Distortions of perceptual judgement in diagrammatic representations

David Peebles (D.Peebles@hud.ac.uk)
Department of Behavioural Sciences, University of Huddersfield,
Queensgate, Huddersfield, HD1 3DH, UK

Abstract

An experiment is reported which investigates the distorting effects of various graphical features in three different diagrammatic representations of the same information. The experiment revealed significant distortions in users' perceptual judgements of distance both between the different diagrams and within each diagram. The results of the experiment are interpreted as indicating the crucial role of anchor points such as axis tick marks along dimensions.

Introduction

Presenting large, complex data sets to a broad and non-specialist audience is a challenging task if one is to be sure that the information is to be interpreted in an appropriate manner. Choosing the most suitable representation for a particular communicative goal is an important aspect of the task and designers of information artefacts must be aware of how the low-level visual features of individual graphical representations can facilitate or hinder the interpretation of information.

Many studies have shown that the perception of one or more graphical elements in a figure can be distorted by the relationships between them (see, e.g., Deręgowski, 1980; Schiffman, 1995). A famous example is the Müller-Lyer illusion shown in Figure 1a which illustrates how perceptual judgements of line length can be distorted by the acuity of angles subtended by connecting lines. In this illusion, the line on the right of Figure 1a is perceived to be shorter than that on the left, although their lengths are actually the same.

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 viewers of two parallel lines can perceive the lengths of the lines to be closer than they actually are (assimilation) or more different than they actually are (contrast), depending on the ratio of the line lengths and the distance between them. An example is shown in Figure 1b in which the perceived lengths of the lines in each pair are distorted by the length of the line next to it so that viewers 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.

Previous research has shown that visual illusions can have a strong effect on people's perceptual judgements in commonly used diagrams (e.g., Poulton, 1985). Zacks, Levy, Tversky and Schiano (1998) studied, among other things, the effect of the length of neighbouring elements on judgements of bar height and magnitude comparison in bar charts and their experiments demonstrated that the accuracy of participants' judgements depended of the relative height of the target bar and neighbouring graphical elements.

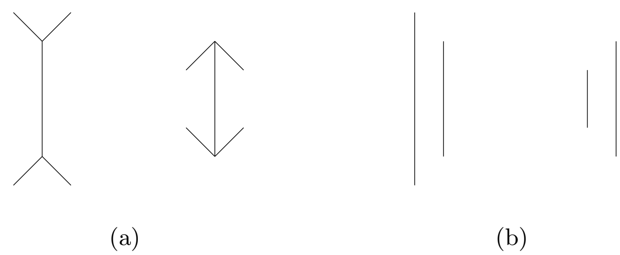


Figure 1: Two visual illusions affecting the perception of line length: (a) The Müller-Lyer illusion, (b) The parallel lines illusion.

In addition to issues relating to visual features, diagram designers must also be mindful of user familiarity. Using a form of diagram that is unfamiliar to a target audience can be problematic as users must expend additional cognitive effort to learn the new representation, something that could discourage engagement with the diagram or result in misinterpretation. Employing a novel representation can be justified, however if it can be demonstrated that the particular representation of information provided by the diagram facilitates a specific set of interpretive tasks. For example, in a series of experiments, Peebles and Cheng (2001, 2002, 2003; Peebles, Cheng, & Shadbolt, 1999) compared the representational and computational properties of two types of Cartesian coordinate (x,y) graph which, according to participant ratings, varied significantly in terms of their familiarity. Our experiments demonstrated that users of the less familiar graph type were able to retrieve information and solve certain problems significantly faster than users of the more familiar form. Eye movement and modelling analysis showed that this was because the graphical format of the unfamiliar representation facilitates certain basic reading and lookup procedures.

These issues are not only of academic concern. They have been brought into the public arena recently by the decision of the UK government to publish its national police performance data in the form of a set of diagrams that are relatively unfamiliar to the general public. With much media attention and at a reported cost of £70,000 (approx. US $128,688), the UK government developed the *performance monitor* (Police Standards Unit, 2003, 2004[1]), a variation of a diagram otherwise known as a *spidergram*, *radar* or *kiviat* chart. The purpose of the performance monitor is to 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) and to allow easy comparison with average performance computed from a set of police forces most similar to the individual force in terms of socio-economic, demographic and geographic makeup.

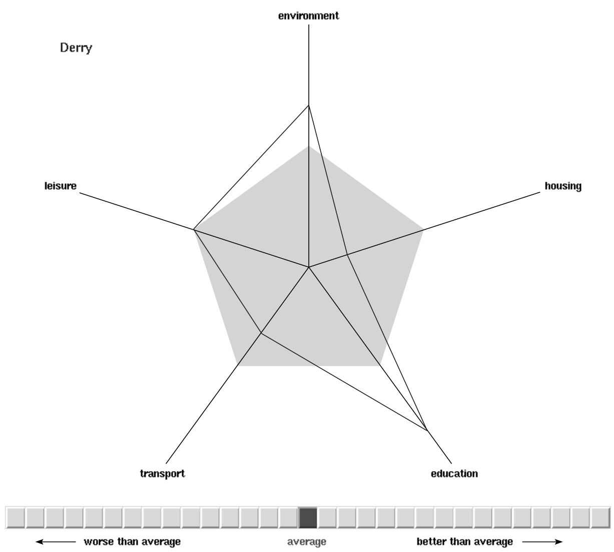


Figure 2: Kiviat chart used in the experiment.

An example of such a diagram is shown in Figure 2. This diagram is taken from the experiment reported here but it is identical in form to the police performance monitor. The subject matter of the diagrams was changed for the experiment to the (fictitious) performance of UK local authorities in five domains (the environment, housing, education, transport and leisure), each of which is indicated by a point on a spoke. 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 centre.

In the most recent version of the police performance report, the central performance monitor diagram has been augmented with five bar charts that illustrate the spread of performance for the most similar police forces in each of the domains. The bar charts are similar in form to the one displayed in Figure 3. In the police performance bar charts, each bar represents the value on that domain of one of the forces from which the average has been computed.

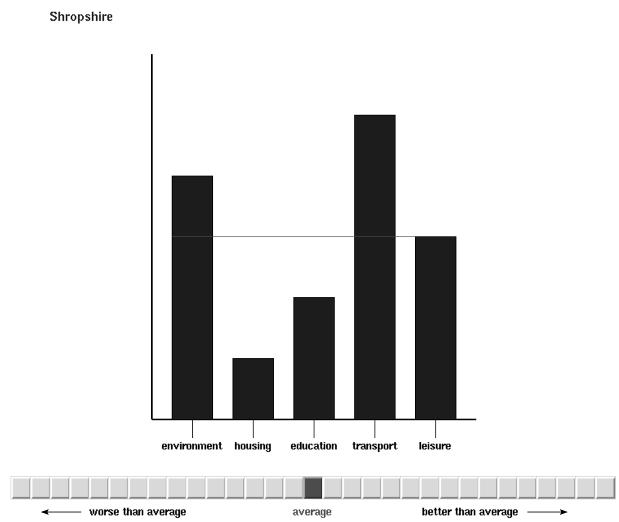


Figure 3: Bar chart used in the experiment.

One striking feature of both the kiviat and bar charts used in the police performance report is the lack of any tick marks on the spokes or axes. Usually the purpose of a scale of numbered tick marks on a chart or graph is to provide numerical values relating to locations in the chart. When numerical values are not deemed necessary, (perhaps because the purpose of the chart is simply to display relative magnitudes), tick marks still provide an objective reference frame within which to compare lengths. Without such a reference frame, it may be the case that perceptual judgements of quantities such as line length become more susceptible to distortion by visual illusions.

For example, in kiviat charts, the two lines connecting a point on a spoke with the two points on the adjacent spokes form a wide range of shapes and angles. In the absence of anchoring tick marks on the spokes, it may be the case that perceptual judgements of distance will be distorted by these angles and shapes by processes analogous to those involved in the Müller-Lyer illusion. Similarly, in bar charts, the lack of tick marks may also permit distortions in perceptual judgements of distance to occur because of the parallel lines illusion.

Another widely used diagram that shares many properties with bar charts is the line graph (see Figure 4), the main obvious difference being that points plotted against the y-axis are joined by lines rather than being represented as the top of a column. In the context of this study, however, an important consequence of this difference is that judgements of distance would not be

[1] A hypertext version of the current Home Office document is available on the Web as `performancemonitors.html` at `http://www.policereform.co.uk/docs/`

susceptible to the parallel lines illusion in line graphs.

Experiment

The primary purpose of the performance monitors and bar charts is to allow a rapid visual comparison of an individual institution's performance with a meaningful average. This could be either at a global level (i.e. to determine how much better or worse than average the institution is overall), or at the level of specific domains. The purpose of the experiment reported here is to determine whether the perceptual judgement of this distance for a particular target domain is affected by the values of the surrounding domains.

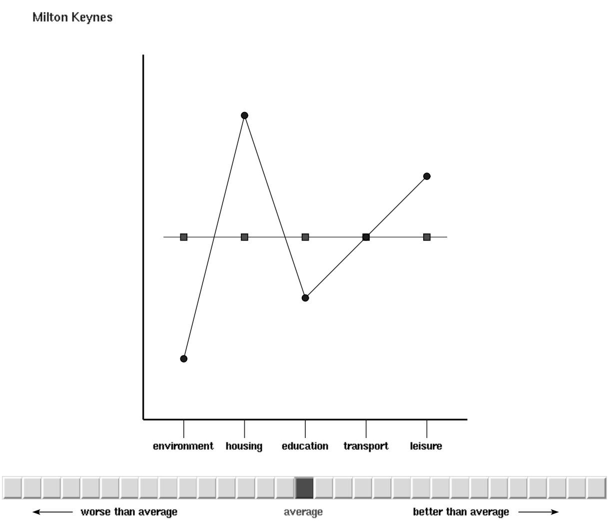


Figure 4: Line graph used in the experiment.

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 chart, 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 took part in the experiment. The occupations of participants varied from academic, clerical and technical positions to graduate students.

Materials The experiment was conducted using three identical PC computers with 17-inch (43-cm) displays. The stimuli used in the experiment were diagrams similar to those in Figures 2–4. The information content of the diagrams was the performance of 150 UK local authorities across five domains.

In order to generate a manageable range of values, the spokes of the kiviat chart and the y axes of the bar chart 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 centre and outermost points of the kiviat spokes respectively. Only the numbers 1 to 5 were used in the experiment and the locations of these on the diagrams can be seen in Figures 2–4. For example in Figure 2, 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 chart and line graph are illustrated in Figure 3, 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.

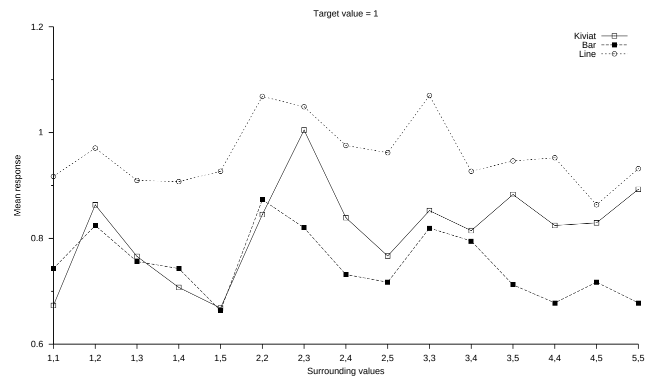


Figure 5: Mean response to target 1, all three diagrams.

The average value was the number 3 located at the centre of the y axes and kiviat spokes. In the bar chart 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 centre points on the five spokes. 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 centre button in the scale was the same red colour 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 centre button allowed the scale to be divided into six equally sized units, each containing four buttons. Below the scale were two arrows indicating that increases in performance were represented by buttons further to the right of the scale.

To test the full range of target and neighbouring 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 spoke has an adjacent domain on either side but in the bar charts and line graphs, two domains (environment and leisure) have only one adjacent domain. To ensure that the target domain on each trial had an adjacent domain on either side, therefore, if the target value was 1, 2, 4, or 5, 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. Responses to the target value of 3 were not to be included in the analysis as they were expected to be rapid and accurate for all graph conditions as this value was marked on both the graph and the scale.

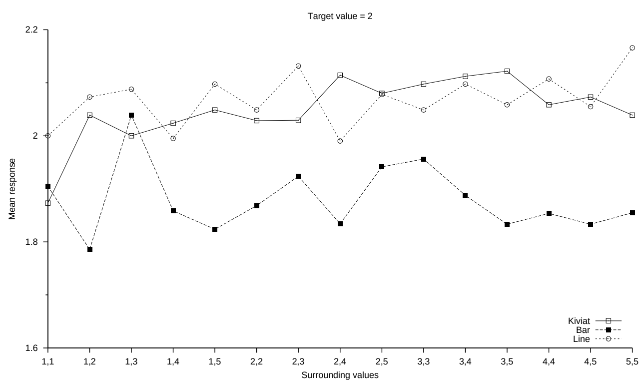


Figure 6: Mean response to target 2, all three diagrams.

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 modelled closely on the format used in the Home Office document. When the participant had finished studying the example, the experimenter then explained the diagram further, explaining the subject matter, highlighting salient points and making sure that they were completely familiar with it. Participants were then 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. Participants were also shown how to enter their judgement by clicking the mouse on the scale below the diagram. It was stressed to participants that they should attempt to respond as rapidly but also as accurately as possible.

On each trial of the experiment, the target domain was first presented in the centre 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

Participants' responses were coded to reflect the underlying scale of the diagrams. A response click on the button to 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 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 large to distort the mean for a specific cell. To reduce the influence of these outliers, the 42 values in each cell were standardised 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 discarded (Tabachneck & Fidell, 2001). From the original set of 9450 data points, this procedure resulted in the removal of 165 cases (1.75%) of the response data and 128 (1.35%) cases of the RT data. Data from the target value = 3 condition were not included in the analysis because, as predicted, responses were almost entirely accurate for all of the diagrams.

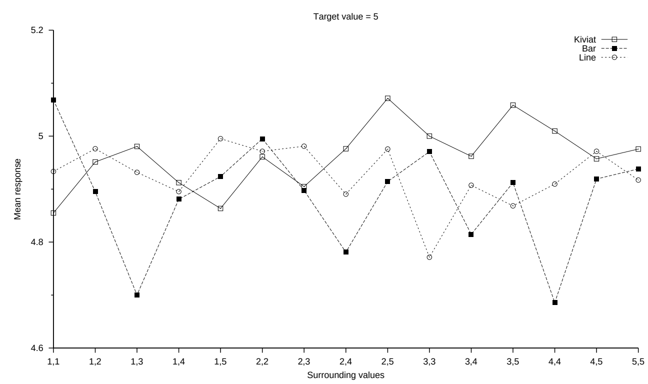


Figure 7: Mean response to target 5, all three diagrams.

An analysis of variance (ANOVA) was carried out on the response data. In the ANOVA, Mauchly's test of sphericity was significant for the target value (Mauchly's $W = .23$, $df = 5$, $p < 0.01$) and adjacent values (Mauchly's $W = .25$, $df = 104$, $p < 0.01$) so the more conservative Greenhouse-Geisser test was used in the analysis. The ANOVA showed that there was a significant main effect of the diagram used, $F(2, 59) = 15.48$, $p < .001$, the target value $F(1.66,97.66) = 4847.11$, $p < .001$ and the adjacent values, $F(9.87,582.49) = 3.96$, $p < .001$, together with significant interactions between the adjacent value and diagram, $F(19.75,582.49) = 2.84$, $p < .001$, and between target value and adjacent value, $F(16.07,974.94) = 2.51$, $p < .01$. The ANOVA also revealed a three-way interaction between diagram, target value and adjacent value, $F(32.13,974.94) = 1.56$, $p < .05$. Although response times were also recorded, due to lack of space, the data are not reported here. The main difference observed was that responses were slower in the kiviat condition than in the other two, probably due to the greater degree of unfamiliarity.

The complex effects revealed by the ANOVA are most

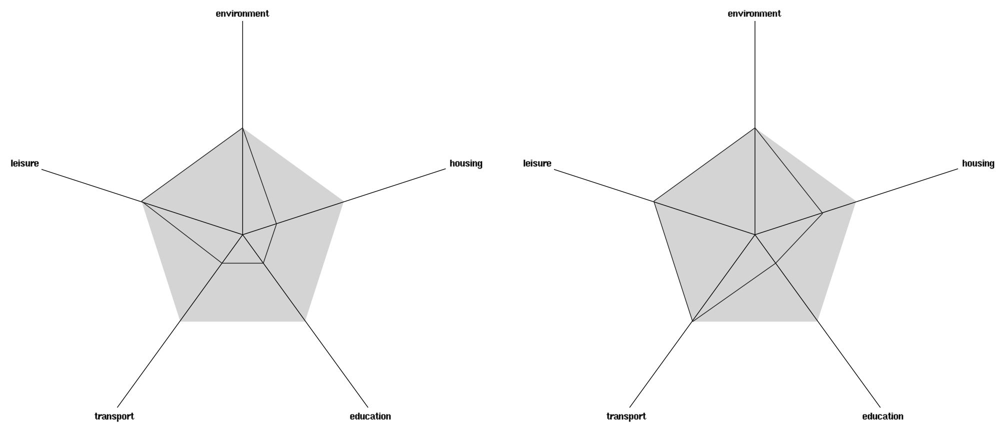


Figure 8: Kiviat charts illustrating a target value of 1 (education) with adjacent values 1,1 (left) and 2,3 (right).

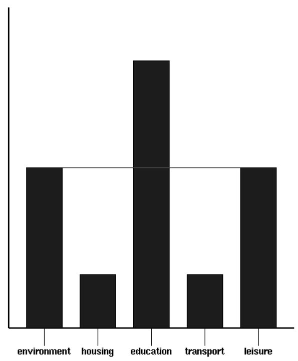
clearly illustrated in Figures 5–7 which present the mean response for target values 1, 2, and 5 respectively as a function of the adjacent values. In these graphs, each labelled tick mark on the y-axis represents a button on the scale. The graphs show a large degree of variation in responses to individual target values both for individual diagrams and between the different diagram types.

One of the most striking differences in responses between the diagrams is shown in Figures 5 and 6, both of which show that participants viewing the line graphs consistently perceive target values of 1 and 2 to be closer to the average than bar chart users, despite the fact that the values are represented at exactly the same locations in the coordinate system in the two diagrams. This marked difference between the graphs is not present for target values of 4 and 5, however, and this may provide a possible explanation. Unlike points in a line graph, bars are attached to and proceed from the x-axis and so form a concrete object. When comparing the distance between the top of a bar with the mean line, therefore, participants' attention is drawn to the length of the bar in comparison to the height of the mean line, (rather than to the distance between them), which may serve to accentuate the perceived difference. In contrast, participants using the line graph may simply attempt the more accurate procedure of judging the distance between the points on the plotted and mean lines.

According to this account, the lack of such a major difference when the target values are 4 and 5 is because users of the bar chart are still judging the length of the bar but, this time are comparing it to the mean line below it. This is very similar to the procedure carried out by the line graph users in that both are judging the same distance. At the moment this is a plausible hypothesis that remains to be tested in a further eye movement study.

Figure 5 also illustrates the wide range of responses that users of the kiviat chart gave to the same target value. It is clear that the perception of the difference between the target value 1 and the mean value 3 is affected by values of the adjacent domains. To provide a clearer demonstration of this effect, two of the diagrams resulting in widely differing ratings are shown in Figure 8. In both diagrams, the target value of 1 is represented on the education domain. The chart on the left has adjacent values of 1,1, for which the mean rating is 0.67, whereas that on the right has adjacent values of 2,3, given a mean rating of 1.0. A t-test shows the difference between these ratings to be significant $t(60.11) = 3.10$, $p < .001$. Figure 5 shows that participants viewing the kiviat charts perceive the target value of 1 to be much closer to the average when surrounded by the values 2 and 3 than for any other combination but that the surrounding values of 1 and 1 were perceived as relatively far from the average. Although a precise explanation for this is still being considered, it is clear that the shape produced by the lines connecting the target value and the adjacent values has a distorting effect on viewers' perception of distance. Figure 5 shows that this effect is not a simple linear function in which greater adjacent values result in a larger perceived target value.

The distortion of perceived distance in the bar chart condition is perhaps best illustrated in Figure 7 which reveals a wide range of responses to the target value 5. As with the kiviat charts, the pattern of results does not conform to a simple analysis. At least one example, however, may be best explained by reference to the parallel lines illusion. The two bar charts in Figure 9 were given ratings at the extreme ends of the range. Both represent a target value of 5 on the education domain and the chart on the left has adjacent values of 1,1, for which the mean rating was 5.07, whereas that on the right has adjacent values of 4,4, given a mean rating of 4.69. A t-test shows the difference between these ratings to be significant $t(76.02) = 3.88$, $p < .001$. The target domain in the chart on the left of Figure 9 was perceived as being further away from the average line than that in right-hand chart. This can be explained in



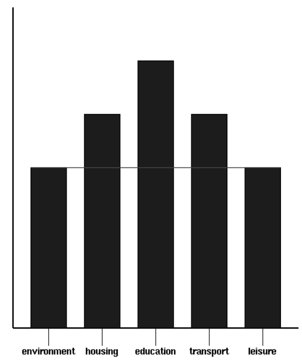


Figure 9: Bar charts illustrating a target value of 5 (education) with adjacent values 1,1 (left) and 4,4 (right).

terms of the parallel lines illusion as it reflects the phenomena of contrast and assimilation described earlier. The target domain in the left-hand chart is seen as being larger because it contrasts with two relatively small adjacent values. The target domain in the right-hand chart, however, was perceived as being smaller because viewers perceived the lengths of the target and adjacent bars to be closer than they actually are (assimilation). It is also interesting to note in Figure 7 that this pattern of responses is not found in the line graph condition.

Discussion

The results of this experiment provide concrete evidence of distortions in perceptual judgements of distance in two graphical representations of a type currently being employed to present public data in the UK. Specifically, the three examples clearly illustrate that simple comparative judgements between two points on a dimension can be significantly affected by the values of adjacent variables. Further work is required before firm conclusions can be drawn but from this initial analysis it seems that, as currently designed, the kiviat and bar charts in the UK government's police monitoring documents may be susceptible to the distorting effects highlighted here.

The use of anchor points, typically tick marks on axes, is generally seen as a way of facilitating the accurate reading of locations relative to a scale. Whether the incorporation of such anchor points into these diagrams reduces the distortions in distance judgements observed in this experiment is to be tested in a future study.

References

Deręgowski, J. B. (1980). *Illusions, Patterns, and Pictures.* London: Academic Press.

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.

Peebles, D., & Cheng P. C.-H. (2001). Graph-based reasoning: From task analysis to cognitive explanation, *Proceedings of the Twenty-third Annual Conference of the Cognitive Science Society*, *23*, 762–767.

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*, 28–46.

Peebles, D., Cheng P. C.-H., & Shadbolt, N. (1999). Multiple processes in graph-based reasoning, *Proceedings of the Twenty-first Annual Conference of the Cognitive Science Society*, *21*, 531–536.

Police Standards Unit. (2003). *Police Performance Monitoring 2001/02*. London: Home Office.

Police Standards Unit. (2004). *Police Performance Monitoring 2002/03*. London: Home Office.

Poulton, E. C. (1985). Geometric illusions in reading graphs. *Perception & Psychophysics*, *37*, 543–548.

Schiano, D. J. (1986). Relative size and spatial separation: Effects on the parallel-lines illusion. *Perceptional & Motor Skills*, *63*, 1151–1155.

Schiffman, H. R. (1995). *Sensation and Perception: An Integrated Approach* (Fourth Ed.). New York: John Wiley & Sons.

Tabachneck, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*, (Fourth edition). Needham Heights, MA: Allyn and Bacon.

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.