Topics in Cognitive Science (2009) 1–19

ISSN: 1756-8757 print / 1756-8765 online
DOI: 10.1111/j.1756-8765.2009.01066.x

Bar and Line Graph Comprehension: An Interaction of Top-Down and Bottom-Up Processes

Priti Shah,[a] Eric G. Freedman[b]

[a]*Department of Psychology, University of Michigan, Ann Arbor*
[b]*Department of Psychology, University of Michigan, Flint*

Received 5 May 2009; received in revised form 31 August 2009; accepted 3 September 2009

Abstract

This experiment investigated the effect of format (line vs. bar), viewers' familiarity with variables, and viewers' graphicacy (graphical literacy) skills on the comprehension of multivariate (three variable) data presented in graphs. Fifty-five undergraduates provided written descriptions of data for a set of 14 line or bar graphs, half of which depicted variables familiar to the population and half of which depicted variables unfamiliar to the population. Participants then took a test of graphicacy skills. As predicted, the format influenced viewers' interpretations of data. Specifically, viewers were more likely to describe *x–y* interactions when viewing line graphs than when viewing bar graphs, and they were more likely to describe main effects and ''*z–y*'' (the variable in the legend) interactions when viewing bar graphs than when viewing line graphs. Familiarity of data presented and individuals' graphicacy skills interacted with the influence of graph format. Specifically, viewers were most likely to generate inferences only when they had high graphicacy skills, the data were familiar and thus the information inferred was expected, *and* the format supported those inferences. Implications for multivariate data display are discussed.

Keywords: Graph comprehension; Display design; Individual differences

1. Introduction

A well-designed graph can be a vivid, memorable, and easy-to-understand depiction of quantitative information (Larkin & Simon, 1987; Shah, Freedman, & Vekiri, 2005; Smith, Best, Stubbs, Archibald, & Roberson-Nay, 2002; Tufte, 2001). For those reasons, graphs are used extensively in textbooks, scientific journals, and the popular print media (Shah et al.,

Correspondence should be sent to Priti Shah, Department of Psychology, University of Michigan, 530 Church Street, Ann Arbor, MI 48109. E-mail: priti@umich.edu

2005; Zacks, Levy, Tversky, & Schiano, 2002). In fact, there has been a dramatic rise in the prevalence of graphs depicting quantitative data. In one analysis, Zacks et al. (2002) found that between 1984 and 1994 the mean number of graphs in academic journals nearly doubled, and the number of graphs in newspapers more than doubled. Corresponding to the increased prevalence of graphs, research on graph comprehension has grown substantially in the last several years (see e.g., Canham & Hegarty, in press; Mautone & Mayer, 2007; Ratwani & Trafton, 2008; Ratwani, Trafton, & Boehm-Davis, 2008).

Despite the fact that graph comprehension is better understood, research has primarily focused on the comprehension of two-variable data (e.g., sex vs. height) with a small number of data points (e.g., 2 or 4). Less is understood about more complex data. However, conclusions from studies of relatively simple graphs and tasks frequently do not scale up to the comprehension of more complex information (Canwan & Hegarty, in press; Ratwani & Trafton, 2008; Shah et al., 2005; Trafton et al., 2000). The current study addresses two as yet unresolved questions about the comprehension of two common graph formats, line and bar graphs, depicting data of medium complexity (see Fig. 1):

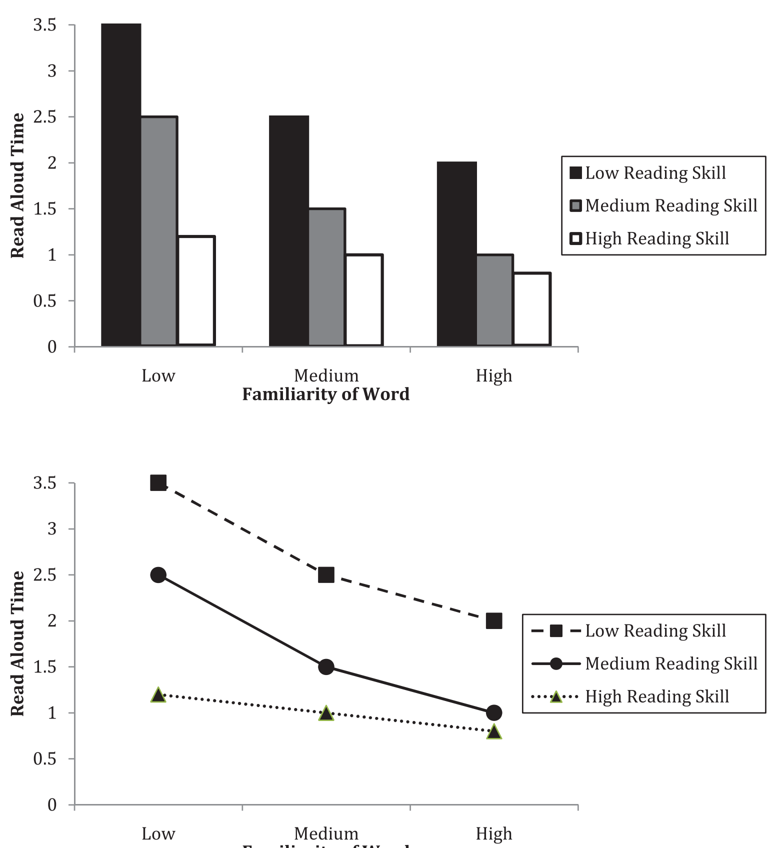

Fig. 1. Sample stimulus graphs. The bar graph (top) depicts the same data as the line graph (bottom).

(1) How does an individual's prior knowledge (both topic familiarity and graphicacy, or graphical literacy skills) interact with format to influence viewers' interpretations?
(2) To what extent does graph format affect fact-retrieval and inference generation in multivariate data comprehension?

An understanding of how prior knowledge interacts with bottom-up features should lead to a better understanding of the design of graphs and other visual displays. Specifically, by focusing on individuals' familiarity with the topic of information in a graph and their graphical literacy skills, we can draw conclusions about what formats benefit different individuals.

We take as our starting point Hegarty's (2005) model of display comprehension. According to this model, a version of which is depicted in Fig. 2 (Kriz & Hegarty, 2007), top-down processes interact with bottom-up information in the comprehension of all external displays. We have proposed that this type of interactive model applies to graph interpretation as it does for other less abstract visual displays (e.g., Freedman & Shah, 2002; Shah, 1997; Shah et al., 2005), yet there is little empirical evidence demonstrating the interaction of top-down and bottom-up processes in the context of graphs. In this study we test an extension of interactive models to consider the effects of two types of prior knowledge: domain-specific familiarity with the content depicted in a graph, and domain-general knowledge about graphs per se (e.g., graphicacy skills, including Pinker's *graph schema*; Pinker, 1990). We expect that these two types of knowledge may have an impact on inference generation processes in graph comprehension.

Inference generation is more likely to be relevant for complex graphs compared to simple graphs. Simple graphs typically require individuals to retrieve a small number of facts that can easily be remembered, whereas complex graphs frequently require making decisions about what information to encode and mental computations and inferences to simplify the information depicted. Consider the graph in Fig. 1. If an individual attempted to encode and remember three values (word familiarity, reading level, and reading time) for each of the nine data points, the information would far exceed his or her working memory capacity. To reduce the amount of information encoded requires selective processing, and, quite likely,

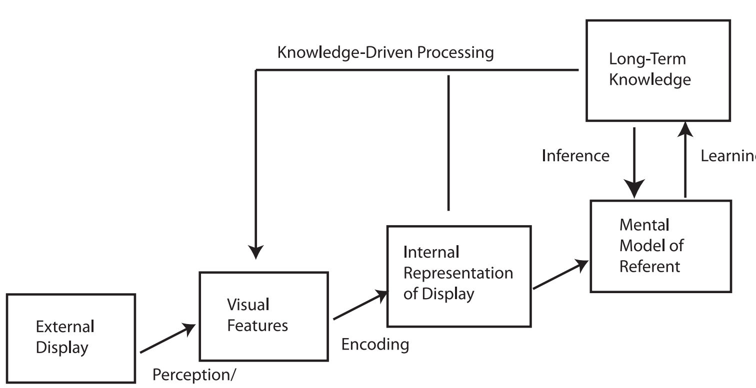

Fig. 2. Model of display comprehension proposed by Kriz and Hegarty (2007).

mentally transforming the information depicted. A related difference between simple and complex data is that complex data allow for a greater number of possible comprehension goals that may be selected by a graph viewer. Each goal might require different component processes, mental transformations, or inferences. For example, a graph viewer may compute average differences between high- and low-ability children because his goal is to see whether, as they expected, high-ability third graders read faster than low-ability third graders. Or she may be more interested in the role of word familiarity, or the possibility that there might be a nonlinear relationship between ability and reading speed. Although the nine data point, medium complexity graphs in our study are not as complex as graphs in other studies (e.g., Trickett & Trafton, 2007), they are complex enough to support multiple goals and to require some simplification or inference generation.

Content familiarity, we argue, will support the likelihood that viewers will generate inferences from the data for several reasons. First, content familiarity will have an influence on the viewers' comprehension goals. For example, viewers may attempt to confirm or disconfirm relationships that they expect to find. When data are complex, those *expected* relationships are more likely to be simpler than the actual data in the graph. In the graphs in Fig. 1, a viewer might expect and wish to confirm that high-ability third graders read faster than low-ability third graders in the data. Although it may seem like a relatively simple inference from the bar graph in Fig. 1, this goal requires mentally averaging the three bars representing words of low-, middle-, and high-familiarity words for high-ability readers and comparing those averages to those for low-ability readers. Or it requires checking and confirming that low-ability readers are systematically slower than high-ability readers for all three groups of bars. That is, identifying the main effect requires mental computation while keeping track of information already encoded such as the fact that the dark bars represent low reading skill individuals. Second, content familiarity knowledge may help viewers keep track of information in the service of the mental computation required in inference generation. Content familiarity may help viewers keep track of that information because they have expectations about which bars are likely to be associated with which condition. Finally, content familiarity may help the viewer identify potential errors. Viewers might double-check results of their computation if the results did not match their expectations.

Graph comprehension skills should support making such inferences for a different reason. Specifically, graph comprehension skills are required to understand how to mentally transform the data to generate the relevant inferences for particular graph formats. High-skilled graph viewers may know how to average across one variable to compute a main effect, for example, but low-skilled graph viewers may not. A consequence of their ability to mentally manipulate information in the graph would be that the presentation format should have *less* impact on comprehension for high-skilled viewers than low-skilled viewers. That is, high graph skilled individuals may have the ability make the appropriate inferences regardless of format. An alternative possibility is that because graphicacy skills include the ability to automatically associate visual patterns with interpretations (e.g., Pinker, 1990), highly skilled/experienced individuals may actually be *more* affected by presentation format. Specifically, viewers may have expectations that multivariate line graphs are intended to convey interactions, and multivariate bar graphs are intended to convey categorical differences.

This line of reasoning suggests a pattern of results that might be the opposite of many interactive models of comprehension: High-skilled graph viewers may be more, rather than less, affected by format. The current study considers both possible interactions of graphicacy and format on viewers' interpretations of data. The nature of the interaction has implications for graph and display design: Does graphic format matter more for novice viewers or for experts?

In sum, the primary goal of the present study is to examine the influence of two kinds of top-down knowledge on graph comprehension: topic familiarity and graph comprehension skills. We predict that both content familiarity and graph comprehension skills may affect inference generation. A secondary goal of our study is to consider how bottom-up processes, too, affect inference generation in graph comprehension.

In graph comprehension, the initial step is that visual elements (e.g., the symbols, colors, types of lines, shape fills) are identified and grouped together into chunks (Pinker, 1990), and these visual chunks influence viewers' interpretations of the data. As discussed below, bottom-up factors such as format (line or bar graph) influence the nature of those visual chunks. Specifically, the display is chunked based on the Gestalt principles of proximity, good continuity, and similarity (Pinker, 1990). We acknowledge that format is not strictly a bottom-up factor because skilled viewers may have knowledge of these formats, but in the context of this study we focus on the visual characteristics of line versus bar graphs and their influence on comprehension.

Previous research establishes that, for simple data, bar graphs and line graphs facilitate the comprehension of different information in fact-retrieval tasks. Viewers are faster at reading individual data points when viewing bar graphs compared to line graphs, and they are faster at making trend judgments when viewing line graphs compared to bar graphs (Simcox, 1984). Likewise, viewers can more accurately identify individual data points from bar graphs than from line graphs (Carswell & Wickens, 1987; Carswell, 1992). Furthermore, individuals are more likely to spontaneously make discrete comparisons (i.e., x_1 is greater than x_2) when viewing data in bar graphs and more likely to describe trends (i.e., as x increases y increases) when viewing line graphs in open-ended description tasks (Carswell, Emery, & Lonon, 1993; Zacks & Tversky, 1999). In fact, even when two discrete data points are plotted in a line graph, viewers sometimes describe the data as continuous. For example, a common interpretation of a line graph depicting height of males and females might be, ''The more male a person is, the taller he/she is'' (Zacks & Tversky, 1999).

Shah, Mayer, and Hegarty (1999) extended studies of bar versus line graphs work to characterize how Gestalt principles might affect comprehension of more complex graphs depicted in high school social studies textbooks. One main conclusion from this work can be illustrated by consideration of the graphs in Fig. 1. In the bar graph, the proximity principle predicts that for bar graphs a viewer would encode the grouped sets of bars representing levels of word familiarity (low, medium, and high). In the line graph, the principle of good continuity suggests that individuals would encode three visual chunks formed by the lines representing reading skill (low, medium, and high). Viewers' descriptions, if based on the visual chunks, would differ depending on format. For the bar graph, a visual chunk-based description may be, ''For the least familiar words, children with high reading skills are much faster than children with low reading skills; for medium familiar words, children with high reading skills are moderately faster than children with low reading skills; and for highly familiar words, children with high reading skills are only a little faster than children with low reading skills.'' Thus, descriptions focus on the effect of the variable plotted on the *z*-axis legend (reading skill) on read-aloud time for each set of bars on the *x*-axis (word familiarity). For ease we refer to this type of description as a *z–y* interaction because it refers to the *z-y* relationship (where ''*z*'' is the variable depicted in the legend) and how it is moderated by the variable depicted along the *x*-axis. For line graphs, a visual chunk-based description might be, ''As word familiarity increases, reading time is slightly faster for high-skilled readers, moderately faster for medium-skilled readers, and much faster for low-skilled readers.'' For simplicity, we refer to this as an *x–y* interaction because it describes the *x–y* relationships and how it is moderated by the ''*z*'' variable. Although these descriptions are, technically, descriptions of interactions, we note that they also require very little processing of the data depicted. As such, for the novice graph viewer such a description does not necessarily imply a deep understanding of an interaction, but it may reflect merely a minimally digested interpretation. Indeed, our previous work suggested that viewers who provided such descriptions were focusing on the surface features and could not recognize the same data when plotted differently (Shah & Carpenter, 1995).

The earlier work on line and bar graph comprehension outlined above primarily examined how format affects fact-retrieval processes in graph comprehension. The current study addresses how format might affect inference generation such as the identification of main effects in the graphs in Fig. 1. In the bar graph, the main effect of familiarity for read-aloud time (the *x–y* main effect) is computed by mentally averaging the three groups of bars representing low, medium, and high reading skill. The grouping of the bars (using the Gestalt principle of proximity) can support comprehension in two ways: First, each group may be encoded as a single entity and one might easily visually extract a trend. Second, the grouped bars may reduce working memory load in mentally computing averages because the three bars to be averaged in each set are together. In a line graph, computing the *x–y* main effect requires ignoring the highly salient interaction depicted in the three lines. Thus, we predict that viewers may be more likely to compute and describe *x–y* main effects when viewing bar graphs than when viewing line graphs.

Generating *z–y* main effect may be less affected by format. For the bar graph in Fig. 1, the *z–y* main effect is computed by mentally averaging the three black bars, the three gray bars, and the three white bars. The visual cues provided by the colors of the bars (Gestalt principle of similarity) may help reduce working memory load in keeping track of the bars to be mentally averaged. Computing the *z–y* main effect for line graphs is supported by the fact that the three sets of points to be mentally averaged are connected in lines. Thus, one might expect that both bar graphs and line graphs, to some extent, support interpretation of this main effect.

In sum, we make the following predictions regarding the influence of bottom-up characteristics on the interpretation of line and bar graphs: Viewers will provide *x–y* interaction descriptions more frequently for line graphs and for bar graphs, and *z–y* interaction descriptions approximately equally for bar graphs and line graphs. We also predict that viewers will be more likely to make inferences about *x–y* main effects for bar graphs than for line graphs.

To examine the influence of format, content knowledge, and graph skills on the comprehension of bar and line graphs, we presented different groups of participants with line graphs or bar graphs depicting the same data. We tested participants on an independent measure of graph comprehension skill. To assess the effect of content knowledge, we varied familiarity of the content depicted in graphs rather than using an individual differences approach. Thus, half the graphs depicted familiar data to this subject population, and half depicted unfamiliar data.

2. Method

2.1. Participants

Fifty-five undergraduates from the University of Michigan, Ann Arbor, participated in the study. Of those 55, 10 participated in exchange for 1 h of credit toward fulfillment of course requirements. The remaining 45 received $15 dollars compensation. All students were proficient in English and had normal or corrected-to-normal vision. They signed up either through an independent website or via experimenter recruitment e-mails. Twenty-six participants were assigned to high-skilled group, and 29 to the low-skilled group, based on their performance on a graph skills test.

2.2. Design and materials

We used a mixed design, with one within-subjects variable (content: familiar vs. unfamiliar) and two between-subjects variables (format: line vs. bar, and graphicacy: high vs. low).

We created 14 data sets for this study; seven of them depicted familiar data and the other seven depicted unfamiliar data. Each data set was plotted in bar graph and line graph form. All data depicted information about psychological content that was typically familiar or unfamiliar to undergraduates in our sample. An example of data with familiar content was the graph with the variables ''read aloud time,'' ''familiarity of word,'' and ''reading ability.'' An example of an unfamiliar data set was one with the variables ''level of monitoring,'' ''race'' and ''punishment.'' To ensure that our initial assessments of familiarity were accurate, an independent group of six raters, similar to our participant pool, rated the familiarity of graphs on a 10-point scale. Although six is a small number of raters, the responses were highly consistent, likely because we predesigned the familiar graph names to be highly familiar and vice versa. We asked participants two questions about the graphs: the degree to which the variable names are familiar to them, and the degree to which they had specific predictions about what the data should look like. Raters judged the familiar graphs as more familiar ($M = 6.07$) than the unfamiliar graphs [$M = 4.45$; $t(1,5) = 4.56$, $p = .006$]. They also claimed to have expectations about what data would look like given familiar variable names ($M = 0.60$) compared to unfamiliar variable names [$M = 0.29$; $t(1,5) = 5.27$, $p = .003$].

Each graph was presented on an individual sheet of paper with a paragraph that put the variables in a context. The paragraph described only the design of a study that led to the data in the graph, but it did not describe the outcome of the study. Following is an example of a paragraph:

> Researchers were interested in how reading skill and the word's familiarity influence children's reading aloud times. Third-grade children were designated as either high-, medium-, or low-skill readers based on their scores on a reading test. Children read-aloud lists of highly familiar, moderately familiar, or low familiar words. Highly familiar words are words that are frequently encountered by children in third grade. Moderately familiar words are words that are encountered by children on a somewhat regular basis. Unfamiliar words are words that are not frequently encountered by children in third grade. Researchers measured the amount of time it took children to pronounce the words. The results of the study are presented in the graph.

Regardless of format condition, each graph contained two independent variables and one dependent variable. The three variables in each of the 14 graphs were logically related to one another and there were three levels for each independent variable. All the variables could be interpreted as continuous (e.g., 10, 20, 30; high, medium, low), making bar or line graphs acceptable formats. The variables, order of presentation, data points, and content descriptions were identical for both line and bar formats, making format the only difference. Familiar and unfamiliar data sets were controlled for the relative magnitude of interactions and main effects so that familiarity would not be confounded by complexity of data.

Based on extensive piloting, we developed a 13-question multiple choice graph skills test for this study. One question was adapted from the individual-differences measure used by Gattis and Holyoak (1996), and one item based on an item in The Test of Graphing in Science (TOGS; McKenzie & Padilla, 1986). We created the other 11 items, and these items required the interpretation of 2–4 variable data plotted in line graphs, bar graphs, and scatterplots (see e.g., Fig. 3). The median on the graph skills test was an 8, with a mean of 8.3 and a range of 4–12. Participants with scores higher than 8 were considered ''high'' in graphicacy skills ($n = 26$), and participants with scores of 8 and below were considered ''low'' in graphicacy skills in our analysis ($n = 29$).

2.3. Procedure

Participants were randomly assigned to format condition (line or bar). For each trial, participants saw a graph on a computer monitor and provided written descriptions on a sheet of paper provided for them. Participants were informed that there were no right or wrong responses, but the experimenters were interested in what the participant believed ''to be the main point or points presented in the graph.'' The experimenter emphasized that each participant should provide approximately 2–4 sentences about each graph and focus on what the subject believed was ''the most important information in the graph.'' The graphs were presented in the same prerandomized order for each participant. Following the graph task, the

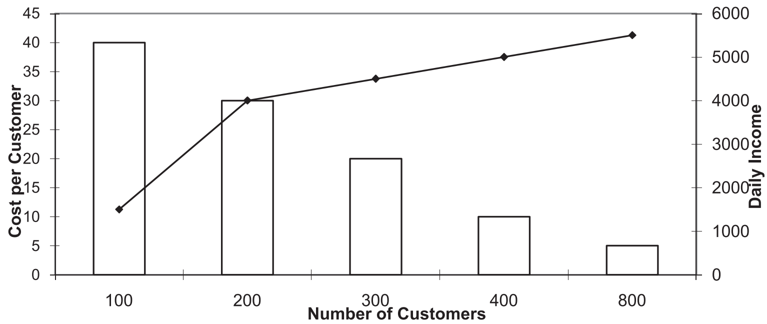

Fig. 3. Sample item from the graph skills test used to classify whether participants are high or low in graphicacy skills.

participants completed the graph skills test and a brief graph usage and mathematical background survey.

Although there was no formal time limit, participants knew that they were participating in an approximately 1½ h study, and that there were questionnaires and a graph skills test to follow the description task. Thus, the experiment orally instructed participants to pace themselves accordingly and not to spend too much time on any one graph.

2.4. Coding verbal descriptions

We coded whether or not each description contained an *x–y* interaction, a *z–y* interaction, an *x–y* main effects inference, a *z–y* main effects inference, and ''other'' information. An *x–y* interaction was a description of the *x–y* relationship and how it is moderated by the third, ''*z*'' variable depicted in the legend. A *z–y* interaction was a description of the *z–y* relationship and how it is moderated by the variable plotted along the *x*-axis. An *x–y* main effect described the *x–y* relationship ignoring the third, *z*-variable. A *z–y* main effect described the *z–y* relationship ignoring the third, *x*-variable. The reader should refer to the introduction for examples of these coding categories. In addition to the categories above, we included an ''other'' category. The other category consisted of comments such as describing colors/shades of bars, making comments about implications of data, making comments about the novelty or ''obviousness'' of data, and so forth. Two raters coded all the data and there was an agreement on 95% of trials. Disagreements were resolved by discussion with a third rater, the first author. It should be noted that it is possible for a description to have more than one type of information in it (e.g., an *x–y* interaction and a *z–y* main effect), in which case both codes were included.

3. Results

Tables 1 and 2 present the results of repeated-measures ANOVA analyses, one for interaction descriptions, and one for main effect inferences. ''Other'' descriptions were made on

Table 1
Interaction description (ANOVA results)

	F Test	MSE	Effect Size (partial η^2)
Format (Fo)	0.25	0.01	0.01
Type of interaction (Int)	52.0**	2.2	0.51
Familiarity (Fam)	3.9+	0.05	0.07
Graph skills (GS)	0.35	0.02	0.01
Format × Int	8.3*	0.36	0.14
Format × familiarity	3.7+	0.04	0.07
Format × graph skills	0.62	0.04	0.01
Int × Fam	7.9	0.16	0.13
Int × GS	1.3	0.06	0.03
Familiarity × graph skill	0.55	0.04	0.07
Fo × Int × fam	3.2+	0.06	0.06
Fo × Int by GS	0.13	0.00	0.00
Fo × fam by GS	0.06	0.00	0.00
Int × Fam × GS	0.49	0.01	0.01
Fo × Int × Fam × GS	2.0	0.04	0.04

$^{**}p < .01$; $^{*}p < .05$; $+p < .1$.

Table 2
Main effect description (ANOVA results)

	F Test	MSE	Effect Size (partial η^2)
Format (Fo)	2.6	0.18	0.05
Type of main effect (ME)	7.7**	0.13	0.13
Familiarity (Fam)	58.7**	1.6	0.54
Graph skills (GS)	5.8*	0.40	0.10
Format × type of main effect	0.61	0.01	0.01
Format × familiarity	4.5*	0.14	0.08
Format × graph skills	1.8	0.13	0.03
ME × Fam	3.2+	0.04	0.06
ME × GS	1.1	0.02	0.02
Familiarity × graph skill	4.9*	0.12	0.09
Fo × ME × fam	1.6	0.02	0.03
Fo × ME by GS	1.1	0.02	0.02
Fo × fam by GS	4.2*	0.12	0.08
ME × Fam × GS	0.27	0.00	0.01
Fo × ME × Fam × GS	0.75	0.01	0.01

$^{**}p < .01$; $^{*}p < .05$; $+p < .1$.

nearly half of the trials (0.46) and typically co-occurred with other codes. However, these ‘‘other’’ comments did not differ as a function of format, graph skills, or familiarity of data and are not discussed further in this article because the focus of our current analysis is to address our empirical questions regarding the bottom-up influence of format, and the top-down influence of familiarity and graphicacy skills on viewers' interpretations of data.

3.1. Interaction descriptions

As discussed in the Introduction, interaction descriptions may be focused on the *x–y* relationship (as a function of *z*, the variable in the legend) or focused on the *z–y* relationship (as a function of *x*, the variable along the *x*-axis). As expected, viewers provided more *x–y* interaction descriptions for line graphs ($M = 0.33$) than bar graphs ($M = 0.23$), and more *z–y* interaction descriptions for bar graphs ($M = 0.10$) than line graphs [$M = 0.04$; $F(1, 51) = 8.3$, $p < .05$]. Furthermore, there was a bigger difference in the proportion of *x–y* interactions and *z–y* interactions for line graphs ($M_{\text{difference between } x\text{-}y \text{ and } z\text{-}y \text{ interactions}} = 0.29$) compared to bar graphs ($M_{\text{difference between } x\text{-}y \text{ and } z\text{-}y \text{ interactinos}} = 0.12$; see Fig. 4). These results demonstrate that the bottom-up features of bar graphs and line graphs have the expected (based on Gestalt principles) influence on the interpretation of the data presented.

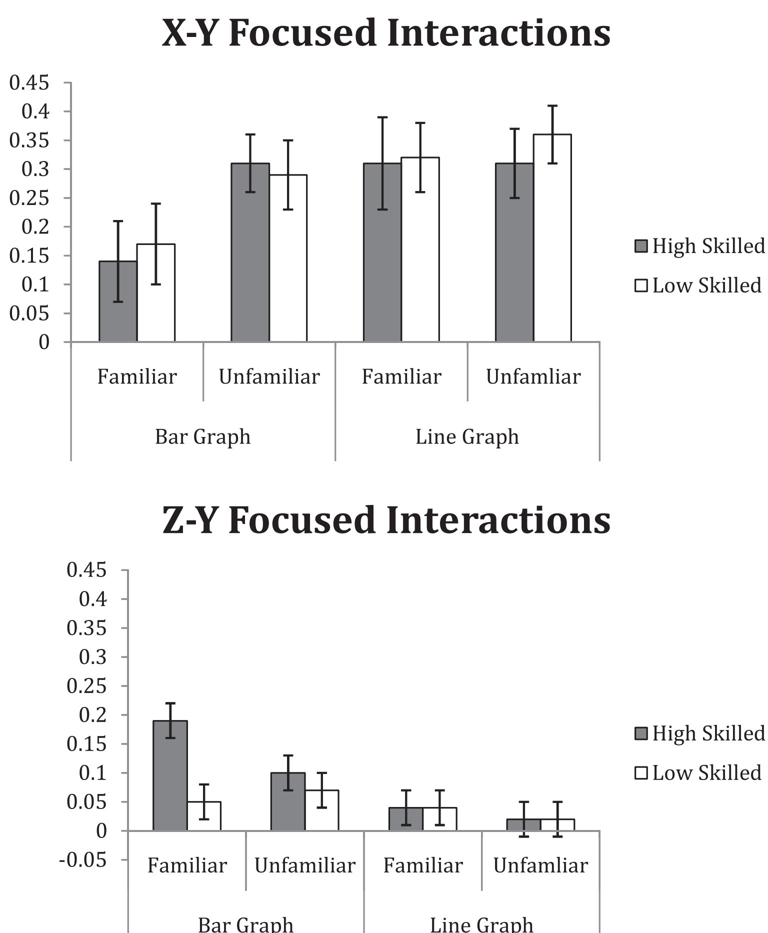

Fig. 4. *x–y* focused (top) and *z–y* focused (bottom) interaction description rate (proportion). Error bars depict standard error.

In addition, the results suggest that line graphs exert a stronger influence on types of descriptions than bar graphs because only the *x–y* interaction is visually salient in line graphs. Our previous research suggests that this type of highly biased description corresponds to a very limited understanding of the data; viewers often cannot even recognize the same data when the *x*- and *z*-variable locations are switched (Shah & Carpenter, 1995). Bar graphs, by contrast, seem to provide a bit more flexibility with respect to emphasis on the *x–y* and *z–y* interactions, perhaps because there are two different Gestalt principles operating in parallel: proximity (supporting *z–y* interaction descriptions) and similarity of bars (supporting *x–y* interactions).

There was a marginal effect of familiarity on the interpretation of interactions consistent with an interactive model of graph comprehension. Overall, viewers were somewhat more likely to describe interactions for unfamiliar graphs (0.19) than for familiar graphs [0.16; $F(1, 51) = 3.9$, $p = .053$]. There was also an interaction such that viewers were most likely to describe *x–y* interactions for unfamiliar data [$F(1, 51) = 7.9$, $p < .05$; see Fig. 4]. This result suggests that when data are unfamiliar, individuals are more likely to rely on bottom-up characteristics of graphs (hence describing *x–y* interactions). The marginal interaction of format by familiarity by type of interaction [$F(1, 51) = 3.2$, $p = .08$] further bolsters this claim; individuals describe *x–y* interactions most of all when data are unfamiliar and when individuals are viewing line graphs that make that interaction most salient.

There were no effects of graph skills on interaction descriptions. This result suggests that for interpretations that require few mental transformations, there are few effects of graphicacy skills.

3.2. Main-effects inferences

A major goal of our current study was to examine the factors that influence inference generation in the context of interpreting multivariate bar and line graphs. Overall, individuals made more *z–y* focused main effects inferences ($M = 0.19$) compared to *x–y* focused main effects inferences [$M = 0.14$; $F(1, 51) = 7.7$, $p < .01$]. This is consistent with the prediction that because both bar graphs and line graphs group data by the variable in the legend, viewers may be more likely to make comparisons across that variable.

As predicted, viewers were more likely to make main effects inferences when viewing bar graphs than when viewing line graphs, but only for familiar data (see Fig. 5). This result suggests that viewers may be most likely to make these inferences when they expect them, and when the format makes it easier for them to make those inferences. Contrary to predictions, there was no interaction such that the effect of format was larger for *x–y* main effects than *z–y* main effects. The original prediction was that bar graphs would support *x–y* main effects inferences compared to line graphs, but both line and bar graphs would support *z–y* main effects inferences. One possible explanation for this lack of interaction is that the most salient information in a line graph is the *x–y* interaction (Carpenter & Shah, 1998; Shah & Carpenter, 1995). Describing the *x–y* interaction requires several sentences or phrases, and thus once described, viewers may judge that they have reported enough ‘‘important’’ information as instructed. If the *x–y* main effect is more salient for bar graphs and is easily

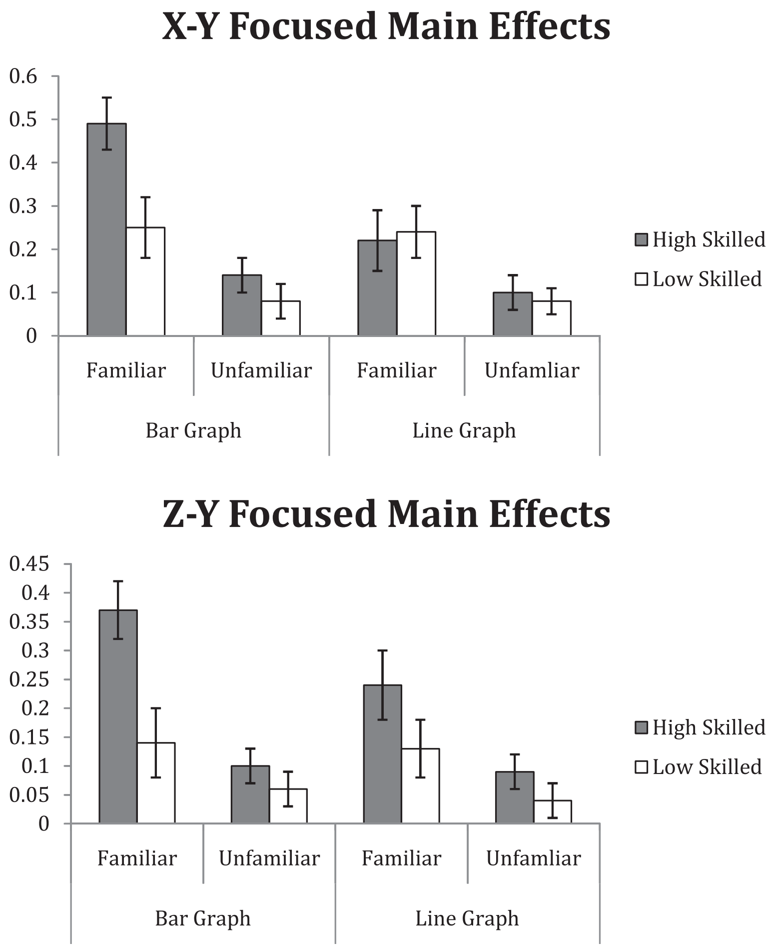

Fig. 5. *x–y* focused (top) and *z–y* focused (bottom) main-effect inference descriptions. Error bars depict standard error.

provided in a single phrase, viewers may be more likely to also report an *x–y* main effect. To test the prediction that both line and bar graphs make computing *z–y* main effects salient might require a task that is not open-ended.

In addition to the interaction of format and familiarity, there was a main effect of familiarity such that viewers made more main effects inferences when viewing familiar data ($M = 0.26$) than when viewing unfamiliar data [$M = 0.09$; $F(1, 51) = 58.7$, $p < .01$]. There was also an interaction of familiarity and graph skills such that only high-skilled graph viewers drive the familiarity effect [$F(1, 51) = 4.9$, $p < .05$]. In fact, there was a three-way interaction: High-skilled graph viewers largely described main effects for only familiar data depicted in bar graphs [$F(1, 51) = 4.2$, $p < .05$]. That is, individuals only made inferences when they had expectations about main effects, they had the requisite knowledge to mentally compute the main effects, and the format provided additional support to make those inferences.

4. Discussion

4.1. Summary of results

The current study has two main results. First, prior content knowledge as well as graphicacy skills had a top-down influence on graph interpretation. Specifically, viewers made more main effects inferences when viewing familiar data compared to when viewing unfamiliar data. Additionally, high-skilled graph viewers were more likely to make main effect inferences than low-skilled graph viewers. Second, these three factors were shown to interact: High-skilled graph viewers were most likely to make main inferences when they expected those inferences based on their prior knowledge/familiarity with the data *and* the format supports them (i.e., bar graph). This result provides evidence that interactive models of visual display comprehension share characteristics with models of text and discourse comprehension (e.g., Hegarty, 2005; Kintsch, 1988).

Our study goes beyond previous studies in that it specifically demonstrates that interactive models of display comprehension apply to graphical displays of quantitative information (e.g., Freedman & Shah, 2002; Hegarty, 2005; Kriz & Hegarty, 2007; Roth & Bowen, 2003; Shah, 2001). Furthermore, this study also suggests that familiarity (prior content knowledge) may have a somewhat different effect on comprehension than graphicacy skills. Neither familiarity, graphicacy, nor format alone supports inference generation in our study. Familiarity may lead individuals to look for main effect trends that they expect, but only individuals with relatively high graphical skills have enough knowledge about graphs to compute and describe those main effects. Furthermore, they only do so when the format supports making those inferences.

The interaction of graph skills, format, and familiarity suggests the possibility that in some cases, graph skills may actually lead to greater effects of format rather than lesser effects of format and is inconsistent with interactive models that predict high skills correspond to less reliance on presentation format. It is possible, as discussed in the Introduction, that high graph skilled viewers may actually differentiate between formats based on their graph schemas, which include knowledge about what kinds of formats convey specific types of data. An alternative, perhaps less interesting, explanation is that even the high-skilled participants in our sample did not have knowledge to compute main effects from line graphs or that the format did not induce them to think that main effects were important in these graphs, but only could or were inclined to do so from bar graphs. We acknowledge that it is difficult to draw strong conclusions about the nature of display and skill interactions and that our results differ from other work showing greater effects of display type for novices (Lowe, 1993). Nonetheless, it is possible that skill may sometimes correspond to greater differentiation between formats rather than less differentiation between formats. If so, this differentiation should most likely occur for open-ended tasks in which viewers are selecting what they consider to be the important information and using their knowledge about format to make such judgments. If viewers were asked to answer specific fact-retrieval questions, it is possible that high-skilled graph viewers may be more likely to do so regardless of format.

A second main result of this study is that the bottom-up features of line graphs and bar graphs have a substantial influence on viewers' interpretations of data, including the nature of their inferences. In our study, viewers were more likely to describe *z–y* interactions when viewing bar graphs than when viewing line graphs, and they were more likely to describe *x–y* interactions when viewing line graphs than when viewing bar graphs. The difference between focus on *x–y* and *z–y* interaction descriptions was greater for line graphs than for bar graphs. Finally, viewers were also more likely to make main effects inferences for bar graphs than for line graphs. The differences between bar graphs and line graphs can be explained by differences in the visual chunks formed by the graphs as predicted by Gestalt principles of proximity, similarity, and good continuity.

4.2. Theoretical implications

The current study adds to the growing body of evidence that display comprehension, like text comprehension, is an interactive process. The differential effect of content familiarity and graphicacy skills suggests a modification of interactive models of display comprehension such as the Hegarty (2005) model. Specifically, prior knowledge should not be considered a single factor, but rather multiple factors. In this study we found that both content/topic familiarity as well as graphicacy skills influenced comprehension, and, further, suggested that these types of knowledge may have had a role in different aspects of the interpretation process. Prior content knowledge may affect a viewer's goals, support a viewer's ability to keep track of mental computations, and help him or her assess the likelihood that his or her interpretation is accurate. Graphicacy skills may support an individual's knowledge about format and how to make relevant inferences from different graphic formats. Future distinctions may also need to be made, for example, between different kinds of graphicacy (e.g., statistical literacy versus knowledge about format), and different kinds of content knowledge (knowledge about experimental design in psychology versus knowledge about content like reading skills). In general, however, diagram comprehension models, like text comprehension models, should begin to make distinctions between broad classes of prior knowledge and their influences on comprehension.

A related theoretical contribution of the current paper is the consideration of how prior knowledge affects different specific component processes of graph interpretation. Prior interactive models of graph interpretation, including our own (Freedman & Shah, 2002; Shah, 2001), make general claims about the top-down influence of prior knowledge on graph comprehension. Future modeling work should specify *how* prior knowledge influences different component processes of graph comprehension, including encoding of visual displays, mapping of visual features to verbal referents, mental calculations of data, and evaluation of one's interpretation of data.

4.3. Implications for the display of multivariate data

The bottom-up influences on graph interpretation have several implications for multivariate data display. When the graph designer wishes to highlight an interaction for the viewer that is present in the data, and in particular has in mind a specific interpretation, it would be useful to present that interaction in a line graph in which the x–y relationship is of interest, and the z-value is a moderator. If, by contrast, a graph designer wishes to highlight main effects, if the x- and z-values have similar status in the interpretation of data, or if the expectation is that different viewers may have different goals in interpreting the data, bar graphs provide greater flexibility. It is interesting to note that in general, viewers prefer bar graphs to line graphs, perhaps in part because of their flexibility and relative ease of interpretation (see e.g., Fortin, Hirota, Bond, O'Conner, & Col, 2001). Math education researchers also propose teaching bar graph comprehension prior to introducing line graphs (Friel, Curcio, & Bright, 2001), a suggestion supported by this research.

Top-down effects further suggest that the choice of how to plot data depends not only on the type of information to be communicated but also on the prior beliefs and expectations of the graph viewer as well as their graphicacy skills. High-skilled graph viewers were able to make main effects inferences when viewing bar graphs that supported their ability to make the necessary mental computations, but not when viewing line graphs. Low-skilled graph viewers, however, could not make such inferences, even when viewing bar graphs. For low-skilled viewers, it may be useful to present not only interactions but also to explicitly represent important to-be-communicated inferences.

Although our study did not address this issue, it is possible that giving explicit inferences is only beneficial to low-skilled graph viewers. Indeed, just as high-skilled readers benefit from forming their own inferences about text (e.g., McNamara, Kintsch, Songer, & Kintsch, 1996), low-skilled graph viewers may benefit from making their own inferences about data. And, it may be that the computational processing required to make inferences could actually support individuals' ability to learn and remember information presented in graphs. Indeed, in another as yet unpublished study conducted in our laboratory, we found that graph formats that required more complex processing to comprehend data (in this case, graphs in which lines were not labeled, but instead, had legends) actually led to better memory and more complete comprehension of data than formats that did not require substantial mental computations (Shah, Freedman, & Miyake, in preparation).

4.4. Future research

Given the powerful impact of Gestalt principles of proximity, similarity, and good continuity on interpretations given to data, it might be valuable to conduct further research on the role of different visual factors on how individuals form visual chunks based on different visual dimensions and the role of that chunking process in viewers' interpretations of data. Cleveland & McGill (Cleveland, 1993; Cleveland & McGill, 1985) provide an excellent analysis of the relative accuracy in using different visual dimensions (e.g., color, shading, area, volume, and so forth) to display quantity, which has been further modified by others

(Carswell, 1992; Simkin & Hastie, 1986; Spence & Lewandowsky, 1991). We argue that similar analyses need to be conducted on visual dimensions that may or may not affect visual chunking during graph comprehension. Does proximity have a greater impact than color similarity, for example, or vice versa? What factors influence the relative salience of proximity and color (e.g., the size of the bars)? Can something be too salient (i.e., bias the viewer and lead him or her to ignore other features)? Without systematic empirical analyses the graph designer is dependent on intuition. With such an analysis, it may be possible to provide more direct advice with respect to the use of different graphic features.

One cautionary note with respect to a systematic analysis of different visual features and how they affect chunking is that the answer may depend on the task of the graph viewer (Carswell & Wickens, 1987). One reason we have focused on open-ended interpretation tasks is that we can make assessments about the salience of information in different graphs based on what viewers describe as the ''most important'' information.

A second cautionary note with respect to future studies is that studies should be conducted with graphs of varying complexity. We have found that interpretation processes differ even between very simple graphs and slightly more complex ones such as those in the current study. Research from other laboratories suggests that the interpretation processes differ even more when data are highly complex. Trafton and Trickett (2001), for example, argue that spatial processing is much more relevant for more complex data than for simple data.

5. Conclusion

The main conclusion from this study is that an individual's interpretation of a bar or line graph is a function of the format, the viewer's familiarity with the content depicted in a graph, and the viewer's graphicacy skills. This study is part of a growing body of research that examines how individuals interpret meaningful data in realistic contexts (e.g., Peebles & Cheng, 2003; Roth & Bowen, 1999, 2003; Trickett, Trafton, Saner, & Schunn, 2007). Such research should lead to further understanding of both bottom-up *and* top-down processes in graph comprehension.

Acknowledgment

This research was supported by Office of Naval Research contract N00014-02-1-0279 (Priti Shah and Eric Freedman).

References

Canham, M., & Hegarty, M. (in press). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*.

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

Carswell, C. M., Emery, C., & Lonon, A. M. (1993). Stimulus complexity and information integration in the spontaneous interpretation of line graphs. *Applied Cognitive Psychology*, *7*, 341–357.

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

Cleveland, W. (1993). *Visualizing data*. Murray Hill, NJ: AT&T Bell Laboratories.

Cleveland, W. S., & McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. *Science*, *229*, 828–833.

Fortin, J. M., Hirota, L. K., Bond, B. E., O'Conner, A. M., & Col, N. F. (2001). Identifying patient preferences for communicating risk estimates: A descriptive pilot study. *BMC Medical Informatics and Decision Making*, *1*, 2.

Freedman, E. G., & Shah, P. (2002). Toward a model of knowledge-based graph comprehension. In M. Hegarty, B. Meyer, & N. Hari Narayanan (Eds.), *Diagrammatic representation and inference* (pp. 8–31). Berlin: Springer Verlag.

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

Gattis, M., & Holyoak, K. J. (1996). Mapping conceptual to spatial relations in visual reasoning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *22*, 231–239.

Hegarty, M. (2005). Multimedia learning about physical systems. In R. E. Mayer (Ed.), *Handbook of multimedia* (pp. 447–465). New York: Cambridge University Press.

Kintsch, W. (1988). The role of knowledge in discourse comprehension. A construction-integration model. *Psychological Review*, *95*, 163–182.

Kriz, S., & Hegarty, M. (2007). Top-down and bottom-up influences on learning from animations. *International Journal of Human Computer Studies*, *65*, 911–930.

Larkin, J., & Simon, H. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, *11*, 65–99.

Lowe, R. C. (1993). *Successful insructional diagrams*. Florence, KY: Routledge.

Mautone, P., & Mayer, R. (2007). Cognitive aids for guiding graph comprehension. *Journal of Educational Psychology*, *99*, 640–652.

McKenzie, D. L., & Padilla, M. J. (1986). The construction and validation of the Test of Graphing in Science (TOGS). *Journal of Research in Science Teaching*, *23*, 571–579.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*, 1–43.

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: Erlbaum.

Ratwani, R., & Trafton, J. (2008). Shedding light on the graph schema: Perceptual features versus invariant structure. *Psychonomic Bulletin & Review*, *15*, 757–762.

Ratwani, R., Trafton, J., & Boehm-Davis, D. (2008). Thinking graphically: Connecting vision and cognition during graph comprehension. *Journal of Experimental Psychology: Applied*, *14*, 36–49.

Roth, W.-M., & Bowen, G. M. (1999). Complexities of graphical representations during ecology lectures: An analysis rooted in semiotics and hermeneutic phenomenology. *Learning and Instruction*, *9*, 235–255.

Roth, W.-M., & Bowen, G. M. (2003). When are graphs worth ten thousand words? An expert-expert study. *Cognition and Instruction*, *21*, 429–473.

Shah, P. (1997). A model of the cognitive and perceptual processes in graphical display comprehension. In M. Anderson (Ed.), *Reasoning with diagrammatic representations* (pp. 94–101). Menlo Park, CA: AAAI Press.

Shah, P. (2001). Graph comprehension: The role of format, content, and individual differences. In M. Anderson, B. Meyer, & P. Olivier (Eds.), *Diagrammatic representation and reasoning* (pp. 173–185). Berlin: Springer Verlag.

Shah, P., & Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology: General*, *124*, 43–61.

Shah, P., Freedman, E., & Vekiri, I. (2005). The comprehension of quantitative information in graphical displays. In P. Shah & A. Miyake (Eds.), *The Cambridge handbook of visuospatial thinking* (pp. 426–476). New York: Cambridge University Press.

Shah, P., Mayer, R. E., & Hegarty, M. (1999). Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *Journal of Educational Psychology*, *91*, 690–702.

Simcox, W. A. (1984). A method for pragmatic communication in graphic displays. *Human Factors*, *26*, 483–487.

Simkin, D. K., & Hastie, R. (1986). An information processing analysis of graph perception. *Journal of the American Statistical Association*, *82*, 454–465.

Smith, L. D., Best, L. A., Stubbs, D. A., Archibald, A. B., & Roberson-Nay, R. (2002). Constructing knowledge: The role of graphs and tables in hard and soft psychology. *American Psychologist*, *57*, 749–761.

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

Trafton, J. G., Kirschenbaum, S. S., Tsui, T. L., Miyamoto, R. T., Ballas, J. A., & Raymond, P. D. (2000). Turning pictures into numbers: Extracting and generating information from complex visualizations. *International Journal of Human Computer Studies*, *53*, 827–850.

Trafton, J. G., & Trickett, S. B. (2001). A new model of graph and visualization usage. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the 23rd annual conference of the Cognitive Science Society* (pp. 1048–1053). Mahwah, NJ: Erlbaum.

Trickett, S. B., & Trafton, J. G. (2007). ''What if...'': The use of conceptual simulations in scientific reasoning. *Cognitive Science*, *31*, 843–875.

Trickett, S. B., Trafton, J. B., Saner, L., & Schunn, C. D. (2007). ''I don't know what's going on there'': The use of spatial transformations to deal with and resolve uncertainty in complex visualizations. In M. C. Lovett, & P. Shah (Eds.), *Thinking with data* (pp. 65–86). New York: Erlbaum.

Tufte, E. R. (2001). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.

Zacks, J., Levy, E., Tversky, B., & Schiano, D. (2002). Graphs in print. In M. Anderson, B. Meyer, & P. Olivier (Eds.), *Diagrammatic representation and reasoning* (pp. 187–206). London: Springer-Verlag.

Zacks, J., & Tversky, B. (1999). Bars and lines: A study of graphic communication. *Memory & Cognition*, *27*, 1073–1079.