Differences in Comprehensibility Between Three-Variable Bar and Line Graphs

David Peebles (D.Peebles@hud.ac.uk)

Nadia Ali (N.Ali@hud.ac.uk)
Department of Behavioural and Social Sciences,
University of Huddersfield,
Queensgate, Huddersfield,
HD1 3DH, UK.

We report an experiment investigating graph comprehension. Verbal protocol data were collected while participants attempted to understand six bar or line graphs representing relationships between three variables. Analysis of the verbal protocols revealed significant differences in the level of comprehension between the two graph types. Specifically, a significant proportion of line graph users was either unable to interpret the graphs, or misinterpreted information presented in them. These errors did not occur in the bar graph condition. The difference is explained in terms of the high salience of the lines in line graphs which hinders the correct or full interpretation of the relationships depicted. The results of the experiment provide a strong rationale for the use of bar graphs to display such three-variable data sets, particularly for a general audience.

Keywords: graph comprehension, diagrammatic reasoning, information graphics.

Introduction

Bar and line graphs are the most commonly used graphical formats for presenting quantitative data, not only for experienced practitioners in science, engineering, and business but also for more a general audience in education and the media (Kosslyn, 2006; Zacks, Levy, Tversky, & Schiano, 1998). Within the space of graphical representations bar and line graphs are very close. Because both utilise the Cartesian coordinate system, knowledge of the representational properties of this system, as a minimum, allows users to understand how the two diagrams 'work' and possibly to extract some basic information from them.

Beyond this underlying similarity in representational framework however, the key difference in how data are represented in the two graphs can have profound effects on how the data are understood and interpreted. Line graphs are typically regarded as a form of *configural* or *object* display because a single line integrates the individual plotted points into a single object. Features of this object—its slope for example—can indicate relevant information about the entire data set (Carswell & Wickens, 1990, 1996). In contrast, bar graphs are an example of a *separable* display as each variable is represented by a single bar.

For these reasons, 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 them (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).

Not only are people's conception and interpretation of bar and line graphs different, their actual perception of values depicted in the two graphs can also vary significantly. In a recent study, Peebles (2008) asked people to compare values plotted in bar and line graphs with an average (represented as a line drawn from the *y* axis parallel to the *x* axis). Despite the fact that the values being compared were plotted at exactly the same locations in the graphs, bar graph users significantly underestimated the size of the plotted value relative to the mean compared to line graph users. This effect was shown to result from a process in which bar graph users' visual attention was drawn via a figure-ground process to the length of the bars as they extend from the *x* axis (cf. Pinker, 1990; Simcox, 1983) rather than to the distance between the top of the bar and the mean line, thereby accentuating the perceived difference between them.

Because of their different representational properties, guidelines recommend bar and line graphs be used for different communicative goals. One such guideline is to use line graphs to display the interactive effects of two independent variables (IVs), each with two levels, on a dependent variable (DV; e.g., Kosslyn, 2006, p. 49). This situation is widely encountered in many scientific and engineering contexts and the use of such *interaction graphs* is taught in a wide range of undergraduate curricula, including psychology.

The rationale for using line graphs in such cases is that the different patterns formed by the lines can be rapidly identified by experienced users as indicating particular quantitative relationships between the variables. So, for example, users familiar with the format should be able to recognise an X pattern as indicating a *crossover* interaction and know that two parallel lines indicate no interaction. By contrast, these patterns are not as salient in bar graphs but must be constructed by the user by mentally connecting the tops of the bars.

Although no doubt useful in such situations, the salience of plotted lines can significantly affect people's interpretation of the data being presented. For example, Carpenter and Shah (1998) showed that for line graphs, the same data presented from alternative perspectives can lead to different interpreta-

Table 1: Dependent and independent variables used to construct the six graphs in the experiment.

Graph No.	Dependent Variable	Independent Variable 1		Independent Variable 2		Scale Range	Scale Increment
		Name	Levels	Name	Levels		
1	Response Time	Task	AA, AB	Stimulus Type	Words, Pictures	50–500	50
2	Fractures	Temperature	Cold, Hot	Stress	Low, High	2–20	2
3	Rainfall	Cloud Cover	High, Low	Cloud Seeding	No, Yes	1–10	1
4	Wellbeing	Gender	Male, Female	Exercise	High, Low	15–150	15
5	Percent Error	Experience	Low, High	Time of Day	Day, Night	10–100	10
6	Weight Gain	Protein Source	Beef, Cereal	Protein Type	High, Low	10–100	10

tions depending on which variable was plotted on the *x* axis and which in the legend (or *z* variable—we will use the two terms interchangeably).

They found that people's understanding of the quantitative relationship between the variables plotted on the *x* and *y* axes was much more comprehensive than that of the relationship between the variables plotted in the legend and on the *y* axis. They argued that this is because readily-interpretable features of the lines (e.g., slope and height) make the *x*–*y* relationship more salient and the quantitative relationship between these variables rapidly and easily available. Their participants tended to uses the markers of the continuous *z* variable as nominal labels but to describe the characteristics of the *x* variable in much greater detail. By contrast, in order to understand the characteristics of the *z*–*y* relationship in similar detail, people are required to note the values of the *z* variable markers identified in the legend and infer the effect the *z* variable is having on the dependent variable.

Given the representational differences between line and bar graphs outlined above, there arises an interesting question whether this effect would be found in bar graphs. Unlike the lines in line graphs, bars are anchored to the *x* axis and draw the user's attention (Peebles, 2008). This may have the effect of reducing the initial focus on the *z* variable and allow the user to generate a more balanced interpretation of the data. One of the aims of the current experiment was to test this hypothesis.

Experiment

Method

Participants Twenty-nine students and members of staff from the University of Huddersfield volunteered to take part in the experiment; 14 men and 15 women. The age of participants ranged from 23.1 to 62.2 years with a mean of 42.8 years (SD = 12.7). The majority (48.3%) were academic staff from various schools in the university with smaller proportions of non-academic staff (20.7%), postgraduate (20.7%) and undergraduate (10.3%) students.

Design The experiment was an independent groups design with two between-subject variables: type of diagram used (bar or line graph) and the allocation of independent variables to the *x* axis and legend. Participants were allocated to graph conditions so as to equalise as much as possible the number of undergraduate and postgraduate students and academic and non-academic staff between graph conditions. 18 participants saw the line graphs and 11 participants saw the bar graphs.

Materials The experiment was carried out using a PC computer connected to a Tobii 1750 remote desktop eye tracker with a 17 inch display. The twenty-four graphs for the experiment were created using the variables shown in Table 1. The data for each graph were selected to create the range of interactions and other relationships between three variables commonly encountered in line graphs (e.g., parallel, crossed and converging lines, one horizontal line and one sloped line, two lines sloping at different angles, etc.). Figure 1 shows four graphs of the type used in the experiment.

Six 'normal' bar and line graphs were created with IV1 on the *x* axis and IV2 in the legend and a further six 'reversed' graphs were then created by switching the IVs on the *x* axis and the legend, thus producing 12 line graphs and 12 bar graphs. Variable names in the graphs were concatenated (e.g., "Weight Gain" was written as "WeightGain") in order to facilitate the eye movement fixation analysis.

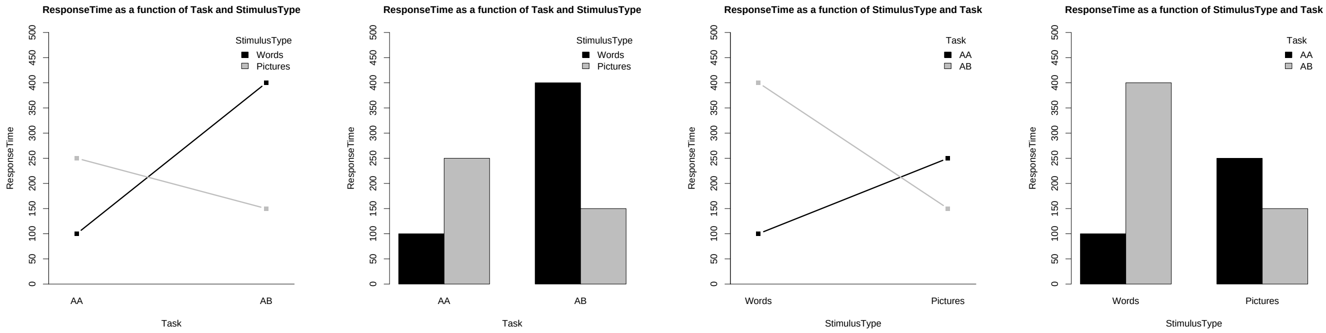
Procedure In both the line and bar graph conditions, half of the participants saw the normal graphs and half the reversed graphs. They were instructed that they would be presented with six three-variable graphs and that their task for each one was to try to understand the information it was conveying as fully as possible while thinking aloud.

Participants studied each graph for as long as they were satisfied that they had understood it mas much as they could, at which point they ended the trial and started the next one until all six had been completed. If a participant went silent during the course of a trial they were prompted by the experimenter to continue thinking aloud. All graphs were presented in random order.

Results

Participants' eye movements were recorded as part of a broader research aim to investigate the cognitive processes involved in the comprehension of bar and line graphs. For reasons of space however, the eye movement data are not reported here.

Participants' verbal protocols were transcribed and anal-



(a) “Normal” line graph 1 (b) “Normal” bar graph 1 (c) “Reversed” line graph 1 (d) “Reversed” bar graph 1

Figure 1: Example bar and line graphs used in the experiment.

ysed according to the three-level model of graph comprehension ability identified by Friel, Curcio, and Bright (2001). According to this model, people at an *elementary* level focus primarily on reading off values but do not elaborate on what similarities or differences in values mean. At an *intermediate* level people interpret the data presented in the graph and, to some extent at least, integrate the information. At an *advanced* level people make inferences beyond what is explicitly stated in the graph by hypothesising based on trends depicted therein.

To ascertain participants' level of comprehension ability, their statements were categorised according to the above criteria and a classification allocated based on the proportion of statements in each category across the six trials. The number of participants in each comprehension category for the two graph conditions is shown in Table 2. The data show that a large proportion in each condition were at an intermediate level, being able to read individual points and integrate the information into coherent statements. There was also a small number of advanced users who were able to extrapolate from the data and read beyond it to a certain extent.

The most surprising aspect of the analysis however, was the large number of line graph users (39%) who failed to understand them at even an elementary level and so were unable to interpret them in any meaningful way. These *pre-elementary* participants were unable to read values of individual points accurately (e.g., in Figure 1a were unable to say something like “when the stimulus type is words and the task is AA, then response time is a hundred”) and made a number of common errors.

Participants were categorised as pre-elementary if after the first two trials they were still unable to point read and were still making mistakes. Participants' transcripts were coded by the second author and a sample (approximately 45% from each graph type) was independently coded by the first author. The level of agreement reached between the two codings was 94% of trials and 100% of pre-elementary categorisations.

According to these criteria no bar graph participants were classified as being pre-elementary. A Fisher's exact test on graph type found a significant difference between number of participants coded as pre-elementary in the bar and line graph condition ($p < .05$). The types of errors produced by the pre-elementary line graph users are outlined below.

Table 2: Number of participants in the four comprehension ability levels in each graph condition.

Comprehension Level	Graph Condition	
	Line	Bar
Pre-elementary	7	0
Elementary	0	1
Intermediate	9	9
Advanced	2	1

Ignoring the *x* variable In line with the results of Carpenter and Shah (1998), a common practice of the line graph users was to focus on the legend variable. Many pre-elementary users however, simply described the effect of this variable and ignored the *x* axis variable altogether. This can be more or less damaging to overall interpretation depending on the nature of the data being depicted. For example, it may be only a partial understanding of the relationships depicted, for example if a participant looking at Figure 1a said “Response time is longer for words than pictures” since the graph shows that this is the case for only one level of the *x* variable. These were classified as *incorrect interpretation (z)*.

Focussing entirely on the *z* variable may not always result in an incorrect interpretation because the nature of the data depicted sometimes makes the resulting statement a correct description of a relationship. Such statements were classed as *limited interpretation (z)* because they did not take into account the effects of both IVs—specifically the *x* axis variable. According to these criteria, pre-elementary line graph users made six *incorrect interpretation (z)* statements and seven *limited interpretation (z)* statements whereas bar graph users made only three *limited interpretation (z)* statements.

Ignoring the *z* variable The inverse of the situation described above is where the participant describes the effect of the *x* axis variable but ignores the legend variable, which was classified as an *incorrect interpretation (x)*. Pre-elementary line graph users made five *limited interpretation (x)* statements, whereas bar graph users made none.

Ignoring a level of the *x* variable Sometimes pre-elementary line graph users would take the *x* variable into account but ignore one of its levels when describing its effect. Typically this occurred when the user identified a level of the legend variable (e.g., "when stress is high"), then described a property of the plotted line but relates this to only one level of the *x* variable (e.g., "fractures increase when temperature is hot"). This type of statement occurs seven times in the line graph condition but not once in the bar graph condition.

Table 3: Type and frequency of errors for the bar and line graphs in the experiment.

Error type	Frequency	
	Line	Bar
Ignoring the *x* variable	16	3
Ignoring the *z* variable	5	0
Ignoring a level of the *x* variable	7	0
Pattern description and missed trials	9	1
Miscellaneous	5	0

General levels of comprehension In general, line graph users had a greater level of difficulty in interpreting the data appropriately or fully, as evidenced by the number of miscellaneous errors, simple pattern descriptions and missed trials.

A statement was categorised as a *miscellaneous error* if participants were relating all three variables together but their interpretation was incorrect, or if they included information in their interpretation that was not depicted on the graph. Miscellaneous errors were less systematic than the other response patterns and each instance typically occurred only once. However this class of error occurred five times in the line graph condition but not at all in the bar graphs.

An example of a miscellaneous error for reversed line graph 2 in Figure 2a is "when it's cold there are fractures. When there's low stress it's leaning towards hot. Fewer fractures when hot". Although all three variables are incorporated into the interpretation, it is incorrect because the graph is showing that when it is cold, whether stress is low or high, fractures are the same. However, when it is hot, fractures are lower at cold temperature and higher at hot temperature.

Often pre-elementary participants would simply describe the pattern at the centre of the graph and make no attempt to relate the pattern back to the variables, or name the variables but not relate them to the pattern. For example, one participant's sole comment for normal line graph 4 (Figure 2b) was "Two level lines—one is at thirty and the other is at ninety".

A *missed trial* was recorded when participants missed a whole trial without saying anything, their verbalisations were incoherent, or not related to the information the graph was displaying (e.g., if a participant said "this is hard", "I've no idea what this means", etc.).

Statement patterns In general, the results of this experiment are consistent with those of Carpenter and Shah (1998) in that they confirm that line graph users focus initially on the legend variable as this is where the plotted lines may be individuated into nominal categories. In bar graphs however, because the bars are anchored to the *x* axis, the visual features vying for the user's attention are more 'balanced' in terms of which IV the user may look to first. Although colour may still attract them to the legend variable, the location of the bars may draw them to the *x* variable. Therefore we should expect the line graph situation to at least be reduced or possibly even reversed.

To determine whether this was the case we used a modified version of the coding scheme devised by Shah and Carpenter (1995) to code participants' statements according to whether they described the *x*–*y* or *z*–*y* relationship. If people mentioned the influence of one level of the *z* variable on the *x*–*y* relations the statement was classified as *metric-x*. A typical metric-x statement for the graph shown in Figure 1a would be "when the stimulus type is words, response time is faster for task AA than for task AB".

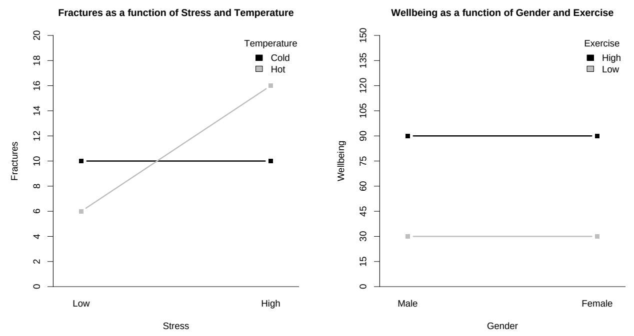
Conversely, if the influence of one level of the *x* variable on the *z*–*y* relations was mentioned, the statement was classified as *metric-z*. An example metric-z statement for the graph shown in Figure 1a would be "when the task is AA, response time is faster for words than for pictures".

As with the previous analysis, all of the participant's transcripts were initially coded by the second author and a sample (approximately 14%) was independently coded by the first author. The level of agreement reached between the two codings was 95%. Figure 3 shows the number of metric-x and metric-z statements for the two graph conditions.

The figure reveals a significant difference in the number of metric-x and metric-z statements made by bar and line graph users ($\chi^2 = 57$, df = 1, $p < .001$). For the line graphs, this pattern of performance is consistent with the results of Carpenter and Shah (1998). However, the pattern is reversed for the bar graphs and indicates that these users tended to focus on the *x* variable, identify a level of that, and then compare the two levels of the *z* variable associated with it. For example people should typically interpret Figure 3 by saying something like "for the bar graphs, people make many more metric-z statements than metric-x statements whereas the situation is reversed for line graphs, where people make many more metric-x statements than metric-z statements".

Discussion

The use of line graphs to present the interactive effect of two two-level variables on a dependent variable is common



(a) "Reversed" line graph 2 (b) "Normal" line graph 4

Figure 2: Two line graphs used in the experiment.

practice and is recommended by authorities in the field (e.g., Kosslyn, 2006). The benefits of doing so are clear: once the patterns representing the various relationships are known, interpretation of at least the general characteristics of the data becomes easier and more rapid.

However, there are specific issues in using this representation that graph producers should be mindful of if they wish to minimise misinterpretation. Apart from the possibility of users overly focussing on the *x–y* relationship noted earlier (Shah & Carpenter, 1995), others have found that users have a strong tendency to interpret lines as representing trends rather than identifying the line ends and interpreting them discretely (Zacks & Tversky, 1999). This can happen to such an extent that users can sometimes describe discrete categories continuously (e.g., "The more male a person is, the taller s/he is", (Zacks & Tversky, 1999)), leading some to advocate the use of bar graphs to present this form of data (Aron, Aron, & Coups, 2006).

The research reported here suggests that the high salience of lines in line graphs can not only distort interpretation, but may actually be making line graphs harder for non-expert users to interpret to a basic level of understanding. Pre-elementary line graph users were unable to integrate the information, primarily because they ignored the *x* variable entirely, or if they did incorporate it, ignored one of its values. This pattern of errors indicates that the salience of the lines is such that it draws the user's attention to them and then—through a process of colour matching—to the legend variable, which they then try to interpret. Because they are focussing on the lines however, they are less able to identify the points at the ends of the lines and interpret them as discrete values associated with levels of the *x* variable.

In the bar graph however, each level of the *z* variable is depicted by a bar projecting from the *x* axis. This allows participants to match bar colour to the appropriate *z* variable level in the legend and, because the bars are located directly above the value labels on the *x* axis, more easily find the associated *x* variable level. The results of this experiment reveal that this balances out attention to the two IVs and promotes a richer understanding of the relationship between them.

The study also shows however, that although attention is more evenly divided in bar graphs on the whole, they do have the opposite effect of line graphs by making *z–y* relations more salient.

This reversal effect can be explained by Gestalt principles of perceptual organisation. In the case of bar graphs, the legend variable values are grouped together as bars on the *x* axis and, by the Gestalt *principle of proximity* (Wertheimer, 1938) the cluster of bars forms a visual chunk. This leads participants to use the *x* variable values as labels and describe the relationship between the *z* and *y* variables.

In the case of line graphs however, data points are connected by the *x–y* lines which, by the Gestalt *principle of connectedness* (Palmer & Rock, 1994), form individual visual chunks. This leads participants to use the *z* variable values as labels and describe the relationship between the *x* and *y* variables.

This finding suggests that people presenting data in graph form should consider plotting the variables differently in the two graphs, even for the same communicative goal. For three-variable line graphs, the variable that the graph producer wishes to emphasize is often plotted on the *x* axis, consistent with the graph's emphasis on *x–y* trends. These results suggest however, that for three-variable bar graphs, the variable to be emphasised should be allocated to the legend to make *z–y* relations more salient.

Retrieving and reasoning about information in graphs and diagrams is a complex skill requiring the interaction of three key factors: the cognitive abilities of the user, the graphical properties of the external representation, and the specific requirements of the task (Peebles & Cheng, 2003). This complex interaction makes the task of balancing the costs and benefits of using a particular representation all the more difficult. In weighing up the evidence Kosslyn (2006, p. 50) suggests that, on balance, the risk and costs of misinterpreting line graphs are outweighed by the benefits provided by the familiar line patterns.

This judgement rests on the assumption that the target user knows how to interpret the patterns, and if this is indeed the case, then the user should no doubt be able to interpret the information rapidly and effectively. If the interpretive skills of the target audience are unknown or are known to be limited however, it may be more beneficial to them to present the data in bar graph form. In our study users with limited knowledge of the specific interpretive schemas and procedures associated with a representation were more readily confused by the line graphs whereas bar graph users were able to benefit from the more balanced representational features of the bar graphs.

Therefore, even though it may require more time and cognitive effort to interpret interactions between variables presented in bar graph form, this is a minor cost to that associated with misinterpreting or being unable to interpret information presented in a line graph.

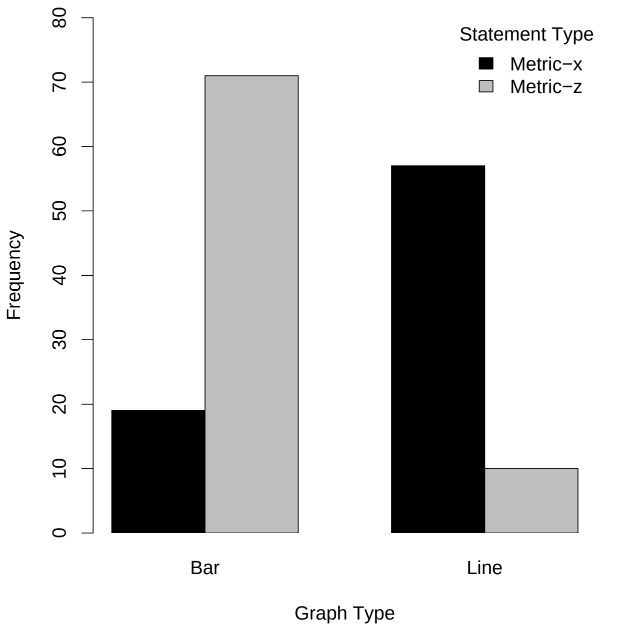


Figure 3: Frequency of metric-x and metric-z statements for bar and line graphs.

Although interpreting three-variable data is not trivial, to find such a large proportion of pre-elementary line graph users was surprising, and we were obliged to recruit additional participants in order to obtain examples of expert line graph users. Furthermore, the pre-elementary participants were not all undergraduate students (with presumably less prior exposure to such graphs) but academic and non-academic staff, some with doctoral-level education.

We are currently conducting further studies to determine the robustness of these findings and to investigate how the interpretation of line graphs by non-expert users may be improved by modifying their design. It seems clear from the current study however, that graph producers must take the abilities of the user more fully into account when considering a format, be mindful that what may be appropriate for experts may actually hinder the interpretation of less experienced users, and consider adopting a more appropriate representation in such cases.

References

Aron, A., Aron, E. N., & Coups, E. J. (2006). *Statistics for psychology* (4th ed.). London: Pearson.

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, *4*, 75–100.

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.

Culbertson, H. M., & Powers, R. D. (1959). A study of graph comprehension difficulties. *Audio-Visual Communication Review*, *7*, 97–110.

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, *32*, 124–158.

Kosslyn, S. M. (2006). *Graph design for the eye and mind*. New York: Oxford University Press.

Palmer, S. E., & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review*, *1*, 29–55.

Peebles, D. (2008). The effect of emergent features on judgments of quantity in configural and seperable displays. *Journal of Experimental Psychology: Applied*, *14*(2), 85–100.

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.

Pinker, S. (1990). A theory of graph comprehension. In R. Freedle (Ed.), *Artificial intelligence and the future of testing* (pp. 73–126). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schutz, H. G. (1961). An evaluation of formats for graphic trend displays—experiment II. *Human Factors*, *3*, 99–107.

Shah, P., & Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology: General*, *124*, 43–62.

Simcox, W. A. (1983). *A perceptual analysis of graphic information processing*. Unpublished doctoral dissertation, Tufts University, Medford, MA.

Wertheimer, M. (1938). Laws of organization in perceptual forms. In W. D. Ellis (Ed.), *A source book of Gestalt psychology.* London: Routledge & Kegan Paul.

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 and Cognition*, *27*(6), 1073–1079.