

Graphs as Aids to Knowledge Construction: Signaling Techniques for Guiding the Process of Graph Comprehension

Priti Shah, Richard E. Mayer, and Mary Hegarty
University of California, Santa Barbara

Graphical displays are frequently used to express quantitative information in texts, but viewers are sometimes unable to comprehend and learn the relevant information. According to a cognitive analysis, graph interpretation involves (a) relatively simple pattern perception and association processes in which viewers can associate graphic patterns to quantitative referents and (b) more complex and error-prone inferential processes in which viewers must mentally transform data. Experiment 1 establishes that graphs can be redesigned to improve viewers' interpretations by minimizing the inferential processes and maximizing the pattern association processes required to interpret relevant information. In Experiments 2 and 3, the researchers isolated one important factor that affects viewers' interpretation (i.e., the perceptual organization of the information in graphs). If relevant quantitative information is perceptually grouped to form visual chunks (because relevant data points are either connected in line graphs or close together in bar graphs), then viewers describe relevant trends. If relevant information is not perceptually grouped, viewers are less likely to comprehend relevant trends.

Graphs are ubiquitous in textbooks, scientific journals, and the popular print media (Kaput, 1987; Kosslyn, 1994; Mayer, 1993b; Mayer, Sims, & Tajika, 1995), but people can have difficulty interpreting and explaining quantitative information depicted in graphs (Culbertson & Powers, 1959; Guthrie, Weber, & Kimmerly, 1993; Leinhardt, Zaslavsky, & Stein, 1990; Mayer, 1993b; Mayer et al., 1995; Shah & Carpenter, 1995; Vernon, 1946, 1950). For example, consider the use of a bar graph in a middle school American history textbook (Armento, Nash, Salter, & Wixson, 1991), as shown in Figure 1. The graph is in a chapter on events leading to the Civil War and is intended to show that in the decades preceding the war, the North was changing from a rural to an urban society, whereas the South remained largely a rural society. The conclusion to be drawn from this graph, albeit implicitly, is that the difference in demographic changes between the two regions was a factor leading to the Civil War. Unfortunately, when we asked people to tell us what this graph says in Experiment 1, they seldom reported

on differences in the rate of urbanization in the North and South. Although the information was presented, they failed to construct a representation of the graph that contained the trends that were intended by the author.

The graph in Figure 1 and other graphs that are used in textbooks and other texts may be difficult to interpret because although they are technically accurate, they are not designed to effectively communicate the main point of the text (Kosslyn, 1994). Implicit in much of graphical design seems to be an *information-telling* perspective (cf. Bereiter & Scardamalia, 1982) in which any technically correct graph is adequate for conveying relevant quantitative information. The graph in Figure 1 does include the relevant information because a reader who compares the urban and rural populations can compute that during the period from 1820 to 1860, the percentage of urban population grew greatly in the North but not in the South. However, in the studies presented here, readers rarely extracted that information from the graph.

When the same data are presented in a way that makes the trends more salient and requires less mental computation by the readers, such as in Figure 2, people are much more likely to spontaneously describe the differences in urbanization trends. In Figure 2, the data plotted in Figure 1 have been reformatted in light of three design principles, which are based on cognitive analyses of graph comprehension (discussed later). First, to organize the data to reflect across-years trends, the graph format has been changed from bars that are grouped by year to lines that connect points representing population across years. Second, to reduce the need for mental computation, the y-axis for the graph on the left has been changed from presenting data on an absolute scale (i.e., population in millions) to a percentage scale (i.e., based on the percentage of rural population). Third, to reduce complexity, one frame containing 12 data points in Figure 1 has been converted into two frames, each contain-

Priti Shah, Richard E. Mayer, and Mary Hegarty, Department of Psychology, University of California, Santa Barbara.

This research was supported by a postdoctoral research fellowship from the James S. McDonnell Foundation. Priti Shah served as the postdoctoral fellow, and Richard E. Mayer and Mary Hegarty served as cosponsors. This research was later supported by a fellowship from the National Academy of Education and Grant N00014-98-0350 from the Office of Naval Research.

We thank Melody Carswell and William Cleveland for comments on an earlier version of this manuscript and Aislinn Evans and Destiny Shellhammer for assistance with verbal description reliability analysis.

Correspondence concerning this article should be addressed to Priti Shah, who is now at the Department of Psychology, University of Michigan, 525 East University, Ann Arbor, Michigan 48109-1109. Electronic mail may be sent to priti@umich.edu, to mayer@psych.ucsb.edu, or to hegarty@psych.ucsb.edu.

Northern and Southern Populations, 1820-1860

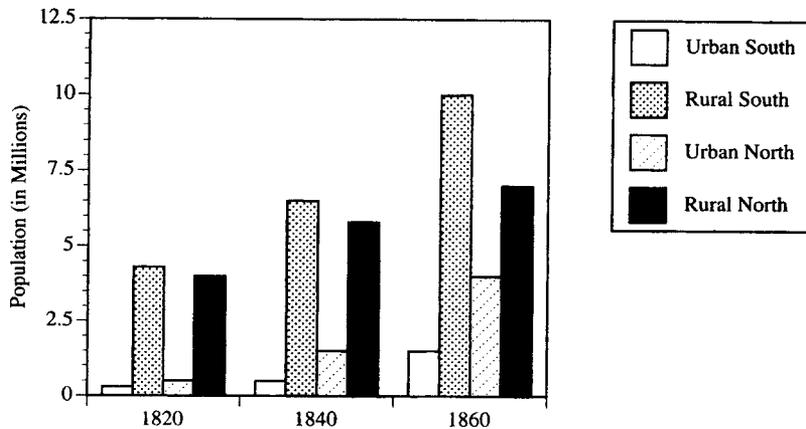


Figure 1. A graph adapted from an eighth-grade history textbook that presents the population in the North and South leading up to the Civil War. We predicted that viewers would have difficulty retrieving the main point of the history lesson from this graph: that before the Civil War, the North was becoming less rural, or more urban, whereas the South continued to have a fairly high rural population. From *A More Perfect Union* (p. 306), by Armento, Nash, Salter, and Wixson, 1991, Boston: Houghton Mifflin. Copyright 1991 by Houghton Mifflin Company. Adapted with permission.

ing 6 data points, in Figure 2; each of these frames represents a single quantitative comparison.

Although Figure 1 and Figure 2 present the same information—thus satisfying the information-telling function—they differ in the cues they give to the reader about how to interpret and build a mental representation of the graph. Just as it is not enough for written text to be technically correct in terms of spelling, punctuation, and grammar (Bereiter & Scardamalia, 1982; Mayer, 1994; Mayer, Bove, Bryman, Mars, & Tapangco, 1996), a *knowledge construction* viewpoint of graph interpretation requires that it is not enough that graphs are merely technically correct in presenting relevant information. Instead, different factors, including the format of the display, influence what knowledge viewers are readily able to construct from the display. The goal of the

current research is to use a cognitive model of the graph interpretation process to identify the factors that influence how well viewers can interpret important information from a graph. One consequence of our empirical analysis will be a set of principled guidelines for the display of quantitative information in instructional contexts (Mayer, 1993a).

Graphs in Social Science

In this article, the graphs we use are taken from social science textbooks. Much of the previous research on graph comprehension has focused on simple interpretation tasks (e.g., fact retrieval) in abstract or arbitrary contexts (for some exceptions, see Carswell, Emery, & Lonon, 1993; Leinhardt et al., 1990). However, people who comprehend

Northern and Southern Populations, 1820-1860

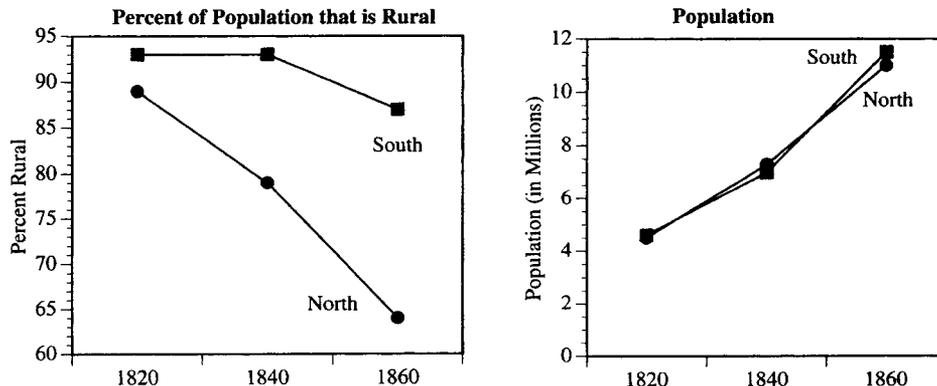


Figure 2. A set of graphs that depict the identical quantitative information as in Figure 1, but that we expected would help people identify the important quantitative information.

and construct graphs in arbitrary contexts may be unable to apply these skills to the interpretation of graphs of meaningful data (Leinhardt et al., 1990). Furthermore, in meaningful contexts, such as in social science, graphs typically depict complex data sets that are meant to convey comparisons of trends and relationships such as in Figure 1 or are intended to provide support for assertions in the text. In this context, the readers' tasks include understanding the link between the graph and the text (Hunter, Crismore, & Pearson, 1987) or understanding what the data imply (Lehrer & Romberg, 1996; Scardamalia, Bereiter, & Lamon, 1994). A major goal of the current research is to understand how people interpret trends and relationships in graphs of meaningful data, such as those in social science textbooks, and how characteristics of the graphical display influence these processes.

Model of Graph Interpretation

According to the dominant cognitive model, graph comprehension entails three major, intertwined component processes (Bertin, 1983; Carpenter & Shah, 1998; Cleveland, 1994; Cleveland & McGill, 1984, 1985; Lohse, 1993; Pinker, 1990; Shah, 1995; Shah & Carpenter, 1995). First, viewers must encode the visual array and identify the important visual features (e.g., a straight line slanting downward). Second, viewers must identify the quantitative facts or relations that those features represent (e.g., a decreasing linear relationship between x and y). Finally, viewers must relate those quantitative relations to the graphic variables depicted (e.g., population vs. year). These three processes are incremental and interactive such that viewers sequentially encode a portion of the visual pattern, identify what quantitative fact or function it implies, and relate it to its graphic referents (Carpenter & Shah, 1998).

In this research, we focus on the second process, that is, identification of the quantitative facts or trends from a graph. Previous models of graph interpretation propose that there are two kinds of processes that are involved in identifying relevant quantitative facts or trends from a graph (Carpenter & Shah, 1998; Casner & Larkin, 1989; Pinker, 1990; Shah & Carpenter, 1995). In one type of process, viewers encode a portion of a visual pattern, or a *visual chunk*, and can automatically associate that visual chunk to a quantitative fact or relationship. For example, a visual chunk in a line graph may be one of the lines. In interpreting a line graph, the viewer may associate each line, such as a straight line or an upwardly curved line, to a specific quantitative relation such as a linear or exponential function (Carpenter & Shah, 1998; Shah & Carpenter, 1995). When relevant quantitative information is directly available in the visual chunks, then pattern perception and association processes are sufficient to interpret relevant quantitative information, and viewers are likely to be able to interpret that information accurately and quickly (Casner & Larkin, 1989; Shah & Carpenter, 1995).

However, there are some situations in which viewers must rely on complex inferential processes. First, a viewer may not have the knowledge to associate the visual chunk to the quantitative referents (e.g., the viewer does not know that a straight line implies a linear relationship). Second, indi-

vidual visual chunks may not be associated with the relevant quantitative referents (e.g., the viewer must compare or relate information in two different visual chunks, say, by computing the relative difference between a number of lines on a graph). Such complex inferential processes involve quantitatively transforming the information in the display (e.g., mentally transforming from a linear to logarithmic scale or calculating the difference between two or more data points; Cleveland, 1984, 1985). When these processes are required to interpret some or all of the information in a graph, viewers have more difficulty, and their interpretations may be inaccurate or incomplete (Casner & Larkin, 1989; Hegarty, Carpenter, & Just, 1991; Shah & Carpenter, 1995).

According to this model of graph comprehension, it may be possible to improve viewers' abilities to construct a representation of the quantitative information in a graph that is consistent with the interpretation intended by the writer by maximizing the pattern perception and association processes and by minimizing the complex inferential processes required to interpret the relevant quantitative information from a graph. Furthermore, to minimize inferential processes, the relevant quantitative information must be explicitly represented in the visual chunks that are encoded by the graph viewers. Finally, viewers' knowledge about how different visual features correspond with quantitative conclusions may influence whether they are able to use pattern perception and association processes to comprehend relevant information and, therefore, influence their interpretations of data.

In this research, we begin to specify how the characteristics of the graphical display influence which visual chunks might be encoded by the graph viewer, and how this might affect viewers' interpretations of, and memory for, information presented in a graph. The graph characteristics that we examine are the graphic format (line graph or bar graph), the scale on the y -axis (absolute or percentage), the complexity of an individual graph, and the perceptual organization of information on a graph (e.g., the grouping of the bars on a bar graph).

In Experiment 1, we redesigned graphs so that viewers would be more likely to comprehend and learn the information that is the relevant point of the text. In this study, we simultaneously altered a number of variables, all of which played the same role: to reduce the amount of complex inferential processes required to comprehend the relevant quantitative comparison by placing the relevant information in the visual chunks. Previous research has examined the role of one or more variables, such as graph format or the use of color, on viewers' identification or confirmation of simple facts or on their interpretations in relatively unfamiliar domains (e.g., Shah & Carpenter, 1995). This experiment goes beyond that previous research to examine how the characteristics of the graph influence viewers' interpretations in the context of familiar, meaningful data and relevant interpretation tasks such as those involved in the interpretation of social science textbook graphs.

In Experiment 2, we systematically examined two properties of the graphs—their format and scale—to further specify how these two factors influence what visual chunks

are encoded by the viewer and, therefore, what kinds of interpretations are given to different graphs. Finally, Experiment 3 tested the idea that the perceptual organization of the data, or where information is placed within a particular graphic format, rather than the graphic format per se, is the major variable that influences viewers' interpretations. Together, these studies demonstrate that if relevant quantitative information is perceptually grouped to form visual chunks (because relevant data points are either connected in line graphs or close together in bar graphs), then viewers describe relevant trends. If relevant information is not perceptually grouped, viewers are less likely to comprehend relevant trends.

Experiment 1

Experiment 1 was designed to establish that the characteristics of a graph can have a profound influence on viewers' interpretations of data, even in the context of complex, meaningful data and relevant tasks such as those involved in interpreting social science textbook graphs. In this study, we examined people's descriptions of graphs taken from middle school history textbooks (original graphs such as the one in Figure 1), as well as informationally equivalent graphs that were created by explicitly depicting the relevant quantitative trends in the visual chunks (revised graphs such as the one in Figure 2).

We predicted that viewers would be better able to construct a representation of the relevant quantitative trends from the revised graphs than from the original graphs. The revised graphs were designed to maximize the role of pattern perception and association processes and minimize the complex, inferential processes by explicitly representing the relevant information in the visual chunks. We tested this hypothesis by examining the kinds of verbal descriptions that viewers spontaneously gave to the original and revised graphs.

Method

Participants, design, and materials. Sixteen undergraduates from the University of California, Santa Barbara participated in this study for course credit. Each participant was tested on three original graphs and three revised graphs so that graph format was a within-subject variable. The three original graphs (including the graph shown in Figure 1) were adapted directly from middle school U.S. history textbooks. The three graphs shared the characteristic that they did not present the relevant quantitative information explicitly, which, in each case, involved understanding how the relative percentages of different categories changed with time. The original graphs were different from one another in several respects because they were adapted directly from textbooks (e.g., two were bar graphs, and one was a divided bar chart; they contained different numbers of data points).

The revised graphs, including the graph shown in Figure 2, were created by making three changes to the original graphs; the changes played the role of minimizing the number of mental computations required to interpret the trends in the graphs. First, for all of the revised graphs, we split the data set into two separate graphs, with fewer variables in each graph so that viewers could focus on the relevant variables (such as trends in the North and South over three

time periods, in the case of Figure 2). Second, we transformed the dependent variable in one graph to a percentage, rather than absolute scale, so that viewers need not mentally compute the relevant percentage trends (such as the percentage of rural population). Finally, for two of the three graphs, we changed the format to a line graph to explicitly represent the relevant trends in the visual chunks.

Two graph sets were created to ensure counterbalancing of the underlying data sets and variables used for original and revised graphs. The two graph sets conveyed the same data but differed in which data sets were presented in the original format and which data sets were presented in the revised format. The data sets for the three original graphs in Set 1 (and the three revised graphs in Set 2) involved the number of urban and rural inhabitants in the North and the South in three time periods ($2 \times 2 \times 3$), the number of immigrants from five nations in three time periods (5×3), and the median income of men and women workers in six time periods (2×6). The data sets for the three revised graphs in Set 1 (and the three original graphs in Set 2) involved the number of tons of domestic and imported grain consumed in the United States and the USSR in three time periods ($2 \times 2 \times 3$), the number of animals on the prairie for five animal types in three time periods (5×3), and the unemployment rate for African Americans and White Americans in six time periods (2×6). The data sets for the original and revised graphs within each set were isomorphic, and they depicted data that were linear transformations of one another (e.g., the number of immigrants from five nations in three time periods is isomorphic to the number of animals on the prairie from five animal types in three time periods); to keep the data isomorphic in the two conditions, the data do not necessarily represent accurate historic information. In addition to the materials already mentioned, we created a brief questionnaire about participants' mathematics and statistics courses, Scholastic Aptitude Test (SAT) scores, and their experience with different graphic formats.

Procedure. Participants were tested individually. At the beginning of the study, participants were briefly instructed in the description and statement-verification tasks. They were given one practice trial in which they were asked to give a brief description of a relatively simple line graph and practice the statement-verification task.

After the instructions, participants were presented with a graph (or graph pair). They described each graph or graph pair aloud while viewing the graph, and their descriptions were tape-recorded. This procedure was repeated for each of the six graphs. We used the same, prandomized order of presentation of the graphs for all participants. At the end of the study, viewers were asked to fill out the brief questionnaire.

Coding verbal descriptions. Each of the six graph descriptions produced by each participant was categorized as expressing (a) a within-year comparison, (b) an across-years trend description, (c) a mixed description, or (d) other. A protocol was classified as a within-year comparison if the student compared the value of two or more variables for a given year, such as "In 1860 there are more people in the rural South than in the rural North." A protocol was classified as an across-years trend description if the student described differences within a single variable across two or more years, such as "The number of people in the rural North decreased dramatically between 1820 and 1860, but the number of people in the rural South did not decrease as much." A protocol was categorized as mixed if it contained both within-year and across-years comparisons. A protocol was classified as other if it did not contain within-year or across-years comparisons. Such descriptions included nonquantitative information such as descriptions of the shading of the different bars. On the basis of this analysis, four scores were computed for both the original and revised trials for

each participant: (a) the proportion of trials that were within-year comparisons, (b) the proportion of trials that were across-years comparisons, (c) the proportion of trials that were mixed, and (d) the proportion of trials that were other. A second rater coded the data for 4 participants (24 trials), and then the two raters discussed any discrepancies. Then, the second rater coded the data for the 12 remaining participants (72 trials). The two raters agreed on the classification of the protocols on 83% of the trials. In the final analysis, classifications made by the first rater were used. Because of the relatively high reliability for verbal description classification, only one rater coded the data for Experiments 2 and 3.

In this analysis, we calculated the proportion of descriptions, for each graph type, that are coded as being in each of the four categories. The average number of descriptions that each participant made for each graph type can be calculated by multiplying the proportion by 3 (because each participant described three original graphs and three revised graphs).

Results and Discussion

Overall, the results are consistent with the knowledge construction hypothesis, which states that the kind of information that is retrieved from graphs is highly dependent on the format and characteristics of the graphical display. As shown in the left panel of Figure 3, people were more likely to describe trends across time when viewing revised graphs ($M = .81$) than when viewing original graphs ($M = .44$), $F(1, 15) = 22.09$, $MSE = .05$, $p < .01$. In contrast, as shown in the right panel of Figure 3, people were more likely to describe facts within a time period when viewing original graphs ($M = .29$) than when viewing revised graphs ($M = .04$), $F(1, 15) = 10.39$, $MSE = .05$, $p < .01$. Finally, a number of descriptions were categorized as mixed ($M = .20$), but the proportions of these types of descriptions were not influenced by the graph format.

An example of a typical across-years trend description of the revised graph in Figure 2 indicates that the revised graph helped people identify and describe the main point of the lesson. As one student described:

The percent of the population that is rural decreases over time, but there is a greater decrease for the northern population that

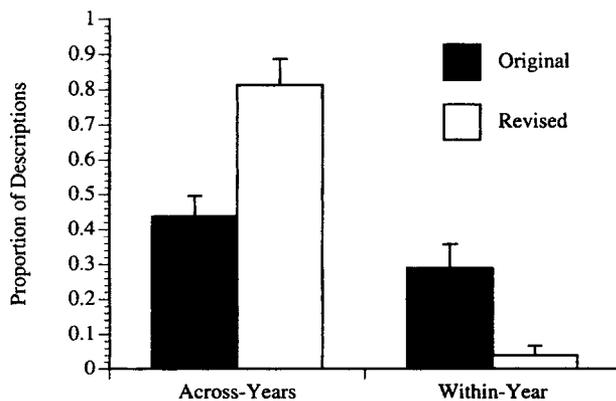


Figure 3. Proportion of viewers generating across-years and within-year descriptions for the original and revised graphs in Experiment 1. Viewers provided more across-years trend descriptions for revised graphs than original graphs, and more within-year descriptions for original graphs than revised graphs.

is rural than the southern population. The total population increases in both the North and the South almost equally over time.

In this case, the student focused on an important difference in demographic trends over time between the North and South. This comparison is central to the author's contention that growing demographic differences between the North and South helped lead to the Civil War. Another description of the revised graph in Figure 2 also indicates an understanding of the different demographic trends in the North and South:

The first graph the southern population was about the rural population was about really high 90%, it went down a little but it was still up like 88%, whereas the northern population started at 90% but it went down till 1860. It gradually decreased all the way down till approximately 65%.

Two examples of a typical within-year comparison, provided for the original graph in Figure 1, indicate that in viewing this type of graph, people often failed to find the author's intended message, instead focusing on comparisons with each time period:

In 1820, the greatest population was the rural South, and then the rural North, and then the urban North, then the urban South. And in 1840 same, but greater amounts in the population. And in 1860, it went rural South, rural North, urban North, and urban South. Same . . . [reads values of the points].

The population in 1820 the rural South is more than the North, and it's the still is the same in 1840, the rural South is larger than the north but the rural North in 1860 the urban South has increased again but not as large amount as the urban North which is a large, has grown up to about 4, 4.5 million when it started out less than .5 million . . . [repeats the same information].

Although descriptions of original graphs were typically longer than descriptions of revised graphs (because they often included descriptions of multiple individual data points), viewers often failed to clearly convey the main point of the lesson.

Summary

In summary, the results of Experiment 1 demonstrate that by replotting the data in graphs, one can guide a viewer's cognitive processing of the graphs so that he or she is more likely to represent the data as the author intended—in this case, as a comparison of trends over time. Although it may appear from the results of this study that our revised line graphs are always better than the original bar graphs or divided bar charts, the benefit of a particular graph format is likely to be dependent on the task of the graph viewer (Carswell & Wickens, 1987). If the goal was for people to remember specific facts about relative differences between categories within a particular time period, then the original graphs from these studies may have been the best format because those facts may be retrieved by means of relatively simple pattern perception and association processes. However, in the history texts from which we took our original graphs, the goal of the lesson was not to teach people about specific quantitative facts but to help them understand the

relationship between certain trends and other historical events. In this case, the revised graphs, by explicitly depicting relevant quantitative trends in the visual chunks of line graphs, greatly improve viewers' understanding of the data.

Experiment 2

Although the results of Experiment 1 demonstrated that people abstracted different kinds of information from revised graphs than from original graphs, there were several differences between the original and revised graphs. In Experiment 2, we attempted to untangle two aspects of the change from original to revised graphs—the change from bars to lines and the change from absolute to percentage scale (and the simultaneous change from a single complex graph to two simpler graphs, necessary to keep the two versions informationally equivalent).

Our goal was to determine what role these two graph characteristics played in influencing people's interpretations of graphs. To separate the role of format and scale (and number of graphs) in the previous study, we compared viewers' interpretations of four kinds of graphs: line graphs and bar graphs that depicted information using an absolute scale on the y -axis, and line graphs and bar graphs that depicted information using percentage scale on the y -axis for one graph. Figures 1, 2, 4A, and 4B depict the four kinds of graphs. We predicted that both format and the scale/number of graphs factor should have an influence on viewers' interpretations because each of these factors influences what information is explicitly represented in the visual chunks. To test this prediction, we analyzed viewers' spontaneous verbal descriptions of graphs, as well as their accuracy in judging from memory whether statements about the data were true or false.

Method

Participants, design, and materials. Forty undergraduates from the University of California, Santa Barbara participated in this study for course credit. The experiment had a $2 \times 2 \times 4$ mixed factorial design with the first (within-subject) variable being format (bar or line) and the second (within-subject) variable being scale (one graph with absolute scale vs. two graphs, one of which had a percentage scale). For the third (between-subjects) variable, we plotted each of the eight data sets from Experiment 1 in each of the four possible forms (a total of 32 graphs) and divided them into four sets, each of which included two versions of graphs in each of the four possible forms. Each of these four sets was administered to 10 different participants. Note that absolute bar graphs correspond to the original graphs in Experiment 1 that contained three variables (such as Figure 1), and the percentage line graphs correspond to the revised graphs in Experiment 1 (such as Figure 2).

Procedure. The procedure was identical to Experiment 1 except that there were eight graphs, and viewers made true–false judgments for individual statements (rather than comparing two statements) in the memory task. The viewers' descriptions were coded as in Experiment 1.

Results and Discussion

The proportions of across-years and within-year descriptions produced for absolute bar (original), percentage bar,

absolute line, and percentage line (revised) graphs are shown in Figure 5. Consistent with Experiment 1, viewers were more likely to describe across-years trends when shown a revised (i.e., percentage line) graph ($M = .86$) than when shown an original (i.e., absolute bar) graph ($M = .52$), $F(1, 39) = 20.57$, $MSE = .11$, $p < .01$. Also, consistent with Experiment 1, viewers were more likely to describe comparisons within years when shown an original graph ($M = .33$) than when shown a revised graph ($M = .04$), $F(1, 39) = 21.90$, $MSE = .08$, $p < .01$.

A new question addressed in Experiment 2 concerns whether the major advantage of the revised graph over the original graph has to do with a change from absolute to percentage scale (and the change from a single graph to two separate graphs) or a change from bar to line format. As predicted, viewers were more likely to describe trends across time when interpreting line graphs ($M = .84$) than when interpreting bar graphs ($M = .46$), $F(1, 39) = 53.9$, $MSE = .10$, $p < .01$. By contrast, viewers were more likely to describe within-year comparisons when interpreting bar graphs ($M = .33$) than when interpreting line graphs ($M = .05$), $F(1, 39) = 29.1$, $MSE = .11$, $p < .01$. As in the previous study, viewers' interpretations of data are shown to be influenced by graphic format. Some graphic formats, such as line graphs, are better for depicting x – y trends, and in this study, the x – y trends are trends across time. Other graphic formats, such as bar graphs, are better for depicting differences within the clusters of bars, and for the graphs in this study, the clusters of bars represent individual years. In addition, line graphs are more biasing in the kind of interpretations they elicit: Almost all line graph descriptions were of historical trends, whereas bar graphs elicited descriptions of historical trends and within-year comparisons equally.

Viewers never mentioned scale in describing the graphs, except in referring to the names of the variables on the y -axis. Because of this, the interpretation of the coding of the descriptions was not sensitive to the difference between percentage and absolute scale.¹

¹ A second task, in which participants answered true–false questions about the graphs from memory, suggested that scale, too, may have had an influence on viewers' understanding of the data. For this task, participants were presented with two statements each about absolute facts (e.g., "In 1860, most people in the United States lived in the rural parts of the South"), absolute trends (e.g., "The number of people who live in the rural parts of the North increases with time"), percentage facts (e.g., "In 1820, a higher percentage of the population lives in the Rural North than in the Rural South"), and percentage trends (e.g., "The percentage of the population that is Rural decreases with time in the South"). Viewers were more accurate in answering questions about absolute information when graphs had absolute scale ($M = .74$) than when graphs had a percentage scale ($M = .64$), $F(1, 39) = 17.6$, $MSE = .03$, $p < .01$. Similarly, viewers were more accurate in answering questions about relative information when graphs had percentage scales ($M = .84$) than when the graphs had absolute scales ($M = .68$), $F(1, 39) = 42.4$, $MSE = .03$, $p < .01$. These results suggest that viewers had difficulty translating between percentage and absolute interpretations of data on the basis of their memory of the graph, perhaps because they did not include the scale information in their mental representations, as suggested by their descriptions.

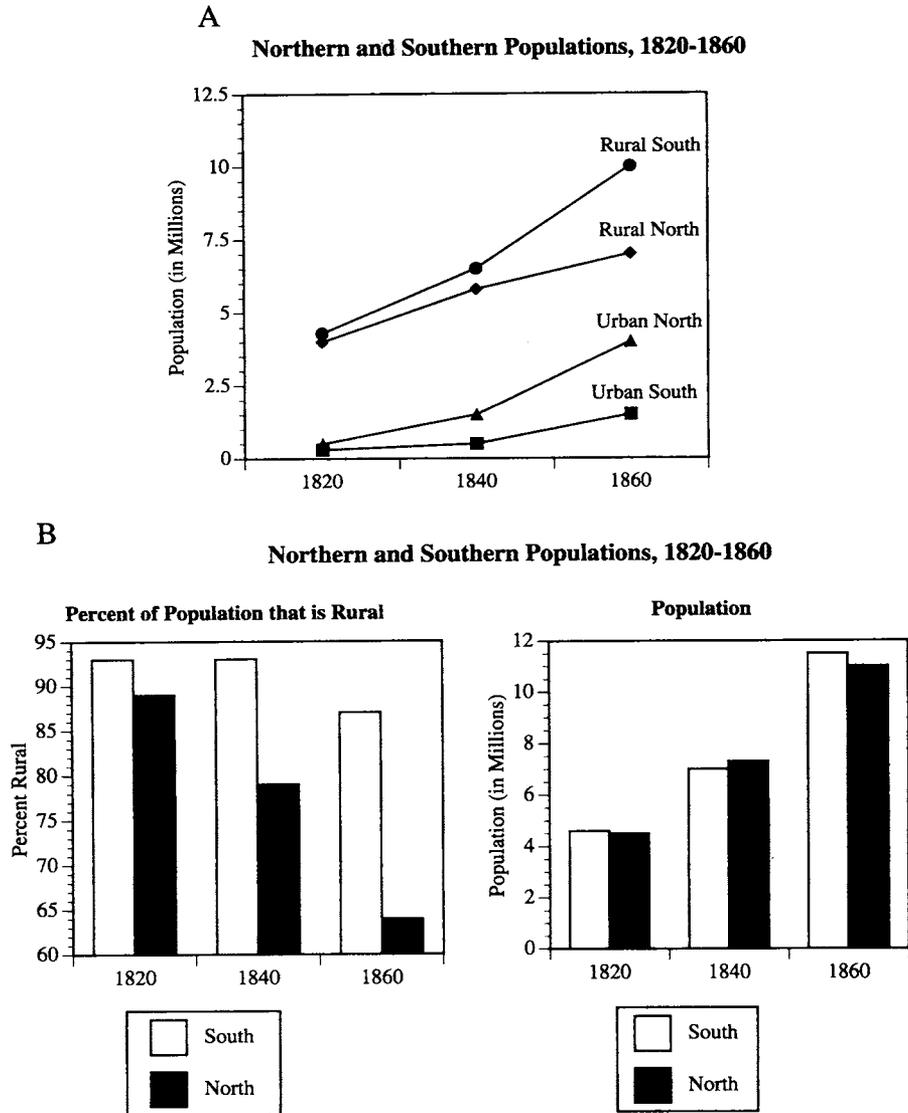


Figure 4. The two new graph types created for Experiment 2. The absolute line (4A) and percentage (4B) bar graph versions of the identical pre-Civil War population data used in Experiment 2.

Summary

Experiment 2 demonstrates that both format and scale influenced the relative ease and difficulty of interpreting different quantitative information from a graphical display. Second, the task of the graph viewer may interact with the role of graph characteristics. Specifically, line graphs may be better when the task is to describe trends, whereas bar graphs may be better when the task is to describe facts.

Experiment 3

Experiments 1 and 2 in this article demonstrate that the format of the graphical display influences the types of interpretations that viewers spontaneously give to data. The goal of Experiment 3 was to distinguish between two possible hypotheses about the role of format in the interpre-

tation of trends and facts from graphs that are consistent with the results of Experiments 1 and 2. One possibility, which we refer to as the *format-only hypothesis*, is that line graphs cue trends, whereas bar graphs cue facts.

A second possibility which is consistent with our cognitive model of graph interpretation, is the *perceptual organization hypothesis*. According to this hypothesis, viewers' interpretations are not influenced by format per se but rather are influenced by which quantitative trends are represented by visual chunks in the graph. We propose that when information is represented by a visual chunk, viewers can rely on pattern perception processes to interpret the trend and do not have to rely on complex inferential processes. Our definition of visual chunks is based on Gestalt principles of perceptual organization, such as connectedness and proximity. In the case of line graphs, the visual chunks are

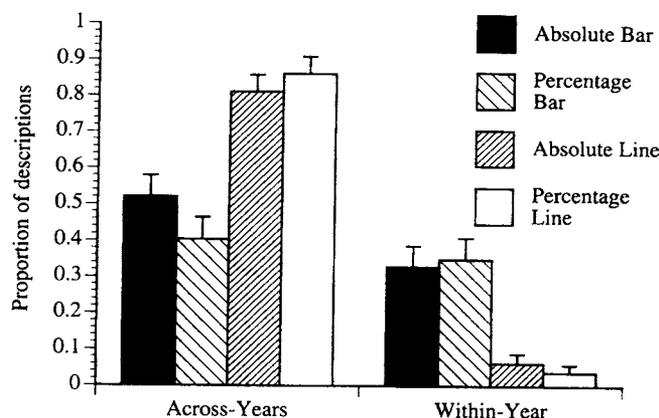


Figure 5. Proportion of viewers generating across-years and within-year descriptions for the four graph types in Experiment 2. Viewers gave across-years trend descriptions more for line graphs than for bar graphs, and more within-year descriptions for bar graphs than for line graphs.

the x - y lines that physically connect data points (Shah & Carpenter, 1995) on the basis of the Gestalt principle of connectedness. In the case of bar graphs, a group of bars that are close together (e.g., the group of bars for each year in Figure 1) is defined as a visual chunk on the basis of the Gestalt principle of proximity.²

In this experiment, we attempted to distinguish between the format-only and perceptual organization hypotheses by asking viewers to describe bar graphs and line graphs in which we varied the perceptual organization of information. According to the format-only hypothesis, line graphs (Figures 6A and 6C) should cue across-years trends, whereas bar graphs (Figures 6B and 6D) should cue within-year comparisons. According to the perceptual organization hypothesis, however, data points that are grouped, either by connectedness (in the case of line graphs) or proximity (in the case of bar graphs), should be described more readily. Thus, the line graph in Figure 6A and the bar graph in Figure 6B should cue across-years trends, and the line graph in Figure 6C and the bar graph in Figure 6D should cue within-year comparisons.

Method

Participants, design, and materials. Twelve undergraduates from the University of California, Santa Barbara participated in this study for course credit. Each participant was tested on two bar graphs with within-year visual chunks (bar-within), two bar graphs with across-years visual chunks (bar-across), two line graphs with within-year visual chunks (line-within), and two line graphs with across-years visual chunks (line-across), so graph format (line vs. bar) and visual chunk (within-year vs. across-years) were within-subject variables. To ensure counterbalancing, four equivalent sets of eight graphs were constructed, with 3 participants receiving each set. Thus, the experiment was based on a $4 \times 2 \times 2$ mixed design with the first (between-subjects) variable being which counterbalanced set was presented, the second (within-subject) variable being format, and the third (within-subject) variable being visual chunk.

The same eight data sets used in Experiment 2 were used in this study. We plotted each of the eight data sets in the four possible

forms, as shown in Figure 6. We then assigned the 32 graphs to each of the four conditions, with the constraint that each condition included two examples of each of the four graph types and one version of each data set. Viewers' descriptions were coded as in Experiments 1 and 2.

Results and Discussion

The proportions of across-years and within-year descriptions generated by the four graph types are shown in Figure 7. As predicted by the perceptual organization hypothesis, viewers were more likely to describe historical trends when across-years trends were explicitly represented in visual chunks ($M = .67$) than when within-year trends were represented in visual chunks ($M = .19$), $F(1, 11) = 17.9$, $MSE = .15$, $p < .01$. Format (bar vs. line), on other hand, did not have a significant influence on the proportion of across-years trend descriptions, as can be seen by comparing the two graphs in Figure 7, $F(1, 11) = 1.1$, $MSE = .10$, $p > .1$. There was, however, a significant interaction between format and visual chunks such that visual chunks defined by connectedness in line graphs had a bigger influence on interpretation than did the visual chunks defined by proximity in bar graphs, $F(1, 11) = 7.5$, $MSE = .09$, $p < .05$. This interaction is consistent with the results of Experiment 2: Overall, line graphs bias viewers' interpretations according to the position of the display more than bar graphs bias viewers' interpretations.

Consistent with the perceptual organization hypothesis, viewers were also more likely to make within-year, across category comparisons when that information was explicitly depicted in the visual chunks ($M = .29$) than when it was not ($M = .04$), $F(1, 11) = 11.0$, $MSE = .07$, $p < .01$. Format per se had only a marginal influence on the proportion of within-year trend descriptions, as can be seen by comparing the left panels of Figure 7A and Figure 7B, $F(1, 11) = 3.66$, $MSE = .05$, $p = .08$, and there was no significant interaction between format and visual chunks, $F(1, 11) = 0.47$, $MSE = .04$, $p > .10$.

A substantial proportion of descriptions were coded as other ($M = .24$) descriptions. There was a larger number of these other descriptions, including merely perceptual information (such as "the bar on the left has stripes") or very minimal content information (such as "the highest value is 90"), than in the other experiments, perhaps because of some viewers were not sure how to describe the relatively unusual new graphs that were designed for this study (e.g., Figure 6B and particularly 6C, in which we violated a standard graphic convention by connecting categorical information). Thus, there were more other descriptions for those two types of graphs ($M = .40$) than for the two commonly used formats ($M = .08$), $F(1, 11) = 19.8$, $MSE = .06$, $p < .01$.

² Overall, connectedness is a more powerful Gestalt principle and takes priority over the principle of proximity. Thus, points that are connected by lines form visual chunks even when other points are close together. The relative strength of the connectedness principle might also explain why line graphs are more biased than bar graphs in Experiment 2.

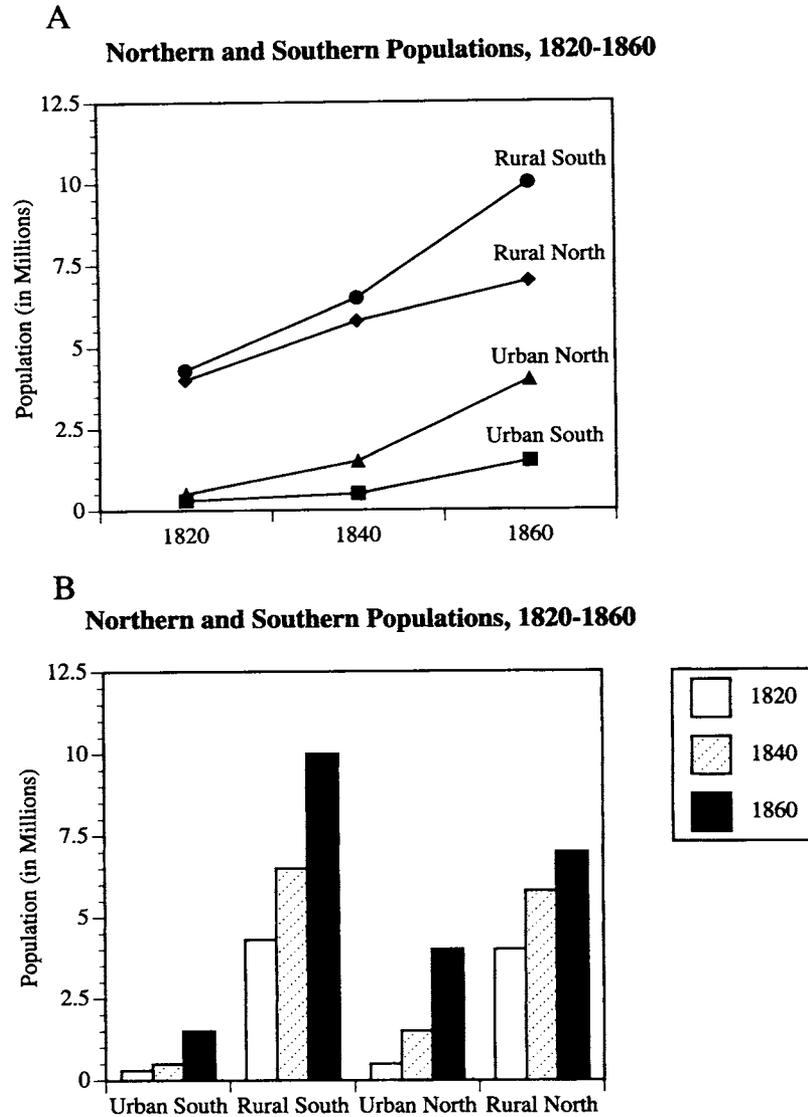


Figure 6. Figures 6A and 6B depict line (6A) and bar (6B) graph versions of the pre-Civil War population data, which, by proximity and connectedness, have visual chunks that emphasize historical trends; these graphs are examples of the across-years visual chunk graphs used in Experiment 3. Figures 6C and 6D (*opposite*) depict line (6C) and bar (6D) graph versions of the pre-Civil War population data, which, by proximity and connectedness, have visual chunks that emphasize within-year comparisons; these graphs are examples of the within-year visual chunk graphs used in Experiment 3. Figure 6D from *A More Perfect Union* (p. 306), by Armento, Nash, Salter, and Wixson, 1991, Boston: Houghton Mifflin. Copyright 1991 by Houghton Mifflin Company. Adapted with permission.

Finally, the proportion of mixed descriptions ($M = .17$) did not differ as a function of the graphic format or perceptual organization, $F(1, 11) = 0.31$, $MSE = .03$, $p > .10$.

Summary

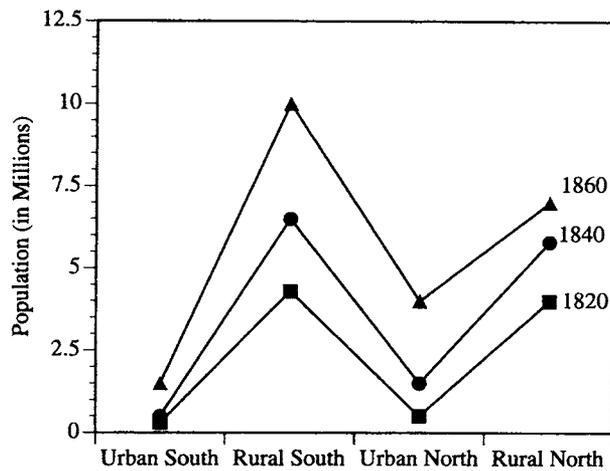
Experiment 3 demonstrated that the kinds of perceptual inferences that can easily be made from graphical displays is

dependent on what information is grouped together in a display (e.g., across-years vs. within-year chunks) rather than graph format (e.g., bar vs. line).

General Discussion

We began with the finding—replicated across three experiments—that people often fail to glean the author's intended message when they are asked to describe what they see in

C
Northern and Southern Populations, 1820-1860



D
Northern and Southern Populations, 1820-1860

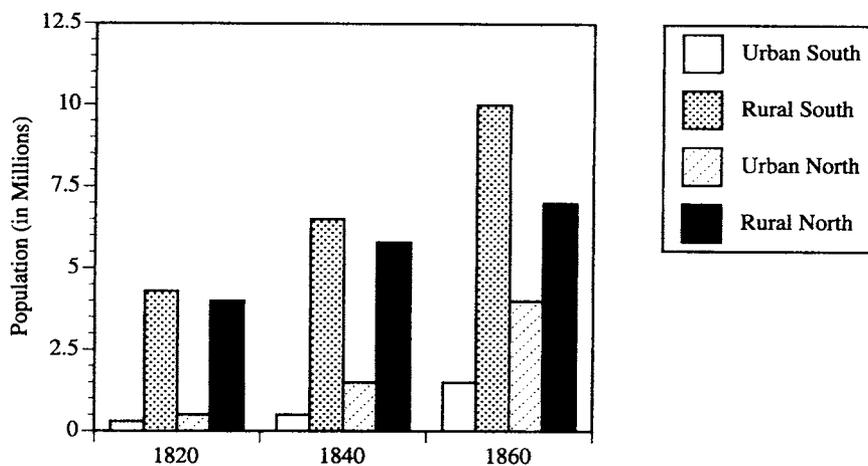


Figure 6 (continued)

social science graphs. Specifically, our participants often failed to describe trends across years when they were given original graphs reproduced from middle school textbooks on U.S. history. However, when we revised the graphs in light of three basic cognitive principles, people were far more likely to describe across-years trends. In Experiment 1, we converted original graphs into revised graphs by altering the graph format from bars to lines, by altering the scale of the dependent measure from absolute to percentage, and by splitting the information into two graphs. This study established that the graphic format can have a profound influence on viewers' interpretations, even in familiar domains and for relatively complex interpretation tasks.

The goal of Experiment 2 was to begin to specify which aspects of the changes to the revised graphs caused changes in viewers' interpretations. To untangle two of the con-

founded variables—format and scale—we conducted a second study. In this study, we examined whether the advantage of revised over original graphs was due to altering the graph format, the scale (and with that, the number of graphs), or both. People were more likely to describe trends across years when given line graphs than when given bar graphs. They did not pay attention to the graphs' scale in their descriptions. Thus, graphic format plays a role in viewers' representations of graphs: The change from bars to lines provides cues to the reader to focus on across-years trends in reading the graph when year is plotted along the x-axis, and the scale of the graph focuses viewers on one type of quantity.

The change from bar to line format in Experiments 1 and 2 confounded two theoretically interesting variables: the format per se (lines vs. bars) and the perceptual organization

of the visual chunks (i.e., the way that bars were clustered into groups or points were connected by lines). To untangle these two aspects of format, we conducted a third study in which we examined whether the advantage of lines over bars was due to format per se or to the perceptual organization of the visual chunks. People were more likely to describe trends across years when given graphs containing across-years visual chunks than when given graphs containing within-year visual chunks; this same pattern occurred for both line graphs and bar graphs, indicating that format per se was not the major factor influencing how people interpreted graphs. These results help to isolate the perceptual organization of visual chunks as an important cue to readers for whether to focus on across-years trends or within-year comparisons in reading graphs.³

Implications for a Model of Graph Interpretation

Overall, these results are consistent with models of graph comprehension that predict that displays that support perceptual processes, such as pattern perception and association, are easier to interpret than displays that support complex logical inferences, such as multiple computational or spatial transformations (Carswell & Wickens, 1987; Casner & Larkin, 1989; Cleveland, 1994; Larkin & Simon, 1987; Pinker, 1990; Shah & Carpenter, 1995; Simkin & Hastie, 1986). In these studies, we established specific features of graphs that make relatively easy, perceptual inferences possible. In particular, we isolated one important variable on viewers' interpretations, namely, the perceptual organization of the visual chunks of graphs. Furthermore, we specified how the perceptual organization of visual chunks influences viewers' interpretations of line graphs and bar graphs. The visual chunks of line graphs are the x - y lines; when relevant historical trends are plotted as x - y lines, viewers tended to spontaneously describe those trends. The visual chunks for bar graphs are bars that are grouped closely together, and viewers are more likely to spontaneously describe trends across bars that are grouped together. Finally, bar graphs are less biasing than line graphs and may be better suited when the relationships of two independent variables are to be equally emphasized.

These results extend the results of previous research on the role of graph format on viewers' interpretations (Carswell & Wickens, 1987; Shah & Carpenter, 1995; Zacks & Tversky, in press). This study goes beyond many of the previous studies comparing graph format, in which the comprehension tasks were simple fact retrieval or verification, and the graphs depicted relatively simple data sets (e.g., two data points) representing information about impoverished or unfamiliar domains. Specifically, this study demonstrates that the perceptual organization of the data is an important factor that influences viewers' spontaneous interpretations and understanding of data even when the data and tasks are relatively complex and the domains are familiar.

Implications for a Model of Document Reading

These results provide further specification of models of document and visual display processing in general (e.g.,

Guthrie et al., 1993; Mosenthal, 1996; Mosenthal & Kirsch, 1991, 1992), which typically predict an interaction of type of task and type of document. According to the Mosenthal (1996; Mosenthal & Kirsch, 1992) model, this interaction occurs because viewers are required to perform relatively simple *locate* processes, such as identifying a particular data point, for certain task-document combinations, and more complex *cycle* and *integration* processes, which require performing a sequence of more simple processes. In this study, we also find an interaction between type of task (within-year or across-years comparisons) and document type (bar graph vs. line graph), which we argue is based on whether relevant information is represented in the visual chunks. When relevant information is represented in the visual chunks, perceptual inferences (locate processes in the Mosenthal model) may be sufficient, but when relevant information is not represented in the visual chunks, complex inferential processes (such as integrate and cycle processes in the Mosenthal model) may be required. Furthermore, in spontaneous description tasks such as those used in this study (rather than question-answer tasks), viewers attempt to use easier locate processes and thus are biased toward describing information that is represented in the visual chunks.

Implications for the Design of Textbook Graphs

These studies contribute to the development of research-based principles for the design of graphs (Kosslyn, 1994). Most generally, the major design principle identified in this set of studies is that **visual chunks in graphs should relate the data points that the author wishes the student to compare.** For example, the data set depicted in Figure 1 consists of 12 data points, with each point represented as a bar. The bars were clustered into three groups of four bars each, such that each cluster (i.e., each visual chunk) consisted of four different categories within a given year. In this case, people have a greater relative tendency to describe comparisons between categories within a given year, in conflict with the author's intended message. In contrast, the data set depicted in Figure 2 consists of 12 data points, with each point represented as a dot. The dots are connected by four lines of three dots each, such that each line (i.e., each visual chunk) consists of a measurement of a single category across three different years. In this case, people are more likely to describe comparisons of a single category across years, as the author intended.

We also identified a number of specific properties of graphical displays that influence what visual chunks are encoded by graph viewers and, therefore, what interpretations viewers can easily make. The influence of these

³ Other perceptual characteristics, such as the aspect ratio of the graphs or the width or color of bars, may also influence the perceptual organization of the information in a graph and, thus, have an influence on viewers' interpretations above and beyond the influence of format and position of variables found in this study (Cleveland, 1994).

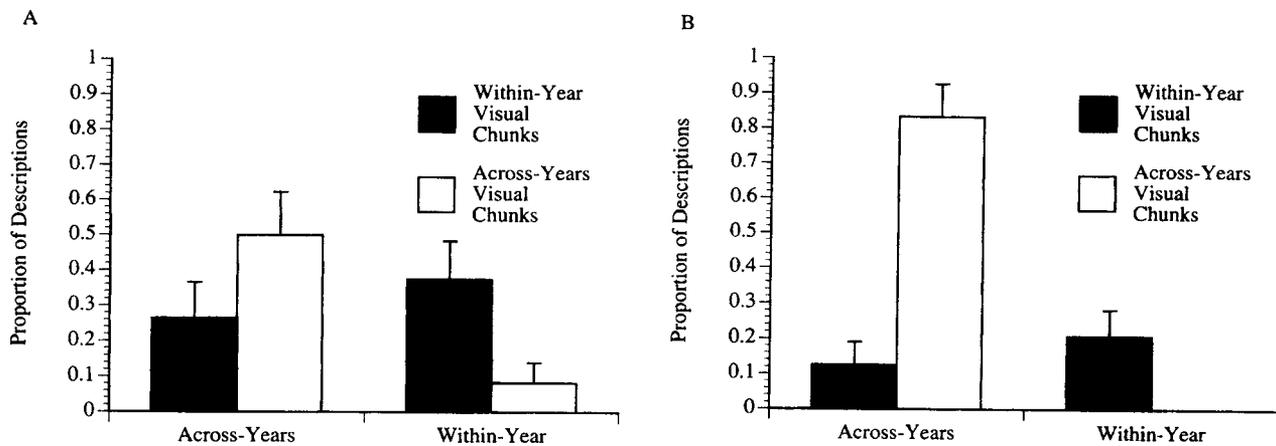


Figure 7. Proportion of viewers generating across-years and within-year descriptions for the four graph types in Experiment 3. For bar graphs (7A) and line graphs (7B), viewers described across-years trends when the across-years trends were plotted closely together or connected (across-years visual chunks). Viewers described within-year trends when they were plotted closely together in bar graphs or connected by lines (within-year visual chunks).

properties leads to a number of specific principles for plotting data:

1. Line graphs emphasize x - y trends. If there are three or more variables in a data set, then the most important relationship should be plotted as a function of the x - and y -axes.

2. Bar graphs emphasize comparisons that are closer together on the display. If there are three or more variables, the most relevant trends should be plotted closer together along the axes when using bar graphs.

3. Line graphs are more biasing (emphasizing the x - y relations), whereas bar graphs are more neutral; thus, if two independent variables are equally important, bar graphs should be used. If a particular trend is the most important information, then line graphs should be used.

4. The scale should reflect whether the goal is to understand relative or absolute information, because people have difficulty translating between different graphic scales.

References

- Armento, B. J., Nash, G. B., Salter, C. L., & Wixson, K. K. (1991). *A more perfect union*. Boston: Houghton Mifflin.
- Bereiter, C., & Scardamalia, M. (1982). From conversation to composition: The role of instruction in a developmental process. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 2, pp. 1-64). Hillsdale, NJ: Erlbaum.
- Bertin, J. (1983). *Semiology of graphics: Diagrams networks maps* (W. Berg, Trans.). Madison: University of Wisconsin Press.
- 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., 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.
- Casner, S. M., & Larkin, J. H. (1989). Cognitive efficiency considerations for good graphic design. *Proceedings of the Eleventh Annual Meeting of the Cognitive Science Society*, 11, 275-282.
- Cleveland, W. (1994). *The elements of graphing data*. Belmont, CA: Wadsworth.
- Cleveland, W. (1995). *Visualizing data*. Murray Hill, NJ: AT&T Bell Laboratories.
- Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 77, 541-547.
- Cleveland, W. S., & McGill, R. (1985, August 30). Graphical perception and graphical methods for analyzing scientific data. *Science*, 229, 828-833.
- Culbertson, H. M., & Powers, R. D. (1959). A study of graph comprehension difficulties. *Audio Visual Communication Review*, 7, 97-100.
- Guthrie, J. T., Weber, S., & Kimmerly, N. (1993). Searching documents: Cognitive process and deficits in understanding graphs, tables, and illustrations. *Contemporary Educational Psychology*, 18, 186-221.
- Hegarty, M., Carpenter, P. A., & Just, M. A. (1991). Diagrams in the comprehension of scientific texts. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 641-668). New York: Longman.
- Hunter, B., Crismore, A., & Pearson, P. D. (1987). Visual displays in basal readers and social science textbooks. In D. Willows & H. A. Houghton (Eds.), *The psychology of illustration: Vol. 2. Instructional Issues* (pp. 116-135). New York: Springer-Verlag.
- Kaput, J. J. (1987). Representation and mathematics. In C. Janvier (Ed.), *Problems of representation in mathematics learning and problem solving* (pp. 19-26). Hillsdale, NJ: Erlbaum.
- Kosslyn, S. M. (1994). *Elements of graph design*. New York: Freeman.
- Larkin, J. H., & Simon, H. A. (1987). Why a diagram is

- (sometimes) worth ten thousand words. *Cognitive Science*, 11, 65–99.
- Lehrer, R., & Romberg, T. (1996). Exploring children's data modeling. *Cognition & Instruction*, 14, 69–108.
- Leinhardt, G., Zaslavsky, O., & Stein, M. K. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. *Review of Educational Research*, 60, 1–64.
- Lohse, G. L. (1993). A cognitive model for understanding graphical perception. *Human-Computer Interaction*, 8, 353–388.
- Mayer, R. E. (1993a). Comprehension of graphics in texts: An overview. *Learning & Instruction*, 3, 239–245.
- Mayer, R. E. (1993b). Illustrations that instruct. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 5, pp. 253–284). Hillsdale, NJ: Erlbaum.
- Mayer, R. E. (1994). Visual aids to knowledge construction: Building mental representations from pictures and words. In W. Schnotz & R. W. Kulhavy (Eds.), *Comprehension of graphics: Advances in psychology* (Vol. 108, pp. 125–138). Amsterdam: North-Holland/Elsevier Science.
- Mayer, R. E., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1996). When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology*, 88, 64–73.
- Mayer, R. E., Sims, V., & Tajika, H. (1995). A comparison of how textbooks teach mathematical problem solving in Japan and the United States. *American Educational Research Journal*, 32, 443–460.
- Mosenthal, P. B. (1996). Understanding the strategies of document literacy and their conditions of use. *Journal of Educational Psychology*, 88, 314–332.
- Mosenthal, P. B., & Kirsch, I. S. (1991). Toward an explanatory model of document literacy. *Discourse Processes*, 14, 147–180.
- Mosenthal, P. B., & Kirsch, I. S. (1992). Types of document knowledge: From structure to strategies. *Journal of Reading*, 36, 64–67.
- 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.
- Scardamalia, N., Bereiter, C., & Lamon, M. (1994). The CSILE Project: Trying to bring the classroom into the world. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice*. Cambridge, MA: MIT Press.
- Shah, P. (1995). *Cognitive processes in graph comprehension*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.
- Shah, P., & Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology: General*, 124, 43–61.
- Simkin, D. K., & Hastie, R. (1986). An information processing analysis of graph perception. *Journal of the American Statistical Association*, 82, 454–465.
- Vernon, M. D. (1946). Learning from graphical material. *British Journal of Psychology*, 36, 145–158.
- Vernon, M. D. (1950). The visual presentation of factual information. *British Journal of Educational Psychology*, 20, 174–185.
- Zachs, J., & Tversky, B. (in press). Bars and lines: A study of graphic communication. *Memory and Cognition*.

Received June 17, 1998

Revision received March 19, 1999

Accepted March 20, 1999 ■