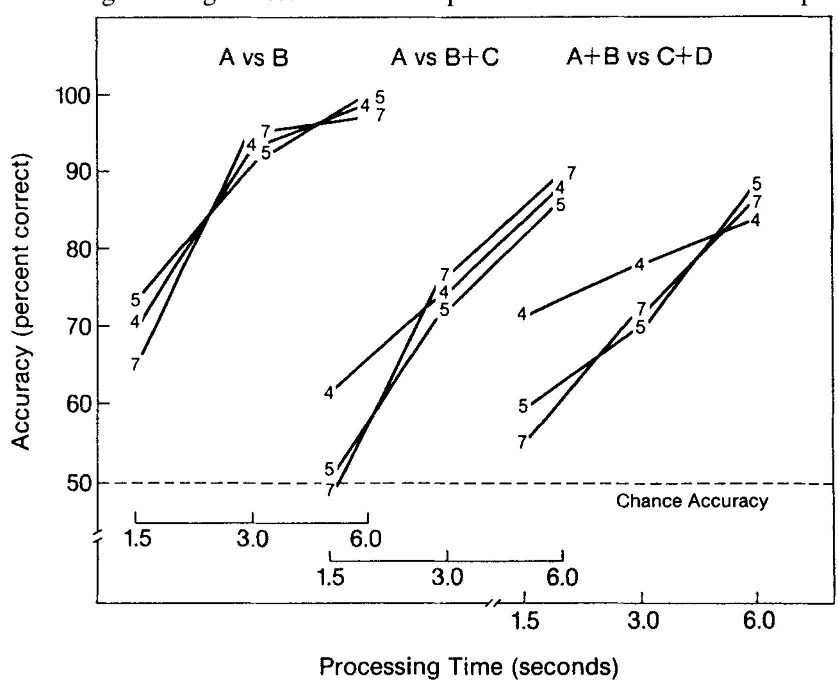

Figure 3. Speed–accuracy trade-off (SAT) functions for different numbers of components. Panels 1, 2, and 3 show SAT functions for the A vs. B, A vs. B+C, and A+B vs. C+D tasks, respectively. The numbers (4, 5 or 7), used as plotting symbols, denote the number of components.

bly perceptually more difficult—than those involving an aligned scale. Notwithstanding, in Experiment 1, subjects performed at comparable accuracy levels with both types of bar chart.

Three additional SAT functions, showing the effects of the number of components in the display, are shown in Figure 3. For the simple A vs. B task the number of components was unimportant, but when pairs of components (Panels 2 and 3) were involved, performance with the four-component display was better than with the more complex graphs. Statistically, components had no effect for group 1, but did for group 2 [$F(2,18) = 2.76$, $p < .10$, $MSe=.028$] and group 3 [$F(2,18) = 4.24$, $p < .05$, $MSe=.024$]. Regardless of the type of display, performance for graphs with either five or seven components was about the same, and differed from graphs with four components.

EXPERIMENT 2

Experiment 1 showed that, except at short processing times, display type has little effect on performance. The two types of bar chart differed little, if at all, from each other. Therefore, the horizontal divided bar chart was omitted and a table substituted to examine possible differences between graphical and non-graphical modes of presentation. Another purpose of Experiment 2 was to investigate the effect of ordering the components of a graph. The segments of a pie chart are frequently ordered in a clockwise fashion such that the largest segment starts in the 12 o'clock position, followed by the next largest segment, and so on, with the smallest segment immediately to the left of the largest one. In a bar chart or table, components are also often ordered by magnitude; from the largest on the left to the smallest on the right. It has been suggested that there may be some advantage to ordering components by magnitude: Ehrenberg (1975), for example, is a strong advocate of ordering table entries to improve the presentation.

Design and procedure

Since at the longest deadline of Experiment 1 (6.0 seconds) no differences were observed among displays, and because the results did not differ much between the two shorter processing times, display duration was fixed at 3.0 seconds. The design was a within-subjects factorial combination of display, components, task, and ordering. Three displays were used (pie chart, bar chart, and table); however, since Experiment 1 showed that five and seven components produce a similar effect, only two levels of components (four and seven) were used. Also, unlike the first experiment, the same group of subjects performed two tasks (A vs. B and A vs. B+C). Half the displays had ordered components, and the other half did not (ordering). Each of the 24 cells in the design was replicated 10 times, producing a total of 240 test trials. Ten subjects participated in sessions, which contained three self-paced break periods (after 60, 120, and 180 trials), and lasted less than an hour.

Results and discussion

Similar to Experiment 1, approximately 3 per cent of all observations (evenly distributed over conditions) exceeded the 1.5 seconds latency cut-off. Somewhat surprisingly, the results did not show an overall advantage of the graphs over the table. Although analysis of variance showed the display variable to be significant [$F(2,18) = 4.08, p < .05, MSe = .018$], the effect was mainly due to the superiority of the bar chart (89 per cent correct) over the pie chart (83 per cent), with the table located somewhere in between (86 per cent).

As before, components [$F(1,9) = 17.9, p < .005, MSe = .014$] and task [$F(1,9) = 39.0, p < .001, MSe = .016$] were highly significant in the expected direction: performance was better using displays with four components (89 per cent) than those with seven (82 per cent), and the simple task (91 per cent) was more accurately performed than the more complicated one involving pairs of components (81 per cent). Moreover, in replication of the between-subjects pattern of Experiment 1, these two variables interacted [$F(1,9) = 12.1, p < .01, MSe = .009$] such that performance on the simple task was unaffected by components (92 and 90 per cent for four and seven components, respectively), whereas for the more complicated task, performance was sensitive to number of components (86 vs. 75 per cent).

The effect of ordering was negligible, with 85 per cent average accuracy for the displays with ordered components, versus 87 per cent for displays with randomly positioned components. There were no display by ordering, task by ordering, or components by ordering effects. This was unexpected, given the popular advice to order components, and also because in an ordered display the location of components coincides with their magnitudes. For the simple task, when told to compare two bars, say, a subject could—knowing that the display was ordered—base the response on the fact that the leftmost to-be-compared bar must necessarily be the greater one. The absence of an ordering effect suggests that subjects did not make use of this potential short-cut, but used the same perceptual strategy in all conditions.

EXPERIMENT 3

The objective was to examine, more closely, performance as a function of the complexity of the task and the type of display. In contrast to the between-groups design of Experiment 1, each subject was required to make comparisons of components and combinations of components, with all three displays.

Design and procedure

The within-subjects design involved three variables: display (pie chart, bar chart, and table); task (A vs. B, A vs. B+C, and A+B vs. C+D); and components (four and seven). The 18 cells of the design were replicated 10 times to form the test sequence of 180 trials. On each trial the stimulus was exposed for 3.0 seconds. Subject-paced break periods were given after 60 and 120 trials. Ten subjects participated.

Results

At 6 per cent, the number of responses exceeding the latency cut-off was slightly higher than in the first two experiments. Analysis of variance showed that all three main effects, and the display by task interaction, were significant. The effect of components [$F(1,9) = 26.30, p < .01, MSe = .0092$], with performance levels of 81 and

73 per cent for four and seven components, respectively, paralleled the finding of the previous experiments; accuracy was better with a smaller number of components. Accuracy declined with the complexity of the experimental task [$F(2,18) = 27.45, p < .01, MSe=.0158$]. Performance was 85, 78, and 68 per cent, for the A vs. B, A vs. B+C, and A+B vs. C+D comparisons, respectively. Of the three displays [$F(2,18) = 7.94, p < .01, MSe=.0150$], the pie and bar chart differed little (80 and 79 per cent respectively), but the table led to poorer overall performance (72 per cent).

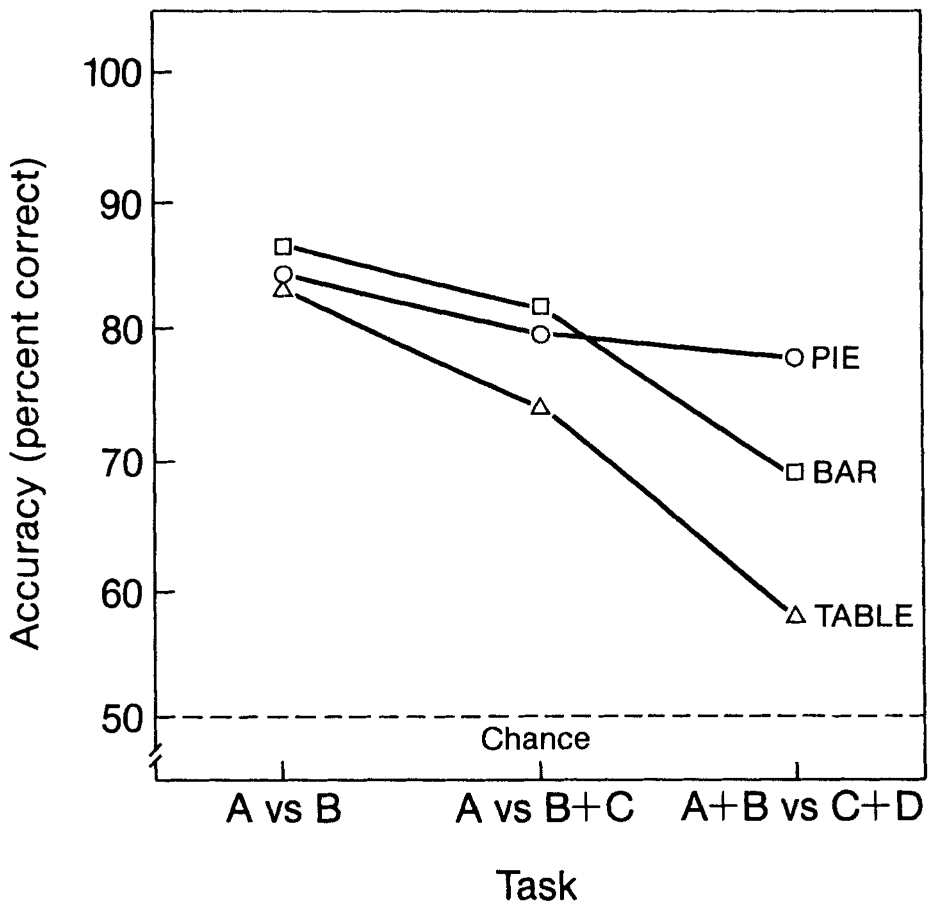

Figure 4. Accuracy observed in Experiment 3 for the three displays and three tasks. Stimuli were presented for 3.0 seconds.

While the main effects of both display and task were significant, it is their interaction [$F(4,36) = 3.54, p < .05, MSe=.0136$], shown in Figure 4, that is most interesting. With the simplest task (A vs. B), as in the previous studies, all three display types led to roughly the same performance. As the task is made more complicated, performance degrades differentially for the different displays. When a single component is compared against a pair, both charts produce comparable performance and both have the edge over the table. Most interestingly, when two pairs of components must be compared, the pie chart outperforms both the table *and* the bar chart. The Fisher least significant difference (at the .05 level) associated with the interaction was 7 per cent, supporting this ordering of the displays.

Discussion

The most interesting result of Experiment 3 is the differentiation of the three displays with the A+B vs. C+D task, and in particular the superiority of the pie chart. One possible reason is apparent: if the components *within* a pair happen to be adjacent, they may function as a single, albeit subdivided, component. With the pie chart the 'complicated' comparison may be reduced to a simple one. With the bar chart, adjacency of components does not change the nature of the task. Two bars do not form a single graphical element, regardless of their proximity.

In this experiment components were randomly selected for each comparison, and the number of adjacent vs. non-adjacent comparisons was therefore not always the same for a given subject–cell combination. With this qualification in mind a *post hoc* analysis of adjacency revealed that, for the A+B vs. C+D comparison, it was the *bar chart* that benefited most from adjacency. For the pie chart virtually no difference in performance (78 vs. 79 per cent) was found between adjacent and non-adjacent components within a pair, whereas the difference for the bar chart (77 vs. 60 per cent) was considerable. Thus, adjacency can be ruled out as the reason for the advantage of the pie chart.

Examination of the experimental task suggests at least two processing stages that may be responsible for the pie chart's superior performance. The task—comparing pairs of components—requires that the components first be located and identified, before they are cognitively combined and compared. With the pie chart it may be easier to locate components than with the bar chart; the labels 'A' and 'B' may be easier to find when arranged in clockwise order in a pie chart than when shown in sequence from left to right in a bar chart. If so, the advantage of the pie chart should be reduced if the observer can locate components more easily, without having to search for an alphabetic label. Alternatively, it may be easier, perceptually or cognitively, to combine and compare the components of a pie chart. Perhaps it is easier to combine the slices representing A and B in a pie chart—even if they are not adjacent—than it is to sum the lengths or areas of two bars. In that case, facilitating the search for the components should not affect the superior performance of the pie chart. Experiment 4 was designed to explore these issues.

EXPERIMENT 4

Whereas the first three studies required subjects to locate components using alphabetical labels, this experiment used colour to code components. Colour is used in many applications of pie and bar charts; it is subjectively compelling and provides a way for the designer of a chart to emphasize particularly important groupings of components. In the present context it also eliminates the need for subjects to search for alphabetical labels.

Method

The method and experimental design were identical to those of Experiment 3, except that the components, or combinations of components, to be compared were coloured either blue or red. As before, each graph was presented as a blue outline on a dark grey background; but the to-be-compared components were filled with blue or red. Thus a pie chart might have two blue slices and a red one, or a bar chart might have two bars of each colour, and so forth. Entries in the table were coded by colour patches above the numbers to be compared. For all comparisons, if the blue component (or components) represented the larger proportion, subjects pressed the left response key. Similarly, the right response key was associated with red components, and was pressed whenever the red component (or components) was the greater. Since colour was randomly assigned to components in the graph, there was no position bias associated with the correct response. To remind subjects of the mapping of the response keys, a blue patch was presented at the lower left of the screen, and a red one at lower right.

Because subjects did not have to search for components by locating labels, this version of the task was much easier than in the foregoing experiments. Consequently, stimuli were presented for 1.0 second only. Eleven subjects were tested.

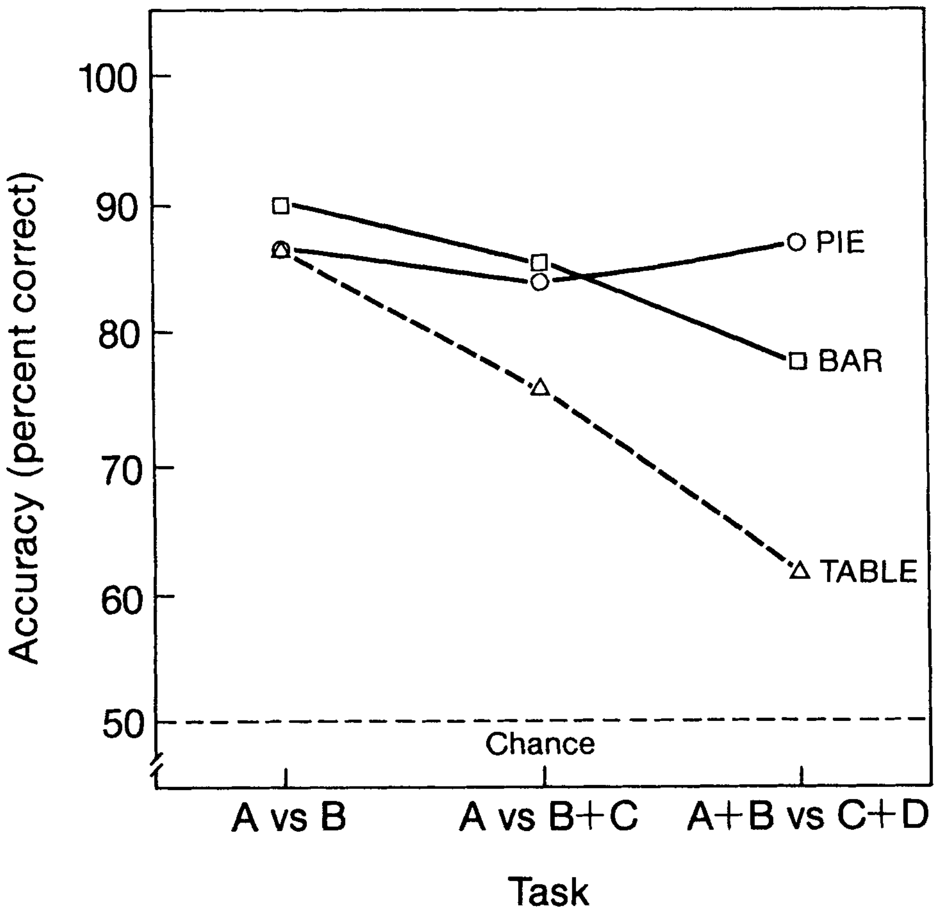

Figure 5. Accuracy observed in Experiment 4 for the three displays and three tasks. Colour was used to identify the components to be compared, and stimuli were presented for 1.0 second.

Results

The proportion of responses exceeding the latency cut-off was again in the vicinity of 6 per cent, as in the preceding study. Slow responses were evenly distributed throughout all conditions. In contrast to the results of the preceding experiment, only the main effects of display and task, and their interaction, were significant. A components effect [$F(1,10) < 1$] was not present. As before, the displays effect [$F(2,20) = 14.05, p < .01, MSe=.0170$] reflected poorer performance on the table (75 per cent) than on the graphs (86 and 84 per cent for pie and bar, respectively), and the task effect [$F(2,20) = 13.01, p < .01, MSe=.020$] reflected a drop in performance as complexity of the comparison increased (88, 82, and 76 per cent, for simplest to most complex).

The interaction between display and task, shown in Figure 5, was again significant [$F(4,40) = 5.20, p < .01, MSe=.018$], and replicates the most interesting finding of the previous study. The Fisher least significant difference (at the .05 level) associated with the interaction was 8.2 per cent, showing that when pairs of components must be compared the pie chart was again significantly superior to the bar chart, which in turn was better than the table. As in the preceding experiment, an analysis of adjacency showed that, for the pie chart, performance did not differ much between adjacent and non-adjacent components within a pair, whereas the bar chart benefited somewhat from adjacency. As in Experiment 3, all displays were about equal for the simplest (A vs. B) task, and for the mixed (A vs. B+C) task the table produced poorer performance than the two charts, which in turn did not differ much.

Discussion

Unlike the first three studies, Experiment 4 did not show a components effect: accuracy was the same, whether there were four or seven components in the display. This suggests that previous manifestations of the components effect were due to the differential ease with which components could be located in the display, rather than to differences in the ability to make perceptual or cognitive comparisons. Once the critical components have been located, further cognitive processing appears to be unaffected by the number of components in the display. This suggests that, if the designer of a graph wishes to draw attention to comparisons of particular combinations, the use of colour may be effective in charts with a large number of segments.

The last two experiments showed that, except for simple tasks, a graph is to be preferred to a table when judging the relative magnitudes of proportions. At least with the present tasks the data do not support Ehrenberg's (1975) recommendation to use a table—even when the number of data points is small (cf. Tufte, 1983). Moreover, if judgements involve pairs of components the pie chart is to be preferred to the bar chart, independent of whether or not subjects need to search for components by their alphabetical labels, and whether or not components within a pair are adjacent.

Both Experiments 3 and 4 show the bar chart to be superior to the pie chart with the two simpler tasks, although the difference is about half the size of that found in Experiment 2 (about 3 per cent rather than 6 per cent). However, the good performance of the table in Experiment 2 was not replicated in either Experiment 3 or 4. The anomaly is due to the table's unusually good performance in the simplest (A vs. B) condition in Experiment 2 (93 per cent).

GENERAL DISCUSSION

As Macdonald-Ross (1977) has observed, practical communicators have a good sense of what works, and does not work, in display graphics. He cautions that one should not lightly ignore this accumulated body of practical wisdom even though it may have little formal experimental base. Part of the motivation for the experiments reported here derived from our observation that the pie chart is widely used in presentation graphics despite the alleged drawbacks catalogued by experts over the years. We note that Eells (1926) was practically the lone statistically expert advocate of the use of the 'circle diagram' (see also Spence and Lewandowsky, 1989).

We have shown that the pie chart is a useful tool for the display of proportions, especially when the observer is required to make comparisons involving combinations of components. Although this work was not designed to test a formal cognitive model that would explain why the pie chart is effective, our experiments do not support a simple visual search explanation: even when components can be visually located with little perceptual effort—when colour is used to code segments—the pie chart outperforms the bar chart for complicated comparisons, suggesting that the perceptual addition and comparison of components is inherently easier with the pie chart than the bar chart.

Few investigators have proposed models of the cognitive processes that subjects may use when viewing graphs. One exception is the work of Simkin and Hastie (1987), who provide an information processing analysis of graph perception. While their analysis is *ad hoc*, the cognitive framework they propose is consistent with the present results, and suggests future experimentation. Simkin and Hastie's (1987) scheme applies previous work in visual cognition (e.g. Kosslyn, 1980; Ullman, 1984) to graph perception: observers are assumed to use four visual routines, which are invoked after the basic elements of the graph have been identified by low-level edge-detecting processes.

The *anchoring* routine allows the observer to segment a selected component of the image in order to provide a standard for a subsequent estimate. For example, the observer may mentally divide a pie into four quarters, thereby providing anchor points at 25, 50, and 75 per cent of the total area. Similar anchors can be created with a bar chart. Subsequently, a *scanning* routine 'sweeps' the image, from the closest anchor to the point which has to be estimated, and the duration of the scan is used to generate the estimate. Consider a slice of a pie chart, representing 37 per cent, extending in a clockwise direction from the 12 o'clock position. To estimate the proportion the anchoring routine would first place an anchor at the 3 o'clock position, and the scanning routine would sweep across the image, from the anchor to the boundary at roughly 4 o'clock. This description is consistent with Simkin and Hastie's (1987) data, which show that magnitude estimations are more accurate, and significantly faster, when the true values are close to the presumed anchor points (at 25 or 50 per cent).

Two other visual routines are proposed for comparing elements. The *projection* routine employs a reference line that runs, for example, horizontally from the top of one bar in a bar chart to intersect another bar. Scanning, from the point of projection, is then used to estimate the height of the bar relative to the second bar. The *superimposition* routine 'moves' elements of the graph to new locations. For example, if two pairs of components are to be compared, the superimposition routine would rearrange the elements of each pair so that they were contiguous, before subsequent comparison of the totals.

Our results may be interpreted within this framework. For the (A vs. B) comparison, a *projection*, from bar A to B, followed by a decision as to whether or not the projection intersected with B, suffices for the bar chart. For the pie chart a simple projection is insufficient, and two separate anchoring-plus-scanning processes—one for each of the segments involved in the comparison—would be necessary prior to the decision. Thus, it is not surprising that the bar chart seems slightly better suited for this simple task.

For the most complicated task (A+B vs. C+D), the final comparison between the pairs presumably differs little from the simple A vs. B comparison. However, this final comparison stage must be preceded by a *superimposition* of components, in order to form sums within a pair. Specifically, for the bar chart, two bars need to be 'moved' on top of their corresponding within-pair counterpart, before a projection (from the resulting imaginary total of the two bars within a pair) can take place. For the pie chart, segments must be rotated around the centre to contiguity, before the anchoring-plus-scanning analysis can take place. If processes like these are operating, the results of our last two experiments suggest that superimposition is inherently more difficult with the bar chart than the pie chart. The present data cannot explain why this is the case, but we surmise that an explanation may involve the centre of the pie chart. Specifically, a superimposition of graphical components may be facilitated if they can be rotated around a static point, or about a fixed axis, as opposed to requiring a free 'movement' from one location to the other. According to this analysis the advantage of the pie chart should be negated if all slices were slightly 'exploded' from a common centre, thereby requiring a free superimposition similar to that needed with the bar chart.

Outside the domain of mental imagery research, little direct evidence exists for the kind of 'visual routines' advocated by Simkin and Hastie (1987). Notwithstanding, their general approach may have some utility, since the predictions it yields are testable. We believe that a more cognitive approach to the study of graphical perception is a necessary complement to purely psychophysical research.

CONCLUSION

We have found that the pie chart compares favourably with the bar chart when tasks other than direct magnitude estimation are required. In most situations there is little to choose between the two, but when the reader of a graph must make comparisons of combinations of proportions, the pie chart enjoys an advantage. Moreover, unless the object is to transmit precise numerical values of proportions or percentages to the reader, tables appear to be inferior to charts as display devices. If the designer of a chart wishes to bring particular groupings of components to the attention of the reader, the use of colour appears beneficial. To conclude, we note that an information-processing analysis of the perception of statistical graphs can generate testable predictions for future research.

ACKNOWLEDGEMENTS

This research was supported by grant A8351 from the Natural Sciences and Engineering Research Council of Canada to I. Spence, and also by grant G1779 from the

Natural Sciences and Engineering Research Council of Canada to R. Baecker, A. Fournier, P. Muter, D. Olson, and I. Spence. The authors wish to thank Jeff Bowers, Paula Servin and Linda Tilley for testing the subjects. Correspondence concerning this article should be addressed to Ian Spence, Department of Psychology, University of Toronto, Toronto, Ontario, Canada M5S1A1.

REFERENCES

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

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

Croxton, F. E. (1927). Further studies in the graphic use of circles and bars II: some additional data. *Journal of the American Statistical Association*, **22**, 36–39.

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

Croxton, F. E. and Stryker, R. E. (1927). Bar charts versus circle diagrams. *Journal of the American Statistical Association*, **22**, 473–482.

Culbertson, H. M. and Powers, R. D. (1959). A study of graph comprehension difficulties. *Audio-visual Communication Review*, **7**, 97–100.

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

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

Ekman, G. and Junge, K. (1961). Psychophysical relations in visual perception of length, area and volume. *Scandinavian Journal of Psychology*, **2**, 1–10.

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

Lewandowsky, S and Spence, I. (1989). Discriminating strata in scatterplots. *Journal of the American Statistical Association* (in press).

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

Mahon, B. H. (1977). Statistics and decisions: the importance of communication and the power for graphical presentation. *Journal of the Royal Statistical Society*, Series A, **140**, 298–323.

Peterson, L. V. and Schramm, W. (1955). How accurately are different kinds of graphs read? *Audio-visual Communication Review*, **2**, 178–189.

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

Spence, I. (1989). The visual psychophysics of simple graphical elements (submitted manuscript, under review).

Spence, I. and Lewandowsky, S. (1989). Graphical perception. In J. Fox and S. Long (Eds), *Modern methods of data analysis.* Beverly Hills, CA: Sage (in press).

Stevens, S. S. and Galanter, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, **54**, 377–411.

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

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

Ullman, S. (1984). Visual routines. *Cognition*, **18**, 97–159.

Von Huhn, R. (1927). Further studies in the graphic use of circles and bars I: a discussion of Eells' experiment. *Journal of the American Statistical Association*, **22**, 31–36.

Wainer, H. and Thissen, D. (1988). Plotting in the modern world: statistics packages and good graphics. *Chance*, **1**, 10–20.