Journal of Experimental Psychology: Applied
1998, Vol. 4, No. 2, 119–138

1076-898X/98/$3.00

# Reading Bar Graphs: Effects of Extraneous Depth Cues and Graphical Context

Jeff Zacks, Ellen Levy,
and Barbara Tversky
Stanford University

Diane J. Schiano
Interval Research Corporation

Manipulating the way a graph is drawn influences viewers' ability to extract information from it. In a series of experiments with simple bar graphs, the authors varied the rendering characteristics and relative heights of the bars and asked participants to estimate the quantities portrayed. The addition of 3-dimensional (3D) perspective depth cues lowered accuracy. This accuracy disadvantage diminished when a short delay was introduced before judgments were reported. The height of the judged bar relative to nearby graphical elements also affected accuracy; this effect was about 1 order of magnitude larger and remained intact when the delay was introduced. Nearby elements also affected viewers' bias (under- or overestimation). These effects do not seem to be due to misestimation of object depth. The results suggest that warnings about accuracy decrements due to 3D shading may be overstated, whereas distortions due to neighboring elements should be of more concern.

Recent advances in computing and printing technologies allow one to produce a dizzying array of different kinds of graphs—and have had a great impact on the kinds of graphs we see in newspapers, magazines, and technical journals. In each particular situation, how can one make a reasoned choice among all the possibilities? There is every indication that these choices matter: The visual characteristics of graphs affect the speed, accuracy, and difficulty of information extraction. They also affect memory for the appearance of graphs and for the information they convey (e.g., Cleveland, 1985; Gattis & Holyoak, 1996; Shah & Carpenter, 1995; Tversky & Schiano, 1989). Therefore, the psychological study of the effects of different rendering techniques seems particularly timely.

Studying graphs also provides an elegant means to study quantitative aspects of perception and conceptual inference. By manipulating features of graph rendering, we can learn about how the visual system combines depth cues and how visual elements interact in forming magnitude judgments.

## Three-Dimensional Renderings of Two-Dimensional Data

Modern graphing programs provide the ability to render graphs with the appearance of three dimensions, using perspective cues. In some cases, the third dimension is used to depict a third variable. Research on three-dimensional (3D) data sets has indicated that 3D rendering is important for understanding the full structure of such data sets (Shah & Carpenter, 1995; Wickens, Merwin, & Lin, 1994). (Throughout this article, we will use *3D* to refer to the addition of perspective cues to give the impression of depth

---

Jeff Zacks, Ellen Levy, and Barbara Tversky, Psychology Department, Stanford University; Diane J. Schiano, Interval Research Corporation, Palo Alto, California.

This work was supported in part by Interval Research Corporation, and by National Science Foundation graduate fellowships. The authors would like to thank Gwo-Ing Lee for her assistance.

Correspondence concerning this article should be addressed to Jeff Zacks, Psychology Department, Stanford University, Stanford, California 94305-2130. Electronic mail may be sent to zacks@psych.stanford.edu.

and *two-dimensional* [*2D*] to refer to the absence of such cues.) However, in what seems to be an increasing number of other cases, the third dimension is not used to convey an additional dimension of the data but rather to enhance the visual appeal of the graphic. Because the addition of the third dimension adds visual complexity without adding information, its use has been decried by many, at least for depicting precise values (Kosslyn, 1985; Tufte, 1983; Wainer, 1984).

## Effects of the Addition of Depth Cues

Adding the appearance of a third dimension not only adds extraneous visual clutter, it also adds conflicting depth cues. Pictorial cues such as linear perspective, shading, and occlusion suggest that the figure has a contour in depth, whereas binocular disparity, convergence, and motion parallax all indicate that the figure is flat. Both clutter and the conflict of depth information could have a deleterious effect on graph perception and comprehension. Several competing methods of depth-cue competition have been proposed (Bülthoff & Mallot, 1988; Johnston, Cumming, & Landy, 1994; Landy, Maloney, Johnston, & Young, 1995; Nakayama & Shimojo, 1990; Young, Landy, & Maloney, 1993). It seems likely that, whatever the depth-cue combination algorithm, it will be less accurate in reconstructing the 3D structure of an object when that object is represented by conflicting depth cues. Misperception of the structure of an object in depth (i.e., along the dimension orthogonal to the image plane) can affect not just judgments of the distance or depth of an object but also of its height or width. For example, given two objects that subtend the same vertical visual angle, the one that is seen as farther away will be perceived as taller than one seen as closer. This means that both inaccuracy and bias in depth perception lead to distorted estimates of the height of an object.

Considerations of depth-cue combination gives a theoretical grounding to the preferences of designers (Tufte, 1983; Wainer, 1984) and psychologists (Kosslyn, 1993) for area graphs over volume graphs, at least for making relative height judgments at the time of viewing. However, there is a simpler explanation for the presumed deficiencies of 3D graphs: Lower accuracy could be the result of distraction due to the irrelevant added graphical elements. This leaves us with two open questions. First, does adding depth cues to a graph lower viewers' accuracy for reading that graph? Second, if there is lower accuracy, is this lower accuracy due to depth-cue combination or is it simply a result of adding extraneous markings to a figure?

A conclusive answer has not been forthcoming from the few studies that have examined effects of depth cues on accuracy with graph-like stimuli. In one study, Spence (1990) assessed accuracy judgments with seven different graph types and tables. He concluded that the apparent dimensionality of the graphs did not affect observers' accuracy. However, the graph types in this study were not selected systematically. One of the graphs in the 2D group was an unusual elliptical pie chart; errors with this graph type were much larger than for the other 2D graphs. Inspection of Spence's figures suggests that if this graph type had been excluded, a reliable disadvantage for the 3D graphs would have been observed. The choice in this study to omit the rectangular frame that typically surrounds a published graph raises another interesting question of interpretation. Two of us (Tversky & Schiano, 1989) have found that participants' interpretation of a figure as a graph, rather than something else (e.g., a map of a location), led to differences in the perceived orientation of a line in the figure. It could be that the inclusion or exclusion of a frame in these experiments will influence judgments by a similar mechanism.

In a similar experiment, Carswell and her colleagues did find extraneous depth cues to be associated with lower accuracy (Carswell, Frankenberger, & Bernhard, 1991). They used line, bar, and pie charts and created versions of each, both with and without depth shading. However, two aspects of the stimulus design cloud interpretation of their results. First, the pie chart with added depth shading was tilted so that the pie surface appeared as an ellipse, whereas the version without depth shading was drawn as a circle. Second, the pie and bar graph stimuli portrayed the data values as either an area (for the 2D versions) or as a volume (for the 3D versions), whereas the line graph stimuli portrayed the data with a simple line (for the 2D version) or a surface line (for the 3D version). (See the taxonomy provided below and Figure 2 for definitions and examples of simple line and surface graphs.) The surface-line graph in particular is an unusual type of graph, and the line-graph comparison contrasts a surface with a simple line rather than comparing a volume to an area, as do the pie and bar graphs. The effect on accuracy was dominated by a large difference between the 2D and 3D line graphs. Comparing just the bar graphs with and without depth shading or the pie graphs with and without depth shading revealed no significant differences associated with added depth cues.

Taking into consideration the choice of graph types in these two experiments, there seems to be a small negative effect of adding extraneous depth cues on height-judgment accuracy. However, other factors in graph-rendering style (e.g., the choice of tilted pie graphs or surface-line graphs) had a more dramatic effect on observers' judgments.

## Effects of the Relationships Between Graphical Elements

It is well known that the relationship between elements in a figure can affect the perception of those elements. The graphical context in which an element occurs can produce distortions in judgments of color, angle, size, and orientation (see, e.g., Goldstein, 1989, especially chap. 7; Howard, 1982). One particularly relevant example of such a distortion is given in Figure 1. This figure demonstrates the parallel lines illusion: When two parallel lines are viewed, the viewer tends to perceive assimilation (the lengths of the lines seem closer than they are) or contrast (the lengths of the lines seem more different than they are), depending on the ratio of the line lengths and the distance between them (Jordan & Schiano, 1986; Schiano, 1986).

The parallel lines figure is a very simple example, but the influence of perceptual assimilation and contrast is presumably at work in more complex figures as well. Bar graphs are examples of richer visual stimuli that encapsulate the key features of the parallel lines illusion. Accordingly, judgments of bar height should be affected by the relative heights of the bars in a figure and by the height of the judged bar or bars relative to the surrounding graphical frame.

This generates a third open question: What is the relative importance of the addition of depth cues compared with the relationships in the data elements? It is important to think of the effects of 3D rendering techniques in the context of the other factors influencing perceptual judgment. Beyond asking whether the addition of extraneous depth cues affects accuracy and bias in judgment, it makes sense to ask whether such effects are likely to be important "in the wild." One way to answer this question is to compare the size of effects due to the addition of conflicting depth cues with the size of effects due to the influence of nearby graphical elements. This allows an assessment of how well the depth-cue combination mechanism performs in the face of noise, relative to ubiquitous distortions due to graphical context.

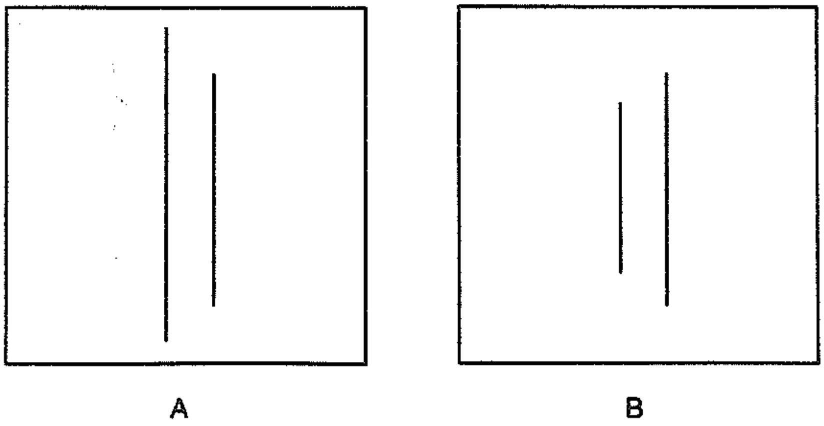


*Figure 1.* The parallel lines illusion. For most viewers, the right line in Panel A appears longer than the right line in Panel B.

This question is also of practical interest. The relative heights of bars in a graph come from the data being depicted, and the designer of the figure has little control over it. On the other hand, a designer has extensive control over the rendering style of the figure, including control over the inclusion of extraneous depth cues. If both factors have an impact on perception (and perceptual distortion), which is larger—the factor that is controllable or the one that is not?

## The Role of Depth Cues in Encoding and Storage

Often, we use graphs to communicate information that is to be used at a later date. However, psychophysical studies of graphical perception have typically examined judgments made while viewing a figure (e.g., Cleveland, Harris, & McGill, 1983; Spence, 1990). This leaves us with a fourth open question: How do effects on perception combine with encoding and storage to influence later judgment?

The experiments reported next were designed to help answer the four open questions described earlier: (a) Does adding depth cues to a graph lower viewers' accuracy for reading that graph? (Experiments 1, 3, and 4); (b) if there is lower accuracy, is this lower accuracy due to depth-cue combination or is it simply a result of adding extraneous markings to a figure? (Experiment 5); (c) what is the relative importance of the addition of depth cues compared with the relationship between the judged data elements and the graphical context? (Experiments 1–5); and (d) how do effects on perception combine with encoding and storage to influence later judgment? (Experiment 2).

## Types of Graphs

To allow for systematic comparisons among graph types, let us introduce a brief taxonomy. A good number of graphs can be classified on two dimensions: rendering style and graph type. Rendering style usually takes on one of four possible values. We call graphs that use lines (without shaded areas) to indicate the data values *simple*. Graphs that use the area of a region to depict the data values are called *area* graphs. Graphs that use a drawing of a volume (e.g., a rectangular box) to indicate data values are called *volume* graphs. Graphs that show the data by drawing floating surfaces are called *surface* graphs. Many common graphs take one of two possible values for graph type: *bar* graphs are figures that use an element oriented relative to the independent variable's axis to show each data point, whereas *line* graphs use a line that connects a set of data points. This two-way classification gives rise to eight (4 × 2 = 8) possible kinds of graphs; examples of each are shown in Figure 2. The experiments described next used stimuli whose rendering style was simple, area, or volume; all were bar graphs. (For a more comprehensive taxonomy that combines visual features of the graph with the implicit task of the viewer, see Cleveland & McGill, 1984.)

## Experiment 1: Perceptual Match

This experiment was designed to examine the effect of extraneous depth cues on height judgments, to measure the effect of neighboring elements on such judgments, and to compare these two effects. Observers made height judgments while looking at bar graphs. We chose a perceptual-match task for two reasons. First, it encapsulates one important use of data graphics: making quick, reasonably accurate quantitative estimates of data values, without necessarily reading an exact value from an ordinate scale. Second, this task has been productively used to study perceptual illusions (e.g., Jordan & Schiano, 1986).

Across trials, the rendering style of the graphs (area or volume) was varied to investigate the effects on perception of adding depth cues. The height of the test bar and the presence of a constant-height context bar were also varied, allowing for a parametric investigation of effects of graphical context, as in the parallel lines illusion.

### *Method*

*Participants.* The 40 participants were undergraduate students at Stanford University. Each took part to fulfill a requirement in an introductory psychology course.

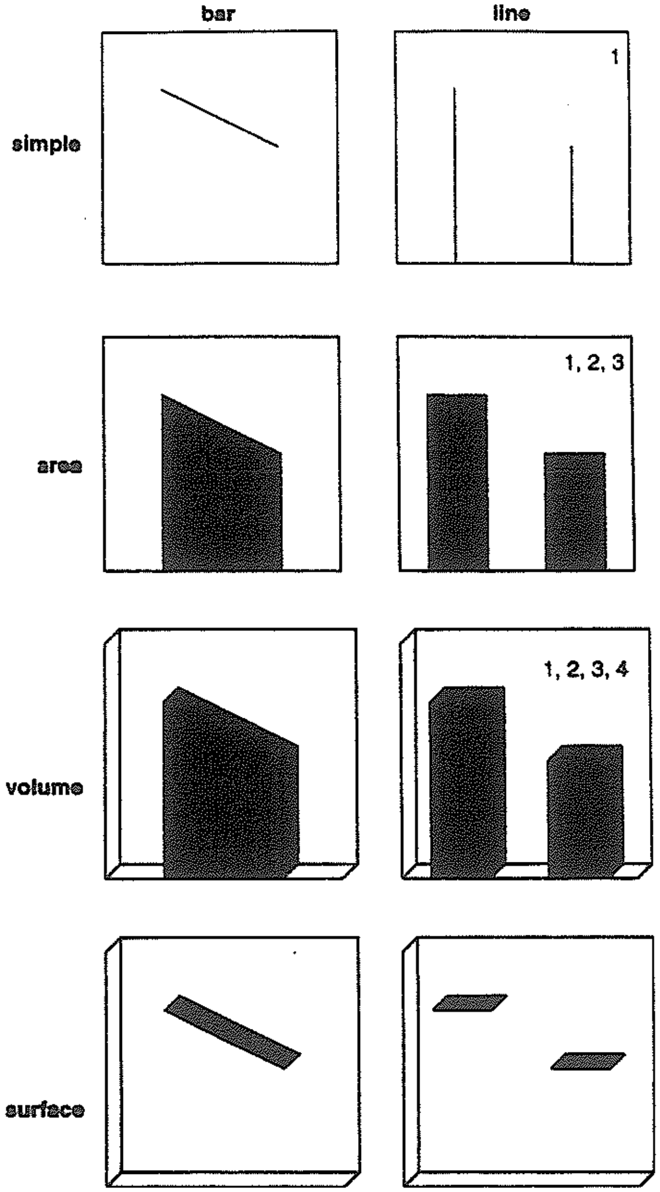


*Figure 2.* A brief taxonomy of some common graph types. Each row shows a different rendering style; each column shows a different graph type. The numbers in the upper right corners give the experiments (if any) in which a given kind of graph appears.

*Stimuli.* Two general types of graphs were prepared as stimuli: ones in which two elements were displayed and ones in which a single element was displayed within a graph frame. The two-element graphs consisted of one element of fixed height (context element) of 20 mm and an accompanying "test" element that varied in height from 20 mm to 100 mm by 20-mm increments, in order to create the ratio relationships between the elements of 1:1, 1:2, 1:3, 1:4, and 1:5. The test elements were placed to the right of the context element in the graph. In addition to these two-element graphs, five distractor graphs with noninteger ratio relationships between elements were created by introducing test elements of 16, 26, 46, 74, and 92 mm. These test elements were placed to the left of a 20-mm context bar in the graph. In both cases, the left edges of the elements were separated by 23 mm. The single-element graphs included only the test elements described earlier without any context element. In these graphs, the test element was placed in the center of the graph frame. In all graphs, the test element was denoted by an asterisk, placed underneath the element and just below the graph frame. The frames were squares 110 mm to a side. See Figure 3 for examples of the stimuli.

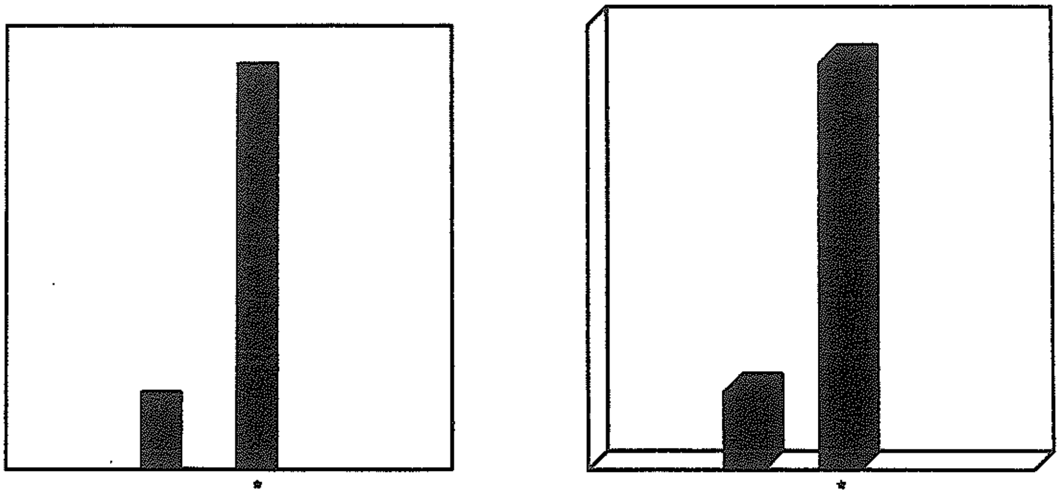
All told, there were five "test stimuli with context" graphs, five "test stimuli without context" graphs, and five "distractor" graphs. Two versions of each of these graphs were created, one using area bars (rectangles) and one using volume bars (boxes). The area bars were 7 mm wide. The 3D elements were created by using the corresponding area bar as the face of the box. The perception of three-dimensionality was created with orthogonal-perspective drawing in which 7-mm lines were drawn at 45° angles from the appropriate corners of the bar and then connected.

Two reference charts, an area version and a volume version, were also created to provide the "match" options available to the participants. Each chart depicted a series of elements increasing in height from 12–112 mm by 2-mm increments. Underneath each element was an identifying label (ranging from A to YY). The charts were pasted on two sides of a piece of tagboard.

*Booklets.* Each graph (2 rendering styles × 3 contexts × 5 ratio relationships) was tested twice, for a total of 60 trials. In addition, six filler pages were included, inserted every nine pages, on which participants were presented with a small data set and asked to draw the graph they felt best represented the information given. Booklets were prepared using one of two random orders of trials, modified only to eliminate the possibility of two trials of the exact same graph type (rendering style, context, and height of test element) occurring consecutively.

In all cases, the graphs were created in the drawing program MacDraw Pro 1.5v2 (Claris Corporation, 1992), and each was printed on a separate 5.5″ × 8.5″ piece of paper.

*Procedure.* Participants were given a booklet and told that they would be making perceptual judgments about elements in the graphs they were about to see. They were instructed to rely



Pick the letter of the bar that best matches the height of the bar marked with an * in the graph above.

Answer: ____

Pick the letter of the bar that best matches the height of the bar marked with an * in the graph above.

Answer: ____

*Figure 3.* Examples of the perceptual-match stimuli used in Experiments 1 and 2. The left panel shows the area rendering style; the right panel shows the volume rendering style.

solely on their visual perception and not to use other means of assessing the goodness of match (e.g., using their fingers to map the height of the element onto the reference chart). The booklet was placed in a cardboard box with one end cut out, and participants were asked to take out one page at a time, complete the judgment, and then place the page face down on the side of the table before proceeding to the next page.

For each graph, participants were instructed to refer to the reference chart (either the area or volume version, whichever matched the element type shown in the graph) and to write down the letter of the bar or box that best "matched" the height of the element in the graph with the asterisk underneath.

The experiment was self-paced, with most participants taking roughly 25 min to complete the entire booklet.

### *Results*

Participants estimated the height of bars by picking a match from a sample array. Two error measures were constructed from their judgments: *raw error,* which was the height of the chosen bar in millimeters subtracted from the correct bar height, and *error magnitude,* which was simply the absolute value of the raw error for a given trial. The aggregated raw errors show any systematic bias in participants' perceptions of the bar heights, whereas the error magnitudes describe how accurate the judgments were.

Bar height judgments were less accurate for the volume graphs than for the area graphs. The mean error magnitude for the area graphs was 4.10 mm ($SEM = 0.150$), whereas for the volume graphs it was 4.62 mm ($SEM = 0.152$). This difference, although small (approximately half a millimeter), was statistically reliable, $F(1, 1553) = 7.57, p = .006$.

Participants were less accurate for taller bars than for shorter bars, $F(1, 1553) = 241, p < .001$. For the shortest bars, the mean error magnitude was 1.88 mm, whereas for the largest bars it was 5.97 mm (see Figure 4).

Under these viewing conditions, the presence or absence of a context bar had little or no effect on error magnitude, $F(1, 1553) = 0.106, p = .744$.

In this experiment, participants tended to slightly overestimate the height of the bar; the mean raw error was 0.879 mm, $t(1599) = 5.82, p < .001$. This overestimation was most pro-

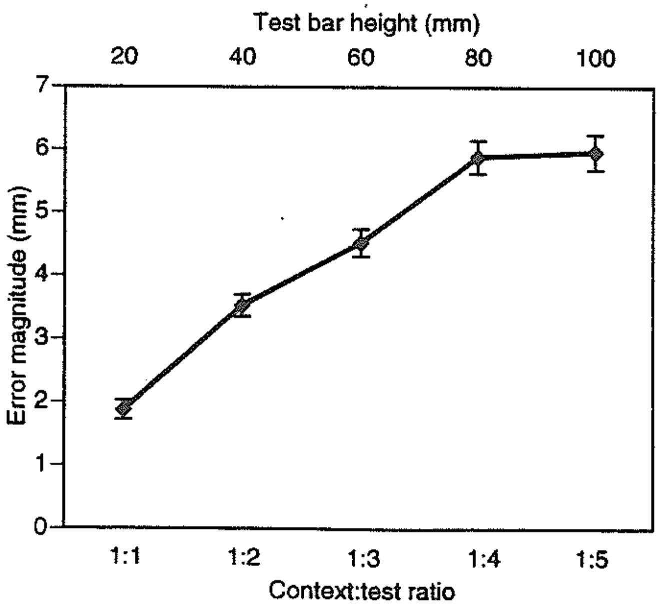


*Figure 4.* For the perceptual-match task, error magnitude depended on test bar height. Participants were more accurate for shorter bars. The figure shows data from Experiment 1. Plotted points represent mean error magnitude for each test bar height, and error bars show 1 standard error of the mean.

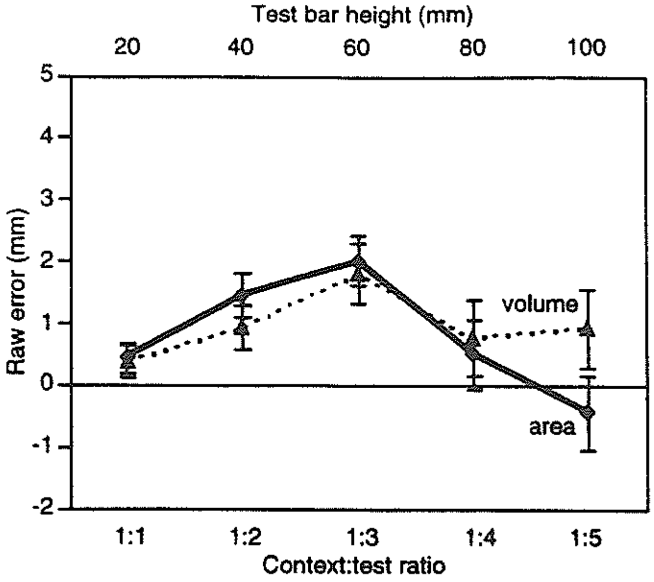


*Figure 5.* For the perceptual-match task, raw errors for the area and volume graphs did not differ significantly except for the tallest bars. For the tallest bars, area graphs were judged shorter than volume graphs. The figure shows data from Experiment 1. Plotted points represent mean raw error broken down by test bar height, and error bars show 1 standard error of the mean.

nounced for the intermediate-height graphs (see Figure 5). The raw errors were relatively insensitive to the experimental manipulations. There was an interaction between the rendering style (area vs. volume) and the height of the test bar that approached significance, $F(1, 1553) = 3.83$, $p = .051$: Despite the general tendency to overestimate the height of the bar, for the tallest bars this was true only for the volume graphs (see Figure 5).

### Discussion

Judgments of bar height were approximately half a millimeter less accurate when 3D depth cues were added to the graphs. This suggests that, as predicted, adding extraneous depth cues does result in lowered accuracy for judgments about the depicted objects. However, neighboring graphical elements also affected judgments: Accuracy depended on the height of the judged bar, and this effect was approximately one order of magnitude larger than the effect of extraneous depth cues. (Another manipulation of neighboring graphical elements—the addition of a context bar—had no reliable effect on accuracy or bias.) There was also a general tendency to overestimate, which was again large relative to the effect of extraneous depth cues. The small relative size of the rendering-style effect suggests that either the visual system's depth-cue combination algorithms are robust in the face of conflicting information or that the effect of depth distortion on perceived height is relatively small. From a practical point of view, it also suggests that we should pause before making strict design recommendations based on the cognitive–visual problems with 3D graphs.

Rendering style, bar height, and the presence of a context bar all had little effect on systematic bias in participants' perceptions of the bar heights. What effects there were might be explained by the role graphical frames (provided by the graph-bounding box and the page) played in generating assimilation and contrast distortions of height judgments.

## Experiment 2: Perceptual Match From Memory

Experiment 1 showed that adding extraneous depth cues lowered accuracy for height judg-