Journal of Experimental Psychology: Applied
2008, Vol. 14, No. 2, 85–100

1076-898X/08/$12.00 DOI: 10.1037/1076-898X.14.2.85

The Effect of Emergent Features on Judgments of Quantity in Configural and Separable Displays

David Peebles
University of Huddersfield

Two experiments investigated effects of emergent features on perceptual judgments of comparative magnitude in three diagrammatic representations: kiviat charts, bar graphs, and line graphs. Experiment 1 required participants to compare individual values; whereas in Experiment 2 participants had to integrate several values to produce a global comparison. In Experiment 1, emergent features of the diagrams resulted in significant distortions of magnitude judgments, each related to a common geometric illusion. Emergent features are also widely believed to underlie the general superiority of configural displays, such as kiviat charts, for tasks requiring the integration of information. Experiment 2 tested the extent of this benefit using diagrams with a wide range of values. Contrary to the results of previous studies, the configural display produced the poorest performance compared to the more separable displays. Moreover, the pattern of responses suggests that kiviat users switched from an integration strategy to a sequential one depending on the shape of the diagram. The experiments demonstrate the powerful interaction between emergent visual properties and cognition and reveal limits to the benefits of configural displays for integration tasks.

Keywords: diagrammatic reasoning, graphs, emergent features, information graphics

Presenting large, complex data sets to a broad nonspecialist audience is challenging if one is to be sure that the information is interpreted in the manner intended. As the power and sophistication of data analysis software increases, the range of information graphics available to practitioners is becoming ever wider (Harris, 1993; e.g., contains a bewildering array of options). Choosing the most suitable diagram for a particular communicative goal is an important aspect of the task, and designers of information displays must be aware of the complex relationship that exists between users' cognitive and strategic processes, the visual features of the external representation, and the requirements of the task being undertaken.

This study seeks to illuminate further this ternary relationship by investigating the representational and computational properties of a diagram variously known as the kiviat, spider, radar, or star graph and comparing these properties with those of more commonly used diagrams. Although relatively unfamiliar to the general public, kiviat charts are commonly used in science, engineering, and business and are available in several forms (see Harris, 1993). Some have radial axes with grid lines or numbered tick marks and labels while others consist solely of one or more polygons plotted on radial axes without tick marks. In some cases, even the radial lines are removed to produce one or more closed polygons that can be compared in terms of area. Whatever the design, the most common purpose of this type of diagram is to allow two or more multivalued objects to be compared, either in terms of an individual dimension or at a global level.

In recent years, kiviat charts have been employed by the U.K. government as the primary vehicle for presenting national police performance data to the public (Police Standards Unit, 2003, 2004). Titled "performance monitors", these diagrams present in summary form performance data for individual police forces in five key areas or "domains" (citizen focus, promoting public safety, resource usage, investigating crime, and reducing crime), together with the average performance computed from a set of police forces most similar to the individual force in terms of socioeconomic, demographic, and geographic makeup.

An example performance monitor and a section of the Home Office document explaining its interpretation are shown in Figure 1. In each kiviat chart, a police authority's performance in a domain is indicated by a point on an axis. The points are connected by straight lines to form a pentagon, and the regular shaded pentagon represents the average performance of a set of most similar local authorities. Better performance is shown further out from the center. Since 2003, the central kiviat chart has been augmented with bar graphs that illustrate the spread of performance for the most similar police forces in each of the domains. In these bar graphs, each bar represents the value on that domain of one of the forces from which the average has been computed.

Kiviat charts are a static form of object or configural display, a form of representation in which multiple variables are combined into a single object, typically a polygon. In contrast, bar graphs are a prime example of a separable display, as each value is represented by an individual bar. The extent to which separate dimensions are

David Peebles, Department of Behavioural Sciences, University of Huddersfield.

I thank Wayne Gray for his help and advice, and David Tyfa, Jeff Zacks, and two anonymous reviewers for their valuable comments on earlier versions of this article. Portions of the data from Experiment 1 were presented in Peebles (2004).

Correspondence concerning this article should be addressed to David Peebles, Department of Behavioural Sciences, University of Huddersfield, Queensgate, Huddersfield HD13DH, United Kingdom. E-mail: D.Peebles@hud.ac.uk

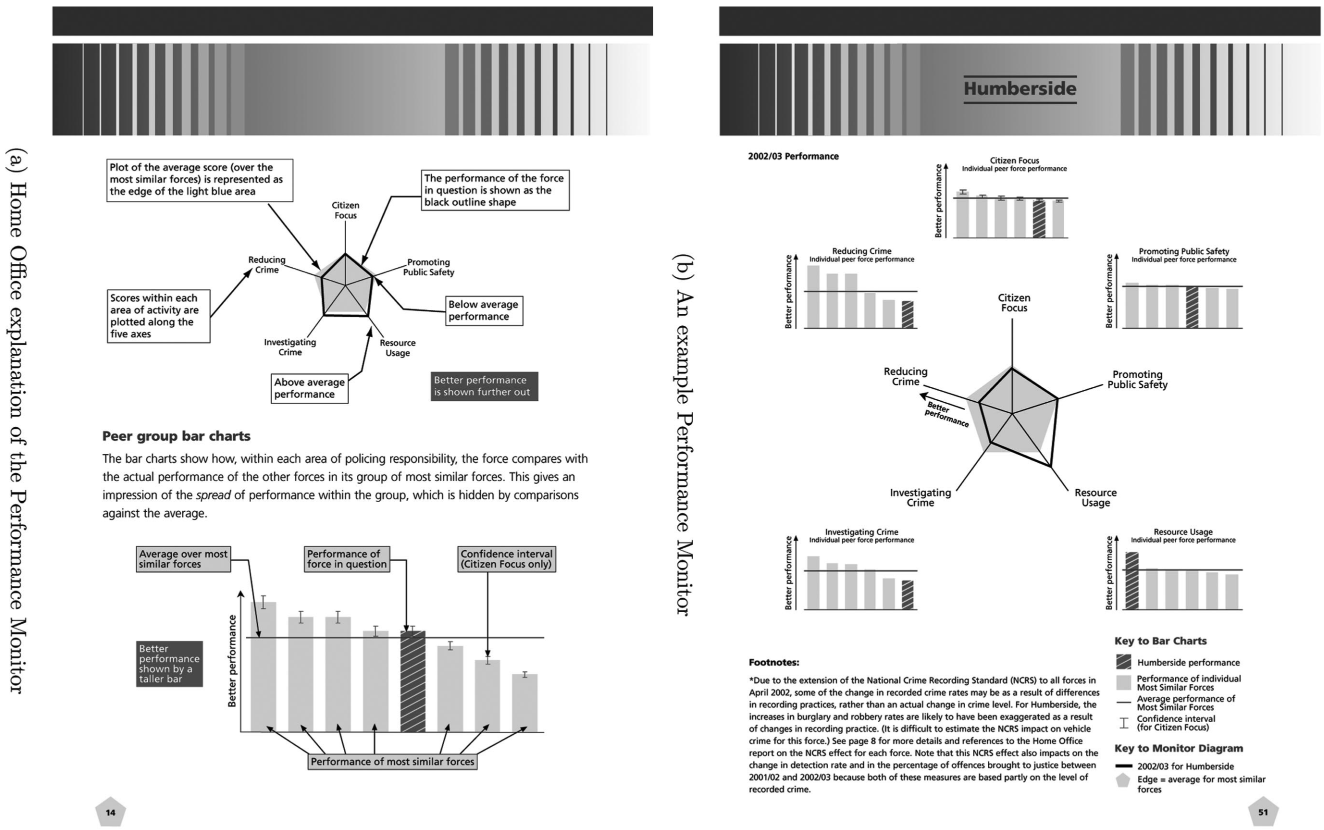


Figure 1. U.K. Home Office Police Performance Monitors, 2003.

perceived as a single object in a display is known as its display proximity (Barnett & Wickens, 1988; Carswell & Wickens, 1987; Wickens & Andre, 1990) and configural displays, such as kiviat charts, are regarded as possessing a high degree of display proximity compared to separable displays, such as bar graphs.

According to the proximity compatibility principle (Wickens & Carswell, 1995), representations with high display proximity facilitate high mental proximity tasks requiring the integration of information from multiple sources. The relationship between display and mental proximity has been revealed in several studies (e.g., Barnett & Wickens, 1988; Bennett & Flach, 1992; Bennett, Toms, & Woods, 1993; Carswell & Wickens, 1987; Wickens & Andre, 1990; Wickens & Carswell, 1995) and it is argued that the advantage comes when emergent features (Pomerantz & Pristach, 1989) of the representation indicate additional information about the domain. For example, in dynamic displays, the shape of a polygon can signal when variables are developing beyond normal boundaries if it is distorted compared to a regular polygon representing normal conditions. Such dynamic polygon displays have been used to represent the fluctuating states of variables in nuclear power plants (e.g., Petersen, Banks, & Gertman, 1982; Woods, Wise, & Hanes, 1981) and to indicate the physiological status of patients undergoing surgery (Gurushanthaiah, Weinger, & Eglund, 1995).

Bar graphs typically facilitate low mental proximity tasks which demand more focused attention to extract information from a single source, (Casey & Wickens, 1986; Wickens & Andre, 1990; although see Bennett & Flach, 1992, for evidence that this is not always the case), but they can also facilitate integration tasks if the emergent features of the bar graph can be mapped onto relevant features of the task (Coury, Boulette, & Smith, 1989; Sanderson, Flach, Buttigieg, & Casey, 1989). Bennett, Nagy, and Flach (1997) have argued however that focused tasks can be performed equally well with object displays because they consist of a hierarchy of features (Treisman, 1996), different levels of which will be attended to depending on whether the task requires integration or focused attention.

The ability to switch between the levels of features in configural displays may not only allow them to be used in specific focused tasks but also may provide alternative strategies for integral tasks. Psychophysical studies have revealed judgments of area to be generally less accurate than other perceptual judgments, such as position along a scale, length, direction, or angle (Cleveland & McGill, 1984, 1985). In situations where shape comparison is difficult or judgmental accuracy is to be preferred over speed, it may be better to compare individual stimulus dimensions sequentially and compute an overall total than to base a judgment purely on the global features of the objects. The facility of switching strategies for a single task is one of the features of object displays that will be investigated in the second half of this study.

Standing between the two extremes of object displays and bar graphs is the line graph. Although line graphs share many visual and representational properties with bar graphs (e.g., a two dimensional Cartesian coordinate system and a similar assignment of variables to axes), comparative studies have revealed marked differences in the ways people represent and interpret them. For example, people typically encode bars in terms of their height, interpret them as representing the separate values of nominal scale data, and are better at comparing and evaluating specific quantities using bar graphs (Culbertson & Powers, 1959; Zacks & Tversky, 1999). In contrast, people typically encode lines in terms of their slope (e.g., Simcox, 1983; reported by Pinker, 1990), interpret them as representing continuous changes on an ordinal or interval scale (Kosslyn, 2006; Zacks & Tversky, 1999), and are better at identifying trends using line graphs (Schutz, 1961).

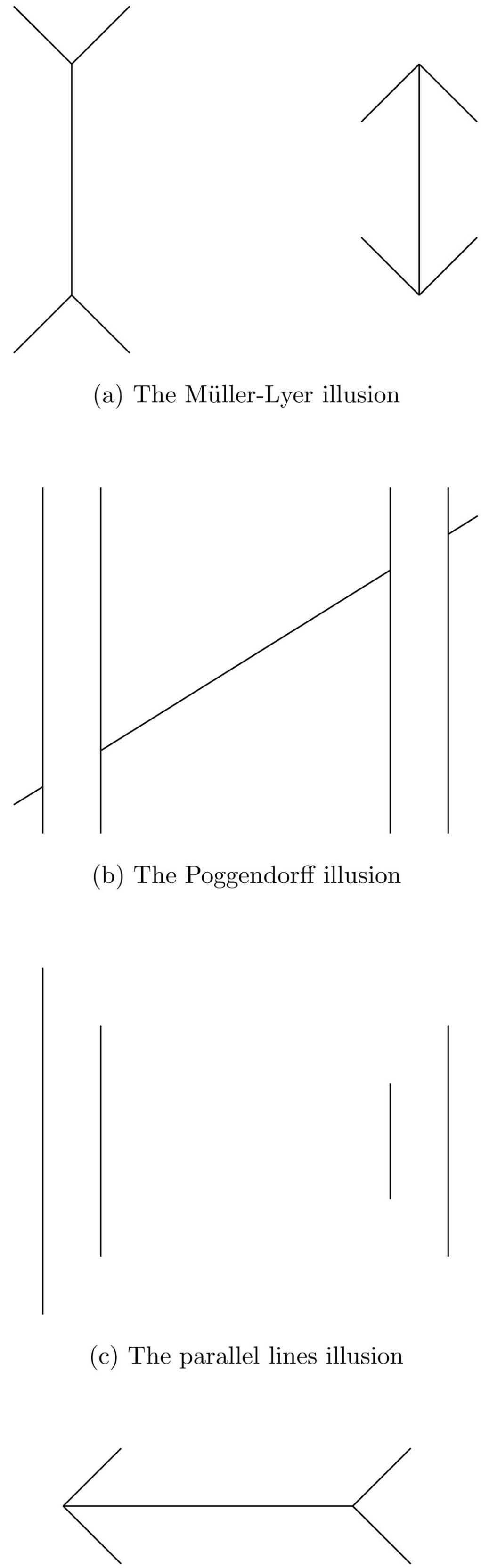
Line graphs can be regarded as object displays because the individual points are integrated into a single line, features of which (e.g., the angle between two points or its overall slope) can indicate important information about the entire data set (Carswell & Wickens, 1990, 1996). However, if the line graph also displays distinct data points (as the example in Figure 3c does), these can also be rapidly identified to facilitate tasks requiring focused attention on individual dimensions. As such, line graphs of this type may be regarded as an intermediate form combining the properties of both configural and separable displays.

Perceptual Distortion in Visual Displays

An important issue for designers is whether the choice of a particular visual display will affect the perception or interpretation of information. The perception of graphical elements in a figure can be distorted by the relationships between them (Deregowski, 1980; Schiffman, 1995) and it has been shown that people's perceptual judgments using different graphs can be affected by geometric illusions.

Figure 2 shows four examples of geometric illusions. The Müller-Lyer illusion shown in Figure 2a is a well-known example of how judgments of line length can be distorted by the acuity of angles subtended by connecting lines. In the Poggendorff illusion (Figure 2b), the diagonal lines are perceived to be misaligned when in fact they are collinear. It has been demonstrated that this illusion can make plotted lines in line graphs appear to be more orthogonal to the reading axis than they actually are (Amer, 2005; Poulton, 1985).

Distortions in the perception of line length can also be caused by a number of so-called contrast illusions, for example the "Parallel Lines" illusion (Jordan & Schiano, 1986; Schiano, 1986) in which the lengths of two parallel lines are perceived to be more similar (assimilation) or more different (contrast) than they actually are, depending on the ratio of their lengths and the distance between them. In Figure 2c, viewers typically see the lengths of the two paired lines as being more similar than they actually are. This has the effect of distorting the perceived length of the right-most line in each pair to make that on the right of the figure seem shorter than that on the left when, in fact, their lengths are the same. Zacks, Levy, Tversky, and Schiano (1998) demonstrated how this illusion can affect judgments of bar height and magnitude comparison in bar graphs.



(a) The Müller-Lyer illusion

(b) The Poggendorff illusion

(c) The parallel lines illusion

(d) The Judd illusion

Figure 2. Four visual illusions.

The distorting effect of geometric illusions may be reduced by the addition of graphical features, such as tick marks or gridlines. Amer (2005), for example, moderated the bias associated with the Poggendorff illusion in line graphs by extending the tick marks on the *y*-axis to form horizontal grid lines. However, in common with many polygon displays currently in use (e.g., Gurushanthaiah et al., 1995; Petersen et al., 1982; Woods et al., 1981), the police performance monitors lack tick marks on the radial axes. This may leave perceptual judgments of quantities, such as line length, prone to distortion by emergent features of the representation. For example, the two lines connecting a point on an axis with the two points on the adjacent axes create two triangles with the axes that may differ widely in terms of the angles and areas produced. It is possible that these emergent features may distort the perception of line length by processes similar to those involved in the Muller-Lyer illusion.

Aims of the Study

Police performance monitors were designed to allow the rapid visual comparison of an individual police authority's performance with average performance, either at a global level (i.e., an integral task to determine how much better or worse than average the authority is overall) or at the level of specific domains (a focused task). The evidence reviewed above suggests that kiviat charts should be the most appropriate representation for both of these tasks; kiviat users should be able to use the emergent properties of the polygon for the integral task and focus on the lower level features of the polygon to carry out the focused task (Bennett et al., 1997).

As with a number of other diagrams, however, it is quite likely that certain emergent features of kiviat charts may serve to distort the perception of represented values in the manner outlined above. In addition, it is possible that some configurations of values may make integration by comparison of emergent features more difficult, forcing users to switch to a focused strategy to integrate information. The two experiments reported here were designed to investigate these two issues for kiviat charts, bar graphs, and line graphs by comparing user performance on integration and focused tasks.

In the focused task of Experiment 1, participants must identify a specific target value and judge its magnitude relative to the mean. It was predicted that these judgments will be affected by perceptual distortions related to emergent features of each diagram—specifically distortions in the perception of individual values will be affected by the context created by the two values adjacent to the target as it is these two values that produce the contrast illusions in bar graphs, the differing angles and areas in the kiviat charts, and the different gradients of line segments in the line graphs.

Although specific effects of perceptual distortions may be anticipated for individual diagrams, predicting relative task performance with the diagrams is not straightforward because evidence that bar graphs facilitate focused tasks (e.g., Casey & Wickens, 1986; Wickens & Andre, 1990) is mixed, with most studies showing no statistically significant difference between display types (Bennett & Flach, 1992).

In the integral task of Experiment 2, participants must combine the information from all values to produce an overall judgment in comparison to the mean. It is in this type of task that the emergent features of configural displays have generally been found to provide a facilitation effect compared to more separate displays (Bennett & Flach, 1992). The purpose of Experiment 2 was to investigate whether this facilitation effect was found in comparison to both bar and line graphs and to determine if the effect was manifest in all task situations or whether it was specific to cases where a judgment based on the emergent features of the representation was relatively straightforward.

For example, in the police performance monitors, if one polygon is larger or more distorted relative to the other, users may be able to make a rapid judgment (i.e., that the police authority is performing generally better than average or is achieving variable levels of performance across domains). This rapid perceptual processing may be relatively easy in situations where the difference is the same for all domains (as this produces a global difference in size or shape) or if differences between domains are not too large or varied.

However, if one polygon is distorted on several dimensions, or if the distortion is sufficiently large (positive or negative) as to give the polygon a very irregular shape, then a simple visual comparison may not be possible and users may be forced to adopt an alternative strategy of performing a sequence of focused tasks comparing the two polygons dimension by dimension. This change from integral to focused strategy should be characterized by an increase in both the accuracy and latency of responses.

Experiment 1

In Experiment 1, the three diagrams were tested on the focused task to determine whether the perceptual judgment of magnitude for a particular target domain is affected by the emergent features created by the values of the adjacent domains.

Method

Design

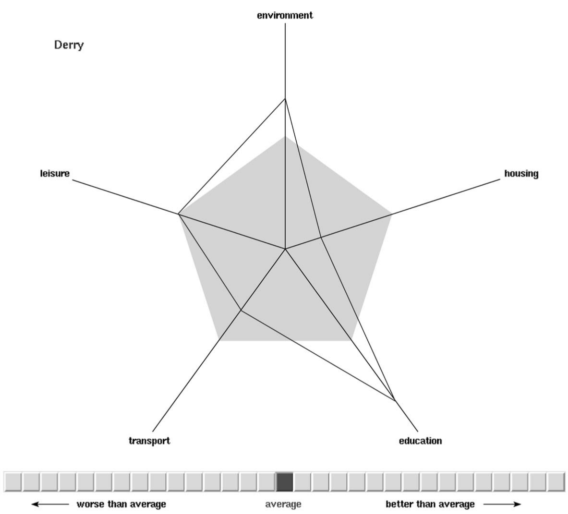
The experiment was a mixed design with one between-subjects variable and two within-subjects variables. The between-subjects variable was the type of diagram used (kiviat chart, bar graph, or line graph). The within-subjects variables were the value of the target domain that subjects were required to rate and the values of the two domains adjacent to the target domain.

Participants

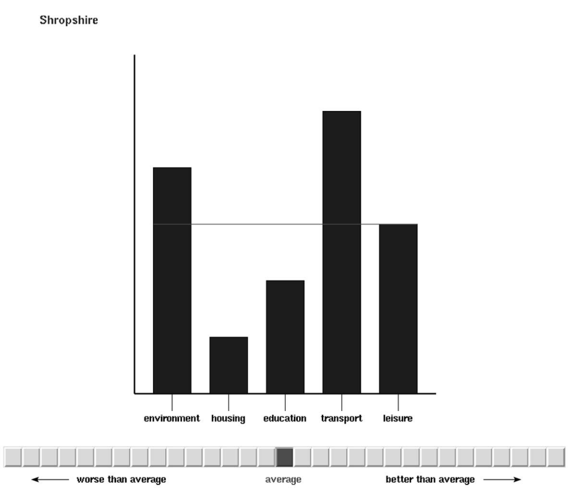
Sixty-three members of staff from the University of Huddersfield were recruited to take part in the experiment. Approximately 60% were women and occupations varied from academic, clerical, and technical positions to graduate students. The majority was from the School of Human and Health Sciences, but approximately 25% came from other schools in the university.

Materials

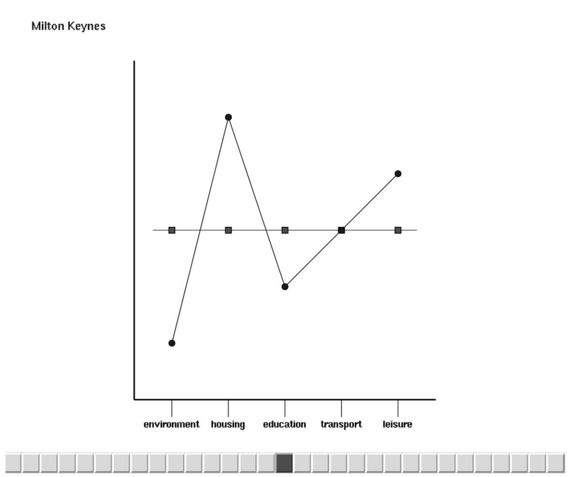
Examples of the diagrams used in the experiment are shown in Figure 3. The subject matter of the diagrams for the experiment was the (fictitious) performance of 150 United Kingdom local government authorities in five domains: housing, education, transport, leisure, and the environment, each of which is indicated by a



(a) Kiviat chart



(b) Bar graph



(c) Line graph

Figure 3. Example Kiviat chart, Bar graph, and Line graph used in Experiment 1. Below each diagram is the scale for participants' responses.

point on an axis. As in the performance monitors, the points are connected by straight lines to form a pentagon and the regular shaded pentagon represents the average performance of a set of most similar local authorities and better performance is shown further out from the center.

In order to generate a manageable range of values, the axes of the kiviat chart and the y axes of the bar graph and line graph were divided into six equally sized sections numbered 0 to 6 (although these divisions or numbers were not visible to the participants). The numbers 0 and 6 were situated at the bottom and top of the y axes and the center and outermost points of the kiviat axes respectively. Only the numbers 1 to 5 were used as target values in the experiment (henceforth referred to as T1, T2, T3, T4, and T5) and the locations of these on the diagrams can be seen in Figure 3. For example, in Figure 3a, Derry council has a housing value of 1, a transport value of 2, a leisure value of 3, an environment value of 4, and an education value of 5. The locations of the values on the y axes of the bar graph and line graph are illustrated in Figure 3b, where Shropshire council has a housing value of 1, an education value of 2, a leisure value of 3, an environment value of 4, and a transport value of 5.

The average value was the number 3 located at the center of the axes. In the bar graph, this was represented by a horizontal red line and in the line graph as the same red line with red squares as markers (to conform to the format of the line graph). In the kiviat chart, the average was represented by a red regular pentagon formed by joining the center points on the five axes. This produced a kiviat chart identical to those used in the police performance document.

Below each diagram was a scale consisting of 31 buttons. The center button in the scale was the same red color as the average marker on the diagrams and underneath it was written the word average in red. The 15 buttons on either side of the center button allowed the scale to be divided into six equally sized units, each containing four buttons. Below the scale were two arrows indicating that decreases and increases in performance were represented by buttons further to the left and right of the scale respectively.

To test the full range of target and neighboring lengths, each of the five target values was combined with the 15 possible permutations of two adjacent values (1,1; 1,2; 1,3; 1,4; 1,5; 2,2; 2,3; 2,4; 2,5; 3,3; 3,4; 3,5; 4,4; 4,5; 5,5) to create a total of 75 triplets.

In the kiviat charts, each domain axis has an adjacent domain on either side but in the bar graphs and line graphs two domains (environment and leisure) have only one adjacent domain. Therefore, to ensure that the target domain on each trial had an adjacent domain on either side, if the target value was 1, 2, 4, or 5, then the target domain was selected randomly from housing, education, and transport, as these had two adjacent values in the bar and line graphs. If the target value was 3, however, the target domain was randomly selected from all five domains. The values of the two remaining domains not adjacent to the target domain were randomly allocated a value of between 1 and 5. The experiment was conducted using three identical PC computers with 17-inch (43-cm) displays.

Procedure

Participants were randomly allocated to one of the diagram conditions. Before starting the task, participants were shown an example of the diagram they were to use and given as much time as they required to become familiar with it. The format of the example was the same for each diagram and was modeled closely on the format used in the Home Office document in Figure 1a. When participants had finished studying the example, the experimenter then explained the diagram further, highlighting the key points until they confirmed that they were sufficiently familiar with it to proceed with the experiment.

Participants were told that on each trial of the experiment their task was to judge how much better or worse than average the performance of a particular authority was on a given domain and to enter their judgment on the scale. Participants were shown the scale, instructed on how to enter their judgment, and requested to respond as rapidly but also as accurately as possible.

On each trial of the experiment, the target domain was first presented in the center of the screen for 1500 ms, after which it was removed from the screen and replaced by a diagram. As soon as the participant had clicked the mouse cursor on one of the scale buttons the diagram was removed from the screen and, after a pause of 500 ms, the next target domain was presented for a new trial. Response times were recorded from the onset of the diagram to the mouse click on a scale button. Participants saw all 75 triplets twice—a total of 150 trials—in random order and were given the opportunity to take a brief, self-terminated break after 50 and 100 trials.

Results

An initial examination of the data revealed the existence of several outlying values that were not associated with a specific participant or condition but were sufficiently abnormal to distort the mean for a specific cell. To reduce the influence of these outliers, the 42 values in each cell were standardized and those cases at the extreme end of the distribution (i.e., with a z score greater than 3.29, $p < .001$, two-tailed test) were replaced by the cell mean (Tabachneck & Fidell, 2001). From the original set of 9450 data points, this procedure resulted in the adjustment of 168 values (1.78%) from the response data and 128 values (1.35%) from the RT data. In the analyses that follow, where Mauchly's Test of Sphericity was found to be significant, the more conservative Greenhouse-Geisser corrected degrees of freedom was used. In addition, for all significant interactions, the effect size η^2 is reported and interpreted according to the conventional criteria of a small effect when $\eta^2 < .06$, a moderate effect when $.06 \leq \eta^2 < .15$ and a large effect when $\eta^2 \geq .15$ (Cohen, 1988).

In T3 trials, the target value was at the same location as the mean to which it was being compared, resulting in responses that were considerably more rapid and accurate than those in the other target conditions for all three diagrams. This demonstrates that if the task is sufficiently simple all three diagram types can be used effectively and any differences in familiarity between the diagram types does not affect performance. This was confirmed by an analysis of variance (ANOVA) on the T3 data that showed that neither response time nor accuracy were significantly affected by the diagram used, $F(2, 60) = 2.44$, $p = .10$, and $F(2, 60) = 2.14$, $p = .13$, respectively, or the values surrounding the target, $F(7.33, 439.56) = .96$, $p = .46$, and $F(6.95, 417.01) = 1.87$, $p = .07$, respectively. Therefore, because of the unique nature of this condition, T3 trials were not included in the analyses that follow.

The eight graphs in Figure 4 show the mean RT and mean difference between participants' judgments and the correct response button for target values 1, 2, 4, and 5 in each diagram as a function of the two values adjacent to the target (represented on the *x*-axis). So, for example, the left-most values in Figure 4a (labeled 1,1 on the *x*-axis) show the mean response error for the three diagrams in the condition where the target value of 1 was surrounded by two values, also 1. In the error graphs, each numbered point on the *y*-axis corresponds to a button on the response scale.

Response Times

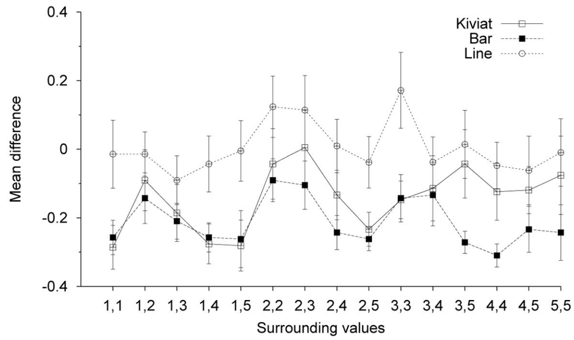
The RT data were analyzed using a three-way mixed ANOVA. The ANOVA revealed significant main effects of target value, $F(3, 180) = 8.40$, $p < .001$, $\eta^2 = .12$, and diagram type, $F(2, 60) = 4.44$, $p < .05$, $\eta^2 = .13$, and significant interactions between target value and diagram type, $F(6, 180) = 7.43$, $p < .001$, $\eta^2 = .20$, and between diagram type and surrounding values, $F(28, 840) = 1.64$, $p < .05$, $\eta^2 = .05$. A Tukey's HSD post hoc test indicated that the significant differences lay between the kiviat chart and the line graph ($p < .05$) and between the kiviat chart and the bar graph ($p < .05$).

Participants in all three diagram conditions typically took between three and five seconds to make a judgment, but kiviat users were on average slower to respond than the other diagram users for all target values. Separate ANOVAs with Tukey's HSD tests revealed, however, that these differences were only significant in the T1 condition for the bar graphs ($p < .005$) and line graphs ($p < .05$) and for the line graphs ($p < .05$) in the T2 condition. Given the similarity in RTs between the diagrams in the T3 condition, the most likely explanation for the overall slower RTs for the kiviat condition is that participants required additional time to a produce a confident estimate of the distance between the target value and the mean, particularly as the values were presented on different axes on the kiviat diagram, with only one being the same straight vertical judgment as in the other two diagrams. This may also be related to the participants' relative unfamiliarity with the kiviat diagram. It can also be seen that the variation in mean difference across the target value conditions is due to fluctuations in RTs for all diagram conditions rather than a simple change in the kiviat condition alone.

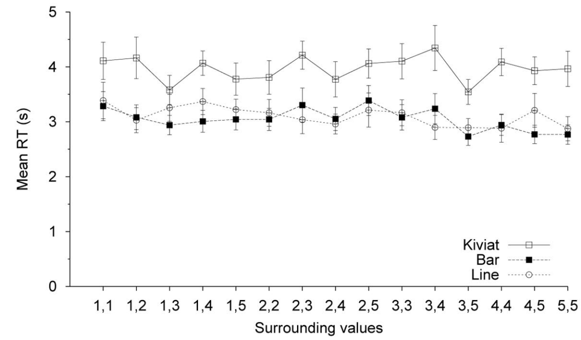
Distance Judgments

To code participants' responses, the response buttons in the experiment were numbered to reflect the underlying scale of the diagrams. A response at the extreme left of the scale was given the value 0 and each successive button was incremented by 0.2 to end at a final value of 6 at the extreme right of the scale. An error score was produced for each response by computing the difference between the participant's response and the correct response.

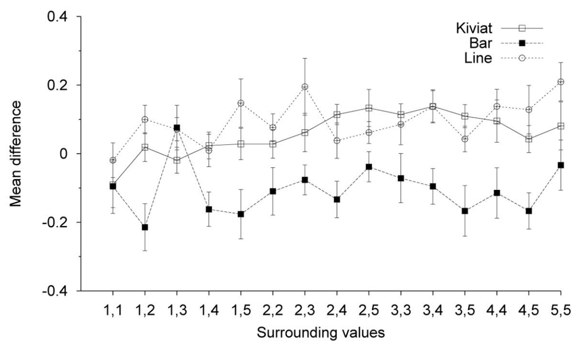
The T1, T2, T4, and T5 error data were then analyzed using a three-way mixed ANOVA. The ANOVA revealed that the diagram used by participants had a large and significant effect on their judgments, $F(2, 60) = 12.00$, $p < .001$, $\eta^2 = .29$. As with the RTs,



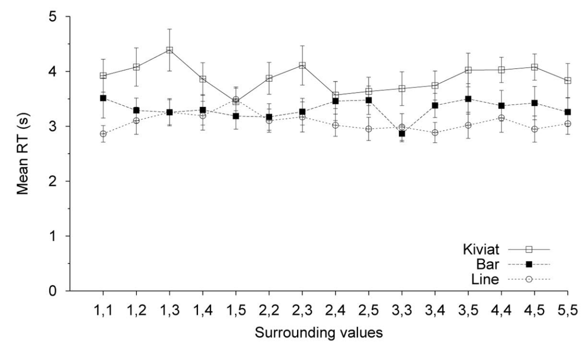
(a) T1 response error



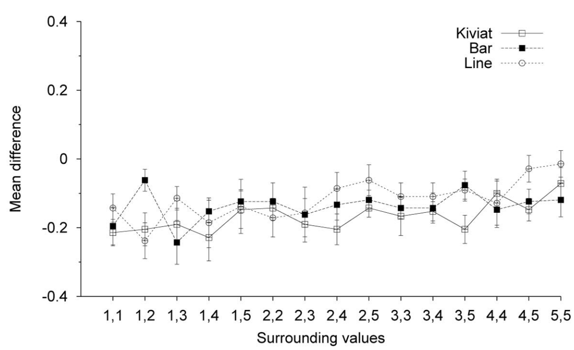
(b) T1 RTs



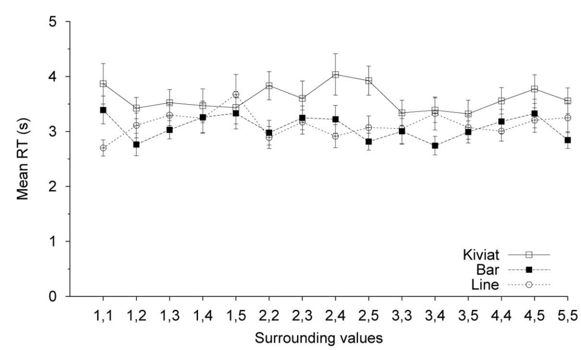
(c) T2 response error



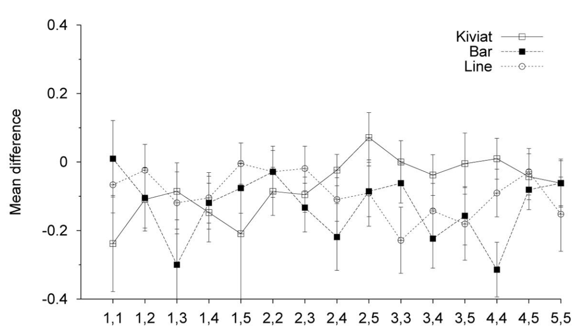
(d) T2 RTs



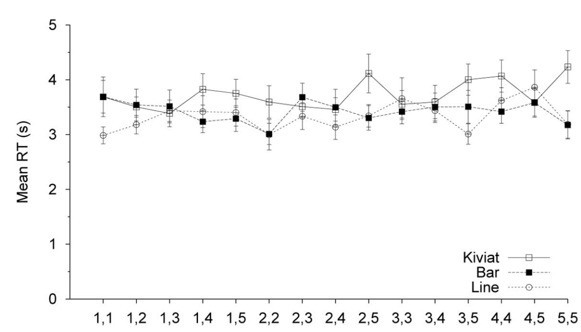
(e) T4 response error



(f) T4 RTs



(g) T5 response error



(h) T5 RTs

Figure 4. Mean response error and RT for targets 1, 2, 4, and 5 in each diagram as a function of the two values adjacent to the target, Experiment 1. Error bars indicate standard error.

this effect was not uniform across all target values but was specific to the T1 and T2 conditions. In both cases line graph users consistently perceived target values to be closer to the average ($M = 1.00$ and 2.09, respectively) than bar graph users ($M = 0.79$ and 1.89, respectively), despite the fact that the values were represented at exactly the same locations in the two diagrams. Separate ANOVAs on the T1 and T2 data confirmed that the effect of diagram type was significant and large in both cases, $F(2, 60) = 5.19$, $p < .05$, $\eta^2 = .15$, and $F(2, 60) = 12.39$, $p < .005$, $\eta^2 = .29$, respectively. Tukey's HSD post hoc tests confirmed that the difference between the bar graph and the line graph for both T1 and T2 was significant ($p < .05$ and $p < .005$, respectively) and that there was a significant difference between the bar graph and the kiviat chart in T2 ($p < .005$).

The ANOVA also revealed significant main effects of target value $F(1.60, 95.95) = 6.92$, $p < .005$, $\eta^2 = .10$, and surrounding values, $F(10.28, 616.90) = 3.39$, $p < .001$, $\eta^2 = .05$. The effect of adjacent values interacted significantly with diagram type, $F(20.56, 616.90) = 1.61$, $p < .05$, $\eta^2 = .05$, and target value, $F(18.34, 1100.14) = 1.66$, $p < .05$, $\eta^2 = .03$. A Tukey's HSD post hoc test indicated that the significant differences lay between the bar graph and the kiviat chart ($p < .05$) and between the bar graph and the line graph ($p < .005$).

To examine these differences in more detail, the minimum and maximum mean response for the four target conditions is displayed for the three diagrams in Table 1 together with their associated surrounding values and the difference between them. An example pair of each diagram is shown in Figure 5. In Figure 5, for consistency, the target value is allocated to a single domain in each diagram and the two values not adjacent to the target have been set to the mean value. Table 1 shows that users of all three diagrams produced judgments for the same target value that were significantly different solely because of the values adjacent to the target.

Discussion

The results present a complex picture. They show that participants' response accuracy and latency were both significantly affected by the diagram used, the particular target value they were attempting to assess, and the emergent feature created by the two values surrounding the target.

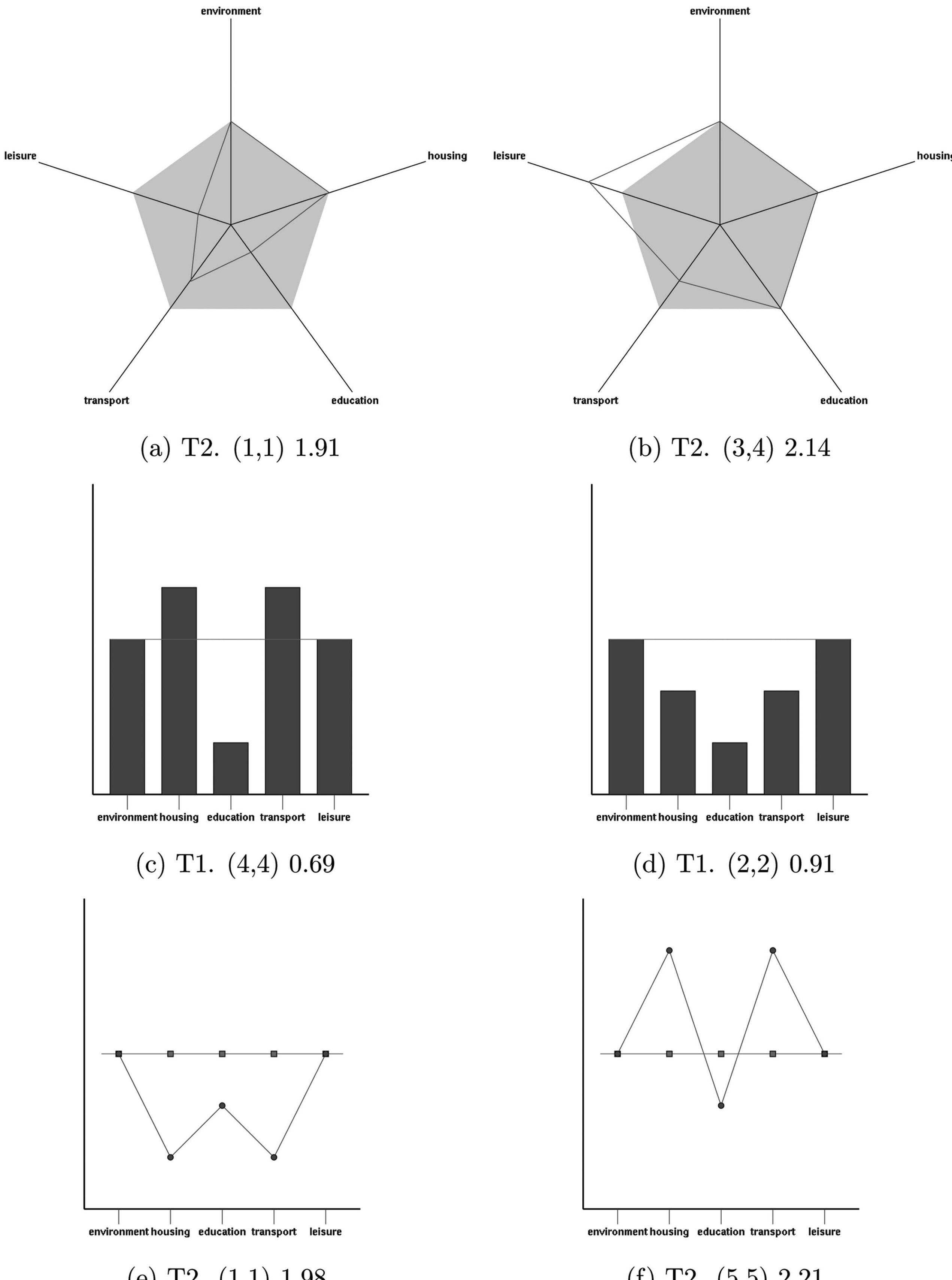
The observed differences between bar graphs and line graphs in these conditions may be surprising as previous results have shown that people tend to overestimate the length of vertical bars (Jarvenpaa & Dickson, 1988; Kosslyn, 2006). For a plausible explanation however, we can look to the visual properties of the two diagrams and how they are perceived. In bar graphs, each vertical bar is identified as a concrete object attached to and proceeding from the *x*-axis and so, when comparing the distance between the top of a bar with the mean line, participants' visual attention is drawn via a figure-ground process to the length of the bar (cf. Pinker, 1990; Simcox, 1983) in comparison to the height of the mean line, rather than to the distance between them, which accentuates the perceived difference between the top of the bar and the mean line. In contrast, the same *y*-axis value on the line graph is marked by a single point and so participants' attention is not focused on the space between the *x*-axis and the point. Line graph users, therefore, are able to compare the spaces above and below the target point more accurately, which has the effect of reducing the perception of the distance between the points on the plotted and mean lines.

This explanation is supported by the T4 and T5 response data, which do not display the same marked differences between bar and line graphs. This was confirmed by separate Tukey's HSD post hoc tests, which showed no significant difference between the graphs either for error rates (T4 $p = .86$, T5 $p = .89$) or RTs (T4 $p = .97$, T5 $p = .93$). In the T4 condition, the responses from all diagram conditions are much more uniform, with all participants judging the stimuli as being closer to the mean than the point defined by the linear response scale. For T5, the mean error of the bar graph condition (−0.13) is considerably smaller than that for the T1 condition (−0.21), although both bars are the same distance from the mean. According to the explanation above, this is because in T5 participants are still judging the length of the bar but are comparing it to the mean line below it. This is very similar to the procedure

Table 1
Extreme Responses to Target Values with Associated Adjacent Values for Each Diagram, Experiment 1

	Target 1			Target 2			Target 3			Target 4		
Response	*M*	*SD*	Adjacent	*M*	*SD*	Adjacent	*M*	*SD*	Adjacent	*M*	*SD*	Adjacent
Kiviat												
Minimum	0.71	0.29	1,1	1.91	0.30	1,1	3.77	0.31	1,4	4.76	0.64	1,1
Maximum	1.00	0.50	2,3	2.14	0.21	3,4	3.93	0.21	5,5	5.07	0.33	2,5
Difference	0.29**			0.23*			0.16**			0.31*		
Bar												
Minimum	0.69	0.91	4,4	1.79	0.31	1,2	3.76	0.29	1,3	4.69	0.37	4,4
Maximum	0.91	0.29	2,2	2.08	0.30	1,3	3.94	0.15	1,2	5.01	0.51	1,1
Difference	0.22*			0.29**			0.18**			0.32**		
Line												
Minimum	0.91	0.33	1,3	1.98	0.23	1,1	3.81	0.33	1,4	4.77	0.44	3,3
Maximum	1.17	0.51	3,3	2.21	0.26	5,5	3.99	0.18	5,5	5.00	0.28	1,5
Difference	0.26*			0.23*			0.18**			0.23**		

Note. * Significant at the .005 level.
** Significant at the .05 level.



(a) T2. (1,1) 1.91

(b) T2. (3,4) 2.14

(c) T1. (4,4) 0.69

(d) T1. (2,2) 0.91

(e) T2. (1,1) 1.98

(f) T2. (5,5) 2.21

Figure 5. Example pairs of each diagram type producing the smallest and largest responses shown in Table 1, Experiment 1. Below each diagram is the mean response and the two adjacent values (in brackets).

carried out by the line graph users in that both are judging the same distance.

Experiment 1 also demonstrated how the values surrounding a target can have a significant distorting effect on participants' perception of distance. The effect is particularly evident in Condition T5, where participants' judgments were far less regular and showed a wide range of responses for all three diagrams, but it can also be seen for other diagrams in the T1, T2, and T4 conditions.

Table 1 reveals interesting regularities in terms of which adjacent values produced extreme judgments across the three diagram types. For example, 54% involve symmetrical adjacent values, approximately equally distributed between the diagram conditions, even though symmetrical values only constitute 33% of the adjacent values possible. In addition, 33% involve the extreme values 1,1 or 5,5, which constitute only 13.3% of the adjacent values. This suggests that symmetrical or extreme adjacent values are more likely to generate emergent features that produce geometric illusions or other distorting effects.

For the kiviat diagrams, the minimum responses are all associated with small adjacent values that form relatively short connecting lines, sharper angles, and correspondingly smaller areas. In contrast, maximum responses are all formed by adjacent values producing longer connecting lines, wider angles, and larger areas. Although this may not be strictly identical to the perceptual effects found in the Muller-Lyer illusion, it is clear that a similar distortion is taking place due to the emergent features produced by the three values.

For the bar graphs, observed differences in the perceived length of target bars of the same length can be explained by the contrast and assimilation effects related to the parallel lines illusion. For example, the smallest response to T1 and largest response to T5 occur when both adjacent values produce a large contrast. The exact opposite occurs, however, for the largest response to T1 and smallest response to T5, which both occur when the two adjacent values are only one value different from the target. This produces an assimilation effect, which reduces the perceived size of the T5 target and increases the perceived size of the T1 target. It is also notable that all four of these instances involve adjacent columns with equal values.

An analysis of the extreme responses produced with line graphs revealed that 62.5% were associated with equal adjacent values, despite these constituting only 33% of the adjacent values. Equal adjacent values generate particular emergent features in line graphs—symmetrical patterns containing one or more isosceles triangles—and there are several instances in the data where the same symmetrical pattern increases or decreases the perceived distance depending on the target value concerned.

For example, the largest response to T1 and the smallest response to T5 (both of which are two units from the mean) occur when the two adjacent values are 3. Similarly, the smallest response to T2 and the largest response to T4 (again involving the same distance from the mean line) occur when both adjacent values are one value from the target and two from the mean. Finally, the largest response to T2 occurs when both adjacent values are 5. In all five cases it seems that the triangles formed by the lines between the target and adjacent points alter the perceived distance between the target and the mean by "pulling" the apex of the triangle toward the base.

These findings are consistent with studies of perceptual distortions in line bisection tasks related to geometric illusions. For example, Fleming and Behrmann (1998) have found that when asked to bisect the Judd visual illusion (Figure 2d) people's bisections are distorted in the direction opposite to which the arrows pointed (i.e., the perceived midpoint is to the right of the true midpoint in Figure 2d). This phenomenon has also been replicated in a task in which participants were shown a single dot on a page and had to draw arrows at the end of an imaginary shaft. Fleming and Behrmann (1998) suggested that these biases were caused by the arrow fins inducing a distorted perception of the length of the shaft—specifically that inward-pointing fins (i.e., the fins extending over the shaft on the left of Figure 2d) pushed the midline away from themselves while outward-pointing fins pulled the midline toward themselves.

A similar phenomenon has also been demonstrated in several studies using variants of the Muller-Lyer illusion (Green & Nelson, 1997). In a related task Shulman, Alexander, McGlinchey-Berroth, and Milberg (2002) asked participants to bisect a line drawn between the base and apex of isosceles triangles and found that the perception of the line's midpoint was distorted toward the base of the triangle. One explanation they suggested for this effect is that preattentive visual processes compute the center of mass of the triangle, which then biases the perception of the midpoint toward the base.

The current findings are consistent with both of these studies in that the direction of the perceptual distortion is away from the apex of the triangle (inward-pointing fins) and toward the base (outward-pointing fins). The data from this experiment suggest, therefore, that the isosceles triangles formed as emergent features by the target and adjacent values in the line graphs are having a similar distorting effect on participants' perception of distance.

In summary, Experiment 1 has provided a clear demonstration that even a relatively basic perceptual task, such as judging the distance between two points, can be significantly affected by the type of diagram being used and the emergent features created by surrounding variables. The size and nature of these effects is not straightforward, however, but depends on a complex interaction of several factors. For example, large and significant differences were found in the perception of the same distance for all three diagrams for particular combinations of target distance and surrounding variables. Although some regularities across the diagrams were found (e.g., the preponderance of equal surrounding values producing symmetrical patterns), the combinations producing extreme responses were often different for each diagram. In addition, the visual features producing the variation were specific to each diagram. For the kiviat charts, extreme responses are related to the lengths of connecting lines, the acuity of angles, and the size of the areas formed, whereas for bar graphs differences can be attributed to the contrast and assimilation effects related to the parallel lines illusion previously identified by others (Zacks et al., 1998). In line graphs, however, extreme responses in a number of cases seem to result from regular triangle emergent features produced by equal adjacent values that shorten the perceived distance between the target value (the triangle's apex) and the base formed by the adjacent values.

Perhaps the most interesting (and surprising) result to emerge from the study, however, is the large and significant differences in judgments found between the bar graph and line graph users, which clearly shows how attention to the graphical elements in a diagram can affect perception of the quantities represented. This has important implications for graph designers, particularly relating to the common practice of combining multiple graphs by superimposing bars and lines in the same Cartesian coordinate system, often referred to as a mixed, composite or overlay graph (Harris, 1993).

Applying the results of this experiment to advocate the use of one particular diagram over another is problematic as all three are susceptible to perceptual distortion given the appropriate combination of factors. However, comparing participants' performance between graph conditions and within each graph, it does seem that for this simple comparison task participants using the line graphs produced the most consistent and accurate judgments overall. This is contrary to the usual practice of using clearly separable displays, such as bar graphs, for focused tasks. However, given the general underestimation of bar height and perceptual distortions revealed by this experiment, it could be argued that bar graphs, particularly without the additional visual cues provided by tick marks, may not be the most effective format for this particular purpose and that, as the line graphs used in the experiment combine separable points with a single integrating line, this format may prove to be appropriate for both focused and integral tasks.

Experiment 2

In this second experiment, participants were required to integrate information from all five domains to compare the overall performance of an individual police authority with the global average. It is in this type of integral task that the benefits of configural displays have previously been most apparent as the emergent features of the representation can be utilized to assess the overall difference based on the polygon shape formed by the variable values (Bennett & Flach, 1992). This is not the only way that this task can be performed with configural displays, however (Harris, 1993). An alternative strategy is to carry out a sequence of focused subtasks comparing each domain in turn while maintaining a running total of the overall difference in working memory.

Given that the primary representational feature of object displays is that of area, one may wonder in which situations this alternative strategy may be more appropriate. The difficulty of comparing the two areas can vary, however, depending on the values concerned. If one pentagon is clearly larger or smaller than the other on all five axes, then perceptual comparison is relatively straightforward. If no individual pentagon is larger than the other on all axes, however, the two polygons will intersect one another, making a global area comparison more difficult. In this case, users may tend to adopt the sequential focused strategy, even though it may be more cognitively demanding. It is likely that the integral comparison strategy will provide a more rapid—but possibly less accurate—judgment than the sequential focused one.

Although it is possible for different strategies to be adopted with all three diagrams, it is likely that the representational properties of each diagram will encourage a particular method. For example, because area is closely related to the domain values in kiviat charts, it is more likely that area comparison will be a prominent strategy adopted by kiviat users—at least in favorable circumstances. In bar graphs, however, because each domain value is represented as a separate unit with a specific height, one can assume that this will facilitate a strategy of systematically comparing each domain in turn. An alternative strategy for bar graphs might involve a comparison of the amount of space filled by the bars above and below the mean line, whereas in the line graphs a similar impression could be formed by comparing the extent of the plotted line or the size of the areas immediately above and below the mean line.

Because the line graphs depict individual data points on the plotted line, identifying these locations should be relatively straightforward (compared to similar line graphs without distinct data points). However, because people are more likely to regard line graphs as configural displays (Carswell & Wickens, 1990; Zacks & Tversky, 1999) and typically encode them in terms of their slope (Culbertson & Powers, 1959; Simcox, 1983), they are required to reorganize the given line pattern into a different set of perceptual units (i.e., a set of points) by perceptually locating each point on the line (Kosslyn, 2006); a process requiring additional attention and cognitive effort compared to bar graphs. We can assume that this process becomes increasingly difficult and time consuming as the complexity of the line graph increases.

The purpose of Experiment 2 was to produce a set of stimuli that would vary in terms of their graphical complexity in order to determine whether this would affect the strategy adopted by users of each diagram type. Based on previous results and the proximity compatibility principle, one should predict that the high display proximity of the kiviat chart will facilitate this integral task as it requires a high degree of mental proximity.

However, if the analysis above is correct, the relative sizes of—and amount of overlap between—the two pentagons in the kiviat charts should vary the appeal and effectiveness of the integral comparison strategy, which will in turn affect the speed and accuracy of users' judgments. Kiviat users should produce more rapid responses as the target pentagons increase or decrease in size (i.e., the global target value increases or decreases) away from the mean. As the global target value approaches the mean, the more the pentagons will overlap and the more likely kiviat users will be required to use the slower but potentially more accurate domain by domain comparison strategy. This pattern of responses should not be seen in the bar graph condition, however, because of the bar graph's facilitation of the sequential comparison strategy, but may be evident with the line graphs because of the additional processing required to extricate the points from the line.

Method

Design and Participants

Experiment 2 was a mixed design with one between-subjects variable (the type of diagram used: kiviat, bar, or line) and one within-subjects variable (the global value of the target domains that subjects were required to judge). Fifty-one students from the University of Huddersfield who volunteered to take part in the experiment were randomly assigned to one of the three diagram conditions. The majority of participants (approximately 80%) were women and, apart from one graduate student, all were studying for undergraduate degrees in the School of Human and Health Sciences.

Materials

The stimuli used were the same as those used in Experiment 1. Stimuli were also coded similarly so that the mean was given the value 3 and domain values ranged from 1 to 5. Based on this value system, a local authority scoring the mean value of 3 on all five domains would have an overall value of 15, whereas one scoring 5 on all domains would total 25. The 3125 permutations of the five possible values of the five domains reduce to 126 unique combinations of five values, each combination summing to a value between 5 and 25. The response scale was therefore reduced to 21 buttons to reflect this new range of responses. As in Experiment 1, the red center button in the scale had the word "average" written underneath it in red and below the scale were two arrows indicating the direction of increases and decreases in performance. For each participant, 150 stimuli were generated by selecting values for the five domains at random.

Procedure

The experiment procedure was identical to that of Experiment 1 with the exception that instead of being presented with a target domain at the beginning of each trial as in Experiment 1, participants were presented with the word "overall" to remind them that they were required to make a global comparison between the local authority's values and the mean.

Results

Response accuracy was measured by computing for each response the absolute difference between the participant's judgment and the true deviation from the average represented in the diagram. As in the previous experiment, the response accuracy and latency data were scanned for extreme outlying values with a z score greater than 3.29, resulting in 50 values (0.65%) from the accuracy data and 98 values (1.28%) from the RT data being replaced by the cell mean. In addition, because the stimuli were generated by randomly selecting values for the five domains, the probability of the different totals occurring in the experiment is not equal but ranges from .00032 for the extreme values 5 and 25 (which only occur by the addition of the five specific values 1,1,1,1,1 and 5,5,5,5,5, respectively), to .122 for the total 15 which can be produced from 12 different combinations (e.g., 1,1,3,5,5; 2,2,2,4,5; 3,3,3,3,3; etc.). Because the values at the extreme ends of the range had been presented relatively infrequently, when analyzing the data, only those target values having a minimum of 47 data points per cell (values 9–21 inclusive) were included in the analysis as this provided a symmetrical set of cells with sufficient data points to produce a stable mean.

Figures 6 and 7 present the mean RT and mean absolute differences respectively for each diagram type as a function of global target size. The two graphs show a clear distinction between the diagram types in terms of the variability in both measures and reveal that participants' responses in all three graph conditions differed depending on the global size of the target. Two-way mixed ANOVAs revealed that global target size had a significant effect on RT, $F(7.68, 368.42) = 12.86$, $p < .005$, $\eta^2 = .21$, and the mean absolute difference between the participant's judgment and the actual target size, $F(2.13, 102.39) = 465.21$, $p < .005$, $\eta^2 = .91$. For the RT data the ANOVA also revealed a significant effect of diagram type, $F(2, 48) = 4.19$, $p < .05$, $\eta^2 = .15$ and a significant interaction between diagram and target size, $F(15.35, 368.42) = 2.56$, $p < .005$, $\eta^2 = .10$. A Tukey's HSD post hoc test confirmed that the significant differences lay between the bar graph and the kiviat chart ($p < .05$).

Discussion

A notable feature of Figures 6 and 7 is the symmetrical nature of the response profiles produced by all three graph conditions. These profiles are related to the graphical patterns that produce the global target values for each diagram. As global target values decrease or increase in size from 15, the proportion of points on the

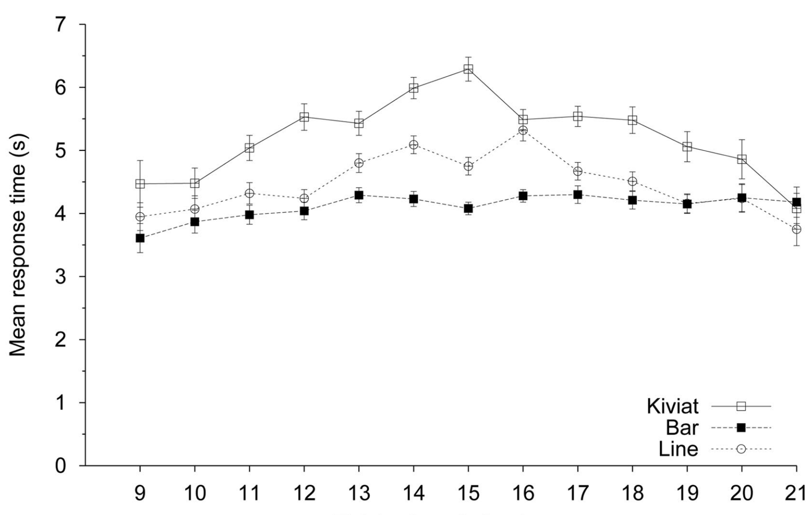


Figure 6. Mean response time (s) plotted as a function of global target size for kiviat, bar and line graphs, Experiment 2. Error bars indicate standard error.

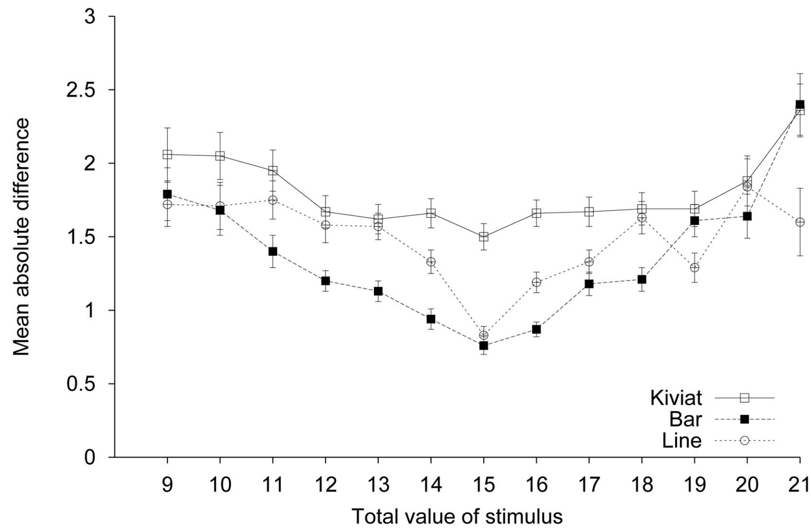


Figure 7. Mean absolute difference between response and global target size plotted as a function of global target size for kiviat, bar, and line graphs, Experiment 2. Error bars indicate standard error.

diagram below or above the mean line (i.e., inside or outside of the mean pentagon shape in the kiviat chart) respectively increases, as does the average distance from the mean of each individual domain value. For example, if the global target is 9, of the 5 possible ways of producing this total, only 2 have a value greater than 3 (1,1,1,1,5 and 1,1,1,2,4) and the majority of domain values are the furthest distance from the mean (i.e., close to the *x*-axis or the center of the kiviat chart). This situation is mirrored for target value 21, although in this case the majority of domain values are outside the mean pentagon. The proportions of values below and above the mean line steadily even out as one moves toward the central value of 15, which can be produced from 12 different combinations but which has an equal number of domain values above and below the mean overall, many more of which are closer to the mean.

This relationship explains the steady increase in RTs toward the 15 total found in the kiviat condition. If kiviat users are making judgments by comparing the areas of the two pentagons, it is easier to do so if one is consistently smaller than another on all or most domains. As the global target value moves to the central value of 15, however, the likelihood that the pentagons will overlap on several domains increases, making a simple area comparison more difficult and requiring users to carry out a slower (but potentially more accurate) domain by domain comparison. A similar, although much less marked, pattern is also found in the line graph condition, most likely for the same reason—with the movement of global target values toward the central value of 15, the plotted line is more likely to cross the mean line more often, creating a more jagged and complex pattern for the user to reorganize into the appropriate perceptual units, thereby increasing the response time.

In contrast, the bar graph Condition RTs are relatively flat and unaffected by global target size (as revealed in the significant interaction between diagram type and target size). Because bar graphs are more readily encoded in terms of their height (Simcox, 1983) and interpreted as representing separate values (Kosslyn, 2006; Zacks & Tversky, 1999), comparison of individual quantities is easier (Culbertson & Powers, 1959; Zacks & Tversky, 1999). In the context of this experiment, therefore, bar graphs facilitate the rapid systematic comparison of each domain in turn, a process that is not made more difficult or time consuming by bringing the values closer to the mean.

These same representational factors also account for the response accuracy data presented in Figure 7. In all three diagram conditions (although to a widely varying extent), response accuracy improves as the global target size moves toward the central value of 15. To explain this, one needs to look at the distribution of domain values for those global target sizes closer to 15. As the global target size moves toward the central value of 15, the proportion of domain values at or close to the mean increases, reducing the average distance to be judged. This can be seen as simplifying the task for all three diagram users if they are carrying out the sequential comparison strategy. The most marked improvement can be seen in the bar graph condition and to a lesser extent the line graph condition, with the least improvement occurring with the kiviat graphs. In the kiviat condition the increased amount of overlap between the pentagons may force users to compare each target domain with the mean in turn which will be slower but more accurate than comparing the pentagons' areas. An additional factor in the response accuracy improvement for target 15 should be noted: that is, if participants judge that there is no difference between the global value of the target and the mean, then the appropriate response button is clearly identified on the scale, reducing the amount of error in the response process.

General Discussion

The comprehension and interpretation of diagrams involves a complex interaction between three components: the cognitive abilities and limitations of the user, the visual and computational properties of the graphical representation, and the nature of the task being undertaken—the so-called "cognition-artifact-task triad" (Gray & Altmann, 2001). In attempting to understand interactive behavior with diagrams, it is essential that all three elements be taken into account as each constrains the others in a reciprocal relationship. This study has revealed the complexity of this relationship in two relatively simple but realistic tasks. Experiment 1 revealed the complex relationship between graphical representation, cognition, and task by demonstrating the significant distortions in distance perception that can occur for focused tasks using all three diagrams. The type of distortion was different for each diagram and related to the emergent features of each. In the kiviat chart, it was related to the size of the angles and areas produced by the lengths of the lines connecting adjacent values, whereas in the bar graph it was related to the contrast and assimilation effects of the parallel lines illusion. In the line graph the perception of the distance between the triangle's apex and its base was reduced when symmetrical triangles were formed by the lines between the target and adjacent points.

In addition to these within-diagram differences, a large global difference was found between the bar and line graphs for some target values. Previous studies have revealed differences in people's conception and interpretation of bar and line graphs. Experiment 1 showed that people's actual perception of the quantities depicted by bars and lines can also differ significantly. Bar chart users systematically underestimated distances compared to line graph users. The data suggest that this is because attention is being drawn to the bar itself, whereas in line graphs attention is directed to the point on the line. This is an important finding that counterbalances previous studies showing that people overestimate the length of vertical bars (Jarvenpaa & Dickson, 1988; Kosslyn, 2006) and suggests that biases in distance perception are being introduced by the object, perhaps by drawing attention to the rectangle's center of mass. The findings also have real-world implications, for example, in diagrams where bars and lines are superimposed on the same axes, where this distortion may have a significant effect on the interpretation of data.

The focus of Experiment 2 was the research over the last two decades into display and mental proximity, the proximity compatibility principal and, in particular, the superiority of configural displays for integral tasks. The general facilitation for integral tasks provided by configural displays has been widely investigated (e.g., Barnett & Wickens, 1988; Bennett & Flach, 1992; Bennett et al., 1993; Wickens & Carswell, 1995) and is generally regarded as coming from the closeness of the mapping between the emergent features of the display, the underlying relationships between the data being represented, and the properties of the data that are relevant for the task (Bennett & Flach, 1992; Wickens & Carswell, 1995).

One such property is the extent to which variable values correlate with themselves over time (auto-correlation) or with the values of other variables (cross-correlation). A previous comparison of configural and separable displays (pentagons and staggered bar graphs respectively) on an integral task (Jones, Wickens, & Deutsch, 1990) found an overall benefit of the configural display but also that this benefit was greatest when the information had low levels of auto-and cross-correlation (i.e., the variables were uncorrelated and varied randomly overtime). Jones et al. (1990) argued that this was because the users' task was simplified by the prior integration provided by the object display and also because pentagons that vary greatly in terms of shape are more perceptually salient than bars of different heights.

The results of Experiment 2 do not support those of Jones et al. (1990), however. The stimuli for Experiment 2 were generated by randomly sampling a wide range of variable values and so have a low degree of auto-and cross-correlation. Although this produced pentagons with a wide range of shapes that may have been more perceptually salient than bars of different heights, the integration they provided was insufficient to produce a performance advantage compared to the more separable line and bar graphs. Rather, there was a general disadvantage found with the configural display and the symmetrical patterns in the RT and error data suggest that performance was affected by the degree of overlap between the target and the mean.

The data show that where there is little overlap (i.e., one pentagon is globally larger than the other), comparison can be made by relatively rapid (but potentially more error-prone) perceptual processes. However, the generally poorer performance of kiviat users is consistent with previous psychophysical studies (e.g., Cleveland & McGill, 1984) showing judgments of area to be less accurate than other perceptual judgments, such as position, along a common scale or nonaligned scales, length, direction, or angle. Where there is greater overlap between the pentagons, accurate judgments may be better made by a slower point-by-point comparison and maintenance of a running total. Although these symmetries are also visible in the RT data for the line and bar graphs, they are much less pronounced (particularly in the latter), suggesting that they both promote a strategy involving a sequence of focused comparison tasks.

The results of Experiment 2 are important, therefore, because they provide strong evidence against the widely held view that object displays are most appropriate for integral tasks. In addition, it is commonly assumed that users of object displays will use their emergent properties for integration tasks and their specific lower level properties for focused tasks. Experiment 2 has provided evidence to suggest that users will strategically switch between the configural and specific properties for the same integral task depending on the nature of the diagram being studied.

It is an interesting question whether, and to what extent, the results of the experiments are determined by differences in user familiarity. This question touches on the more general issue for graph designers concerning the adoption of relatively unfamiliar representation formats. Using familiar representations allows the target audience to employ previously learned procedures and strategies to retrieve information and reduces the cognitive effort required to learn how a new diagram represents information and the procedures for extracting that information.

It may be argued that a novel display should only be considered if its particular representational properties or emergent features provide the best mapping to the data or if its computational properties are most appropriate for the interpretive task. Previous studies have shown that if an unfamiliar graph is most appropriate for a task, once having learnt the representation, users are able to retrieve information and solve certain problems significantly faster than users of the more familiar form (Peebles & Cheng, 2001, 2002, 2003). Although kiviat charts are relatively infrequently used, it does not seem that they are too unusual or difficult for their properties to be learned rapidly and to be used effectively (anecdotally, participants in the kiviat condition in both experiments showed no difficulty in initially comprehending the diagram). It may be the case that relative lack of familiarity had some effect on performance (particularly response times), but the lack of a general difference in error scores across all target conditions in Experiment 1 suggests that unfamiliarity was not a dominant factor.

A second issue to be considered in relation to these findings is whether the distortions found would be significantly reduced by the addition of tick marks. In contrast to many other diagrams, there seems to be no established best practice with regard to the inclusion of tick marks in object displays as many studied in the literature (e.g., Carswell & Wickens, 1987; Casey & Wickens, 1986; Coury et al., 1989; Hughes & MacRae, 1994; Jones et al., 1990; Sanderson et al., 1989) or currently in use do not contain them, including the two dynamic object displays for nuclear power plants (Petersen et al., 1982; Woods et al., 1981) and anesthesia machines (Gurushanthaiah et al., 1995) mentioned earlier.

In many situations, perfect representational accuracy may not be the purpose of diagrams (Cleveland & McGill, 1984; Ehrenberg, 1975) and it may be the case that performance monitors were designed primarily to support "quick and dirty" perceptual comparisons. Given the psychophysical evidence of the relative inaccuracy of judgments of angle or area compared to judgments of length (Cleveland & McGill, 1984) however, there are good reasons to include tick marks, even if no numerical values are associated with them, to demarcate the distances to be compared.

It is likely that the anchoring effect of tick marks would reduce significantly, if not eliminate entirely, the perceptual distortions found in all three diagrams because comparing values using tick marks requires a different set of perceptual and cognitive operations from those involved in these experiments (e.g., Lohse, 1993; Peebles & Cheng, 2003).

Although bar or line graphs without tick marks are less common in everyday situations (the performance monitor bar charts being an interesting and informative exception), several previous studies have employed these diagrams without tick marks (e.g., Cleveland & McGill, 1984; Schiano & Tversky, 1992; Tversky & Schiano, 1989; Zacks et al., 1998; Zacks & Tversky, 1999). Most relevant to this study is the previous demonstration of contrast effects in bar graphs that used sparse, content-free bar graphs without tick marks (Zacks et al., 1998). The results of the current study are important as they provide strong support for the contrast effects found by Zacks et al. (1998) using a realistic task and more naturalistic bar graphs adapted from a real-world example.

Given the goal of presenting police performance data to a general audience, it is an interesting question whether this study can provide any guidance as to which representation would be most appropriate. If users are comparing individual domains, then Experiment 1 would suggest that the line graph provides a representation that is least susceptible to perceptual distortion and provides the most consistent and accurate judgments over the range of contexts. This goes against the general consensus, however, that line graphs should not be used to present values of separate discrete entities, such as the domains in the performance monitors (Kosslyn, 2006; although Zacks and Tversky, 1999, argue that this rule may be as much a result of communicative convention as of any cognitive or perceptual biases). One important factor in these line graphs that makes them particularly suitable for focused tasks is the plotting of distinct data points on the lines as these can be rapidly identified, giving line graphs features of both configural and separable displays.

If, however, users are comparing the global target value with the mean, then Experiment 2 suggests that bar graphs provided the most consistent and accurate judgments over the range of contexts, although the large degree of variance in responses produced by the different global target values in all three conditions reduced the statistical significance of the differences between the diagrams. If the distortions found with the bar graphs in the focused task are reduced or eliminated by tick marks, it is likely that the most consistent and accurate comparisons for both individual and global measures would be provided by bar graphs with tick marks.

Although the weight of empirical evidence supports the use of configural displays for the goal of comparing multivariate data, the decision of the U.K. Government to present police performance results in this form was still a bold one given their relative unfamiliarity. This study has taken the opportunity provided by this decision to address important questions concerning the distorting effects of emergent features and the extent of the benefit of configural displays for integral tasks. Taken together, the results of the two experiments reveal the complex nature of emergent features and the powerful interactions that occur between the visual and computational properties of information graphics, the specific requirements of the task being undertaken, and the perceptual and strategic characteristics of users. The results suggest that although all three diagrams significantly distorted perception, taken as a whole, kiviat charts were in fact the least successful diagram to use for both tasks. In addition, the results suggest that the general consensus of the superiority of configural displays for integration tasks is in need of further analysis and reconsideration.

References

Amer, T. S. (2005). Bias due to visual illusion in the graphical presentation of accounting information. *Journal of Information Systems, 19*(1), 1–18.

Barnett, B. J., & Wickens, C. D. (1988). Display proximity in multicue information integration: The benefit of boxes. *Human Factors, 30*(1), 15–24.

Bennett, K. B., & Flach, J. M. (1992). Graphical displays: Implications for divided attention, focused attention and problem solving. *Human Factors, 34*(5), 513–533.

Bennett, K. B., Nagy, A. L., & Flach, J. M. (1997). Visual displays. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (3rd ed., pp. 1191–1221). New York: Wiley.

Bennett, K. B., Toms, M. L., & Woods, D. D. (1993). Emergent features and graphical elements: Designing more effective configural displays. *Human Factors, 35*(1), 71–97.

Carswell, C. M., & Wickens, C. D. (1987). Information integration and the object display: An interaction of task demands and display superiority. *Ergonomics, 30*(3), 511–527.

Carswell, C. M., & Wickens, C. D. (1990). The perceptual interaction of graphical attributes: Configurality, stimulus homogeneity, and object interaction. *Perception & Psychophysics, 47,* 157–168.

Carswell, C. M., & Wickens, C. D. (1996). Mixing and matching lower-level codes for object displays: Evidence for two sources of proximity compatibility. *Human Factors, 38*(1), 1–22.

Casey, E. J., & Wickens, C. D. (1986). *Visual display representation of multidimensional systems: The effect of information correlation and display integrality* (Tech. Rep. No. CPL-86-2/MDA903-83-K-0255). Cognitive Psychophysiology Laboratory, Department of Psychology: University of Illinois, Champaign.

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association, 79,* 531–554.

Cleveland, W. S., & McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. *Science, 229,* 828–833.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Erlbaum.
Coury, B. G., Boulette, M. D., & Smith, R. A. (1989). Effect of uncertainty and diagnosticity on classification of multidimensional data with integral and separable displays of system status. *Human Factors, 31*(5), 551–569.
Culbertson, H. M., & Powers, R. D. (1959). A study of graph comprehension difficulties. *Audio-Visual Communication Review, 7,* 97–110.
Deregowski, J. B. (1980). *Illusions, patterns, and pictures.* London: Academic Press.
Ehrenberg, A. S. C. (1975). *Data reduction: Analyzing and interpreting statistical data.* New York: Wiley.
Fleming, J., & Behrmann, M. (1998). Visuospatial neglect in normal subjects: Altered spatial representations induced by a perceptual illusion. *Neuropsychologia, 36*(5), 469–475.
Gray, W. D., & Altmann, E. M. (2001). Cognitive modeling and human-computer interaction. In W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors* (pp. 387–391). New York: Taylor & Francis Ltd.
Green, E., & Nelson, B. (1997). Evaluating Muller-Lyer effects using single fin-set configurations. *Perception & Psychophysics, 59*(2), 293–312.
Gurushanthaiah, K., Weinger, M. B., & Eglund, C. E. (1995). Visual display format affects the ability of anesthesiologists to detect acute physiologic changes. *Anesthesiology, 83,* 1184–1193.
Harris, R. L. (1993). *Information graphics: A comprehensive illustrated reference.* Atlanta, GA: Management Graphics.
Hughes, T., & MacRae, A. W. (1994). Holistic peripheral processing of a polygon display. *Human Factors, 36*(4), 645–651.
Jarvenpaa, S. L., & Dickson, G. W. (1988). Graphics and managerial decision making: Research-based guidelines. *Communications of the ACM, 31*(6), 764–774.
Jones, P. M., Wickens, C. D., & Deutsch, S. J. (1990). The display of multivariate information: An experimental study of an information integration task. *Human Performance, 3*(1), 1–17.
Jordan, K., & Schiano, D. J. (1986). Serial processing and the parallel-lines illusion: Length contrast through relative spatial separation of contours. *Perception & Psychophysics, 40,* 384–390.
Kosslyn, S. M. (2006). *Graph design for the eye and mind.* New York: Oxford University Press.
Lohse, G. L. (1993). A cognitive model for understanding graphical perception. *Human Computer Interaction, 8,* 353–388.
Peebles, D. (2004). Distortions of perceptual judgement in diagrammatic representations. In K. Forbus, D. Gentner, & T. Reiger (Eds.), *Proceedings of the twenty-sixth annual conference of the cognitive science society.* Mahwah, NJ: Erlbaum.
Peebles, D., & Cheng, P. C. H. (2001). Graph-based reasoning: From task analysis to cognitive explanation. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the twenty-third annual conference of the cognitive science society* (pp. 762–767). Mahwah, NJ: Erlbaum.
Peebles, D., & Cheng, P. C. H. (2002). Extending task analytic models of graph-based reasoning: A cognitive model of problem solving with Cartesian graphs in ACT-R/PM. *Cognitive Systems Research, 3,* 77–86.
Peebles, D., & Cheng, P. C. H. (2003). Modeling the effect of task and graphical representation on response latency in a graph reading task. *Human Factors, 45,* 28–45.
Petersen, R. J., Banks, W. W., & Gertman, D. I. (1982). Performance-based evaluation of graphic displays for nuclear power plant control rooms. In J. A. Nichols & M. L. Schneider (Eds.), *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 182–189). Gaithersburg, MD: ACM Press.
Pinker, S. (1990). A theory of graph comprehension. In R. Freedle (Ed.), *Artificial intelligence and the future of testing* (pp. 73–126). Hillsdale, NJ: Erlbaum.
Police Standards Unit (2003). Police Performance Monitoring 2001/02. London: Home Office.
Police Standards Unit (2004). Police Performance Monitoring 2002/03. London: Home Office.
Pomerantz, J. R., & Pristach, E. A. (1989). Emergent features, attention and perceptual glue in visual form perception. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 635–649.
Poulton, E. C. (1985). Geometric illusions in reading graphs. *Perception & Psychophysics, 37,* 543–548.
Sanderson, P. M., Flach, J. M., Buttigieg, M. A., & Casey, E. J. (1989). Object displays do not always support better integrated task performance. *Human Factors, 31*(2), 183–198.
Schiano, D. J. (1986). Relative size and spatial separation: Effects on the parallel-lines illusion. *Perceptual and Motor Skills, 63,* 1151–1155.
Schiano, D. J., & Tversky, B. (1992). Structure and strategy in encoding simplified graphs. *Memory & Cognition, 20,* 12–20.
Schiffman, H. R. (1995). *Sensation and perception: An integrated approach* (4th Ed.). New York: Wiley.
Schutz, H. G. (1961). An evaluation of formats for graphic trend displays—Experiment II. *Human Factors, 3,* 99–107.
Shulman, M. B., Alexander, M. P., McGlinchey-Berroth, R., & Milberg, W. (2002). Triangular backgrounds shift the bias of line bisection performance in hemispatial neglect. *Journal of Neurology, Neurosurgery & Psychiatry, 72,* 68–72.
Simcox, W. A. (1983). *A perceptual analysis of graphic information processing.* Unpublished doctoral dissertation, Tufts University, Medford, MA.
Tabachneck, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th Ed.). Needham Heights, MA: Allyn & Bacon.
Treisman, A. M. (1996). Properties, parts, and objects. In K. Boff, L. Kaufmann, & J. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2, pp. 1–70). New York: Wiley.
Tversky, B., & Schiano, D. J. (1989). Perceptual and conceptual factors in distortions in memory for graphs and maps. *Journal of Experimental Psychology: General, 118,* 387–398.
Wickens, C. D., & Andre, A. D. (1990). Proximity compatibility and information display: Effects of color, space and objectness on information integration. *Human Factors, 32*(1), 61–77.
Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors, 37*(3), 473–494.
Woods, D. D., Wise, J. A., & Hanes, L. F. (1981). An evaluation of nuclear power plant safety parameter display systems. In R. C. Sugarman (Ed.), *Proceedings of the 25th annual meeting of the human factors society* (pp. 110–114). Santa Monica, CA: Human Factors Society.
Zacks, J., Levy, E., Tversky, B., & Schiano, D. J. (1998). Reading bar graphs: Effects of extraneous depth cues and graphical context. *Journal of Experimental Psychology: Applied, 4,* 119–138.
Zacks, J., & Tversky, B. (1999). Bars and lines: A study of graphic communication. *Memory & Cognition, 27*(6), 1073–1079.

Received February 7, 2007
Revision received December 20, 2007
Accepted December 20, 2007 ■