

ABSTRACT

Title of Dissertation: COMPARING SMALL GRAPHS: HOW
DISTANCE, ORIENTATION, AND
ALIGNMENT AFFECT THE
COMPARABILITY OF SMALL MULTIPLE
BAR GRAPHS

 Benjamin Keniray Smith,
 Doctor of Philosophy, 2012

Dissertation Directed By: Professor Kent L. Norman,
 Department of Psychology

Small multiples play a vital and growing role in the display of complex information. They are particularly useful for depicting spatiotemporal data, for which more traditional graphs and maps are inadequate. However, the scientific investigation of the usefulness of small multiples has been limited and often misdirected. In five experiments, small bar graphs are used to investigate several factors that could influence the comparability of the small graphs that comprise a small multiples graph. These factors include the distance between the graphs, the alignment of the graphs, the orientation of the bars, the length of the bars, and whether the graphs contain a single bar or multiple bars. In all cases, the most important factor affecting the comparability of the graphs was the difference in lengths, or difference in the increase of lengths, that the participants were asked to compare. The effects of distance were greater when the bars were closer to each other than when they were farther apart, suggesting that the bars are

compared using central vision. For pairs of graphs with a single bar each, comparability decreased as the distance between the graphs increased, although this effect was more prominent measured by accuracy than response time. Graph arrangements with horizontal alignments and vertical orientations were more comparable, although these effects were more subtle than the distance effects. For pairs of graphs with two bars each, the distance between the graphs had no effect on the accuracy of the comparison, and only a slight effect on the response time. Alignment and orientation had no effect on the comparability of graphs with two lines. The similarity of the lines in each graph, including but not limited to the critical length increase, significantly affected the comparability of the graphs.

Part of a graph difficulty principle for small multiple graphs is offered as advice for graph creators.

COMPARING SMALL GRAPHS:
HOW DISTANCE, ORIENTATION, AND ALIGNMENT AFFECT
THE COMPARABILITY OF SMALL MULTIPLE BAR GRAPHS

By

Benjamin Keniray Smith

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2012

Advisory Committee:
Professor Kent L. Norman, Chair
Professor Kenneth R. Fleischmann
Professor Jimmy Lin
Professor Catherine Plaisant
Professor Thomas S. Wallsten

© Copyright by
Benjamin Keniray Smith
2012

Dedication

For Andi.

For Daria.

Acknowledgements

I blame the following people for making this dissertation possible. They know what they did.

Madelyn R. Smith, Leah A. Smith

Jeffrey K. Smith, Lisa Smith

Andréa D. L. Smith, Daria R. Smith, Julia Livi

Evelyn Rose, Mike Rose, Bettie Smith, Ralph Smith

Susan G. Campbell, Walky Rivadeneira

Thomas S. Wallsten, Michael R. Dougherty, Catherine Plaisant, Avis Cohen,

David Huber, Jimmy Lin, Kenneth R. Fleischmann, William S. Hall, Bill Idsardi

Pamela Komarek, Sandy Davis, Carol Gorham, Tony Chan

Michael Bunting, Catherine Doughty, Jared Linck, Meredith Mislevy Hughes,

Scott Jackson, Amira Allen, and many other people at CASL

A horde of undergraduate assistants

Steve Wozniak, Steve Jobs, Susan Kare, Bill Atkinson

And last but certainly not least, Kent L. Norman

Table of Contents

Dedication.....	ii
Acknowledgements.....	iii
Table of Contents.....	iv
List of Tables.....	vii
List of Figures.....	viii
Chapter 1: Introduction.....	1
Spatiotemporal Maps and the Need for Complex Graphs.....	1
Terms.....	2
Tasks.....	5
Types.....	7
Static Maps Showing a Limited Number of Times.....	8
Static Maps Showing and Unlimited Number of Times.....	12
Dynamic Maps.....	15
An Example: 2008 Presidential Polling Data.....	19
The Case for Small Multiples.....	22
Small Multiples and Theories of Graph Reading.....	25
Controversy Over Small Multiples.....	28
Graph Difficulty Principle for Small Multiples.....	31
Judgments of Size and Line Length.....	35
Psychophysics.....	37
Two theories of size perception.....	39
Judgment Time as a Measure of Difficulty.....	41
Eye Movement in HCI research.....	43
Research Overview.....	45
Research Questions.....	46
General Research Questions.....	47
Chapter 2: Experiment 1.....	48
Research Questions.....	48
Hypotheses and Research Questions.....	49
Method.....	51
Participants.....	51
Apparatus and Materials.....	51
Stimuli.....	52
Conditions.....	54
Procedure.....	56
Results.....	58
Distance.....	60
Accuracy.....	60
Response Time.....	62
Alignment.....	64
Accuracy.....	64
Response Time.....	66
Alignment and Orientation.....	69
Accuracy.....	69

Response Time.....	71
Extra Elements	73
Accuracy	74
Response Time.....	75
Discussion	78
Distance.....	78
Alignment and Orientation	79
Extra Elements	80
General Discussion	81
Chapter 3: Experiment 2	83
Research Questions.....	83
Method	84
Participants.....	84
Apparatus	85
Materials	87
Design	89
Procedure	91
Results.....	93
Accuracy	95
Response Time.....	97
Absolute and relative differences.....	100
Discussion	105
Chapter 4: Experiment 3	108
Research Questions.....	108
Method.....	110
Participants.....	110
Apparatus and Materials	110
Design	111
Procedure	113
Results.....	113
Accuracy	113
Response Time.....	116
Absolute and relative differences.....	118
Discussion	120
Chapter 5: Experiment 4	123
Research Questions.....	124
Method	125
Participants.....	125
Apparatus and Materials	125
Design	125
Procedure	128
Results.....	128
Accuracy	128
Response Time.....	131
Baseline Heights	134
Discussion	136

Chapter 6: Experiment 5	138
Research Questions	138
Method	139
Participants	139
Apparatus and Materials	140
Design	140
Procedure	141
Results	142
Accuracy	142
Response Time	144
Discussion	146
Chapter 7: General Discussion	148
Summary of Findings	148
Distance	149
Alignment and Orientation	150
Complexity of Graphs	151
Limitations	152
Practical Advice for Graph Makers	152
Directions for Future Research	154
Break the response time into its component parts	154
Maximize the comparability of graphs like the ones studied here	154
Expand empirical studies to more complex graphs	155
Comparisons of more than two graphs, and of complex graphs	155
Comparisons of other types of graphs	156
Comparisons of Spatiotemporal Maps	156
Final Thoughts	157
References	158

List of Tables

Table 1. Conditions of Experiment 1.....	55
Table 2. Repeated-Measures ANOVA for accuracy (Experiment 1 distance conditions).	61
Table 3. Repeated Measures ANOVA for time (Experiment 1 distance conditions).....	63
Table 4. Repeated-measures ANOVA table for Accuracy (Experiment 1 alignment conditions).....	65
Table 5. Repeated-measures ANOVA table for response time (Experiment 1 alignment conditions).....	67
Table 6. Repeated-measures ANOVA for accuracy (Experiment 1 alignment/orientation conditions).....	70
Table 7. Repeated-measures ANOVA for response time (Experiment 1 alignment/orientation conditions).....	72
Table 8. Repeated-measures ANOVA for accuracy (Experiment 1 extra elements conditions).....	74
Table 9. Repeated-measures ANOVA for response time (Experiment 1 extra elements conditions).....	76
Table 10. Conditions of Experiment 2.....	91
Table 11. Factors for Repeated-Measures ANOVA, Experiment 2.....	94
Table 12. Repeated-Measures ANOVA table for accuracy (Experiment 2).....	95
Table 13. Repeated-Measures ANOVA table for response time (Experiment 2).....	97
Table 14. Conditions for Experiment 3.....	111
Table 15. Repeated-measures ANOVA table for accuracy (Experiment 3).....	114
Table 16. Repeated-measures ANOVA table for response time (Experiment 3).....	117
Table 17. Conditions for Experiment 4.....	127
Table 18. Repeated-measures ANOVA table for accuracy (Experiment 4).....	130
Table 19. Repeated-measures ANOVA table for response time (Experiment 4).....	132
Table 20. Conditions for Experiment 5.....	140
Table 21. Repeated-measures ANOVA table for accuracy (Experiment 5).....	143
Table 22. Repeated-measures ANOVA for Response Time (Experiment 5).....	145

List of Figures

Figure 1. A map of Earth at night.	4
Figure 2. A choropleth map.	5
Figure 3. A change map.	9
Figure 4. A map pair.	10
Figure 5. Habitat loss map.	11
Figure 6. A bar graph map.	13
Figure 7. A map array.	14
Figure 8. Minard's map.	15
Figure 9. An interactive, animated weather map.	16
Figure 10. A cube map.	17
Figure 11. A multi-component interactive map.	18
Figure 12. Maps showing 2008 Presidential polling and results.	20
Figure 13. Line graph predicting the electoral college outcome of the 2008 U. S. Presidential election.	21
Figure 14. A chartmap of the state-by-state polling data.	22
Figure 15. Small multiples as part of a complex interactive interface.	23
Figure 16. A screen from an earlier experiment on comparing small choropleth maps. ..	34
Figure 17. Reaction Time as a function of Line Length Difference, data from Henmon, 1906.	42
Figure 18. Standard and Comparison Bars for Experiment 1.	52
Figure 19. Graphs used in Experiment 1.	53
Figure 20. Trial Procedure for Experiment 1.	58
Figure 21. Accuracy by Distance, absDif, and sLength (Experiment 1).	62
Figure 22. Response Time by Distance and absDif (Experiment 1).	64
Figure 23. Accuracy by Alignment and absDif (Experiment 1).	66
Figure 24. Response Time by Alignment, absDif (Experiment 1).	68
Figure 25. Accuracy by Alignment and Orientation (Experiment 1).	71
Figure 26. Response time by Alignment and Orientation (Experiment 1).	73
Figure 27. Accuracy by Presence of Extra Elements (Experiment 1).	75
Figure 28. Response Time by Presence of Extra Elements (Experiment 1).	77
Figure 29. Logitech® Dual Action™ Gamepad.	86
Figure 30. Welcome screen (Experiment 2).	88
Figure 31. Trial screen from Experiment 2.	89
Figure 32. Screens from Experiment 2.	93
Figure 33. Accuracy by distance, absDif, and sLength (Experiment 2).	96
Figure 34. Response time by distance and absDif. (Experiment 2).	98
Figure 35. Response time by distance and absDif, distance plotted on a log scale (Experiment 2).	100
Figure 36. Response Time by relative and absolute differences in length (Experiment 2).	102
Figure 37. Response Time for two pairs of lines at each distance (Experiment 2).	103
Figure 38. Accuracy by relative and absolute differences in length (Experiment 2). ...	104
Figure 39. Accuracy for two pairs of lines at each distance (Experiment 2).	105
Figure 40. Alignment and Orientation Conditions for Experiment 3.	112

Figure 41. Accuracy by Alignment and Orientation (Experiment 3).	115
Figure 42. Accuracy by alignment and absDif (Experiment 3).	116
Figure 43. Response Time by Alignment and Orientation (Experiment 3).	118
Figure 44. Accuracy by relative and absolute differences in length (Experiment 3).	119
Figure 45. Response Time by relative and absolute differences in length (Experiment 3).	120
Figure 46. Distance Conditions for Experiment 4.	127
Figure 47. Length Conditions for Experiment 4.	128
Figure 48. Accuracy by Distance and csIncAD (Experiment 4).	131
Figure 49. Response Time by Distance and csIncAD, (Experiment 4).	133
Figure 50. Accuracy by sBase, cBase (Experiment 4).	135
Figure 51. Response Time by sBase and cBase (Experiment 4).	136
Figure 52. Alignment Conditions for Experiment 5.	141
Figure 53. Accuracy by Alignment and csIncAD (Experiment 5).	144
Figure 54. Response Time by Alignment and csIncAD (Experiment 5).	146

Chapter 1: Introduction

We live in a complex and changing world, and to describe that world visually, we often need complex graphs. In recent years, there has been an explosion in the variety of complex graphs available in media such as newspapers and the World Wide Web.

One aspect common to many complex graphs is known as *small multiples*. This term, coined by Tufte (1983), refers to a collection of small, similar graphs that are component parts of a larger, complex graph. A viewer may look at one small graph at a time to examine the data it represents, or compare different small graphs to one another to determine how the underlying data sets differ.

This project is focused on identifying and quantifying how certain choices made by the graph designer can assist or hinder these comparisons, using bar graphs as the underlying graph type. These choices include how the small graphs are arranged in the larger graph (distance and alignment), and how the bars are oriented. Other factors such as the complexity of the bar graphs and similarity of the bars are also examined.

In this chapter, I first describe why complex data visualization is necessary, and why I believe small multiples are an essential tool for data visualization, using the motivating example of spatiotemporal data. Next, I describe the literature about judgments and comparisons of line length, and explain why the research described here focuses on bar graphs. Finally, I give an overview of the experiments I conducted.

Spatiotemporal Maps and the Need for Complex Graphs

Some of the most difficult, yet essential kinds of data to represent visually are those that vary over both space and time. Spatiotemporal maps are a particularly

challenging type of high-dimensional graph (Andrienko & Andrienko, 2006; Tufte, 1983). Like any thematic map, spatiotemporal maps show how particular data differ geographically, but spatiotemporal maps also show how that data differ over time. Most graphs that display time do so using the horizontal dimension, but most maps use both the horizontal and vertical dimensions to display geographic space. Thus the map designer can't simply rely on these conventions in order to show how an attribute varies over time and space without making some kind of concession (Bertin, 1967/1983). There are many approaches to displaying data in spatiotemporal maps, and some of these are discussed in the *Types* section below.

Terms

Spatiotemporal maps can be either *static*, meaning they do not move and could be printed in a book or newspaper, or *dynamic*, meaning they move or change. Dynamic maps include *animated* maps, which change without requiring input from a user, and *interactive* maps, which respond to input. Some maps are both animated and interactive. To be a spatiotemporal map, information from more than one *time* must be shown. Peuquet (2002) discusses that the question of how to segment space and time has been unsettled for thousands of years. Here, *time* refers to a particular point in the space-time continuum; other terms like *moment* (which Andrienko & Andrienko, 2004, use), *epoch*, or *time slice* are other choices for this term. *Time* is intentionally somewhat vague here, and could refer to a *time interval* of a year or a single instant, but must be defined clearly for any particular spatiotemporal map. I will use the convention of referring to distinct times with numbers, for instance, *Time 1*, *Time 2*, etc.

The data being represented in a map are *attributes*, so if we imagine, as Andrienko and Andrienko (2004) do when they used the terms below that I will use here, a map of crime rates in the U. S., crime rate would be the attribute. The entire U. S., which is the total geographical space being represented, is the *territory*. Within the territory some maps predefine several *locations*, areas smaller than the territory, which in this example are the states. In some maps, the data do not need to be aggregated beyond the resolution of the overall map. Figure 1, a map of the world at night (Mayhew, C. and Simmon, R., 2008), does not have predefined regions; there are as many locations in the map as there are pixels to show them. (Note that although the images that Figure 1 is based on were taken at different times, the map does not depict data from more than one time at a particular location, and therefore it is not a spatiotemporal map.) In a map of crime rates, for which data are less abundant and precise than those available to NASA for their map, the data are likely to be based on predefined locations, so each location has a calculable rate. Figure 2 shows a map with data aggregated at the state level.



Figure 1. A map of Earth at night.

This map, created by NASA from satellite images, shows light around the world at night.

Mayhew, C. and Simmon, R. (2008).

Spatiotemporal map readers want to know how the value of the attribute varies across space and time. Any variation constitutes a difference, but Andrienko and Andrienko (2004) use particular terms for certain kinds of differences. A *change* is a difference between two times in a single location, and a *trend* is a pattern of differences in a single location over three or more times. The *distribution* is the pattern across the territory at a particular time. Of course, it might be reasonable to talk about the change or trend in a *region* of the territory, or the distribution within one.

Data may be represented on maps in many ways. Areas of the map may be colored differently to show different values of an attribute. When this coloring is done on the basis of predefined locations, the result is a *choropleth map*, like the one in Figure 2. (Figure 2 and similar maps have distorted state boundaries to allow space for the bar graphs that appear in later figures.) Data may also be represented with *signs* (Bertin, 1967/1983) placed on the map, which can be in various shapes, sizes, colors, orientations,

textures, etc. Bertin calls these *retinal* variables to distinguish them from the *planar* variables of the X and Y dimensions. Brewer (2008) discusses effective color schemes for choropleth maps.

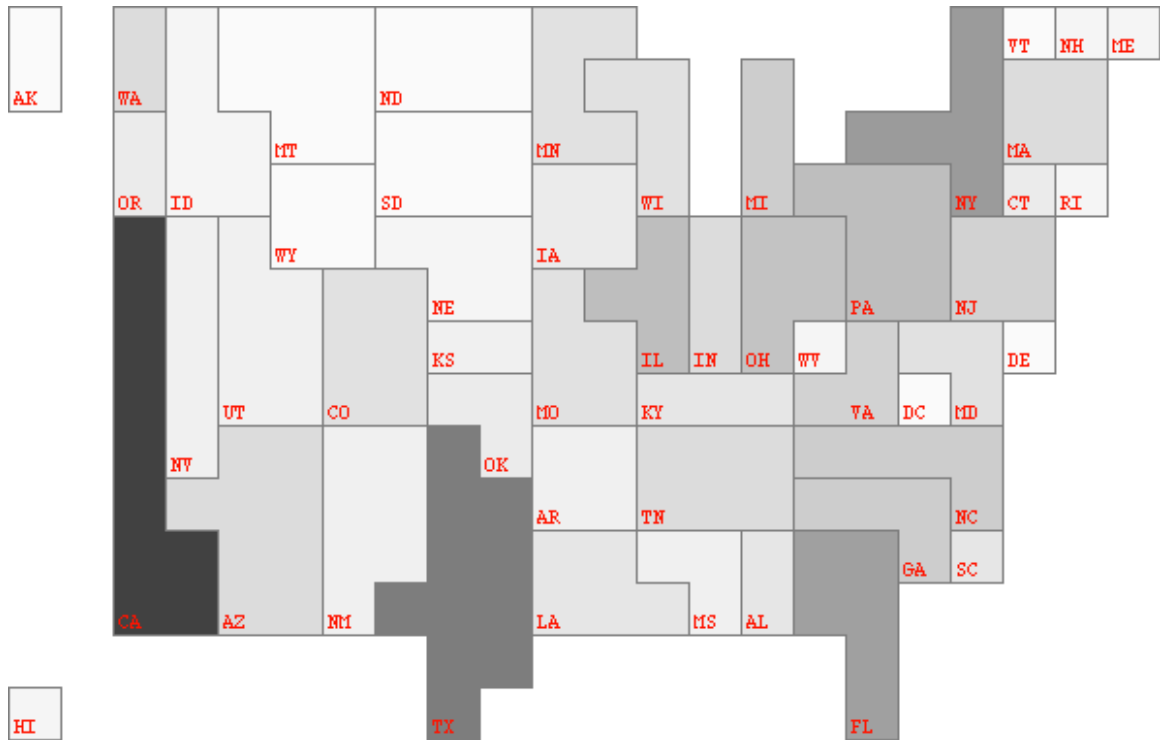


Figure 2. A choropleth map.
Darker colors represent higher total population in the represented State, according to the 2010 Census.

Tasks

Andrienko and Andrienko (2006) point out that there are a nearly infinite number of tasks that can be performed using spatiotemporal maps. However, they do organize them into nine rough questions that can be asked of the data, which are paraphrased below (Andrienko & Andrienko, 2004).

- 1a. What is the value of the attribute at one location at Time 1?
- 1b. What is the distribution of the attribute across the territory at Time 1?
- 2a. What is the change in the attribute from Time 1 to Time 2 at one location?
- 2b. What is the distribution of changes in the attribute from Time 1 to Time 2 across the territory?
- 2c. What is the change of the distribution of the attribute from Time 1 to Time 2 across the territory?
- 3a. What is the trend in the attribute at one location?
- 3b. How do the trends in the attribute at Location 1 and Location 2 compare?
- 3c. What is the distribution of trends in the attribute across the territory?
- 3d. What is the trend of distribution of the attribute across the territory?

Questions 1a and 1b can be asked of any thematic map, as they do not involve more than one time. Question 2a can be asked of any time series graph that shows georeferenced data. Questions 2b and 2c require the same data to answer, but are different questions. Imagine a map of crime data in the U. S. (as Andrienko & Andrienko, 2004, use). If crime dropped in Maryland from 1990 to 2000, but remained steady elsewhere, the answer to 2b would be that the change was in the negative direction in Maryland, and there was no change elsewhere. The answer to 2c might be that crime, which had been concentrated in Maryland, was now evenly distributed across the country. But it might be that crime was still concentrated in Maryland, or was now lowest in Maryland, depending on what the distribution was in 1990 and the severity of the drop. Another example is that if crime dropped uniformly across the territory, this

would not change the overall pattern. So while we would answer question 2b by saying that there were negative changes everywhere, we would answer 2c by saying there was no change.

Trends are inherently more complex than changes, and are thus more difficult to describe. Lee, Butavicius, and Reilly (2003) found it difficult to formulate and evaluate questions that involved more than a few pieces of information. Question 3a, like question 2a, can be asked of any time series graph. An answer to this type of question might be that crime is decreasing more slowly in Maryland than it had been. Question 3b is a comparison of two time series. An answer to this type of question might be that crime is leveling off faster in Maryland than in Virginia. Question 3c makes this even more general, and requires a more complex answer. An answer to this type of question might be that crime is leveling off faster in the Northeast than in the South, is beginning to rise in the West, and is just beginning to drop in the Midwest. Finally, question 3d requires a similarly complex answer, such as that while crime was moving from high concentrations in the Northeast and South to the Midwest, it is now starting to move to the West. The added complexity of these later questions sometimes makes it difficult to express the answers precisely in words, making the use of information graphics even more beneficial.

Types

There are many kinds of spatiotemporal maps, and I have listed some of them here. All spatiotemporal maps show at least one attribute at least two times and at least two locations, although the word map suggests more than two locations. The map types described below are organized by the number of times they depict, and whether they are

static or dynamic. The number of attributes and locations depicted are also important, but are generally less useful for classifying maps. It is possible to create new, more complex visualizations of spatiotemporal data by combining these maps with other graphs like time series line graphs or scatterplots. This list is not intended to be exhaustive, only to show many of the spatiotemporal maps in common use and those being developed by researchers.

Static Maps Showing a Limited Number of Times

These maps show how some attribute changes over time, but their design inherently limits how many times they depict.

Change Map. A change map is a single map that shows a single attribute: the change in some data from Time 1 to Time 2 (Figure 3). It is good for answering the questions about change (questions 2a, 2b, and 2c), but can not be used to answer the simplest question (1a) about values at a particular time and place. Thus, change maps are not true spatiotemporal maps, because they only represent one time, namely the time between Time 1 and Time 2. Change maps are discussed in Andrienko, Andrienko, and Gatalsky (2003).

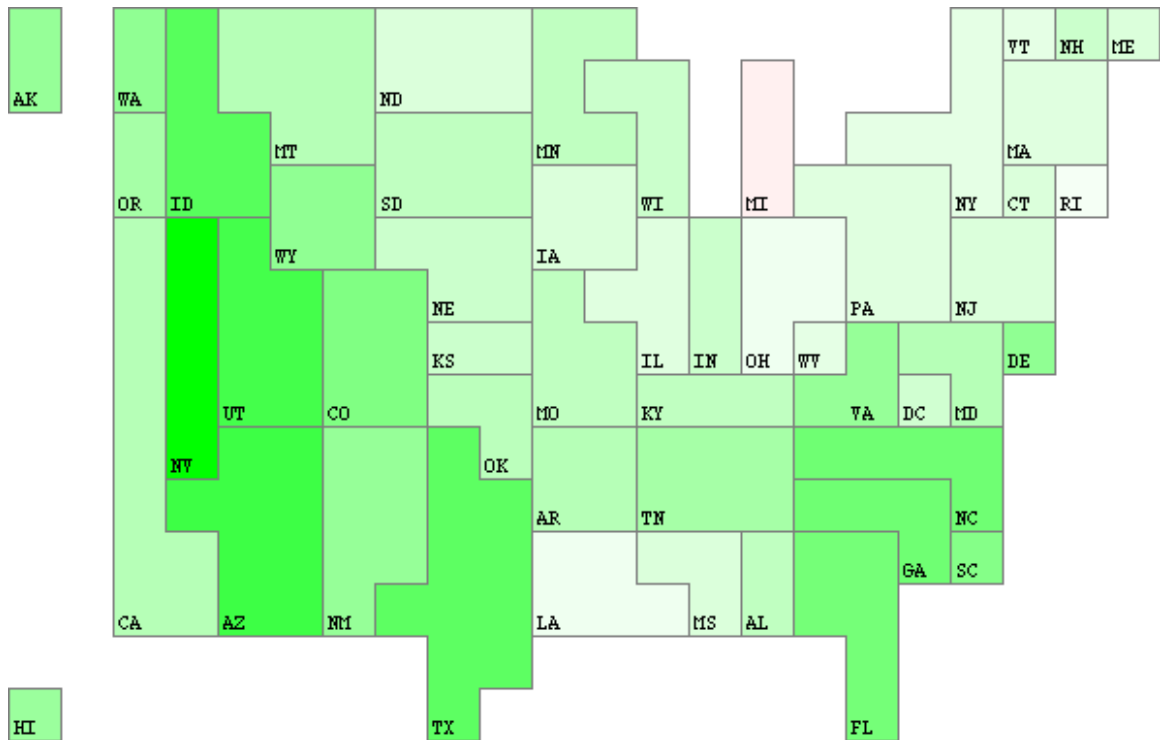


Figure 3 A change map.

This change map depicts the population growth rate of each State from 2000 to 2010, measured as a percentage of current population. Darker greens indicate greater percentage growth. The light pink of Michigan represents a decrease in population. There is no way to determine from this map which States have large populations.

Map Pair. A map pair is two maps of the same territory, each depicting the same attribute at different times, arranged next to one another. Any thematic map depicting an attribute at one time can be made into a map pair by adding a second map representing a second time. It is the simplest form of small multiple map. Map pairs can be used to show “before and after” information, as in Figure 4.

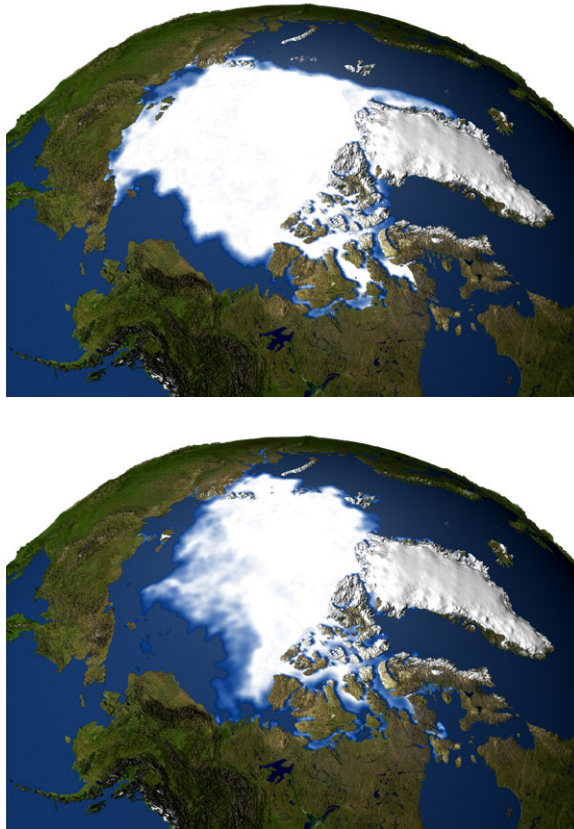


Figure 4. A map pair.

This map pair shows the Arctic ice cap in 1979 (left) and 2003 (right). NASA (2003).

Habitat Loss Map. This is a choropleth map in which the colors represent the extent of some attribute at different times (Figure 5). As more times are added, more colors become necessary, making the map increasingly difficult to interpret, especially because it is difficult to remember an ordering of variables like color (Bertin, 1967/1983). These maps are fine for depicting areas that shrink or grow monotonically, such as natural habitats or urban sprawl, but fail when depicting more nuanced data. If a wild animal population recovers, or part of a city is destroyed, the map must become more complex or less precise. Note the ambiguity in the map in Figure 5; we do not know whether there are any locations that had tigers in 1990, but not 1900. A fourth color

would be necessary to depict these areas. If a third time were added, up to four more colors would be necessary to show which areas the tigers lived in at various times. One way to alleviate this confusion is to use different retinal variables (outlines, patterns, etc.) for different times (Bertin, 1967/1983; MacEachren, 1995). See Maps With More Than One Way of Showing the Attribute below.

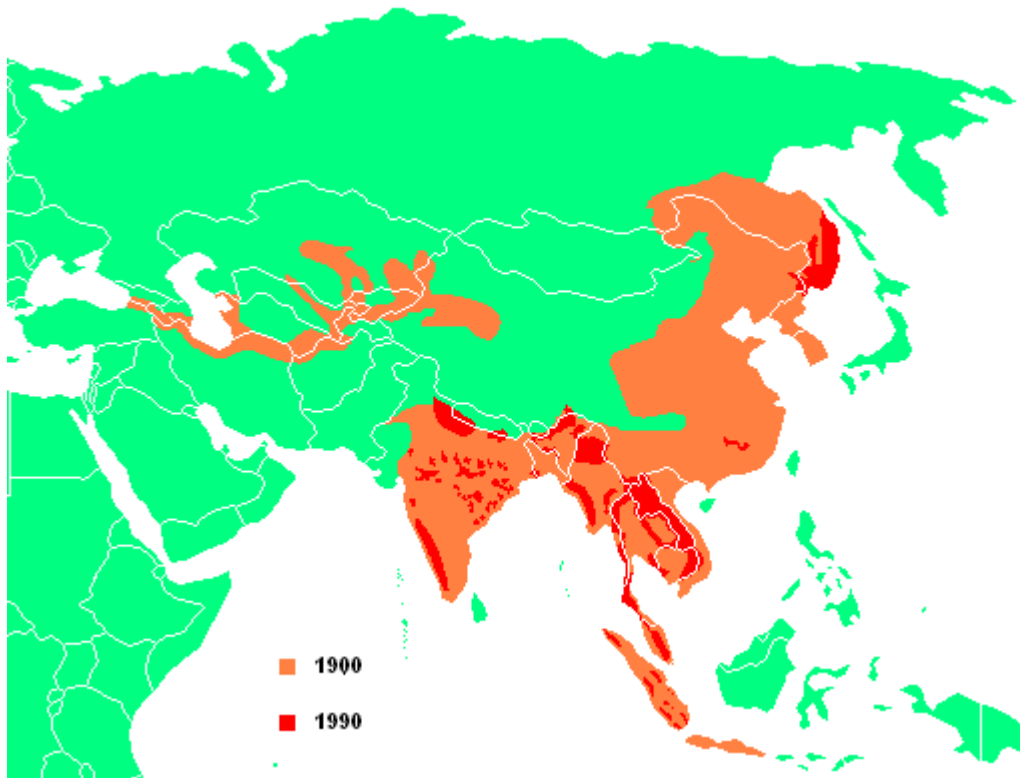


Figure 5. Habitat loss map.

In this map of tiger habitat, areas that were not tiger habitat in 1900 or 1990 are depicted in green. Areas that were tiger habitat in 1990 are depicted in red. The other areas, depicted in orange, represent habitat lost to tigers between 1900 and 1990.

<http://en.wikipedia.org/wiki/Tiger>

Static Maps Showing and Unlimited Number of Times

These maps do not necessarily depict continuous or data, or infinite amounts of data, but it is relatively easy to add new times to one of these types of maps, by adding new bars, lengthening line graphs, adding small maps, etc.

Bar Graph Map. A bar graph map is a map with small bar graphs embedded in it at each location. Each bar in the small graph represents a different time. The bar graphs can have as few as one bar in them, but the map is spatiotemporal when they have two or more (Figure 6). Line graphs or other time series graphs can also be used. Bertin (1967/1983) calls these *chartmaps*, that is, maps embedded with small charts. Chartmaps can also have pie charts, or any other charts, not just time series.

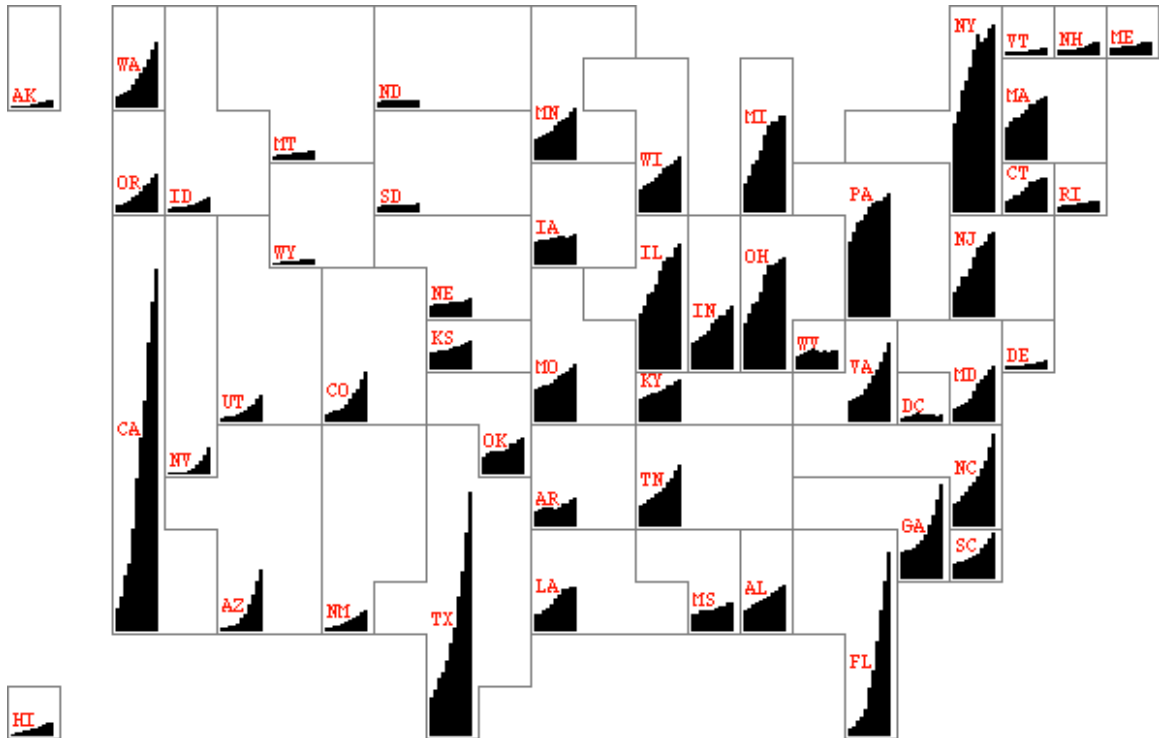


Figure 6. A bar graph map.

The population of each State, measured by decennial Census data, is shown using bar graphs.

Map Array. A map array is a series of maps, each showing a different time, arranged near each other, usually in a row or several rows (Figure 7). This is the extension of the map pair that allows the user to answer questions about more than two times. If there are many maps, they may need to be small to fit within a defined space.

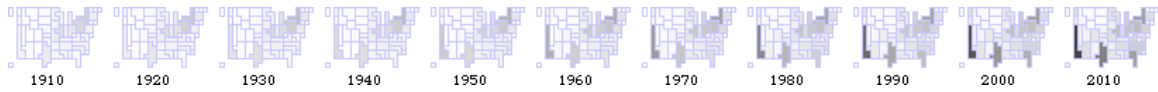


Figure 7. A map array.

The population of each state is shown for each census year from 1910 to 2010. Darker grays indicate higher population.

Maps With More Than One Way of Showing the Attribute. A map could show the values of an attribute at Time 1 as a choropleth map, with bars or other signs indicating changes from Time 1 to Time 2 and beyond. Andrienko and Andrienko (2006) discuss maps with arrows showing the migratory patterns of birds. Football diagrams show the starting position and how each player is supposed to move and change direction when the play begins. There are many possible maps of this type that may be useful for specific tasks. Maps showing historical battles often use shaded areas or lines to show where military forces began the battle, with arrows or other illustrations of the movements of those forces. The most famous of these is Minard's map (Figure 8) of Napoleon's disastrous campaign in Russia (Tufte, 1983), which shows the army getting smaller as time goes on using a band that gets thinner. The orientation of the band, and several geographic and other graphical elements included in the map allow it to depict six variables (Tufte, 1983). Although the attention this map has gained is well deserved, it is not possible to create a similar map for most kinds of spatiotemporal data.

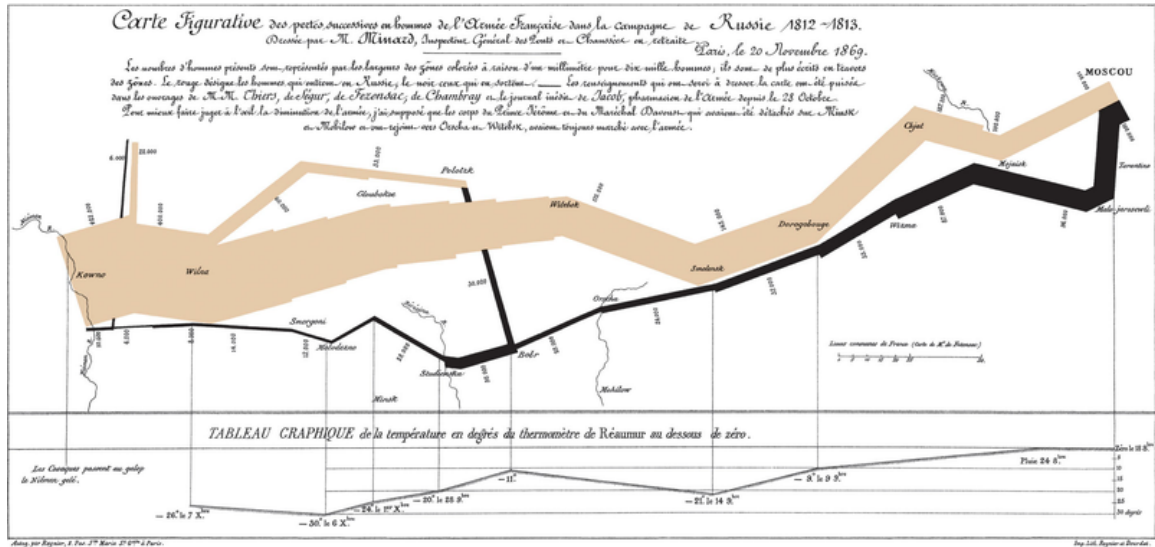


Figure 8. Minard's map.

http://en.wikipedia.org/wiki/Charles_Joseph_Minard

Dynamic Maps.

These maps require an electronic medium such as a computer or television. These maps convey some of their information by changing their state over time or in response to input from the user.

Animated Map. Just as the name implies, an animated map is a map in which the image changes to show the attribute at different times. Weather maps on television or the web use animation to show changing weather patterns. Animated maps are often promoted as particularly useful because they represent the dimension of time with changes in the depiction over time, albeit at different scales. Tversky, Morrison, and Betrancourt (2002) call this the congruence principle. Without some level of interactivity, they say, animations have limited applications. Controllable animations are common on the web. Figure 9 shows a web-based weather map that allows the user to control the speed of the animation, among other functions.

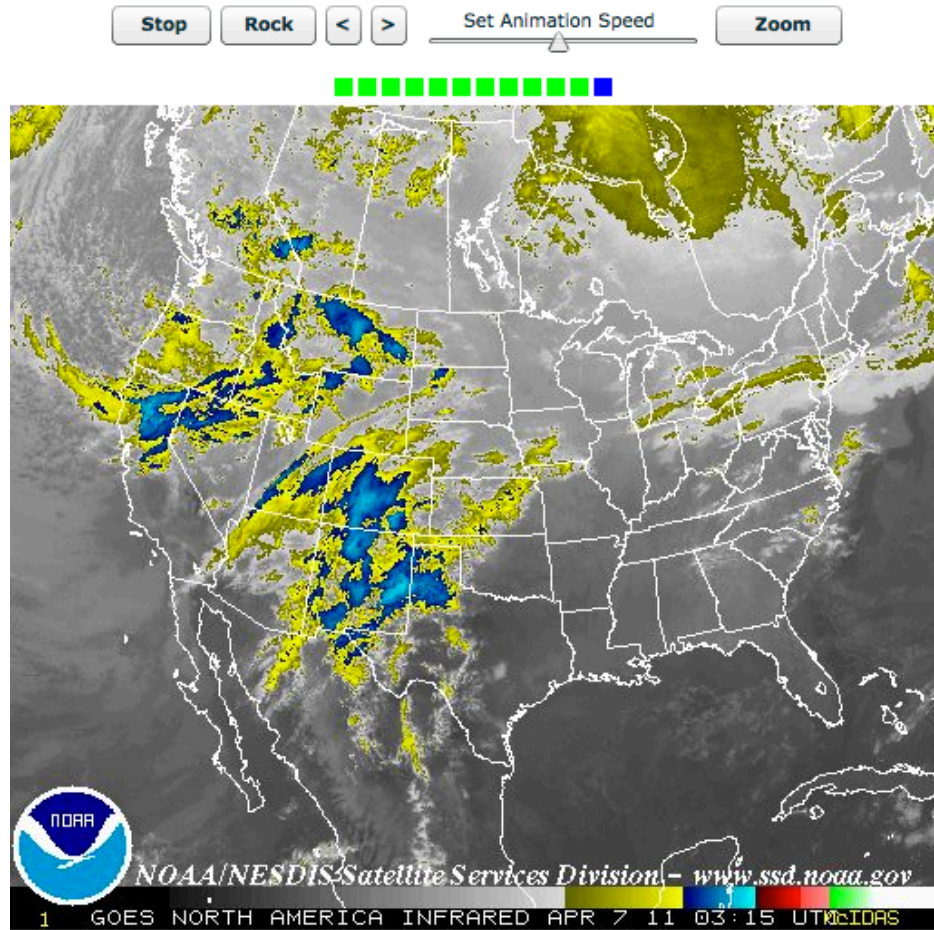


Figure 9. An interactive, animated weather map.

www.weather.gov/sat_loop.php?image=ir&hours=24

Cube Map. A cube map is a 2-D projection of a 3-D cube in which two dimensions represent space, and one, time, and the attribute is represented as signs, such as spheres in this 3-D space (Andrienko et al., 2003; Andrienko & Andrienko, 2006). Cube maps are usually used for point data, not for data with predefined locations. Rotating the cube makes it possible to see the signs in their virtual space, and clusters become readily apparent. As a static image, (Figure 10) the cube map is uninterpretable.

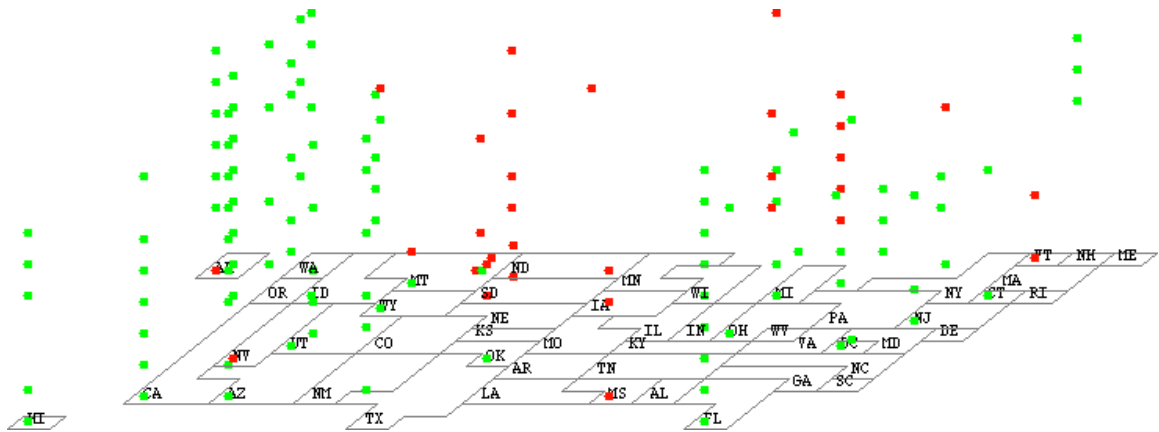


Figure 10. A cube map.

The dimension perpendicular to the map surface represents time. Each dot corresponds to a particular point on the map, here, a particular State. Dots higher in the vertical space represent more recent data. Green dots represent a growth rate of over 20% from the previous Census. Red dots represent any negative growth rate from the previous Census.

Multi-Component Interactive Maps. Visualizations can have related components that are linked such that each reacts when any of them are manipulated by the user. These can include line graphs, scatterplots, tables, or any other component (Figure 11). Many such systems have been developed and described, including by MacEachren, Dai, Hardisty, Guo, and Lengerich (2003), Andrienko and Andrienko (1999) Andrienko and Andrienko (2004), Guo, Chen, MacEachren, and Liao (2005), and Carr, Chen, Bell, Pickle, and Zhang (2002).

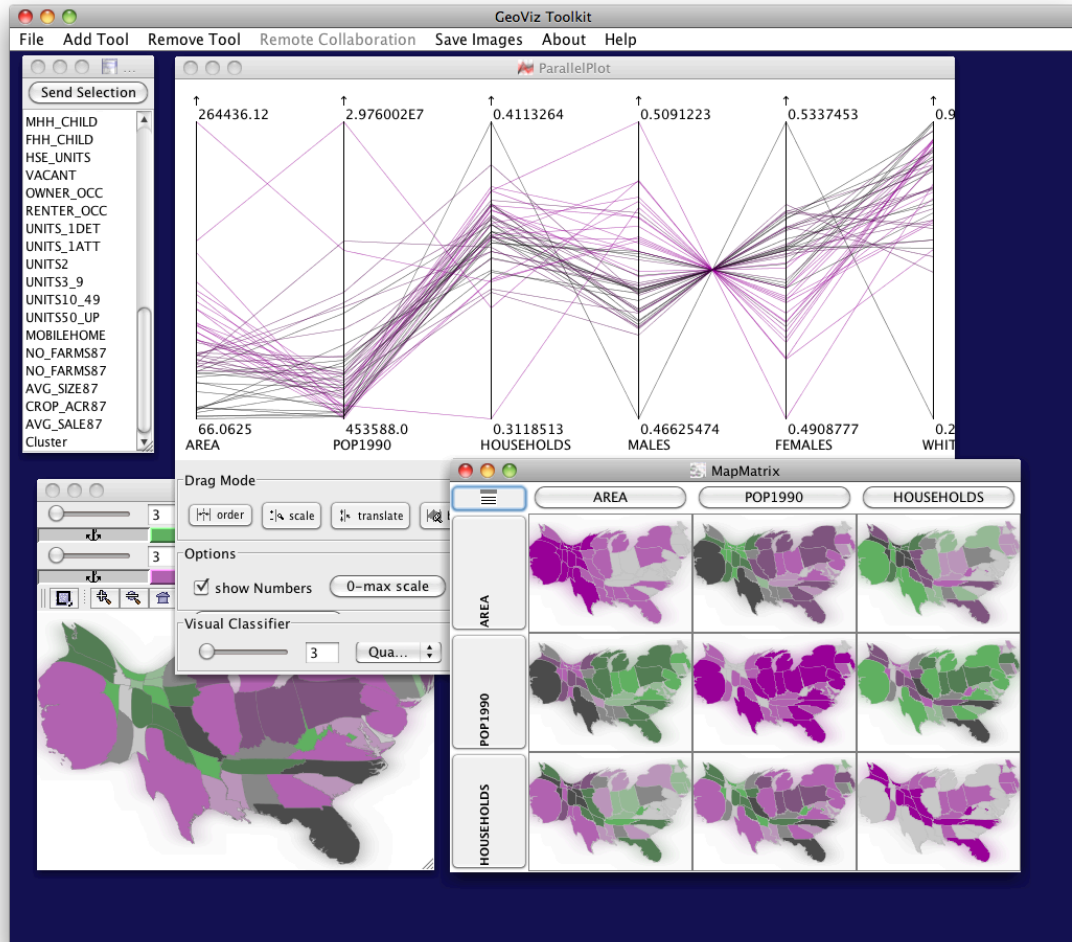


Figure 11. A multi-component interactive map.

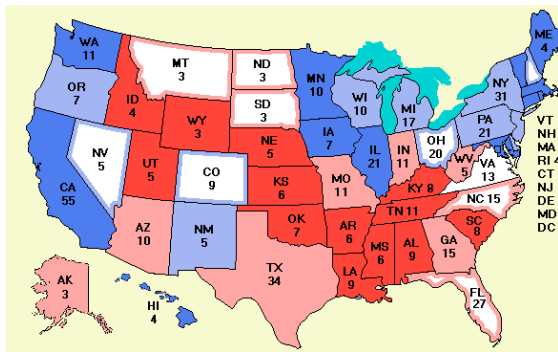
This still image is from GeoViz Toolkit, a GIS program that includes linked interactive maps and graphs. Hardisty, Myers, & Liao (2007).

There is a broad variety of ways of displaying spatiotemporal information. Each new method has some kind of drawback, and is usually best suited for looking at particular kinds of data or answering particular kinds of questions. Another common feature is that each new method builds on a simpler method (or more than one). Thus the presence of new and exciting visualizations should encourage research into how people can make use of more basic visualizations.

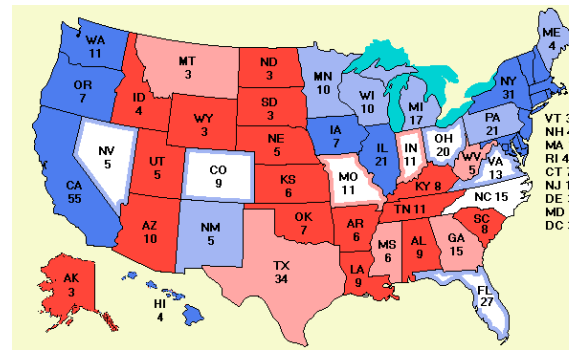
An Example: 2008 Presidential Polling Data

Presidential elections in the United States are decided by the Electoral College, the members of which, called “electors,” are chosen according to which candidate they are pledged to support by the voters of each state and the District of Columbia (which is like a state in this process, and will be treated as one for the remainder of this example). Because different electors are chosen by each state, close followers of presidential politics often want to predict which candidates will win each state, based on past trends, demographics, and polling data. During a presidential election cycle, it is common to see maps of the U. S. colored in various shades of red, white, and blue, depicting which candidate led in each state, and the strength of his lead, according to recent polls. These maps appear in newspapers and magazines, and on web sites. Some web sites include animated versions of these maps, showing how the lead changes over time (Figure 12 shows some examples from the 2008 presidential election cycle). Comparisons of these maps, either simultaneously or sequentially, allows the viewer to see changes, like Obama’s surprise win in Indiana, or trends, like Obama’s gradual increase in support in several Southeastern States, three of which he eventually won. Another common visualization of poll data is a line graph showing the level of support for each candidate with a colored line. The reader can observe the relative positions of these lines at various times on the X axis to see which candidate led, and by how much. The most common versions of these graphs show estimates of the overall popular vote, but some show estimates of the final tally of the Electoral College instead (Figure 13). The chartmap in Figure 14 shows a new way to display this information. The standard two-line graph is shown for each state, with a vertical bar representing each actual poll. The bar is color

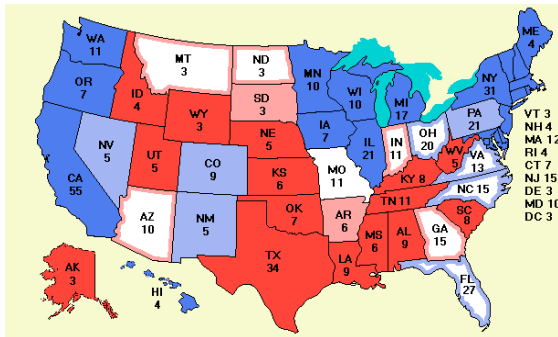
coded to represent the leading candidate. Local trends and the distribution of these trends are made clear by showing small graphs arranged as a map. A drawback is that the distribution at any one time is not displayed prominently, although this could be solved by highlighting (Andrienko & Andrienko, 2006), or another technique added to an interactive version of the electoral chartmap.



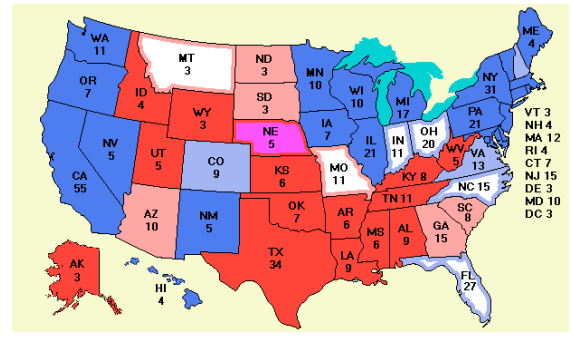
September 4, 2008



October 4, 2008



November 4, 2008 (final prediction)



December 31, 2008 (final results)

Figure 12. Maps showing 2008 Presidential polling and results.

Blue represents Obama, and red, McCain. Darker colors indicate stronger levels of support for the leading candidate in each state, while white indicates a toss-up. The red-and-purple coloring of Nebraska in the final map represents that state splitting its electoral delegation.

<http://electoral-vote.com/evp2008/Pres/Maps/Sep04.html> (and Oct04.html, Nov04.html, Dec31.html)

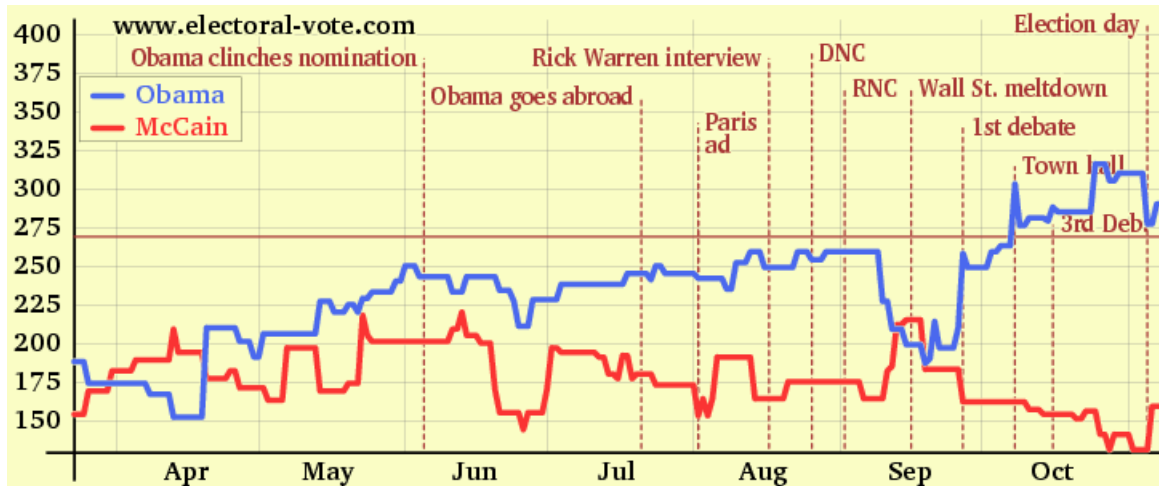


Figure 13. Line graph predicting the electoral college outcome of the 2008 U. S. Presidential election.

This graph does not count toss-up States, which is why the lines are not mirror images of one another.

http://electoral-vote.com/evp2008/Pres/ec_graph-2008.html

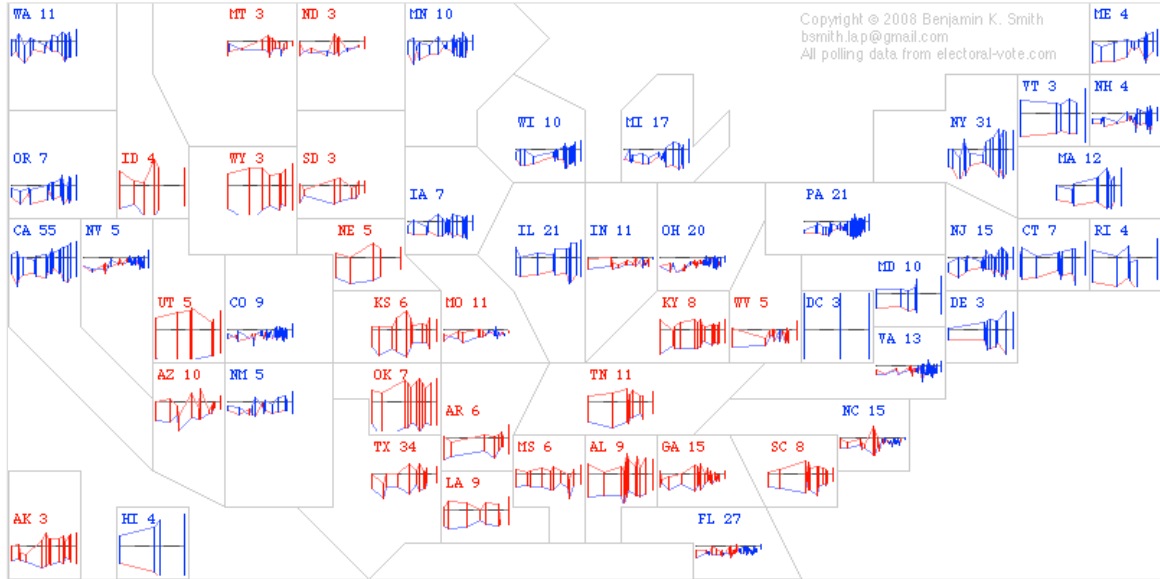


Figure 14. A chartmap of the state-by-state polling data.

Each state has its own graph. Each colored vertical bar represents a poll. The color of the bar represents which candidate led, and the length of the bar indicates the size of that candidate's lead. The gray horizontal line through each graph represents 50% of the poll, so if a bar is completely under the line, neither candidate polled 50% or more. The shapes and sizes of the states have been distorted to allow all the graphs to fit on one page, without overlapping. The numbers indicate the electoral vote totals for that state. The colors of the state abbreviations indicate the final winner of the state, as does the rightmost bar of each graph. The pink and light blue lines are interpolations.

http://lap.umd.edu/LAP/People/benjamin_smith/Polls.html

The Case for Small Multiples

With the many kinds of graphs available to depict spatiotemporal information, why should we pay particular attention to small multiples? First of all, small multiples are easy to create because they are made up of familiar forms. Our knowledge about how to draw effective bar graphs or choropleth maps can be reused to make effective small multiples of these forms. Second, along the same lines, it is easy to comprehend small

multiples of existing forms. Third, although small multiples do not require interactivity, they can be an integral part of an interactive system, even in a small space, as illustrated in Figure 15 (Apple, Inc. 2009).



Figure 15. Small multiples as part of a complex interactive interface.

Small multiples show a timeline for movies captured on an iPhone. (Apple, Inc. 2009).

A fourth reason is that despite their increasing popularity, small multiples remain somewhat controversial. Few scientific studies of the usefulness of small multiples have been conducted until fairly recently (Bauer, Geurlain, & Brown, 2010), and some of those have only attempted to show how other visualizations are better for a specific purpose (Griffin, MacEachren, Hardisty, Steiner, and Li, 2006; MacEachren, 1995). See Controversy Over Small Multiples, below.

The purpose of graphs is to assist people in acquiring and comprehending information by taking advantage of our considerable perceptual resources. This is true whether or not the graph is created for the purpose of promoting a particular interpretation of the data. Much of the discussion on effective visual communication centers on how to frame graphs, what labels to add, what information to include or not include on a graph. Tufte (1983) divided this information into *data ink*, meaning only those marks on the page that directly convey and identify the actual values of the underlying attributes, and *chartjunk*, which is everything else. Tufte cautions that chartjunk can distort the data, and confuse the graph reader. Monmonier (1996), describes the same phenomena from a different angle in *How to Lie With Maps*, instructing the reader in how to identify, and, if necessary, employ, graphic techniques to promote a particular viewpoint. Tufte discusses the importance of not lying with the data ink, but rather keeping the data and the ink used to depict it proportional. Bertin (1967/1983), however, cautions that the ink used and how that ink is interpreted are not always proportional. He describes how, when using black and white shading, proportional increases in the percentage of shading will be interpreted as more significant near total blackness or whiteness, and thus proposes a nonlinear system of shading to portray equal classes of information. The desire to avoid the slippery slope of disproportional depiction, along with its broad utility, led me to choose length as the retinal variable in the experiments described below. Different lengths are generally judged to be proportional in accordance with their objectively measurable proportions, something that is rare in the human perceptual system (Gescheider, 1997).

Small Multiples and Theories of Graph Reading

Several theorists have considered how graphs help us take advantage of the abilities of our visual processing system to better understand data. All of these theorists prefer simpler graphics to more complex ones when the added complexity does not assist the graph reader in comprehending or interacting with data. However, they reach different conclusions about the practicality of small multiples.

Bertin's *Semiology of Graphics* (1967/1983) describes principles of map and graph making inspired by Gestalt principles. Color, for example, can be used to group objects in different places, allowing the graph reader to attend to only one group at a time, an ability called *selective perception*. Some other retinal variables, such as shape, do not afford this ability. But others, such as size, do, and offer the added ability of ordering, meaning that the reader can easily determine how different objects vary on some scale. Although there are many ways to display the same information, Bertin argues, the best methods maximize the efficiency with which a graph reader can answer questions about the information displayed. The most efficient type of graphic depends on the kinds of questions it must answer. Bertin draws a distinction between two types of graphics: an *image*, in which all of the information can be comprehended in a single "instant of perception," and a *figuration*, in which multiple instants of perception are necessary, and the comparison of the information from those instants requires some mental cost. Images are not necessarily smaller, and do not necessarily contain less information than figurations, although this is often the case. Bertin gives the example of a large, detailed map that contains only one kind of information as an image, and a basic pie chart, which Bertin finds unreadable, as a figuration. Although Bertin prefers images

to figurations generally, when a graphic needs to convey more than a small number of variables, figurations are necessary. These figurations include both graphs with multiple kinds of signs or with signs that differ in more than one quality, and collections of smaller, simpler graphics. Bertin considers most chartmaps and most small multiples to be figurations. Figurations are useful for graphics that record lots of information in a fairly raw state, and Bertin gives guidelines for how to construct these. Graphics that convey too many kinds of information may end up unusable for any purpose, because the signs are too complex to comprehend. Ultimately, Bertin argues, in order to use graphics to express a particular message, the information must be simplified to the point where it can be drawn in an image, because people can only remember images.

Pinker (1990) proposed a *Graph Difficulty Principle*, that information is harder to retrieve when top-down encoding and inferential processes are necessary to comprehend it. Pinker's preference was to have the quantitative information readable directly from the graph, and for the reader to then apply that information to the relevant conceptual question. He gave the example of two variables, both increasing over time. In a table, it is easy to determine the exact value of either variable at any given time, by reading the number. If the variables are presented in a bar graph, it is harder to determine the exact value of a variable at a given time; the comparison of the bar height to the values on the axis is a kind of top-down processing. However, the bar graph makes it easier to compare the ratios of the values at a given time, because the comparison of the bars is a visual process. Replacing the bars with a line graph of the variables makes it harder to compare the ratios, but easier to compare the relative trends by looking at how the lines differ (Simcox, 1984). By similar logic, Pinker (citing Schutz, 1961a, 1961b) prefers line

graphs with multiple lines are preferred to multiple graphs with a single line each, for a small number of lines.

Tufte (1983, 1990), coined the term small multiples, and promoted them as an efficient way to display highly dimensional data. He notes that they allow a reader to find interesting features, compare one multiple against another, use spatial dimensions for more than one variable without causing confusion, and in fact take advantage of more than one spatial dimension to find complex patterns. Tufte calls the ability afforded to the reader of examining local or global patterns *micro/macro readings*, similar to Bertin's (1967/1983) concept of the image. By contrast, some graphs and most tables can not be comprehended as wholes, which Tufte calls *flatland*, similar to Bertin's figuration.

Problems comparing small multiples, Tufte argued, can be alleviated by reducing lines or other non-data ink that provide visual interference, leaving only thin lines as guides when necessary to help the reader read the graph. He describes how color, used sparingly, can make graphs easier to read by focusing the reader's attention on the key details of each graph.

Shneiderman's (1996) Visual Information Seeking Mantra of "overview first, zoom and filter, then details-on-demand," is aimed at the designers of interactive information visualization systems, with the goal of supporting the tasks that will help users acquire the knowledge they are after. Shneiderman and Plaisant (2005) give several examples of interfaces that use small multiples as a starting point for this process (the overview), with the component signs and pictures moving and changing as the user interacts with them (zoom and filter).

Andrienko and Andrienko (2006) use and recommend small multiples, chartmaps, and complex interactive systems including the cube map. They have a set of ten principles (based on those of Shneiderman, 1996) for visualizing spatiotemporal data, that begin with “see the whole,” and work their way to “attend to particulars.” These principles also include “simplify and abstract,” and, “zoom and focus.” These kinds of operations are made easier through the use of small multiples in the forms described earlier, and different forms of small multiples are better for particular tasks. The cube map is for “looking for the recognisable,” precise graphs for “attending to the particulars.”

There seems to be agreement on a few points. First, small multiples are useful as a way of displaying large amounts of complex data. Second, a desirable use of a complex display is for the reader to easily discover both global and local patterns. Where the theorists differ is their belief about small multiples fulfilling this second use in addition to the first.

Controversy Over Small Multiples

Efforts to create new kinds of spatiotemporal maps have increased with the development of computer graphics, but researchers disagree over which basic approaches are the most useful, and research in this area has been inconsistent. Morse, Lewis, and Olsen (2000), and Chen and Yu (2000) found that not enough of the discussion of map reading (including spatiotemporal maps and small multiples) was based on objective research. Exceptions include Purchase (2000), who tested people using graphs designed using competing principles.

Tufte (1983, 1990) argued for the usefulness of “small multiples” for conveying multidimensional information, and for small multiple maps in particular as one useful way of showing changes in space over time. Monmonier (1990) concurred, showing how small graphs can simplify complex comparisons. These maps were criticized, however, by MacEachren (1995), on the grounds that he and his colleagues did not like them for this purpose, preferring animated maps. MacEachren, et al., (2003), argued later that small multiples were useful as part of an interactive system.

Griffin, MacEachren, Hardisty, Steiner, and Li (2006) attempted to show that animated maps were better than small multiple maps for detecting cluster movement. The clusters were groups of nearby dark hexes on a uniform hex grid, with the center of the cluster moving from one time, represented by a frame of the animation, or small map, to the next. They set up equivalent animated and small multiple maps and asked people to find moving clusters. They did in fact find that people were better at identifying moving clusters when viewing an animation than when viewing the small multiples. This was not a surprising finding, because the apparent motion created by animation attracts people’s attention (Ware, 2004), and the maps had no fixed features that someone could use to help them make comparisons. Furthermore, in the small multiples condition, the maps were on two rows, forcing the participants to move diagonally while trying to find clusters that were also moving diagonally. Unfortunately, no attempt was made to understand why the difference existed or whether it could be alleviated, allowing people to be more effective at using small multiples for this particular task. Koussoulakou and Kraak (1992) found that animated maps produced faster responses than static maps that showed different times as layers in a single space, but said that to be useful, these

animated maps required interactivity. Tversky, Morrison, and Betrancourt, (2002) described the key advantage of animation as the fact that it follows the Congruence Principle, that time is represented by time, and space by space. When animations fail, they say, it is because they do not follow the Apprehension Principle, that the structure and content should be easy to understand. Animations, they say, may be hard to perceive, and users may think of time, in some cases, as discrete events as opposed to a smooth flow. Another issue is *change blindness* (Simons, 2000), the phenomenon of people not being able to detect changes in what they see if those changes are very slow or if the view is interrupted.

Although the discussion to this point has focused on spatiotemporal maps, small multiples are often used for other purposes, which have produced more experimental research. One such purpose is to assist people monitoring complex mechanical processes. The most common way to do this is with an instrument panel, or a computer monitor showing various critical values at various locations on the screen. A criticism of this approach is that it takes time for people to move their eyes to various parts of the screen, wasting time that could be critical when the state of a system changes rapidly. A proposed alternative called RAP COM or RSVP uses a single display location and changes the content of that display periodically, once a second, or faster. Initial studies of RAP COM systems (Payne & Lang, 1991) showed people reacting to them quickly, but less accurately than with multiple displays arranged on a screen. Later studies (Konrad et al, 1996) showed that the accuracy penalty may have been related to response mode.

Spence, et al., (2002) tested people's memory for pictures using either a static grid of pictures; a large, rapidly changing picture; or a mixed mode of four pictures that changed at a slower pace. They found the lowest error rates and highest user preference for the mixed mode, followed by the static mode, particularly when the total time allowed for viewing the images was short. They found that people tended to look at the center of the screen where the four pictures met in the mixed mode, reducing the need for eye movements.

There are two important points to consider for our discussion of small multiples as a way of displaying spatiotemporal information, or similar complex datasets, as opposed to monitoring. First, as Payne and Lang (1991) point out, it is difficult for a person being presented with rapidly changing information to absorb all of that information before the display changes. Second, because these systems are automated, it is not easy to make detailed comparisons between two nonconsecutive graphs. Overcoming these obstacles requires an interactive, not merely a dynamic system, and ultimately, a small multiple might be best for the final detailed comparison. Bederson, Shneiderman, and Watenberg (2002) used an interactive system with small multiples arranged hierarchically to allow people to find images quickly.

Graph Difficulty Principle for Small Multiples

The goal of this research is to provide a Small Multiple Graph Difficulty Principle that addresses concerns specific to small multiples. Research will help to test the competing theories of small multiple reading and provide guidance to the graph maker. One concern of any small multiple visualization is the layout of the various graphs, maps,

or pictures. If graphs are close together, fewer and smaller eye movements are needed to examine them. But if we are limited to showing only graphs that are near one another, the potential benefits of graphs with many displays are lost to us. Tufte (1990), Bertin, (1967/1983), and Andrienko and Andrienko (2006) all showed graphs made up of hundreds of multiples, and all discussed the benefits of interpreting a complex graph as a whole. A second concern is how the signs used to convey attributes can affect how well the graph can be read. Bertin suggests that individual graph readings must be fast and easy, based on retinal variables, to enable these kinds of readings. A third concern is how a too-complex graph, or one laden with too much chartjunk or too few guides to allow comparisons, can impair graph readers.

Hollands and Spence (1998, 2001) tested people reading various graphs, basing their experiments on the idea that perceptual variables are the easiest visual information to comprehend. They found that for judgments of proportion, the whole must be represented as a perceptual variable for optimal readability. Jessa and Burns (2007) added redundant orientation information to a bar graph and found that it assisted users in noticing changes, whereas another kind of redundant information did not. Van den Berg, et al. (2008) found that orientation was useful, but has a limited usable range of values, and was often dominated by other cues.

To begin investigating a graph difficulty principle for small multiples required some decisions to be made about which aspects of small multiples to study, including which to vary experimentally. One aspect that did not vary was the type of sign used; all experiments used bar graphs.

Why bar graphs? I initially studied small choropleth maps (see Figure 16), and found that people were able to make detailed comparisons of small maps at different scales with reasonable accuracy. There were some drawbacks to using small maps as the component graphs of an investigation of small multiples visualizations. First, the inherent complexity of maps allows for a great variety of aspects to manipulate, and a great many tasks, including complex tasks, for graph readers to perform. There were too many to pick from. Second, as a result of this complexity, items took a fairly long time, on the order of ten seconds per item. This would make it impractical to run the hundreds of trials necessary to investigate more than a very small number of conditions.

Bar graphs were chosen as a simple alternative for these experiments, so the focus could be on small multiples, rather than map reading. It is therefore left to future research to develop a graph difficulty principle for small multiple maps.

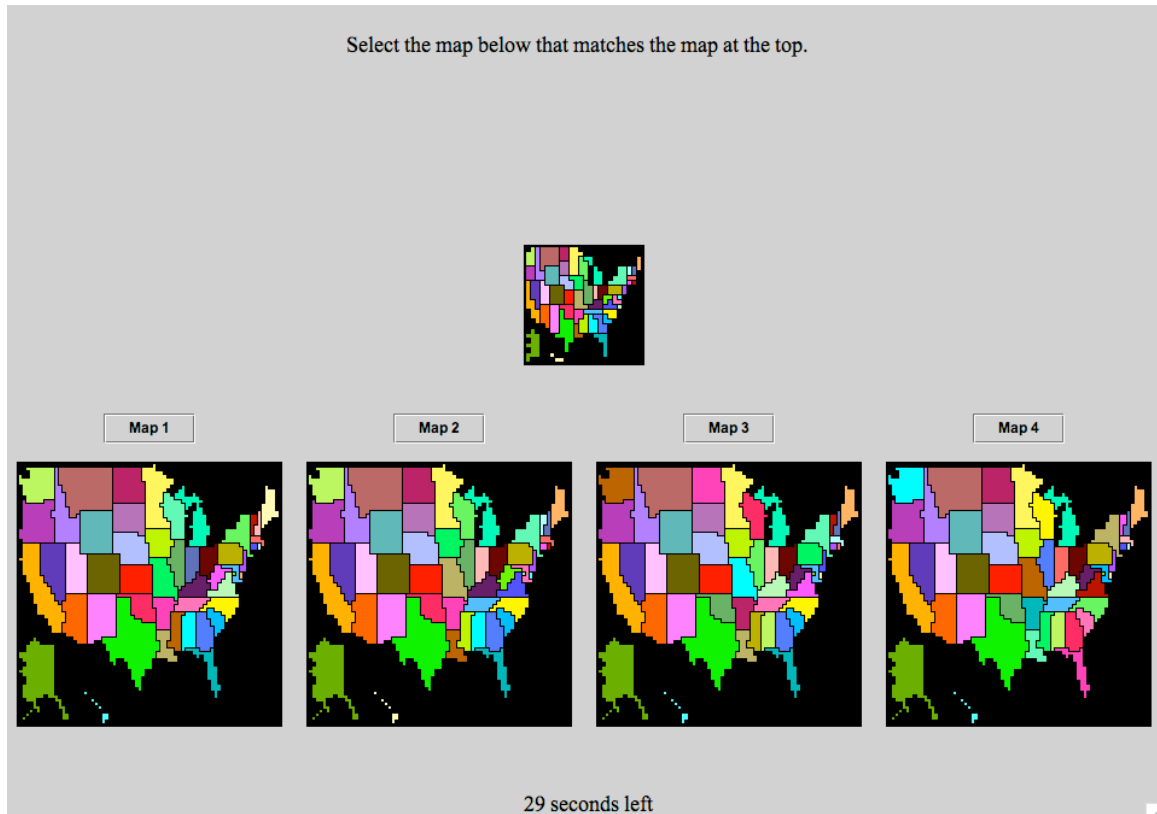


Figure 16. A screen from an earlier experiment on comparing small choropleth maps.

Three aspects of how small multiples can differ were investigated. First is the arrangement of graphs on the page, including the alignment of the graphs and the distance between them. These differences are critical to understanding small multiples, because constructing even the simplest small multiple visualization (compare Figure 4 to Figure 7, which represent time with different dimensions) requires a decision to be made about how to arrange graphs. A poor choice can limit the usefulness of the visualization.

A second aspect was how a difference in the sign used can affect the comparability of graphs. Rather than investigate the many kinds of signs that can be made into a small multiple visualization, I picked a simple difference, the orientation of the bars in the graphs. Both horizontal and vertical bar graphs are common, including as part of small multiples. I wanted to know whether the orientation of the bars interacted

with the alignment of the graphs they were part of. Culbertson and Powers (1959) found that vertical bars were more somewhat more comparable than horizontal bars under some circumstances, but not how the orientation and alignment interacted.

A third aspect was the number of bars in the graph. A fourth was the presence of non-data ink, including additional graphs that were not relevant for the task, and graph axes. Finally, although not least, I considered different kinds of tasks that a graph reader might want to perform using small multiple graphs.

Judgments of Size and Line Length

Judging the size of a distant object is a basic task of the human visual system. It is one we perform whenever our eyes are open and the light from distant objects enters them. Philosophers and scientists have been trying to figure out how we do it since before even this basic fact of the visual system was understood. Euclid (Smith, 2001), for instance, believed that the eye emits rays, known as visual flux, and that people determine the size of an object by the number of rays touched by the object. Objects that are distant enough are invisible because they fall into the gaps between these rays. Close objects are perceived better because more rays touch them. Whether Euclid believed these rays to be a literal mechanical explanation of vision (as Smith believes), or simply a tool for explaining its geometry, the model does address the most important factor for determining the apparent size of objects: the degree of visual angle that those objects subtend in our field of vision. It does not, however, explain the paradox of how a small, close object, and a larger, farther object, which subtend the same degree of visual angle, are not perceived to be the same size.

Ptolemy refined Euclid's model somewhat, changing flux from tightly packed rays to a continuous cone of vision (Smith, 2001). That cone's vertex was in the center of the eye, and the flux was strongest in the center of the cone. Ptolemy understood light to be important in the visual process, something that Euclid's model does not address. Rather than the light of distant objects entering the eye, however, light was necessary for the distant object to color the visual ray. Ptolemy's model added an important ability to the visual system: the innate sense of how far away an object is, an ability provided by the cone of visual flux. Using this distance, the degree of visual angle subtended by a distant object, and geometry, the observer is able to determine the size of objects. Ptolemy, with his focus on color and geometry, was also aware of other clues that our visual systems use for this task. The dimmer of two equally sized objects, for instance, will be perceived as both farther away and larger than the brighter of the two, a trick used even then by visual artists.

Alhacen, writing circa 1038 (Smith, 2001) was one of the first thinkers to understand that light entering the eye was the true basis of vision. He described the parts of the eye, and how they interact, including the transmission of visual information through the optic nerve. Alhacen understood that the angle subtended by the light from distant objects was part of how we determine the size of those objects, but that there were other factors that could override this angle. Familiar objects, for instance, are understood to be a certain size, and moving them forwards or backwards relative to the eye will not change this perception under ordinary circumstances. Alhacen understood that the distance between the eye and the object, used in combination with the angle subtended by the object in the visual field, allowed the viewer to determine the size of the object.

Alhacen also understood, as did Ptolemy, that the viewer was resistant to the changes in angle and size of lines that occur when objects are tilted.

Psychophysics

The study of comparisons of line length as an experimental, rather than philosophical, exercise began with Weber (1834/1978), who drew lines of varying lengths on pieces of paper, and had people compare pairs of these to determine which line was longer. The shortest line was 100 mm, the others ranged from half a mm longer to several mm longer. Observers who were skilled artists, or who had practiced the task extensively, could discriminate reliably between the 100 and 101 mm lines, although they sometimes made mistakes when tired. Other observers could reliably discriminate between the 100 and 105 mm lines. Weber measured people's perceptual abilities by what ratio of stimuli they could reliably discriminate between, and thus noted that the first group could discriminate the lines that differed by one percent, and the second group by five percent. Weber also discussed the issue of comparing two lines that are at different distances from one another, noting that it is easier to compare the lengths of two parallel lines that have the same baseline, if they are closer together than if they are farther apart. Weber's hypothesis is that we imagine a line that connects the tops of the two visible lines, and that the larger the angle of this line to the horizontal, the easier the comparison is. He was also aware that the center of the retina is much stronger than the rest of the retina, and that lines closer together can be seen simultaneously with the retina. (This argument, oddly enough, is part of a larger argument that things are easier to compare if presented sequentially, rather than simultaneously. Weber argues that almost

all vision is done sequentially, because the most sensitive part of the retina is so small, and we need to turn our eyes to see different objects in detail.)

Fechner (1887/1987) formalized and expanded on Weber's principle of constant ratios of discriminability, calling it Weber's Law. Using the example of comparing line lengths, the artist/practiced group could notice a difference of 1%, and the other group could notice a difference of 5% of the length of one of the lines. Fechner, like Weber, believed that the noticeable differences varied with the magnitude of the stimulus, resulting in a consistent ratio. Fechner treated these Just-Noticeable Differences, or JNDs, as a countable measure of sensation. He believed that these ratios would hold for any kind of sensation. Thus a small JND between two short lines or soft sounds was equal to the large JND between two long lines or loud sounds.

A distinction that Fechner described was that of outer psychophysics and inner psychophysics. Outer psychophysics describes mathematical relationship between outside stimuli that can be objectively measured, and the reports people make of their sensations of these stimuli. Inner psychophysics describes the specific details of how this process is achieved, such as the workings of the eyes and ears, neural pathways, and so on. Fechner knew that he was limited to studying outer psychophysics, but that inner psychophysics would be studied by future scientists. This distinction is still used as scientists from different branches try to reconcile findings about sensation and perception.

Fechner's theory stood for some time despite numerous challenges (Boring, 1950). Stevens (1975) showed that JNDs were not a constant function of the intensity of a stimulus. Increasing levels of electric shock, for instance, are perceived to grow more

quickly than the objectively measurable intensity level of the stimulus. Stevens proposed the power law as a substitute for the logarithm-based JND. The perception of each kind of stimulus can be described by an equation $\psi = k\phi^\beta$, where ψ is the perceived intensity, ϕ the objectively measured intensity, k a constant related to units of measurement, and β the exponent that gives the power law its name. For many kinds of stimuli, such as sound, the exponent is less than one, producing curves that are similar to the logarithm curves of Weber's law. For line length, the exponent is 1, meaning that perception of line length is not distorted at large or small values, making it an excellent basis for comparison to other stimuli. There has been wide acceptance that the power law is more generally accurate than the logarithmic law, although there have been debates about Stevens's measurement methods and what sort of data analysis is appropriate (Wagenaar, 1975, Parker, Schneider, and Kanow, 1975, Billock and Tsou, 2011).

Two theories of size perception

Holway and Boring (1941) finally addressed the topic of just how we compare distal and proximal images that subtend equal degrees of visual angle upon our retinas. Observers (including the authors) viewed circles that were projected onto screens in a dark hallway in the middle of the night. There was a comparison stimulus 10 feet away, and a standard stimulus projected onto a screen that was farther away, down a long hallway. The hallways were orthogonal to one another, so the observers had to turn to look at each one. The observers had to adjust the comparison stimulus until they believed that the physical sizes of the two stimuli were the same by communicating with the experimenter, who adjusted the apparatus. Time was not a factor. The standard stimulus

was always 1° of visual angle, so it was larger when projected onto a screen farther away from the observer.

Holway and Boring wanted to test two theories of how people perceive the sizes of objects at a distance. Under the law of the visual angle, the apparent sizes of objects depend on the amount of visual angle that they subtend. So under this theory, the observers would not adjust the size of the comparison stimulus when the standard stimulus was farther away and physically larger. Under the law of size constancy, objects of equal physical size should be perceived as equally sized, regardless of distance. Under this theory, observers would increase the size of the comparison stimulus as the standard stimulus gets farther away, which would be graphed with a slope of $\tan 1^\circ$. Thus the stimuli would actually be the same size.

What Holway and Boring found was their observers (who included themselves) increased the size of the comparison stimulus to be very close to the actual physical size of the standard stimulus, closely resembling the law of size constancy. In some cases, the comparison stimulus was actually made larger than the standard stimulus, by about 12%, which they attribute to some unidentified experimental error. This ability depended on observers picking up on faint light sources, such as reflections off of the floor. Titchener and Pyle (1907) had studied the effects of faint light on line length perception using a Müller-Lyer illusion, although without showing a reliable effect. Holway and Boring tried blocking the already existing faint light sources, and found that these caused their observers to adjust the comparison stimulus more in line with the law of visual angle than the law of size constancy.

Judgment Time as a Measure of Difficulty

The earliest measure of the difficulty of a psychophysical judgment was introspection, as used by Weber (1834/1978), for example. Weber also considered the accuracy of judgments, as described above, but mostly as a means to finding the absolute limits of discriminability, not as a measure in and of itself. Holway and Boring (1941) measured accuracy, not in terms of percent correct in the study described above, but by measuring a subject-controlled stimulus with a ruler and comparing it to an experimenter-controlled stimulus. Another measure of the difficulty of a judgment is the time required to make that judgment. Henmon (1906, 1911), claimed that judgment time was a better measure of the difficulty of a judgment than the accuracy of the responses. Henmon noted that judgment time had been used by previous researchers as well, such as by using a metronome (Martin and Müller, 1899) in an experiment about judging lifted weights. Henmon measured judgment time using precise timers that were connected to a mechanical apparatus that exposed stimuli to the observers quickly, and recorded responses quickly. The judgments in these experiments included comparisons of the lengths of horizontal lines. Judgments were made more quickly when the differences between the lines were greater. For instance, in one experiment (Henmon, 1906), a standard line was 10 mm, and comparison lines were 10.5 mm to 13 mm, in .5 mm increments. I have plotted the results of this experiment in Figure 17, taking the data from Henmon's tables 13 and 14. Reaction times were shorter when the difference between the lines was small than when this difference was large. The difference was most pronounced between a 5% length difference and a 10% length difference, with little or no difference between a 25% and a 30% length difference. Henmon also asked for

confidence judgments after each response. Correct responses were faster, overall, than incorrect ones. Confidence was also strongly correlated with accuracy. Within each confidence category, however, incorrect answers were shorter than correct ones, indicating that in many cases people got the answer wrong because they answered too quickly, but in other cases they spent a long time trying to make a judgment, and then made the wrong decision. Henmon designed his experiment expecting to get accuracy near 84%, and got accuracy just less than that overall.

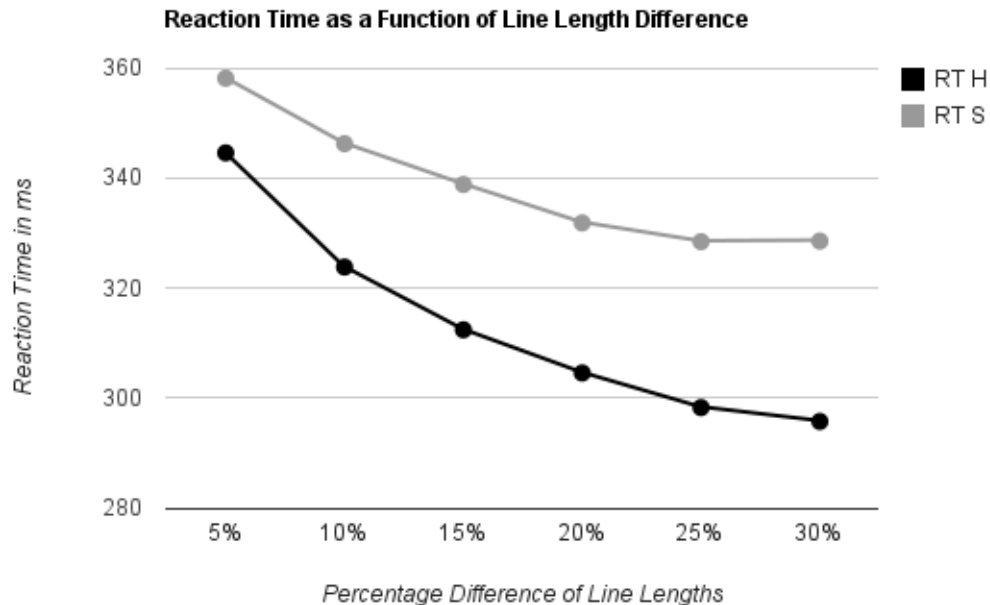


Figure 17. Reaction Time as a function of Line Length Difference, data from Henmon, 1906.

H and S were two subjects. A 5% difference was equal to .5 mm, or about .057 degrees of visual angle.

Link used a computer-controlled oscilloscope to sequentially present pairs of lines to participants, who judged their lengths to be either the same or different by pressing one of two switches (Link, 1971, Link & Tindall, 1971). The lines were horizontal, with a standard of either 20 mm or 16 mm, and comparison lines of 16, 17, 18, 19, and 20 mm,

with the comparison the same as the standard on half of all trials. These judgments were made under 3 time constraint conditions: 260 ms limit, 460 ms limit, and no limit.

People were able to adjust their reaction time to the time condition, although under a 260 ms time limit this caused a serious loss of accuracy. Under 460 ms time constraints or no time constraint, there were large differences in accuracy between lines differing by 1 mm and those differing by 2 mm, and less as the differences became larger. In the speeded 260 ms condition, this difference did not exist; instead, accuracy steadily improved as the two lines became more different.

Eye Movement in HCI research

The part of the retina in the center of the visual field, known as the fovea, has the most photoreceptors, the highest concentration of the cones used for fine and color vision, and the clearest optic pathway of any part of the eye. The concentration of cones drops rapidly from the very center known as the foveola. Until about 1 degree of visual angle away from the center, the concentration of cones is greater than the concentration of rods. To take advantage of the finer vision available at the fovea, we routinely move our eyes towards interesting features in the environment, in what are called saccades. Although saccades are very fast, they disrupt vision for a short period. Furthermore, because visual acuity decreases away from the fovea, it sometimes takes several saccades, each relying on peripheral vision, to acquire a visual target. Thus the distance in the visual field between two objects affect the amount of time it takes to look from one to another (Bailey, 2002; Goldberg, 2000; Tessier-Lavigne, 2000).

For as long as people have been studying interface design, these eye movements have been an important topic (MacKenzie, 1992). Fitts, Jones, and Milton (1950) studied how pilots look from one instrument to another, finding that pilots looked longer at instruments that were more difficult to interpret. The arrangement of instruments also mattered; it was more efficient to have instruments that were related positioned close to one another. This is due to the decrease in visual acuity from central to peripheral vision, and the time it takes to move the eye to reposition central vision at the new target. This time has received different treatment in HCI literature. Card, Moran, and Newell (1983), treated this time as a constant. More recent researchers have measured the importance of minimizing the number of long eye movements to help people find information in a computer interface (Goldberg and Kotval, 1999). One common technique is to treat eye movements the same way as other movements in HCI, namely pointing devices. Miniotas (2000) treated eye movements explicitly as a pointing device and found that Fitts' law (Fitts, 1954) described them accurately.

The experiments described here include several in which the stimuli are separated by different distances, so it is informative to consider Fitts' Law (1954) and its implications. Although researchers argue over which version of the formula to use (MacKenzie, 1992), in all versions, pointing time increases with the log of the distance, (perhaps with some adjustments). This means that as the distance increases, the pointing time increase decreases. Thus if response time increases faster than the distance, we would be able to conclude that the time taken to look at the different stimuli was not only due to Fitts' Law-like factors, but to other factors.

For instance, the accuracy of comparisons may be affected by the extra time taken to make comparisons, because the memory used to make visual comparisons degrades over time. Phillips (1974) showed that visual comparisons were more accurate when the stimuli were seen within 100 ms of one another, and less accurate as the stimuli were separated by longer periods. Baddeley and Hitch (1974) proposed a multi-component system of short term memory called Working Memory, versions of which remain the predominant model today. Both of these papers proposed a system with very short term, very accurate iconic memory, and a longer-term, but limited capacity visual working memory that is prone to interference (Baddeley and Andrade, 2000). It will be important to determine how the extra time or other interfering factors affect the accuracy of comparisons.

Research Overview

A number of experiments were undertaken to replicate the findings about comparisons of line length, and to measure how those comparisons are affected by conditions that could arise as those lines are drawn as part of a visualization containing multiple small bar graphs.

Experiment 1 was a pilot study in which participants compared single lines. Some of the independent variables in this experiment were further studied in later experiments, and some were dropped, so that those experiments could focus on other variables. This experiment is described in Chapter 2.

Experiments 2 and 3 focused on the comparison of single lines. In Experiment 2, these lines were at various distances from one another, but had in all conditions the same

orientation and alignment. In Experiment 3, the lines were at the same distance in all conditions, but were oriented and aligned differently. These experiments are described in Chapters 3 and 4.

In Experiments 4 and 5, participants were no longer comparing single lines, but small bar graphs made of pairs of lines. The participants indicated which of the two graphs had the larger absolute difference between the two lines. In Experiment 4, the two graphs were separated by different distances in different conditions. In Experiment 5, the two graphs were separated by the same distance, but the graphs were aligned differently in different conditions. These experiments are presented in Chapters 5 and 6.

Chapter 7 is a general discussion of the findings of this research.

Research Questions

I did not have an *a priori* working hypothesis for every possible question that could be asked of the data I collected. I use the term *research question* to refer generally to issues I hope to resolve with my experimental analysis, and *hypotheses* to specific expectations about how these research questions will be resolved. Sometimes different authors' findings or theories suggest different hypotheses, and I have tried to indicate which hypotheses were based on my own suspicions and which were presented without my specific endorsement. Rather than present an exhaustive list of research questions and hypotheses here, I describe the general research questions here (denoted with a G) and longer lists of research questions and hypotheses about each experiment in its respective chapter, denoted with the number of the experiment.

General Research Questions

Research Question G-1. How is the comparability of lines affected by the distance between the lines?

Research Question G-2. How is the comparability of lines affected by the alignment of the lines?

Research Question G-3. How is the comparability of lines affected by the orientation of the lines?

Research Question G-4. How is the comparability of lines affected by the interaction of the alignment and orientation of the lines?

Research Question G-5. How is the comparability of small bar graphs affected by the factors described in questions 1, 2, 3, and 4?

Chapter 2: Experiment 1

This chapter describes a pilot study in which small bar graphs were placed on the screen separated by some distance, and participants were asked to choose the graph containing the longer bar. Different conditions from one block of trials to another were designed to investigate how the comparability of graphs is affected by various factors that are present in real visualizations that include small multiples. These were the distance between graphs, the alignment of graphs on the page, the orientation of the lines, and the presence of lines on the screen that were not those being compared: other bars in a graph, axes drawn around the graph, and task-irrelevant graphs.

The basic design of the experiment was to have a participant look at two graphs containing bars of different lengths, decide which of two bars was larger, move the mouse cursor to the location of the larger bar, and click on it. Performance was measured in two ways: accuracy, meaning the proportion of the trials for which the correct, larger bar was selected, and speed, measured as how quickly responses were made, expressed in responses per second. The term *comparability* is used below to refer to both higher accuracy and speed.

Research Questions

This experiment was designed to test a few basic assumptions about how well bars could be compared at a distance, and to provide answers to questions for which no *a priori* hypothesis was made. I call the former hypotheses and the latter research questions.

Small multiples require the user to compare graphical elements at a distance. This is more difficult than comparing elements that are near each other. This experiment was designed to explore how the comparability of graphical elements (small bars of varying lengths) changes as the elements are moved in space relative to one another. This includes both varying the distance between the elements and the angle at which they are arranged on the page, which is here called *alignment*. Two elements may be aligned horizontally, vertically, or on some diagonal. It also explores how changes to the elements, in this case, the orientation of the bars, interacts with these differences in alignment. A further component looks at how extra elements, such as additional data bars or visual guides (axes) affect the comparability of elements.

Hypotheses and Research Questions

Hypothesis 1-1a. As the distance between two bars increases, the comparability of those bars will decrease.

Hypothesis 1-1b. There should be a drop in comparability as the bars move far enough away from each other that they can not both be seen with central vision at the same time.

Research question 1-1c. After this drop, will comparability decline at a faster rate, or decline only steadily?

Hypothesis 1-2a. Vertically oriented bars aligned diagonally should be more comparable than those aligned vertically, but less comparable than those aligned horizontally. The graph reader can use the mental resources available for a two-

dimensional image, but can not simply compare the vertical position of the tops of the two bars.

Hypothesis 1-2b There will be an interaction between the alignment of two graphs and the orientation of the elements within those graphs. Comparability will be better when the alignment and orientation are *cross-dimensional*, that is, one is horizontal and one vertical. The cross-dimensional bars will form a trapezoid when the illusory contour (Ware, 2004) of lines connecting the end points of the bars are created by the graph reader, allowing a richer representation to be created than for the collinear bars.

Research question 1-2c. Will collinear bars be easier to compare when they are aligned and oriented horizontally or vertically?

Research question 1-2d. Will cross-dimensional bars be easier to compare when the alignment is horizontal and orientation is vertical, or when the alignment is vertical and the orientation is horizontal?

Hypothesis 1-3a. The presence of extra bars similar to those being compared, widely spaced on the screen, should decrease comparability. The extra bars should increase visual interference. (An alternative possibility is that they will increase comparability by giving the user something else to compare stimuli to.)

Hypothesis 1-3b. The presence of extra bars similar to those being compared, closely spaced to the bars being compared, should decrease comparability. The extra bars should increase visual interference. (An alternative possibility is that they will increase comparability by giving the user something else to compare stimuli to.)

Hypothesis 1-3c. The presence of axes should increase the accuracy of comparisons, and may speed them up. (An alternative possibility is that they will interfere with perception, decreasing comparability.)

Method

Participants

Students in University of Maryland psychology classes participated for credit in April and May 2009. Thirty-two participants took part in the experiment. Nine of these participants were excluded from these analyses because their response times were under an average of 150 ms for at least one condition, leaving an n of 23. Many of the excluded participants also had accuracy rates below chance for a number of conditions, but this was not used to select which participants to exclude. Of the 23 subjects whose data are analyzed here, 16 were male and 7 female, and the mean age was 20.2 years.

Apparatus and Materials

The experiment was programmed in Java and embedded in a web page. Participants used Mac or Windows desktop computers with the Safari or Firefox browsers, which did not change the experiment beyond minor points of font rendering. Participants used the mouse, not a trackpad or other input device. No attempts were made to standardize the viewing angle.

Stimuli

The graphical elements to be compared by the participant were black bars set against a white background. The bars were two pixels (.5 mm) in the shorter dimension (usually width) and between 5 and 21 pixels (2.5 to 10.5 mm) in the longer dimension (usually height). Figure 18 shows the bars to scale with one another, with the bars aligned horizontally and oriented vertically. In the conditions with horizontally oriented bars, the bars were two pixels in height and between 5 and 21 pixels in width. The size of the bar along its longer direction will be called *length* avoid confusing the vertical size of a bar with its vertical position of the screen. and to make comparisons across the two orientations of bars easier. Each trial had one of two *standard* bars, either 10 or 16 pixels in length, and one *comparison* bar, which was 5, 7, 9, 11, 13, 15, 17, 19, or 21 pixels in length. Each standard bar was matched with each comparison bar four times in each block of trials. On two of those trials, the positions of the standard and comparison bars were switched. For instance, on trials with the bars aligned horizontally, the standard bar was on the left twice and on the right twice. Figure 19 shows how graphs were aligned and oriented in the various conditions.



Figure 18. Standard and Comparison Bars for Experiment 1.

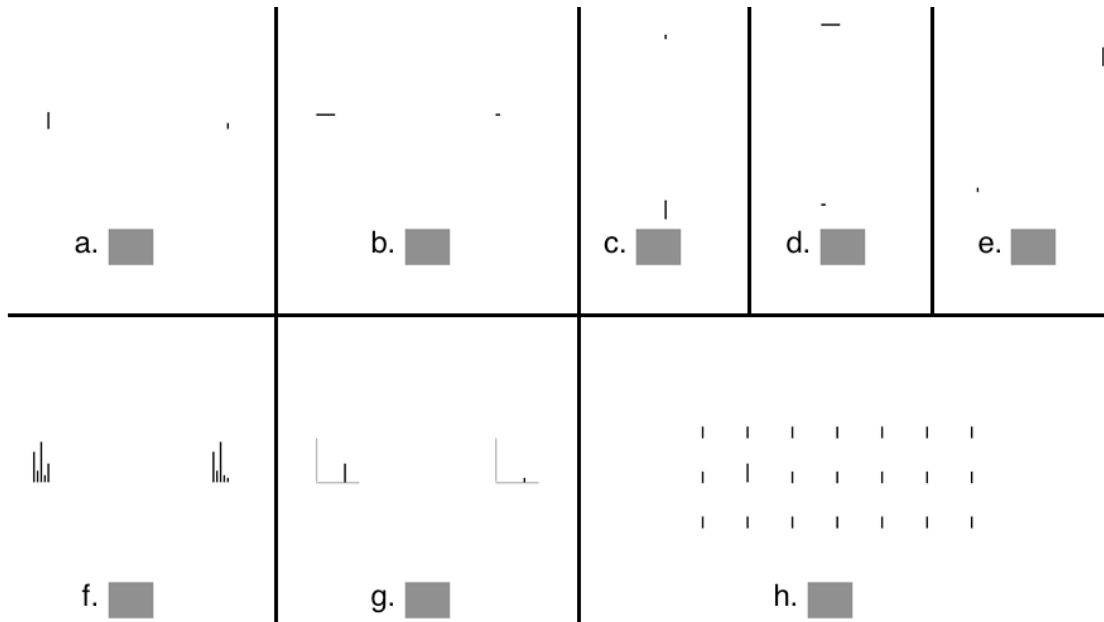


Figure 19. Graphs used in Experiment 1.

The gray box is the ready box, where the mouse pointer began each trial. All graphs in this figure are shown at 200 pixels distance. a. vertically oriented bars aligned horizontally (Block 3). b. horizontally oriented bars aligned horizontally (Block 8). c. vertically oriented bars, aligned vertically (Block 6). d. horizontally oriented bars, aligned vertically (Block 9). e. vertically oriented bars, aligned at a 45 degree diagonal (Block 5). f. graphs made up of closely-spaced vertically oriented bars, for which participants were asked to compare the rightmost bars of each graph only (Block 10). g. vertically oriented bars with thin X and Y axes (Block 11). h. vertically oriented bars with other vertically oriented bars around them (Block 12).

The two bars were separated by some distance, which varied by block. This distance was either horizontal, vertical, or a 45 or -45 degree diagonal, also varying by block. Note that this distance was measured from the bottom left corner of a 50 pixel by 50 pixel rectangle (not visible to the participants) that touched the vertically oriented bars at their bottom point, and the horizontally oriented bars at their leftmost point. The bars are meant to be simple graphs showing one datum represented by the length of the bar.

In three conditions, there were other graphical elements on the page (see Figure 18).

These were other bars, widely spaced and representing other graphs, or other bars, closely spaced to the bars of interest, representing other data in the same bar graph, or thin (1 pixel) gray vertical and horizontal lines representing the axes of the bar graph. These lines were along the bottom and left edges of the 50-pixel by 50-pixel rectangle that contained the bars.

Conditions

The 12 blocks were randomly ordered, as were the trials within each block. Each block contained four trials with each combination of one of the two standard bars and one of the nine comparison bars, for a total of 72 trials. For two of the four trials, the standard bar was in one of the positions, and in two, the other position. The conditions are shown in Table 1.











#	Picture	Distance	Alignment	Orientation	Extras	Analysis
1		50	0° horizontal	vertical	none	distance
2		100	0° horizontal	vertical	none	distance
3		200	0° horizontal	vertical	none	distance, alignment, orientation, extras
4		400	0° horizontal	vertical	none	distance
5		200	45° diagonal	vertical	none	alignment
6		200	90° vertical	vertical	none	alignment, orientation
7		200	-45° diagonal	vertical	none	alignment
8		200	0° horizontal	horizontal	none	orientation
9		200	90° vertical	horizontal	none	orientation
10		200	0° horizontal	vertical	closely-spaced bars	extras
11		200	0° horizontal	vertical	x and y axes	extras
12		200	0° horizontal	vertical	widely-spaced bars	none (would have been extras)

Table 1. Conditions of Experiment 1.

Pictures are simplified and not to scale.

Four blocks had vertically oriented bars aligned horizontally (0 degrees), at distances of 50 (Block 1), 100 (Block 2), 200 (Block 3), and 400 (Block 4) pixels. These blocks were analyzed to test the effects of distance on comparability.

Four blocks had vertically oriented bars at a distance of 200 pixels, with alignments of 0 degrees (Block 3), 45 degrees (Block 5), 90 degrees (Block 6), and -45 degrees (Block 7). These blocks were analyzed to test the effects of alignment on comparability.

Two blocks had horizontally oriented bars at a distance of 200 pixels. One was aligned horizontally (Block 8) and one vertically (Block 9). These were analyzed along with Blocks 3 and 6 to test the effects of orientation and the interaction of orientation and alignment on comparability.

Three blocks had visual elements other than the critical standard and comparison bars. All had the critical bars at a distance of 200 pixels, aligned horizontally. Block 10 had four vertical bars of various widths close to, and to the left of, the critical bars. The bars were spaced 2 pixels apart. Block 11 had 1-pixel gray x and y axes placed to the left and bottom of each graph. The x axis touches the bottom of the critical bars. Block 12 had bars spaced out at 50-pixel intervals between and around the critical bars. Due to a programming error, the results of Block 12 could not be analyzed. Blocks 10 and 11 were analyzed along with Block 3 to test the effects of extra elements on comparability.

Procedure

The instructions were followed by practice items, then twelve blocks of regular items. The two graphs appeared in the same location for the duration of each block so the participants knew where to look for them. The other conditions of the block described above were also consistent within a block; the only things that changed from one trial to the next were the heights of the standard and comparison bars. Each block began with

four practice trials, with bars of 5 and 21 or 7 and 19 pixels. If the participant answered any of these practice trials incorrectly, they were instructed to try again, beginning with the first practice trial. Thus each participant began each block by making four correct comparisons. This was done to familiarize the participants with the conditions of each block and to make sure they understood what they were supposed to do. No participants expressed confusion or asked any questions about the procedure to the experimenters.

Each trial began by the participant clicking a ready box on the screen located in the lower middle part of the screen, in a spot below where the graphs are displayed (See Figure 20a). Two fixation crosses were displayed in the locations of the two graphs, and the ready box turned gray (Figure 20b). After half a second delay, the crosses were replaced by the two graphs (Figure 20c). The participant chose one of the two bars as being longer, and moved the mouse pointer towards the graph they chose. Once the mouse pointer left the ready box, the ready box disappeared (Figure 20d), and the two graphs were replaced by masks made of a gray, white, and black pattern. The moment the mouse pointer left the ready box was the end of the trial for purposes of determining how long the trial took, and thus the speed of the trial (this was not explained to the participants). Note that the participant had not clicked the mask yet. The participant then moved the mouse pointer to, and clicked on, the mask that was in the same location as the graph of their choice (Figure 20e). (In blocks 10, 11, and 12, the extra elements also disappeared when the mouse pointer left the ready box. Masks did not replace the widely-spaced extra bars in block 12.) Once the participant had clicked on one of the masks, feedback about accuracy was given. The feedback is not shown in Figure 16, but consisted of the words “Correct” or “Incorrect” being displayed near the top of the

experiment screen. At this point the participant could begin the next trial by clicking the ready box, which was now visible again (Figure 20f).

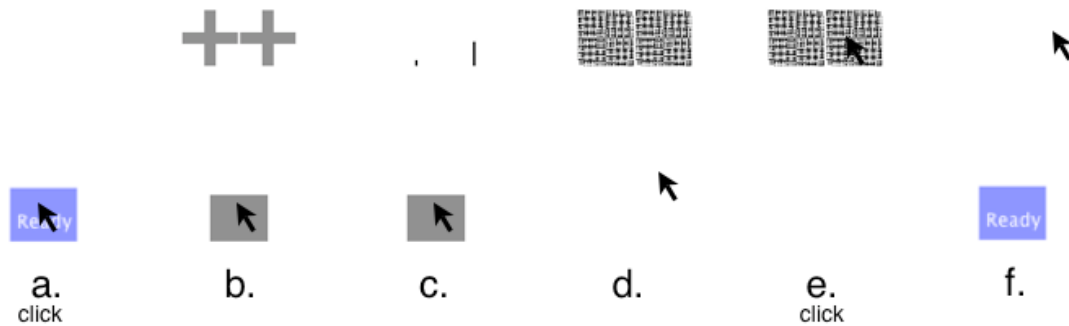


Figure 20. Trial Procedure for Experiment 1.

Results

The data were grouped into four (overlapping) analyses to answer different hypotheses and research questions: distance (1-1a, 1-1b, and 1-1c), alignment alone (1-2a), alignment and orientation (1-2b, 1-2c, and 1-2d), and extra elements (1-3a, 1-3b, and 1-3c). For each analysis, two measures of comparability were evaluated: accuracy and response time. (The results were also analyzed using speed, or 1/time. This measure had the advantage that more comparable conditions were more positive, just as with accuracy. However, it became unwieldy to convert speed results back to time when discussing the meaning of the results, so the speed measure was dropped in favor of the time measure.)

Several methods of data analysis were performed, but the most useful ones involved eliminating some conditions, and creating some new variables. The first of these new variables is the difference between the lengths of the standard and comparison lines, abbreviated *cDiff*. I used *cDiff* as a factor in the ANOVAs I performed when

initially analyzing Experiment 1. To maintain consistency with later analyses, however, the analysis here does not include *cDiff*, but instead two new variables: *absDif*, the absolute difference between the lengths of the lines, and *longerSC*, which is the sign of *cDiff*, in other words -1 if the standard bar is longer, and 1 if the comparison bar is longer.

Trials for which *absDif* was greater than 5 were excluded from this analysis. This represents 24 of the 72 trials for each condition, which had *absDif* values of 7, 9, and 11. The outcomes of these trials were very similar to the trials in their conditions for which *absDif* was 5. These conditions were generally easy, and the larger differences did not affect the response measures much.

The analyses were repeated measures ANOVAs. The factors were: 1. the length of the standard bar, abbreviated *sLength*, 2. *absDif*, 3. *longerSC*, and 4. and 5. factors of interest for each hypothesis. These factors were distance, alignment, orientation, and presence of extra elements.

Data were averaged for each combination of standard bar length and comparison bar length, for a total of 18 conditions per block (of which the present analysis uses 12 for reasons described above). The accuracy measure is the percent correct, averaged over the repetitions. Response time refers to the time between when the bars became visible to the participants, and when the mouse cursor left the gray ready box, and does not depend upon the time it took to click on one of the two gray response boxes (see Figure 20). The response time measure is the mean of the response times for all repetitions. Note that while a median time was used for later experiments, the mean is presented here, so that

the results presented represent those that were used to design the later experiments. All times were measured at the millisecond level, and are expressed in milliseconds.

Distance

These analyses compared two vertical bars at 50, 100, 200 and 400 pixels of horizontal distance, and tested hypotheses 1-1a and 1-1b and research question 1-1c.

Accuracy

Table 2 shows the results of the repeated-measures ANOVA for accuracy. The abbreviated names of the variables are used in the tables and graphs, including *dist* for distance, to save space. In this, and subsequent ANOVA tables, DF_n and DF_d refer to the numbers of degrees of freedom in the numerator and denominator of the F test. The letters [GG] indicate that the p value has been adjusted using the Greenhouse-Geisser correction. An alpha level of .05 was used. Partial omega squared is shown as a measure of effect size for significant effects.

No significant differences in comparability, as measured by accuracy, were found across the four distances. There were also no significant interactions involving distance. The only significant effects were sLength, absDif, and their interaction. These are plotted in Figure 21. The absolute difference in length proved to be the most significant predictor of comparability, as measured by accuracy, with 1-pixel differences much more difficult to distinguish than 3 or 5-pixel differences. This pattern holds for all experiments. The effect of the length of the standard was more subtle, with comparability better for shorter standard bars under some conditions. This effect, and its implications for Weber's Law, are discussed in later chapters.

Effect	DFn	DFd	F	<i>p</i>	<i>p</i> <.05	Partial ω^2
dist	3	66	0.99	.388 [GG]		
sLength	1	22	25.53	<.001	*	0.0217
absDif	2	44	187.64	<.001 [GG]	*	0.2527
longerSC	1	22	0.25	.623		
dist x sLength	3	66	1.12	.345 [GG]		
dist x absDif	6	132	0.72	.520 [GG]		
sLength x absDif	2	44	15.91	<.001 [GG]	*	0.0263
dist x longerSC	3	66	1.89	.153 [GG]		
sLength x longerSC	1	22	3.30	.083		
absDif x longerSC	2	44	0.23	.700 [GG]		
dist x sLength x absDif	6	132	1.54	.204 [GG]		
dist x sLength x longerSC	3	66	0.06	.977 [GG]		
dist x absDif x longerSC	6	132	0.91	.445 [GG]		
sLength x absDif x longerSC	2	44	1.30	.279 [GG]		
dist x sLength x absDif x longerSC	6	132	1.73	.150 [GG]		

Table 2. Repeated-Measures ANOVA for accuracy (Experiment 1 distance conditions).

In this and subsequent ANOVA tables, “[GG]” next to a *p* value indicates that the *p* value shown represents the Greenhouse-Geisser correction. Partial omega squared is shown as a measure of effect size for significant effects.

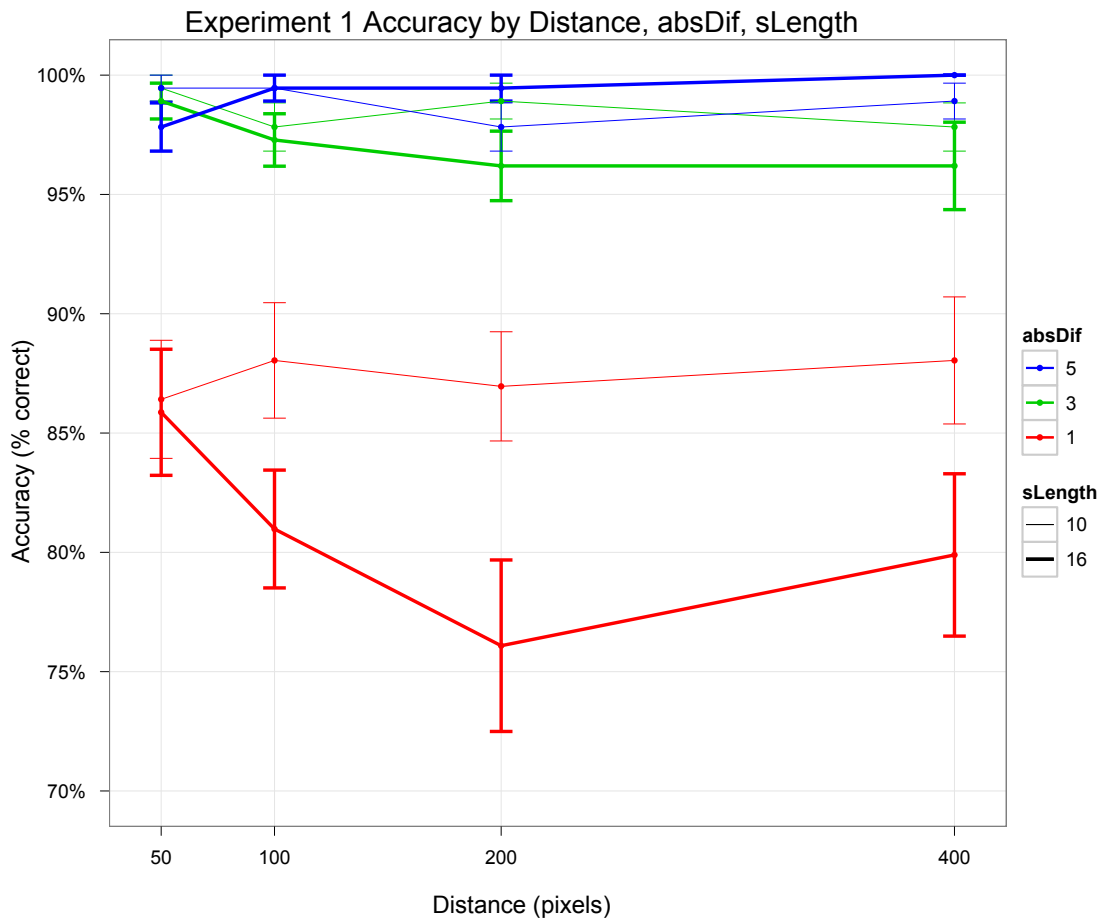


Figure 21. Accuracy by Distance, absDif, and sLength (Experiment 1).

Error bars represent the standard error of the mean.

Response Time

The distance between the bars had a significant effect on the response time (Table 3), with longer distances leading to longer response times. Participants also took significantly more time to respond to pairs of lines with more similar lengths (absDif), and these were the only significant factors.

Figure 22 shows the differences found in the distance and absDif conditions. Note that the initial response time increases in a roughly linear fashion as a function of the distance between the two lines. There is a bend in the line when absDif is 1 between

the distances of 50 and 100 and 100 and 200. This bend indicates a particular advantage for comparing very similar lines when those lines are very close. Although the interaction of absDif and distance is not significant here, this was an interesting enough potential finding that it influenced the design of subsequent experiments.

Effect	DFn	DFd	F	<i>p</i>	<i>p</i> <.05	Partial ω^2
dist	3	66	20.52	<.001 [GG]	*	0.0504
sLength	1	22	3.54	.073		
absDif	2	44	28.37	<.001 [GG]	*	0.0472
longerSC	1	22	0.47	.500		
dist x sLength	3	66	1.22	.305 [GG]		
dist x absDif	6	132	2.62	.057 [GG]		
sLength x absDif	2	44	0.98	.347 [GG]		
dist x longerSC	3	66	0.78	.466 [GG]		
sLength x longerSC	1	22	0.58	.453		
absDif x longerSC	2	44	0.35	.637 [GG]		
dist x sLength x absDif	6	132	1.65	.195 [GG]		
dist x sLength x longerSC	3	66	1.10	.345 [GG]		
dist x absDif x longerSC	6	132	1.18	.321 [GG]		
sLength x absDif x longerSC	2	44	0.71	.441 [GG]		
dist x sLength x absDif x longerSC	6	132	0.54	.621 [GG]		

Table 3. Repeated Measures ANOVA for time (Experiment 1 distance conditions).

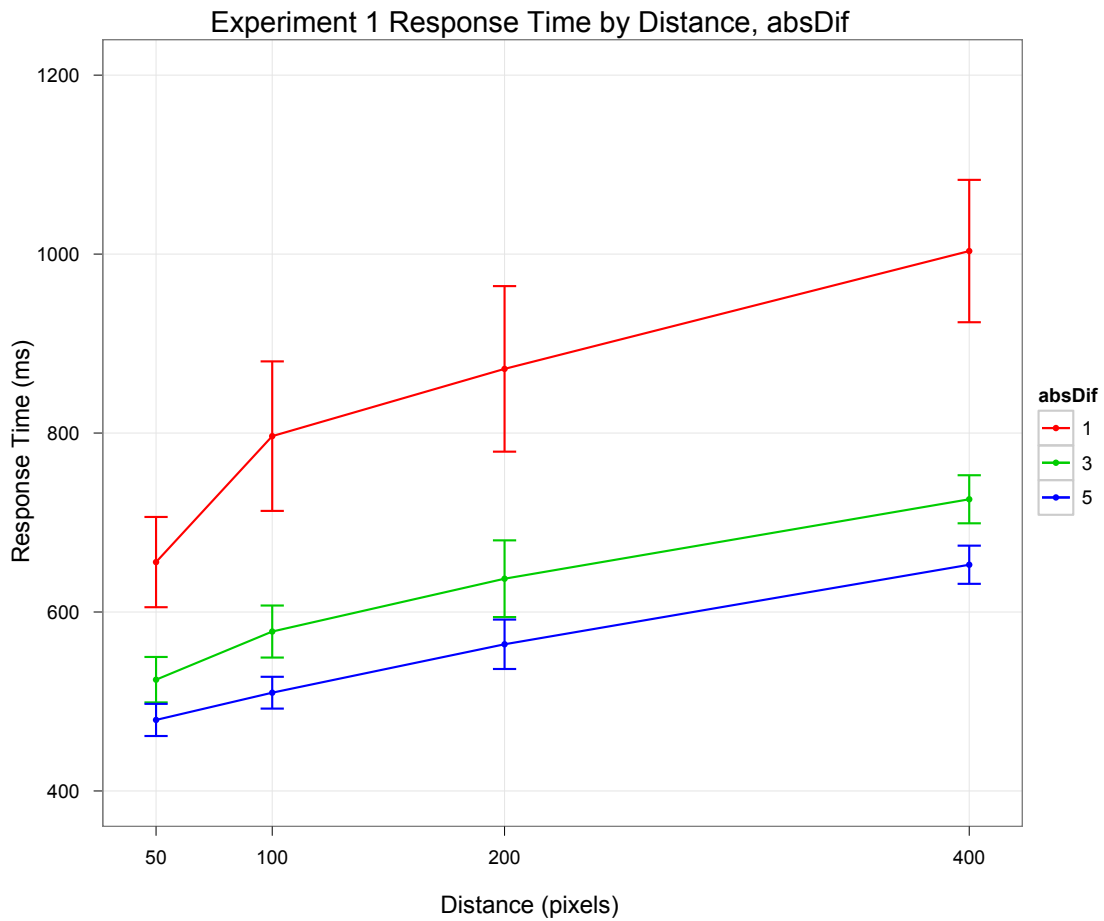


Figure 22. Response Time by Distance and absDif (Experiment 1).

Alignment

These analyses compared two vertical bars at 200 pixels distance, arranged in four alignments, and tested hypothesis 1-2a.

Accuracy

Table 4 shows the results of the repeated-measures ANOVA. No significant differences were found in accuracy across the four alignments. The only significant factors were absDif, sLength, and their interaction. Figure 21 shows that responses were

very accurate (above 95%) for conditions in which absDif was 3 or 5, and above 80% when it was 1 pixel.

Effect	DFn	DFd	F	<i>p</i>	<i>p</i> <.05	Partial ω^2
align	3	66	0.21	.858 [GG]		
sLength	1	22	34.74	<.001	*	0.0297
absDif	2	44	99.50	<.001 [GG]	*	0.1514
longerSC	1	22	0.02	.892		
align x sLength	3	66	0.74	.488 [GG]		
align x absDif	6	132	0.20	.923 [GG]		
sLength x absDif	2	44	18.89	<.001 [GG]	*	0.0314
align x longerSC	3	66	0.66	.557 [GG]		
sLength x longerSC	1	22	1.64	.214		
absDif x longerSC	2	44	0.65	.445 [GG]		
align x sLength x absDif	6	132	1.07	.374 [GG]		
align x sLength x longerSC	3	66	0.55	.627 [GG]		
align x absDif x longerSC	6	132	0.26	.878 [GG]		
sLength x absDif x longerSC	2	44	0.24	.713 [GG]		
align x sLength x absDif x longerSC	6	132	1.49	.210 [GG]		

Table 4. Repeated-measures ANOVA table for Accuracy (Experiment 1 alignment conditions).

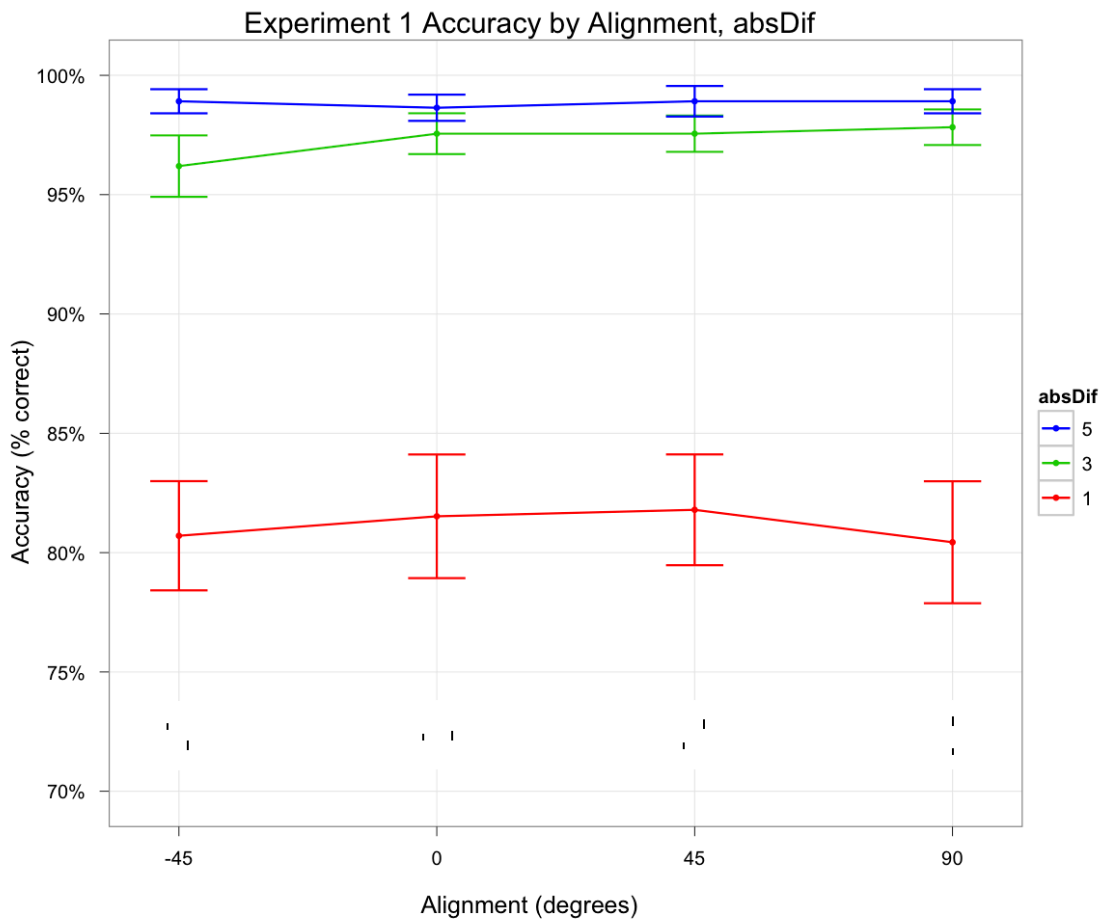


Figure 23. Accuracy by Alignment and absDif (Experiment 1).

Small pictures of each condition have been added for reference.

Response Time

Table 5 shows the results of the repeated-measures ANOVA for response time. Each factor was significant, but none of the interactions were. Again, absDif has the largest effect, but alignment has the second largest effect, with horizontal alignments eliciting the fastest responses, and vertical alignments the slowest. Diagonal alignments were in-between (Figure 24).

Effect	DFn	DFd	F	<i>p</i>	<i>p</i> <.05	Partial ω^2
align	3	66	7.32	.001 [GG]	*	0.0169
sLength	1	22	9.11	.006	*	0.0073
absDif	2	44	33.57	<.001 [GG]	*	0.0557
longerSC	1	22	5.38	.030	*	0.0040
align x sLength	3	66	0.74	.512 [GG]		
align x absDif	6	132	0.90	.432 [GG]		
sLength x absDif	2	44	0.60	.490 [GG]		
align x longerSC	3	66	0.56	.592 [GG]		
sLength x longerSC	1	22	1.92	.180		
absDif x longerSC	2	44	3.54	.055 [GG]		
align x sLength x absDif	6	132	0.69	.555 [GG]		
align x sLength x longerSC	3	66	0.83	.461 [GG]		
align x absDif x longerSC	6	132	0.86	.466 [GG]		
sLength x absDif x longerSC	2	44	3.49	.052 [GG]		
align x sLength x absDif x longerSC	6	132	1.10	.356 [GG]		

Table 5. Repeated-measures ANOVA table for response time (Experiment 1 alignment conditions).

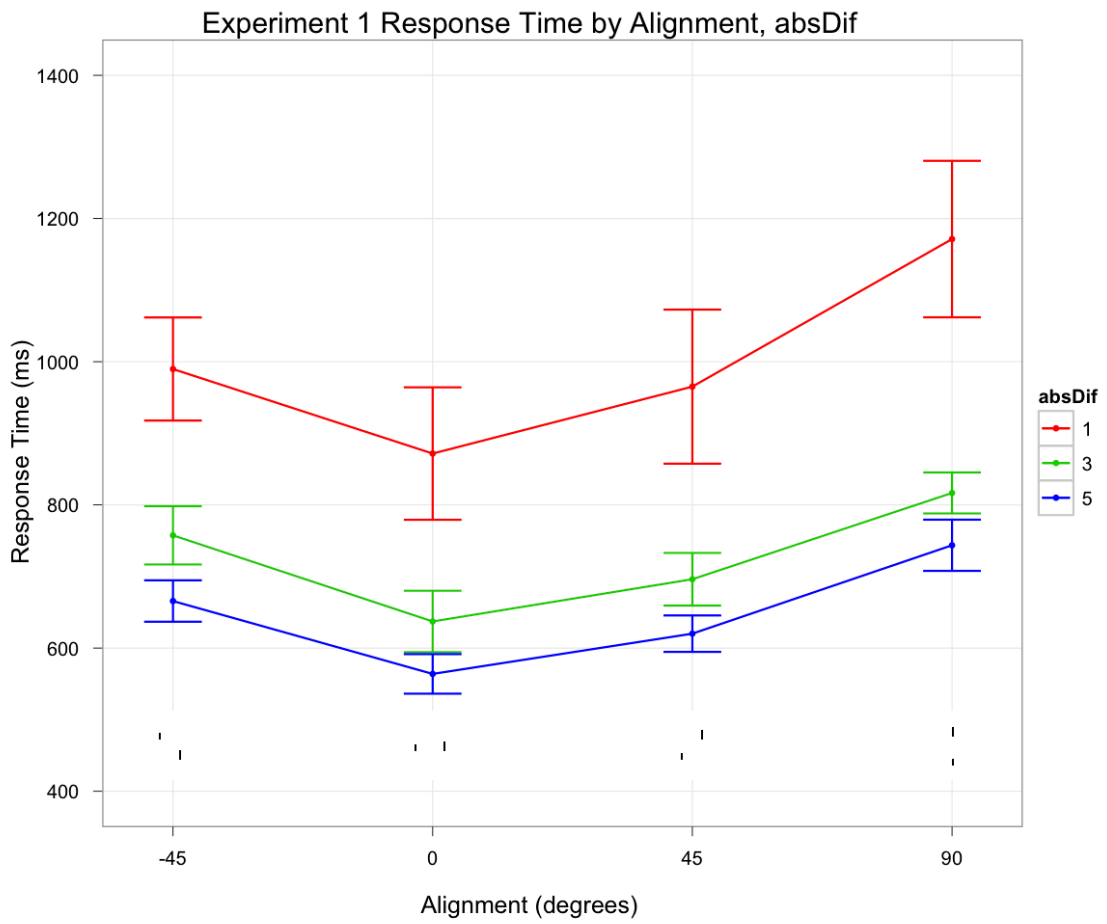


Figure 24. Response Time by Alignment, absDif (Experiment 1).

Small pictures of each condition have been added for reference.

Alignment and Orientation

These analyses compared bars of two orientations (vertical and horizontal bars) crossed with two alignments (vertical and horizontal positioning), all at 200 pixels distance. The diagonal alignment conditions were not included in this analysis.

Accuracy

Table 6 shows the repeated-measures ANOVA for accuracy for the alignment and orientation conditions. Alignment was a significant factor, with horizontally aligned graphs resulting in more correct responses. This effect was very small, however (Figure 25). The orientation of the bars was not a significant factor, nor was the interaction of these factors. Other significant effects were sLength and absDif, along with a few interactions, but none that included both alignment and orientation.

Effect	DFn	DFd	F	p	p<.05	Partial ω^2
align	1	22	5.25	.032	*	0.0038
orient	1	22	1.97	.175		
sLength	1	22	47.24	<.001	*	0.0402
absDif	2	44	133.49	<.001 [GG]	*	0.1936
longerSC	1	22	2.23	.150		
align x orient	1	22	3.21	.087		
align x sLength	1	22	0.76	.392		
orient x sLength	1	22	6.25	.020	*	0.0047
align x absDif	2	44	1.32	.273 [GG]		
orient x absDif	2	44	0.2	.718 [GG]		
sLength x absDif	2	44	12.1	.001 [GG]	*	0.0197
align x longerSC	1	22	7.15	.014	*	0.0055
orient x longerSC	1	22	3.29	.083		
sLength x longerSC	1	22	0.41	.527		
absDif x longerSC	2	44	0.15	.774 [GG]		
align x orient x sLength	1	22	0.91	.349		
align x orient x absDif	2	44	0.72	.457 [GG]		
align x sLength x absDif	2	44	1.52	.234 [GG]		
orient x sLength x absDif	2	44	0.68	.460 [GG]		
align x orient x longerSC	1	22	0.19	.663		
align x sLength x longerSC	1	22	1.99	.172		
orient x sLength x longerSC	1	22	0.03	.862		
align x absDif x longerSC	2	44	0.37	.614 [GG]		
orient x absDif x longerSC	2	44	0.86	.389 [GG]		
sLength x absDif x longerSC	2	44	0.06	.911 [GG]		
align x sLength x absDif x longerSC	2	44	4.13	.041 [GG]	*	0.0056

Table 6. Repeated-measures ANOVA for accuracy (Experiment 1 alignment/orientation conditions).

All non-significant 4-way and 5-way interactions have been omitted from this table to save space.

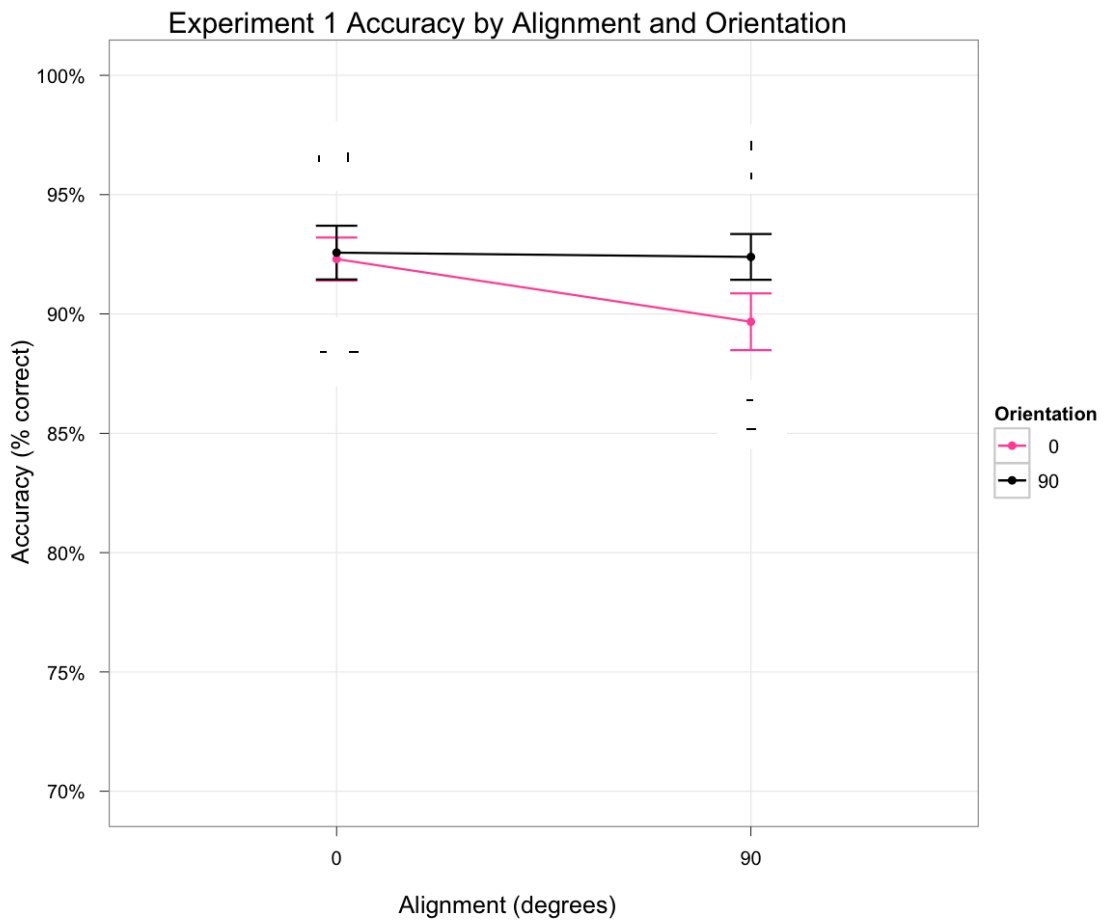


Figure 25. Accuracy by Alignment and Orientation (Experiment 1).

Small illustrations (not to scale) of each condition have been added for reference. The effect of alignment, represented here as the left and right sides of the graph, was significant, but the effects of orientation (represented by the black and pink lines) and the interaction were not.

Response Time

Table 7 shows the repeated-measures ANOVA for the alignment and orientation conditions. Again alignment was a significant factor, with people responding faster to horizontally aligned graphs. Orientation was not a significant factor, but the interaction

of alignment and orientation was significant, with cross-aligned graphs eliciting faster responses than co-aligned graphs.

Effect	DFn	DFd	F	<i>p</i>	<i>p</i> <.05	Partial ω^2
align	1	22	30.57	<.001	*	0.0261
orient	1	22	0.99	.330		
sLength	1	22	3.97	.059		
absDif	2	44	37.92	<.001 [GG]	*	0.0627
longerSC	1	22	1.71	.204		
align x orient	1	22	6.61	.017	*	0.0051
align x sLength	1	22	0.06	.803		
orient x sLength	1	22	5.69	.026	*	0.0042
align x absDif	2	44	4.03	.047 [GG]	*	0.0055
orient x absDif	2	44	2.50	.121 [GG]		
sLength x absDif	2	44	1.25	.286 [GG]		
align x longerSC	1	22	0.13	.723		
orient x longerSC	1	22	1.94	.178		
sLength x longerSC	1	22	0.20	.661		
absDif x longerSC	2	44	0.38	.626 [GG]		
align x orient x sLength	1	22	2.83	.107		
align x orient x absDif	2	44	0.34	.588 [GG]		
align x sLength x absDif	2	44	0.35	.604 [GG]		
orient x sLength x absDif	2	44	2.12	.149 [GG]		
align x orient x longerSC	1	22	0.00	.951		
align x sLength x longerSC	1	22	2.85	.106		
orient x sLength x longerSC	1	22	0.45	.510		
align x absDif x longerSC	2	44	0.11	.863 [GG]		
orient x absDif x longerSC	2	44	4.16	.037 [GG]	*	0.0057
sLength x absDif x longerSC	2	44	1.23	.297 [GG]		

Table 7. Repeated-measures ANOVA for response time (Experiment 1 alignment/orientation conditions).

All non-significant 4-way and 5-way interactions have been omitted from this table to save space.

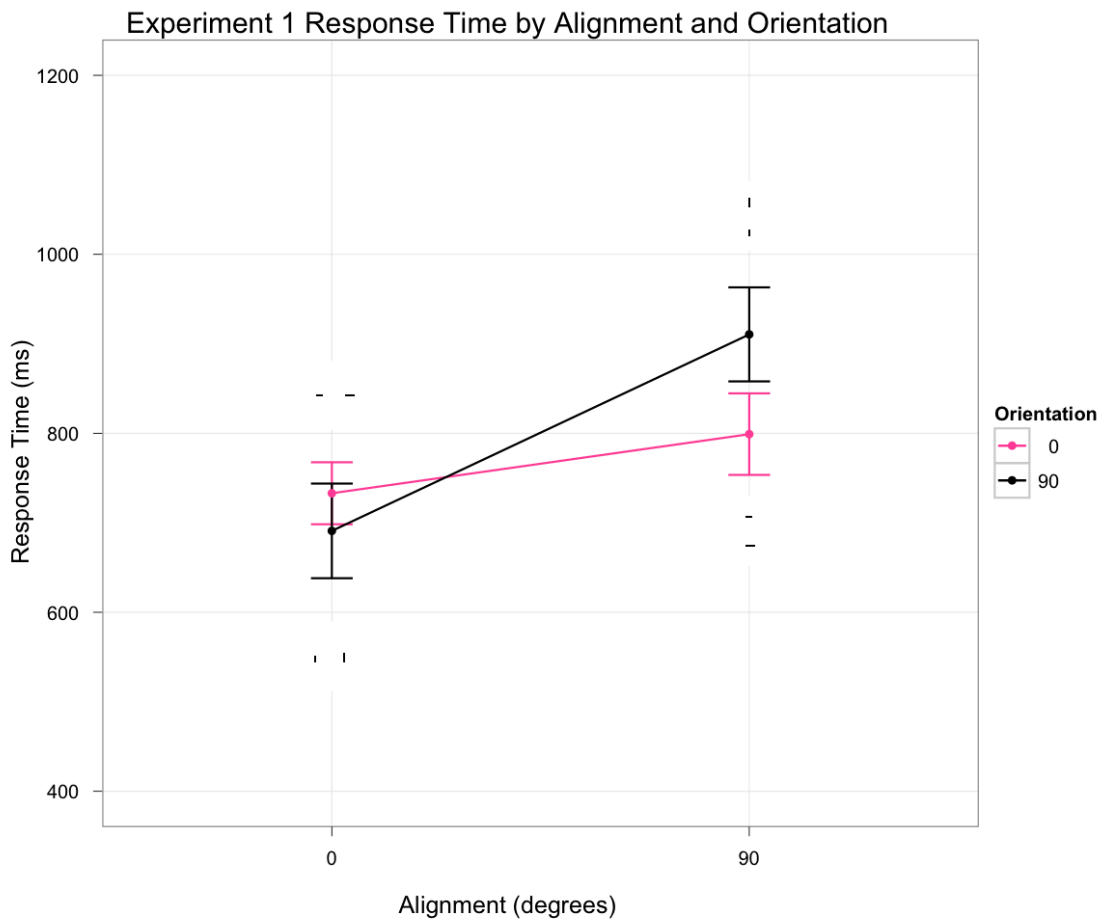


Figure 26. Response time by Alignment and Orientation (Experiment 1).

Small illustrations (not to scale) of each condition have been added for reference. The alignment and interaction effects are significant.

Extra Elements

These analyses compared two vertical bars at 200 pixels distance, aligned horizontally, with or without other elements that might distract or assist the user in making comparisons. These conditions were no extra elements (block 3), close-set

vertical bars (block 10), thin gray x and y axes (block 11), and extra bars representing other graphs (block 12).

Accuracy

Due to a programming error, the accuracy of the responses in block 12, the widely-spaced extraneous bars, could not be analyzed. The presence of extra elements, including other bars in the bar graphs, or thin, gray X and Y axes, did not affect the accuracy of responses (Table 8, Figure 27).

Effect	DFn	DFd	F	p	p<.05	Partial ω^2
extra	2	44	0.38	.645 [GG]		
sLength	1	22	21.78	<.001	*	0.0245
absDif	2	44	152.01	<.001 [GG]	*	0.2673
longerSC	1	22	2.92	.102		
extra x sLength	2	44	0.11	.874 [GG]		
extra x absDif	4	88	0.47	.630 [GG]		
sLength x absDif	2	44	8.07	.004 [GG]	*	0.0168
extra x longerSC	2	44	3.98	.034 [GG]	*	0.0071
sLength x longerSC	1	22	1.09	.307		
absDif x longerSC	2	44	2.45	.110 [GG]		
extra x sLength x absDif	4	88	0.98	.393 [GG]		
extra x sLength x longerSC	2	44	1.62	.212 [GG]		
extra x absDif x longerSC	4	88	1.69	.189 [GG]		
sLength x absDif x longerSC	2	44	0.91	.378 [GG]		
extra x sLength x absDif x longerSC	4	88	1.32	.276 [GG]		

Table 8. Repeated-measures ANOVA for accuracy (Experiment 1 extra elements conditions).

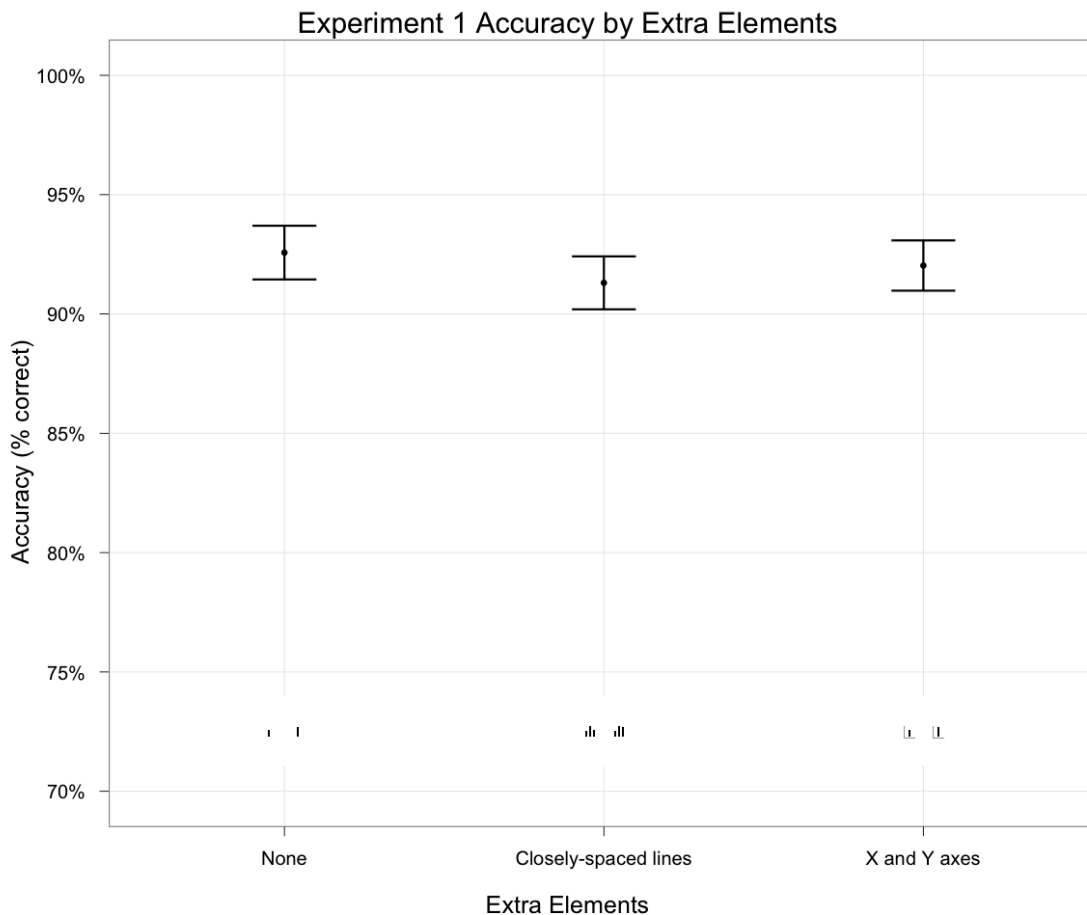


Figure 27. Accuracy by Presence of Extra Elements (Experiment 1).

Small illustrations (not to scale) of each condition have been added for reference. The differences are not significant.

Response Time

The presence of extra elements increased the response time (Table 9). Figure 28 shows that the presence of additional bars in a bar graph had a larger effect than the presence of thin X and Y axes. The closely spaced bars slowed the responses more than the axes did. Block 12, with widely-spaced bars, was not included in either ANOVA, but responses were slow, about the same as the block with closely-spaced bars.

Effect	DFn	DFd	F	<i>p</i>	<i>p</i> <.05	Partial ω^2
extra	3	66	13.56	<.001 [GG]	*	0.0330
sLength	1	22	0.44	.515		
absDif	2	44	25.46	<.001 [GG]	*	0.0424
longerSC	1	22	3.61	.070		
extra x sLength	3	66	4.91	.015 [GG]	*	0.0105
extra x absDif	6	132	1.16	.333 [GG]		
sLength x absDif	2	44	1.15	.314 [GG]		
extra x longerSC	3	66	1.05	.373 [GG]		
sLength x longerSC	1	22	3.42	.078		
absDif x longerSC	2	44	1.28	.284 [GG]		
extra x sLength x absDif	6	132	2.79	.050 [GG]	*	0.0096
extra x sLength x longerSC	3	66	3.48	.028 [GG]	*	0.0067
extra x absDif x longerSC	6	132	2.35	.074 [GG]		
sLength x absDif x longerSC	2	44	2.13	.139 [GG]		
extra x sLength x absDif x longerSC	6	132	1.31	.275 [GG]		

Table 9. Repeated-measures ANOVA for response time (Experiment 1 extra elements conditions).

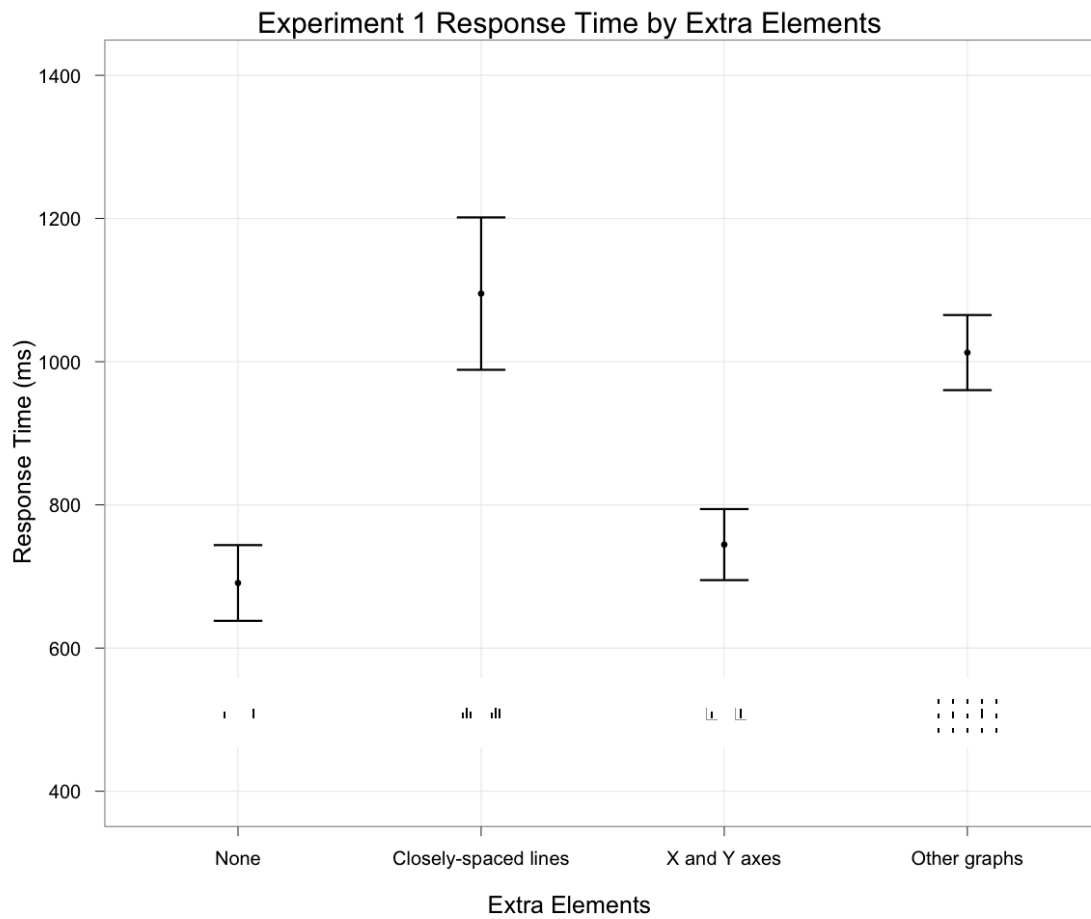


Figure 28. Response Time by Presence of Extra Elements (Experiment 1).

Small illustrations (not to scale) of each condition have been added for reference.

Discussion

Distance

The distance between two bars did not affect how accurately the bars were compared in this experiment. The response time did increase as the distance increased, so Hypothesis 1-1a is supported.

The initial response time increases more quickly, as a function of distance between the bars, between the shortest distances (50 and 100 pixels) than between the other distances (100 and 200 pixels, 200 and 400 pixels). The increase in response time increases with the distance from 100 to 400 pixels. It is not possible to draw a firm conclusion about Hypothesis 1-1b or Research Question 1-1c, but there is no evidence that accuracy decreased with distance or that response time increases any faster than the distance between the bars.

There is a possibility that some of the difference in initial response time was due to differences in how responses were made to different item types. Specifically, in the 50-pixel condition, the participants knew that they would move the mouse towards the middle of the screen. The farther apart the bars, the farther apart the participants would eventually need to move the mouse. This anticipation may have also contributed to some data being lost when people made unusually fast responses.

Alignment and Orientation

It matters how bars are arranged if their lengths are to be compared. Hypothesis 1-2a was supported by the data, which showed that a horizontal alignment of bars was superior to the two diagonal alignments, which in turn were superior to the vertical alignment.

The direction in which bars are drawn matters as well. There is support for Hypothesis 1-2b, which said that cross-dimensional bars would be easier to evaluate than collinear bars. The evidence for this in terms of speed. The unexpected finding was that people responded to collinear horizontal bars a bit *faster* than the vertically aligned horizontal bars. The collinear horizontal bars were faster than the collinear vertical bars, answering Research Question 1-2c. Of the two cross-dimensional conditions in Research Question 1-2d, the horizontally-aligned vertical bars drew faster (and more accurate) responses than the vertically-aligned horizontal bars. One thing to consider, though, is that most conditions (9 of the 12) featured horizontally aligned vertical bars, so familiarity may have been a factor as well. The worst-performing of these conditions, the vertically-aligned horizontal bars, were the only condition in the entire experiment that differed in more than one way from Condition 3, the horizontally-aligned vertical bars.

Other than familiarity, what might explain the better performance for horizontal alignments? One possibility is that, because people's eyes are aligned horizontally, the visual field that is wider than it is tall. Fine vision is much better near the fovea than the periphery of the visual field, and one eye or the other might well find the target lines faster moving sideways than up and down. A second possibility is that people are better at scanning side-to-side because it is useful in our everyday environment. Reading, for

one, involves many more side-to-side than vertical eye movements. If we are looking at similar objects, these objects are more likely to be at different horizontal positions in our visual field than vertical positions. A related explanation is the bias we have to interpret things that are higher in our visual field as farther away, and thus larger for the amount of visual angle we see.

Extra Elements

It is not surprising that the presence of extra bars near the standard and comparison bars slowed down responses, as predicted by Hypothesis 1-3b. They required the participant to determine which bars to compare, and although these were always in the same place, the extra bars provided some visual interference. It is important to note that this delay caused these responses to be slower than in any of the other conditions. Thus having graphs to compare, and not just single items arranged on a page, will cause people to take more time to make a close comparison. The responses were not significantly less accurate, and were still made in an average of less than a second.

The thin axes did not help people make more accurate responses, and did slow people down significantly, refuting Hypothesis 1-3c. It seems that even thin, light-colored lines interfered more than they helped. What we do not know is whether more detailed guidelines, ones with horizontal bars at regular intervals, would help people make accurate responses, or what role the time pressure played on participants' willingness to make comparisons based on visual information other than the bars themselves.

Hypothesis 1-3a could not be evaluated due to a programming error that affected block 12, but the slow responses suggest that many graphs can also slow down the process.

General Discussion

The research participants were able to determine which one of two bars was longer with a high degree of accuracy, over 90%, in all but one of the contexts examined here. Speed, therefore, was the primary measure by which demonstrated the differing comparability of the different contexts. The experimental design could have been changed to make accuracy the primary variable, by decreasing the length differences between the bars, or fixing the amount of time that the participants could view the lines. I rejected this approach for a few reasons. First, I wanted the outcome of this project to have practical application for the graph maker, so a more naturalistic experimental design, in which people look at graphs until they understand how they differ, is preferable to a more artificial one, in which people view graphs for a split second and then make this decision. The general research question for any context is, "how quickly can people make a decision when graphs are presented in this context?"

Second, the question of the utility of small multiples is not just about comparing graphs with slight differences, but also moderate or even large differences. For example, Bertin's (1967/1983) criticism of a group of pie charts was not that the pie charts were too similar, but that the reader could not integrate the information from one segment to another. Having already restricted my experiments to bar graphs, I do not wish to restrict them only to very similar bar graphs.

The design of the subsequent experiments was similar to that of Experiment 1, but the procedure was changed. There was a potential source of bias or noise when comparing the distance and alignment conditions, because the participants moved the mouse cursor in different directions in different conditions. The new procedure, described in the next chapter, was also faster.

The focus of the later experiments was put on the factors of distance, alignment, and orientation. The conditions based on extra elements were cut, for a few reasons. First, how to arrange the small graphs by distance, alignment, and orientation are fundamental decisions that a graph maker must make. The presence of axes or other guidelines are a less fundamental decision; although they may prove helpful or distracting based on the particulars of the visualization or user preference, they can be added or removed, or even switched on and off in an interactive visualization, without making any other changes. Second, it is already clear from that small multiple bar graphs, without any enhancement, do not provide a pop-out answer to the question of how a variable was distributed at any particular time (see Figure 14 and the discussion surrounding it), and this is consistent with the considerable slowdown in the comparison of bars when there are other bars nearby that are not under comparison (Figure 28), although accuracy is not affected (Figure 27). Third, although the effect of surrounding graphs on graph comparison has some potential for a worthwhile experiment, it did not seem compelling enough to collect more pilot data or expand the focus.

Chapter 3: Experiment 2

This chapter describes an experiment in which simple graphs of one line each were placed at a distance from one another. The participant compared the graphs to identify the one with the longer bar. In Experiment 2, the graphs were aligned horizontally, and placed at different distances. These included the four distances from the pilot study, as well as smaller and longer distances, for a range of 25-800 pixels.

The procedure was similar to that of Experiment 1, although the input device used by the participants was changed to a game controller to reduce the error associated with having to move the mouse in different directions depending on the condition. The old computers and chairs used in Experiment 1 were replaced with new computers and chairs for Experiment 2 and later experiments, to reduce any experimental error associated with the old equipment. Furthermore, the distance from the participants' eyes to the stimuli was standardized. The experiment was also removed from the web page to reduce any possible error associated with different browsers.

Research Questions

This experiment was designed to answer the following questions about comparing small lines at various distances:

Research Question 2-1. How does the distance between lines affect their comparability?

In Experiment 1, comparability was lower for more distant lines than for closer ones, as measured by response time, but not accuracy. I expected that with a larger sample size, a wider variety of distances, and a more controlled procedure, some effect of

distance would be evident in both measures, or else that it would be clear that accuracy was not affected by distance under these conditions.

In Experiment 1, there appeared to be a larger slope in the response time measure between 50 and 100 pixels and 100 to 200 to 400 pixels. This suggested that distances within the cofoveal range were greater than distances outside of this range.

Research Question 2-2. How does the difference in length interact with the distance effect?

This interaction was not significant in Experiment 1 for either comparability measure. This finding, if it held up, would be surprising, because it would mean that there is no particular advantage, when comparing similar lines, to have them close together, beyond the advantage gained when lines of any size are close together.

Research Question 2-3. Is the absolute or relative difference in lengths a better predictor of comparability? Psychophysical theories such as Fechner's and Stevens' suggest that the relative length of lines should predict their comparability, all other factors being equal. Another possibility is that the absolute difference in the size of the bars, and thus the proportion of the field of vision subtended by them, is a stronger predictor of comparability than the relative difference.

Method

Participants

Sixty-four students in psychology classes participated for partial class credit as part of the psychology participant pool. There were 35 males and 29 females. The age

range was 18 to 26, with a mean of 19.8. Participants were required to have normal or corrected-to-normal vision. No vision tests were performed on the participants, although none complained that they could not see the stimuli.

Apparatus

All tests were administered on Apple iMac computers with Intel Core 2 Duo processors and 2 gigabytes of RAM, running Mac OS X 10.6. The screen resolution is 1680 horizontal by 1050 vertical pixels. The overall diagonal screen size was 20 inches. Each pixel was a .25 mm square, corresponding to approximately 0.057 degrees of visual angle.

The participants completed the consent forms and questionnaires using the mouse and keyboard. For the graph comparison task, participants used a Logitech® Dual Action™ Gamepad (Figure 29), a game controller which was plugged into the iMac by USB. The game controller output was mapped to the keyboard using USB Overdrive (Montalcini, 2009).



Figure 29. Logitech® Dual Action™ Gamepad.

The hat switch, left thumbstick, and “2” buttons are labeled.

All consent forms and questionnaires were administered on a web browser. I wrote the experimental software used in the graph comparison tasks in the Java programming language. The experimental program was positioned at the top middle of the computer screen.

Participants sat in adjustable office chairs, and were encouraged to find a comfortable sitting position. The chairs were positioned such that the distance between the participant’s eyes and the top of the iMac screen was one meter. An experimenter checked this distance for each participant using a one meter long piece of string. Participants were asked not to move around or lean in their chairs while completing the experiment, and an experimenter was on hand to observe the participants and make sure that they followed the rules.

Materials

The experimental program showed a white rectangle, 980 pixels wide by 780 pixels tall. A 10-pixel wide light gray border was drawn around the outside of this rectangle, making the entire window 1000 pixels wide by 800 pixels tall (about 10 inches wide and 8 inches tall). Note that this gray border was always present, even when the screen is described below as going blank, and that references to the center or edges of the screen refer to the experiment screen unless stated otherwise. The program contained the instructions for how to complete the experiment. These instructions were presented one page at a time, without any scrolling. Participants used the “hat switch” (see Figure 29) on the game controller to move through the instructions. All text was written in Lucida Grande, a sans-serif font. The title text of each page was written in 36 point size, and the body of the instructions was written in 22 point size. Figure 30 shows the first instruction page; subsequent instruction pages had a similar appearance. No participants expressed difficulty reading or understanding the instructions.

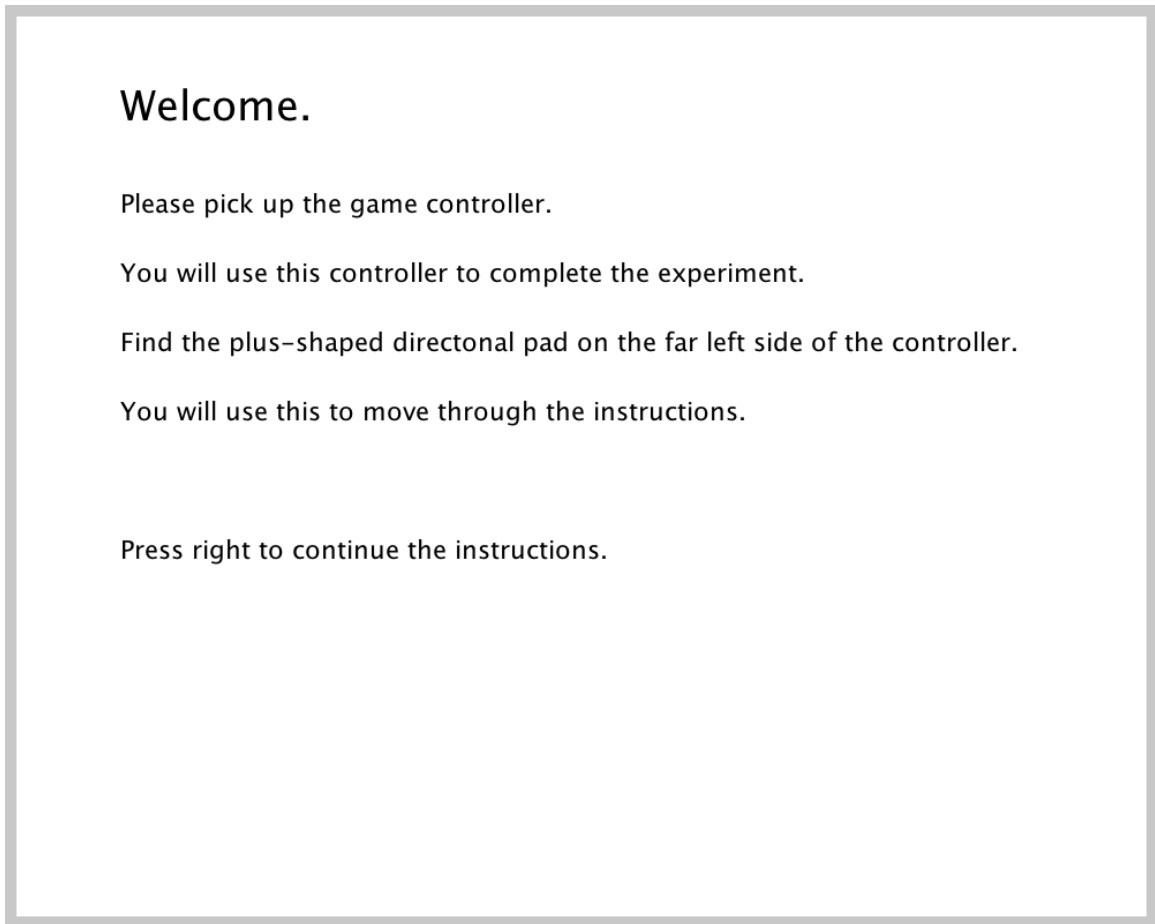


Figure 30. Welcome screen (Experiment 2).

During each trial, participants were shown two small graphs. Each graph consisted of a single black vertical line, 2 pixels wide, and between 5 and 21 pixels tall. The backgrounds of each graph, as well as the background of the experiment screen, were white, with only the same thin gray border around the outside of the experiment screen. Figure 31 shows a typical trial screen. A light blue rectangle containing the words, “Press and release 2 to begin”, and “READY”, in white, was shown before each trial, centered on the screen. Before each trial, medium gray crosses were shown in place of the graphs, and at a certain point in the trial, the graphs were replaced by a medium gray pattern as a mask. Both the crosses and mask were 50 pixels tall by 50 pixels wide,

which was the maximum size of the graph. The feedback shown after each trial was either the word “Correct!” in 36 point black, or “Incorrect” in 36-point red.



Figure 31. Trial screen from Experiment 2.

The lines are 100 pixels apart. The line on the left is 11 pixels long; the line on the right is 16 pixels long. Both lines are 2 pixels wide.

Design

Table 10, below, summarizes the conditions of this experiment, excluding practice trials. The trials were grouped into blocks, with the positions of the two graphs always consistent within a block, and always changing from one block to the next. There

was one practice block, followed by 18 test blocks. Excluding the practice block, there were 6 position conditions, each with the two graphs separated in the horizontal direction by a particular number of pixels, as measured from the center of the graph. The distances were 25, 50, 100, 200, 400, and 800 pixels. In the practice block, the distance was 300 pixels. There were three blocks of trials for each of the 6 positions. The order of the test blocks was randomized, with counterbalancing that ensured that the same block condition would not appear twice before another appeared once, nor three times before another appeared twice, and that no block condition would appear twice within a run of three blocks.

Each block consisted of four practice trials followed by 48 test trials in a random order. In those trials, each of the two standard length lines was placed either in the left or right position (this is called “position of standard” in the table), and paired with each of its 6 comparison bars. Each of these combinations was shown 2 times. This made 48 test trials per block, for a total of 864 test trials over the course of the experiment.

Between block conditions	Number of levels	Levels
Distance	6	25, 50, 100, 200, 400, 800
Block repetitions	3	
Total number of blocks	18	
Within block conditions	Number of levels	Levels
Standard line length	2	10, 16
Comparison line length	6 per standard	<i>10</i> : 5, 7, 9, 11, 13, 15 <i>16</i> : 11, 13, 15, 17, 19, 21
Position of standard	2	left, right
Trial repetitions	2	
Number of trials per block	48	
Total number of trials	864	

Table 10. Conditions of Experiment 2.

Procedure

Each block began with a short reminder about the instructions. This was followed by four easy practice trials. If the participant successfully completed all four of these trials, the test trials followed immediately. If the participant responded incorrectly to a practice item, they were immediately notified of their error, and shown the instructions again. They then repeated the practice trials until they answered all four correctly. This procedure was meant to ensure that the participants understood the trial procedure and to reduce the impact of the longer response times that often occurred on the early trials in a block.

The trial procedure is illustrated in Figure 32. Each trial began with a blue rectangle, called the “Ready Box” in the instructions, shown in the center of the screen (Figure 32a). When the participant was ready, he or she pressed and released the “2” button (see Figure 29) with his or her right thumb. When the “2” button was released, the

screen went blank for 100 ms. Then the two crosses appeared, in the two graph positions of the current block, for 500 ms (Figure 32b). Then, the screen went blank for another 100 ms. Then the two graphs, (each consisting of a single line) appeared on the screen (Figure 32c). These graphs remained on the screen until the participant pressed the “2” button again. Then the graphs disappeared and were immediately replaced by gray square masks (Figure 32d). These masks remained until the participant indicated which graph contained the longer line by moving the left thumbstick in the direction of their choice. Once the thumbstick was moved, feedback about the accuracy of the response was shown, in the form of the words “Correct!” or “Incorrect”. If the response was correct, the “Correct!” message was shown for 500 ms (Figure 32e). If the response was incorrect, the “Incorrect” message was shown for 2000 ms (Figure 32f). The time discrepancy was a mild punishment to discourage careless responding. After the feedback message, the screen went blank for 200 ms, and was replaced by the ready box for the next trial. At the end of a block, an instruction screen was shown, indicating how many blocks had been completed, and the total number of blocks in the experiment.

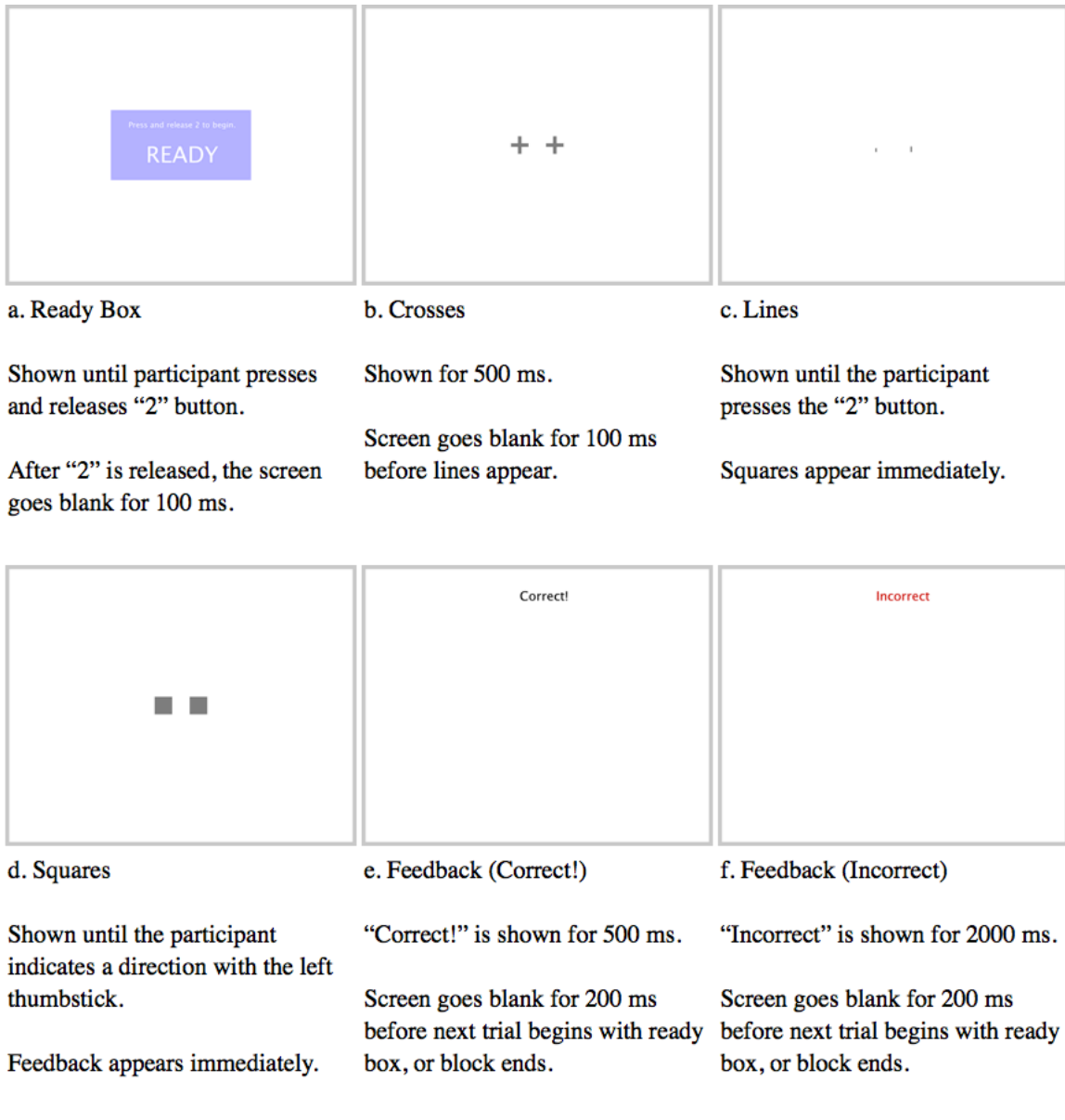


Figure 32. Screens from Experiment 2.

Results

Comparability is measured by the accuracy of comparisons and the time it takes to complete them. Repeated measures ANOVAs were performed for both of these measures. The factors, listed in Table 11, were the distance between the lines (“dist”),

the length of the standard line (“sLength”), the absolute difference in length between the two lines (“absDif”), and whether the standard or comparison line was longer (“longerSC”). These last two factors are derived from the difference between the standard and comparison lines (“cDiff”). For instance, if the length of the standard line is 10, and the length of the comparison line is 7, then the cDiff is -3, which in these analyses would be treated as an absDif of 3, and a longerSC of -1, meaning that the standard line was longer. In other words, absDif is the absolute value of cDiff, and longerSC corresponds to the sign of cDiff. The reason for this change was to illustrate that, as the analysis shows, the absolute difference matters much more than whether the standard or comparison line is longer.

Condition	Number of levels	Levels
dist	6	25, 50, 100, 200, 400, 800
sLength	3	10, 16
absDif	3	1, 3, 5
longerSC	2	-1 (standard line is longer), 1 (comparison line is longer)

Table 11. Factors for Repeated-Measures ANOVA, Experiment 2.

All other conditions were combined for these analyses, so in the accuracy ANOVA, the score was a proportion of the 12 trials that were correctly responded to. In the time ANOVA, the score was the median of these 12 times. The median was chosen, rather than the mean, to diminish the impact of outliers in the long response direction. Although I considered alternative methods of controlling for outliers, the median was both the simplest and the most robust. In later experiments, for instance, there were cases in which all of the trials that made up a score were responded to incorrectly, so it would not be possible to only count accurate trials. Furthermore, because there were only 2 possible responses, many guesses would have ended up influencing these scores anyway.

Accuracy

Table 12 shows the results of the repeated-measures ANOVA for Experiment 1, using accuracy (proportion correct) as the dependent variable. The conditions of the three blocks, two repetitions, and two standard positions were summed, for a total of 12 trials per cell.

Effect	DFn	DFd	F	p	p<.05	Partial ω^2
dist	5	315	134.75	<.001 [GG]	*	0.1267
sLength	1	63	91.46	<.001	*	0.0193
absDif	2	126	832.48	<.001 [GG]	*	0.2652
longerSC	1	63	1.19	.279		
dist x sLength	5	315	7.85	<.001 [GG]	*	0.0074
dist x absDif	10	630	49.21	<.001 [GG]	*	0.0947
sLength x absDif	2	126	28.25	<.001 [GG]	*	0.0117
dist x longerSC	5	315	0.69	.590 [GG]		
sLength x longerSC	1	63	7.69	.007	*	0.0014
absDif x longerSC	2	126	0.31	.644 [GG]		
dist x sLength x absDif	10	630	1.85	.090 [GG]		
dist x sLength x longerSC	5	315	1.71	.152 [GG]		
dist x absDif x longerSC	10	630	0.50	.785 [GG]		
sLength x absDif x longerSC	2	126	0.95	.361 [GG]		
dist x sLength x absDif x longerSC	10	630	2.71	.016 [GG]	*	0.0037

Table 12. Repeated-Measures ANOVA table for accuracy (Experiment 2)

The distance between the two lines was a significant predictor of comparability, with accuracy decreasing as the distance increased. The largest effect was that of the absolute difference in size between the two lines. The interaction of these variables was also significant, as were a few other variables.

Figure 33 shows accuracy graphed against the distance between the two graphs, with the different conditions of absDif and sLength drawn as separate lines. The thinner bar in each color represents the condition in which the standard bar was 10 pixels long; the thicker bar, 16 pixels. The variable representing which bar was longer (longerSC) is not shown. Red bars represents an absDif value of 1, meaning the lines differed in length

by only a single pixel. These were the hardest conditions, and the graph shows it.

Accuracy was near 90% when the two lines were close enough to be easily seen without an eye movement, but accuracy quickly dropped off when the lines grew further apart, eventually falling below 75%. For bars that differed by 3 or 5 pixels, shown as green and blue bars, respectively, the distance effects were mild. Even when the bars were separated by 800 pixels, accuracy remained above 90%.

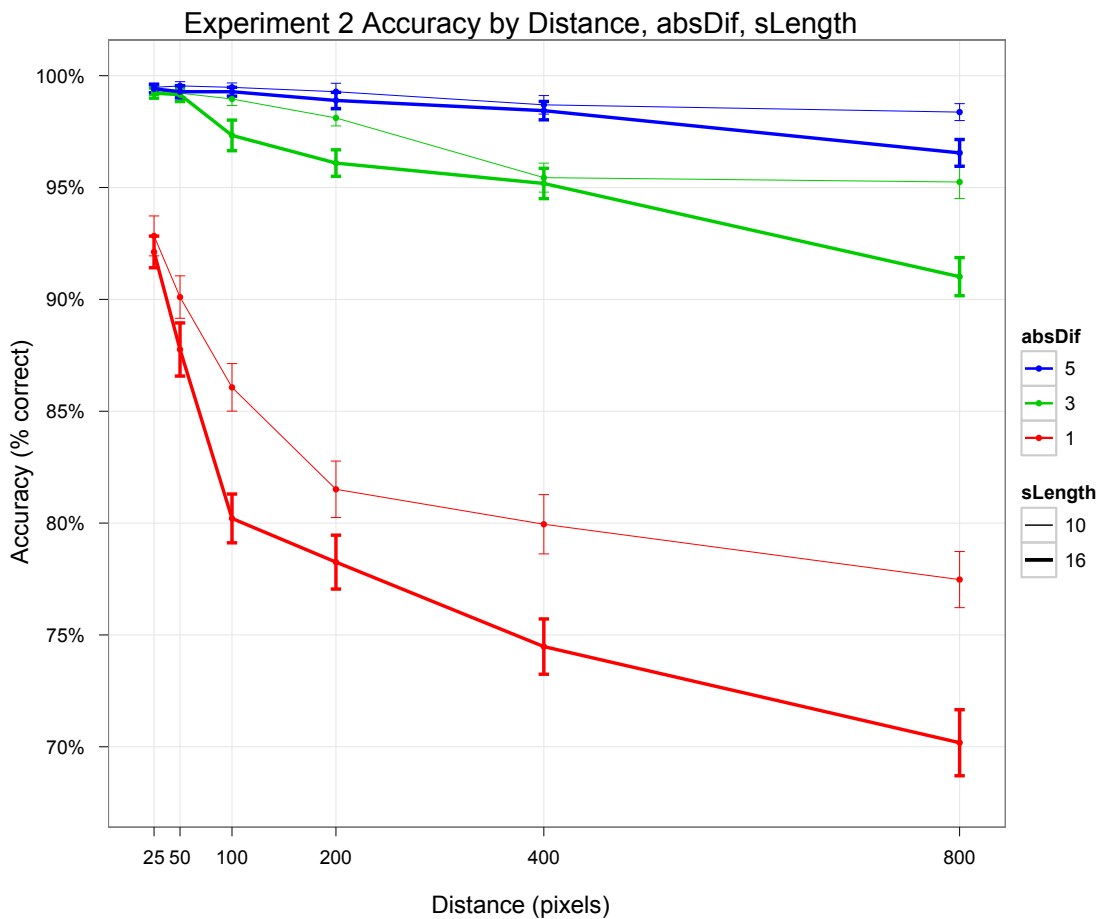


Figure 33. Accuracy by distance, absDif, and sLength (Experiment 2).

There was a significant effect of absDif, meaning lines of similar length (smaller absDif) were less comparable than lines that were more different in length (larger

absDif). The comparability difference was more pronounced between 1 and 3 pixels than between 3 and 5 pixels, in line with Henmon's (1906) finding (see Figure 17).

Response Time

Table 13 shows the ANOVA table for response time. Recall that response time refers only to the time between the appearance of the two graphs and when the 2 button is pressed. Because of the very short response times, an interruption of a second or two could have added quite a bit of noise. The average of these times was calculated using a median, not a mean, of the 12 trials, which reduced the noisiness of the data.

Effect	DFn	DFd	F	p	p<.05	Partial ω^2
dist	5	315	253.43	<.001 [GG]	*	0.2150
sLength	1	63	2.20	.143		
absDif	2	126	86.84	<.001 [GG]	*	0.0359
longerSC	1	63	0.23	.636		
dist x sLength	5	315	4.93	.001 [GG]	*	0.0042
dist x absDif	10	630	5.27	<.001 [GG]	*	0.0092
sLength x absDif	2	126	14.58	<.001 [GG]	*	0.0059
dist x longerSC	5	315	1.90	.114 [GG]		
sLength x longerSC	1	63	0.64	.425		
absDif x longerSC	2	126	0.95	.360 [GG]		
dist x sLength x absDif	10	630	3.35	.005 [GG]	*	0.0051
dist x sLength x longerSC	5	315	2.14	.078 [GG]		
dist x absDif x longerSC	10	630	1.58	.147 [GG]		
sLength x absDif x longerSC	2	126	2.45	.106 [GG]		
dist x sLength x absDif x longerSC	10	630	1.30	.260 [GG]		

Table 13. Repeated-Measures ANOVA table for response time (Experiment 2).

Distance is a significant predictor of response time, as is absDif, and some interactions. Here, sLength is not significant, and neither is longerSC. The average response time is plotted in Figure 34, with separate lines for each absDif condition. The distance conditions are laid out on the x axis in proportion to the distance between the lines, in such a way that if the RT increases linearly with distance, the lines will be

straight. Instead, there is a slight increase in RT in each cDiff condition when the distance is increased from 25 to 50 pixels, then a sharp increase from 50 to 100 pixels, then smaller increases to 200, 400, and 800 pixels. Although the RT increases with the distance between the lines, this increase is less than linear for distances greater than 100 pixels.

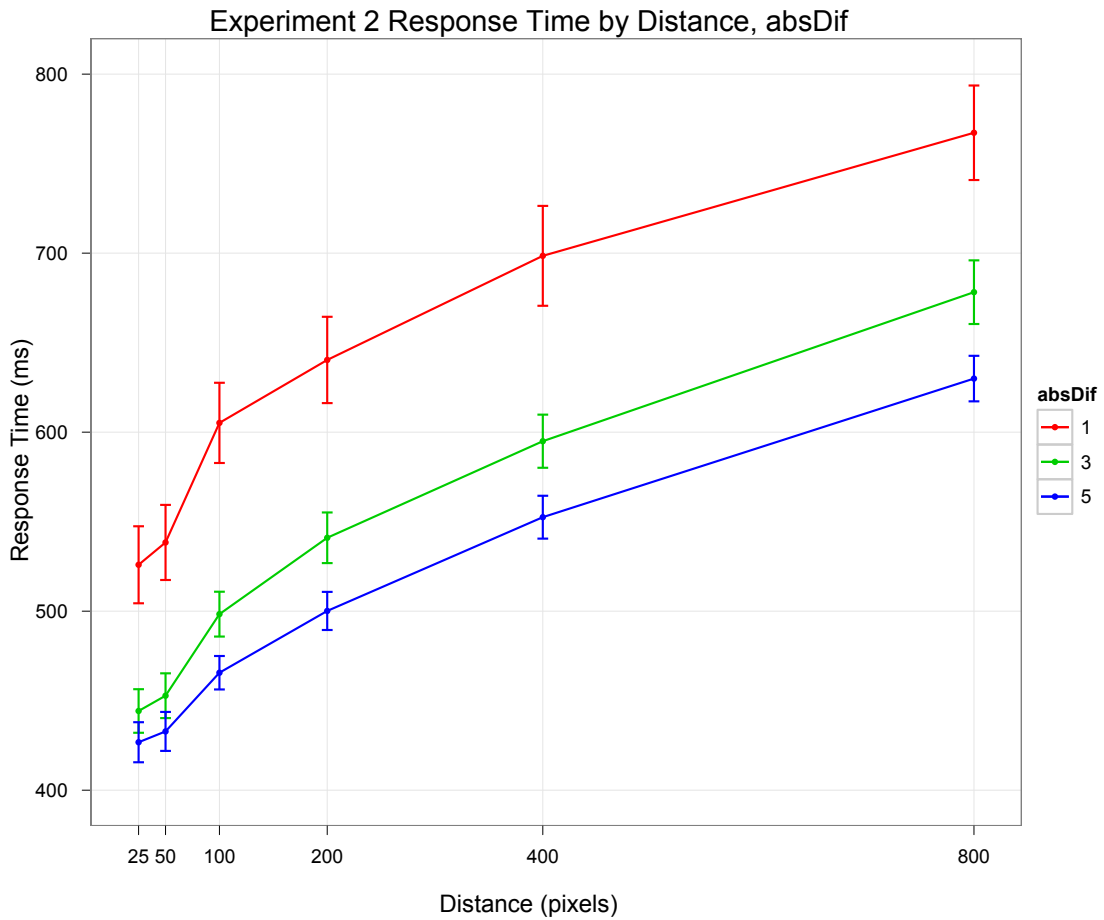


Figure 34. Response time by distance and absDif. (Experiment 2).

The effect of absDif on comparability holds for this measure as well, and again, there is a larger difference between absDif of 1 and 3 than between 3 and 5. There is little (although significant) interaction between the distance and absDif. People take

longer to respond to more similar bars, but this is not much magnified by the distance between the bars.

Figure 35 shows the same data plotted with the distance on a scale of log base 2. If the effect of distance on comparison time is due to looking between the two lines, Fitts' Law (Fitts, 1954, Miniotas, 2000) predicts that the response time should increase linearly with the log of the distance, so the lines should be straight as plotted. This comes close to describing the graph from 50 pixels to 800 pixels, although as noted earlier, there is very little difference between the response times at 25 and 50 pixels distance. This is not surprising, because Fitts' Law often describes movement times poorly at very short distances, and particularly so for eye movements (Goldberg & Kotval, 1999, Miniotas, 2000), and because at 25 pixels, it would have been possible to compare the lines without moving the direction of gaze at all. Even after 50 pixels, there is a slight upward curve for absolute differences of 3 and 5 pixels, indicating that people are either taking longer to look at the lines when they are more distant, or are looking back and forth more often. It is not possible to distinguish between these possibilities because eye movements were not recorded. Nevertheless, the deviations from the predictions of Fitts' Law (above 50 pixels) are not great, and in particular, they are not greater for the more difficult items ($\text{absDif} = 1$).

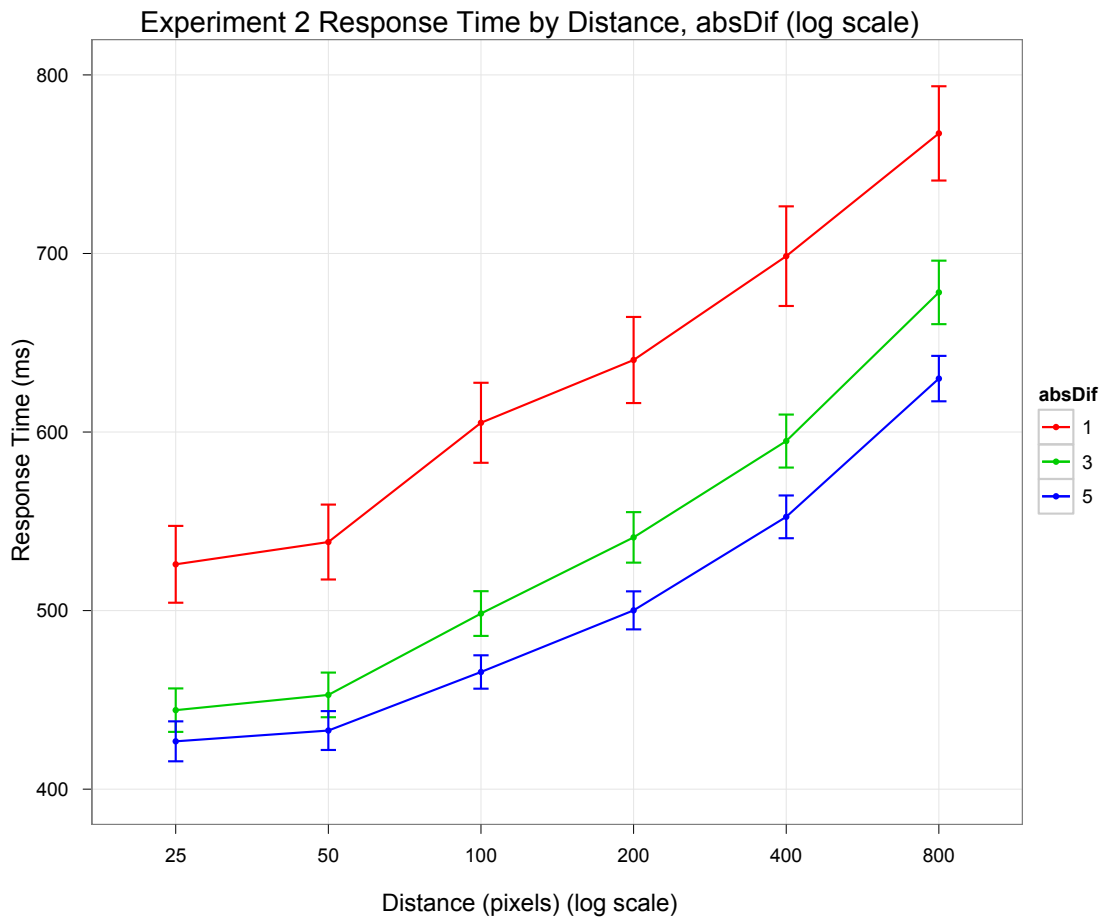


Figure 35. Response time by distance and absDif, distance plotted on a log scale (Experiment 2).

Absolute and relative differences

The six different values of cDiff (-5, -3, -1, 1, 3, 5) were chosen to make some comparisons easier and some more difficult. The different values of sLength (10, 16) were chosen to encourage participants to look at both lines, not just one. Because of these different values, the length difference between the lines is not the same, depending on whether it is measured as an absolute or relative difference. Thus it is possible to use these results to test the psychophysical theory that relative length, not absolute length, is the best predictor of comparability. This was not the goal of this design, so there are not

for instance, two conditions that have the same relative difference, but different absolute differences. There were, however, three sets of four conditions with the same absolute length difference, and different relative length differences, and one pair of conditions for which the absolute and relative differences in length make different predictions about the comparability of the lines.

Figure 36 shows response times in Experiment 2 using the relative difference between the lines on the X axis, and the three colors used earlier for the absolute difference. There are 12 data points, one for each combination of cDiff and sLength. Relative difference is calculated as the absolute difference divided by the length of the shorter line. Response times are much more closely related to absolute difference than relative difference.

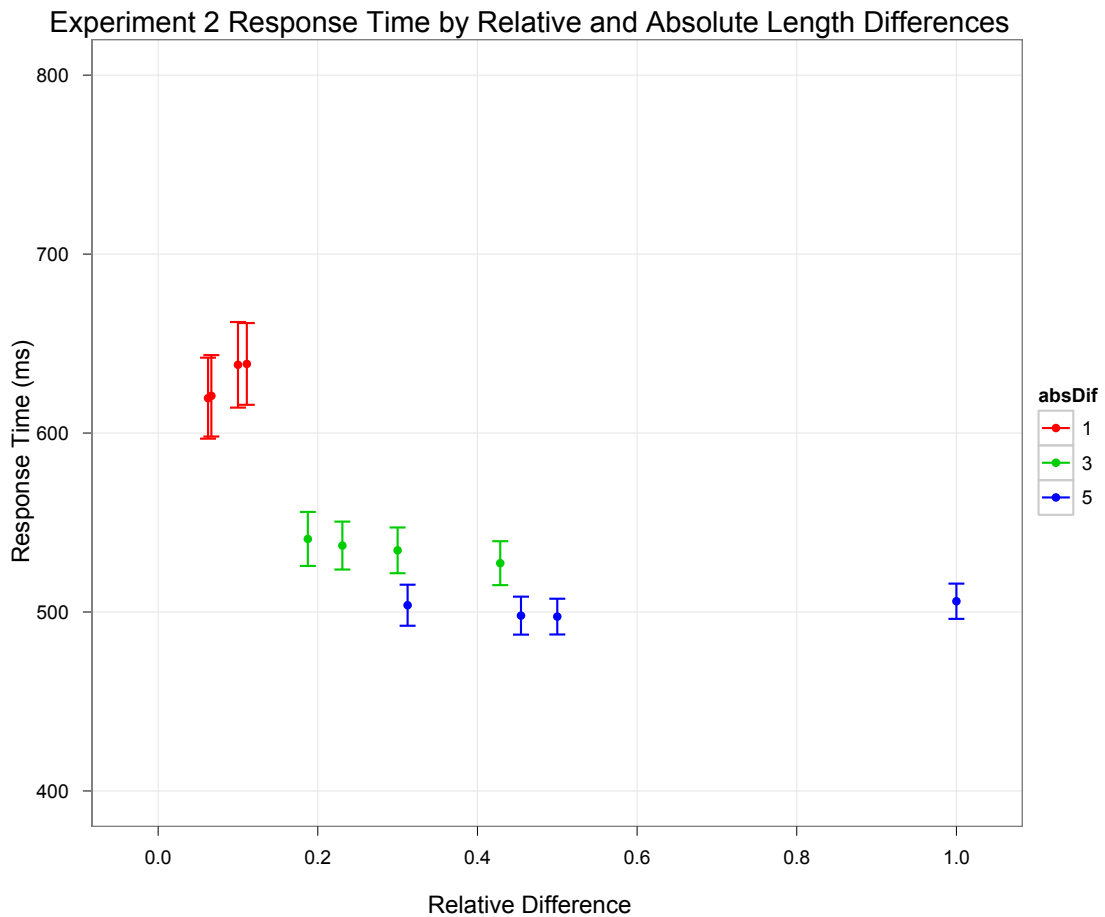


Figure 36. Response Time by relative and absolute differences in length (Experiment 2).

The most interesting points on this graph are the two on either side of .4 on the x axis. These represent pairs of lines that are 7 and 10 pixels (the green point at about .43) and 16 and 21 pixels (the blue point at about .31). The psychophysical law predicts that the lines with the larger relative difference should be more comparable, but we find faster comparisons for those with the larger absolute difference. This difference is fairly stable across different distances, as seen in Figure 37, which plots just these two points at each distance.

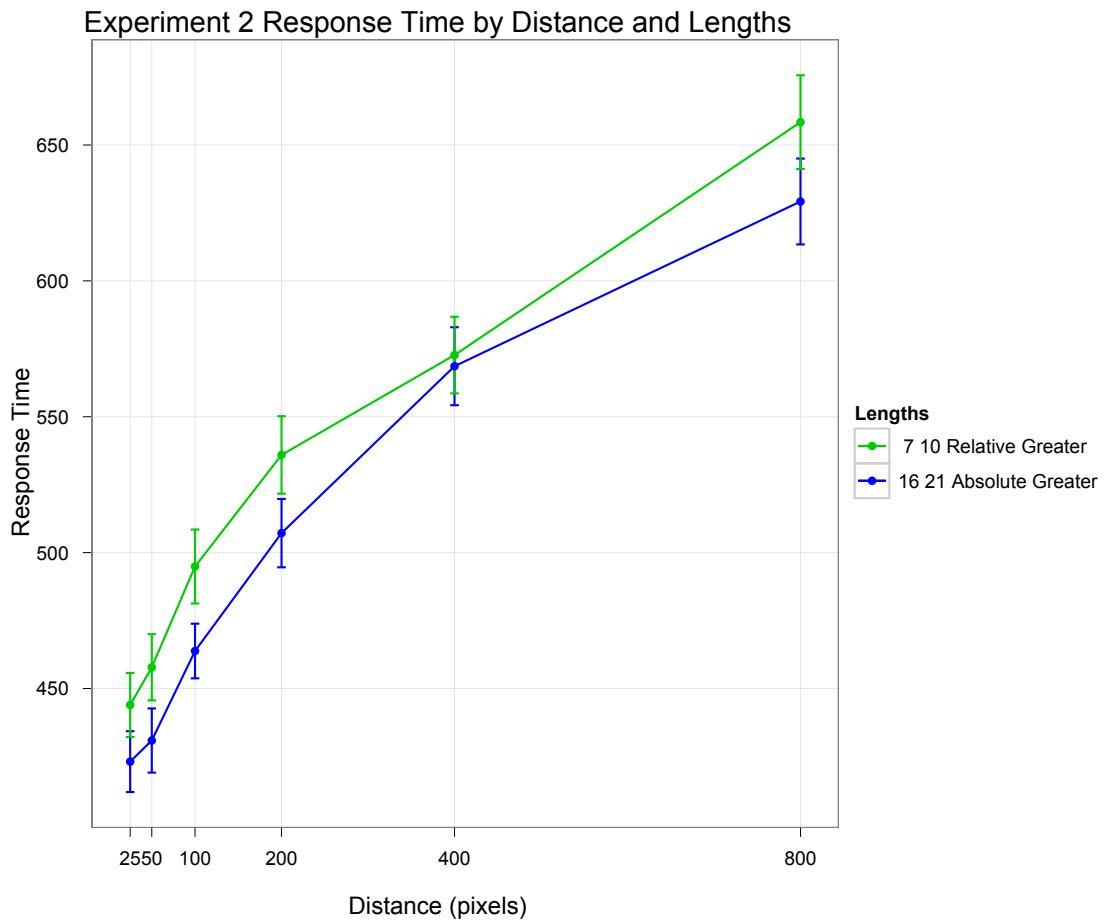


Figure 37. Response Time for two pairs of lines at each distance (Experiment 2).

Similar graphs for accuracy (Figure 38, Figure 39) show smaller differences, but nothing that suggests that relative difference is a better predictor than absolute difference.

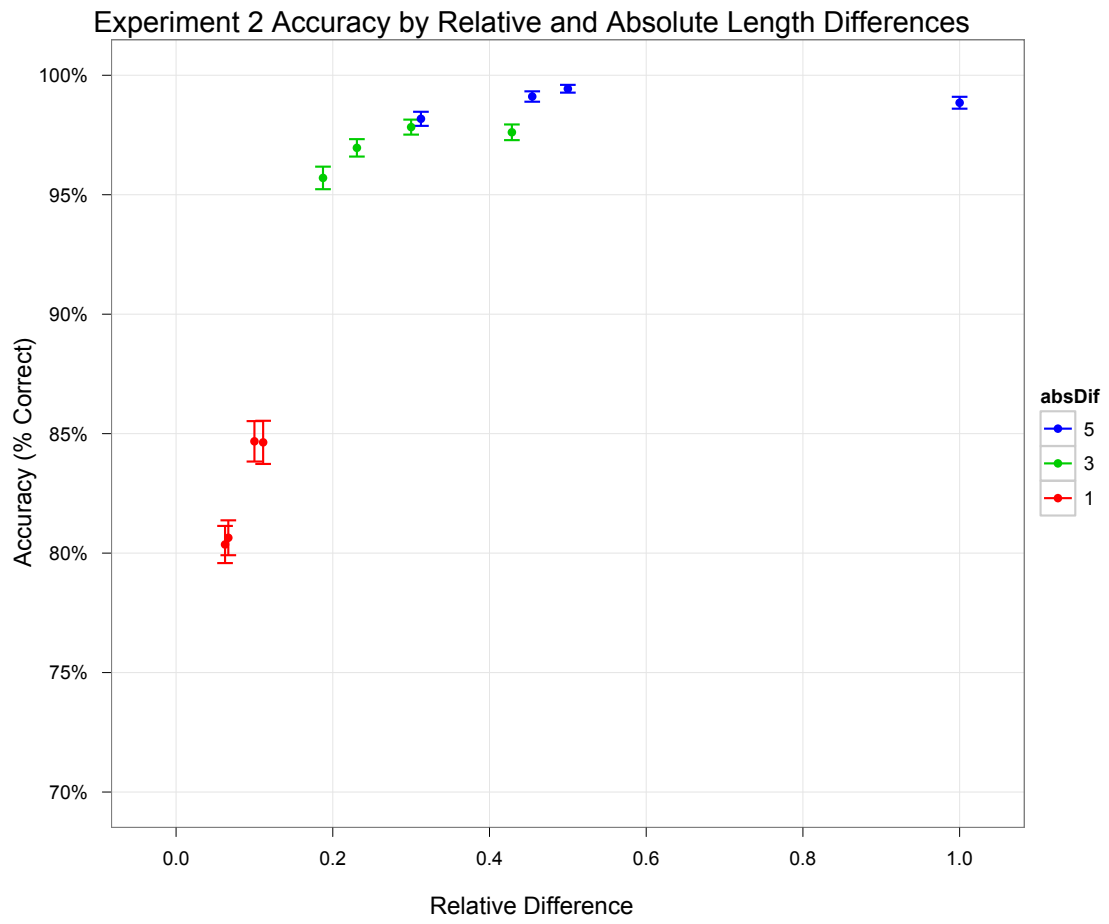


Figure 38. Accuracy by relative and absolute differences in length (Experiment 2).

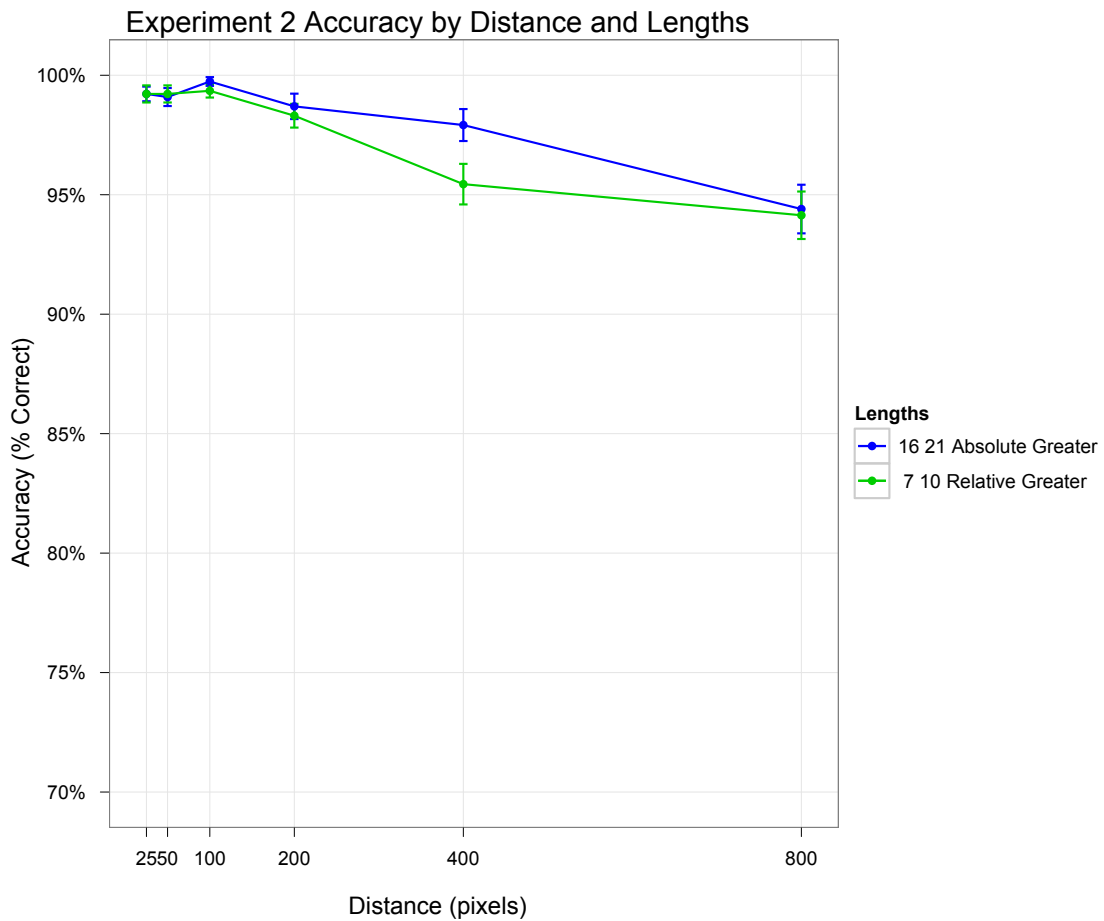


Figure 39. Accuracy for two pairs of lines at each distance (Experiment 2).

Discussion

We can now answer the research questions for this experiment:

Research Question 2-1. How does the distance between lines affect their comparability?

Lines that were farther apart were less comparable, both in terms of accuracy and response time. This is in contrast to the findings in Experiment 1, but in line with what Weber (1834/1978) predicted.

Research Question 2-2. How does the difference in length interact with the distance effect?

Using the time metric, there was fairly little interaction between these effects, and the amount of time taken increased about linearly with the distance between the lines, as that distance increased from 100 pixels to 800 pixels. The increase was greater between 25 and 100 pixels.

Using the accuracy metric, lines that differed by only a single pixel became less comparable rapidly when the distance between them increased from 25 to 100 pixels, and continued to decline until the furthest distance measured, 800 pixels, at which point accuracy was under 75%. Accuracy also declined for comparisons of lines that differed by 3 or 5 pixels, but much less so than the pairs of 1 pixel difference.

Note that for both measures of comparability, the pattern changes around the 100-pixel mark. This distance is about 1.4 degrees of visual angle, which is near the limit of central vision.

Research Question 2-3. Is the absolute or relative difference in lengths a better predictor of comparability?

The absolute difference in length between two lines was a better predictor of the comparability of the lines, than was the relative difference, using both the accuracy and response time measures. The psychophysical theories of Weber, Fechner, and Stevens, which all state that it is the relative difference, not the absolute differences, that affect comparability, do not hold in this situation. Why? One possibility is that this experiment, like the situation depicted in Weber (1834/1978) used lines that were aligned at the bottom (the bottom of each line was at the same y-axis coordinate), with the length

differences detectable as different heights of the lines (different y-axis coordinates). Thus only the tops of the lines, not the bottoms, were relevant for length judgments, which may have caused people to ignore the bottoms of the lines. Stevens (1975) gives an example of lines at various orientations, presented in nonaligned pairs, for which this kind of judgment is impossible. Another possibility is the constant width of the lines, 2 pixels, caused the participants to view what I have been calling “lines” as thin rectangles and that the shape of that rectangle is influencing the judgments.

Chapter 4: Experiment 3

This chapter describes an experiment in which simple graphs of one line each were placed at a distance to one another, and the participant identified the graph with the longer bar, just as in Experiment 2. In Experiment 3, the two graphs were always 200 pixels apart, but their alignment and orientation differed from one condition to another.

Research Questions

This experiment was designed to answer the following questions about small multiple bar graphs:

Research Question 3-1. How does the alignment of lines affect their comparability?

Most studies of comparison of line length use stimuli that are side by side. An exception is the study of optical illusions of perspective, for which visual cues are included to encourage viewers to interpret the lines as nearer or farther parts of a naturalistic scene. Because this study is focused on small multiple graphics, I did not include any such cues, and I expected any perspective-based illusions to have minimal impact.

For the diagonal alignments, there were two hypotheses. The diagonal alignments might be worse than the horizontal or vertical alignments, because people are unaccustomed to comparing visual objects that are not aligned either vertically or horizontally. Alternatively, diagonally aligned lines might be somewhere between vertical and horizontal lines in terms of comparability, for reasons explained below under Research Question 3-3.

Research Question 3-2. How does the orientation of lines affect their comparability?

I had no strong theory about whether horizontal or vertical lines would be easier to compare, independent of their alignment. Horizontal and vertical lines have both been used as research stimuli, as have diagonal lines, which I did not include here. (I excluded diagonals for the sake of controlling the number of conditions, because of the imprecision of length of diagonal lines rendered on a computer screen, and because diagonal lines are not commonly used to draw bar graphs. Even in line graphs, which do use diagonals, the length of the line does not represent a particular attribute.)

Research Question 3-3. How do alignment and orientation interact?

I expected, as per Weber (1834/1978), that it would be easier to compare vertical lines aligned horizontally, or horizontal lines aligned vertically, because the ends of the lines make an angle under these cross-aligned conditions, and there is no angle under the co-aligned condition.

If the lines are more comparable in the cross-aligned conditions than the co-aligned conditions, the diagonally aligned conditions might split the difference, allowing the viewer to form a trapezoid using the two endpoints of each line, and compare the angles of the imagined lines to determine the longer line.

Research Question 3-4. How does the difference in length interact with these effects?

I did not have an a priori theory about how the difference in length would interact with the alignment and orientation conditions.

Research Question 3-5. Is the absolute or relative difference in lengths a better predictor of comparability?

The alignment conditions allow a test of the hypothesis that absolute length differences trumped relative length differences in Experiment 2 because the viewers could ignore the bottom half of each line.

Method

Participants

Forty-two students in psychology classes participated for partial class credit as part of the psychology participant pool. There were 18 males and 24 females. The age range was 18 to 25, with a mean of 19.7. Participants were required to have normal or corrected-to-normal vision. No vision tests were performed on the participants, although none complained that they could not see the stimuli.

Apparatus and Materials

The apparatus and seating procedure were exactly the same as those used in Experiment 2.

The instructions and test program were similar to those in Experiment 2. The only changes were the different stimuli, described in the design section below, and instructions that explained that the bars might be horizontal or vertical.

Design

Table 14, below, summarizes the conditions of this experiment, excluding practice trials. The trials were grouped into blocks, with the positions of the two graphs always consistent within a block, and always changing from one block to the next. There was one practice block, followed by 16 test blocks. Excluding the practice block, there were 4 alignment conditions and 2 orientation conditions, which were completely crossed. The alignment conditions (how the two graphs were arranged on the page) were -45° (upper left/lower right), 0° (side by side, as in Experiment 2), 45° (upper right/lower left), and 90° (one over the other). The orientations were 0° (horizontal bars) and 90° (vertical bars, as in Experiment 2). The distance between the graphs was 200 pixels. In the practice block, the distance was 300 pixels, the alignment was 65° , and the orientation was 90° . There were two blocks of trials for each of the 8 combinations of alignment and orientation. The order of the test blocks was randomized, with counterbalancing that ensured that the same block condition would not appear twice before another appeared once, and that no block condition would appear twice within a run of five blocks.

Between block conditions	Number of levels	Levels
Alignment	4	$-45^\circ, 0^\circ, 45^\circ, 90^\circ$
Orientation	2	$0^\circ, 90^\circ$
Block repetitions	2	
Total number of blocks	16	
Within block conditions	Number of levels	Levels
Standard line length	2	10, 16
Comparison line length	6 per standard	10: 5, 7, 9, 11, 13, 15 16: 11, 13, 15, 17, 19, 21
Position of standard	2	left, right
Trial repetitions	2	
Number of trials per block	48	
Total number of trials	768	

Table 14. Conditions for Experiment 3.

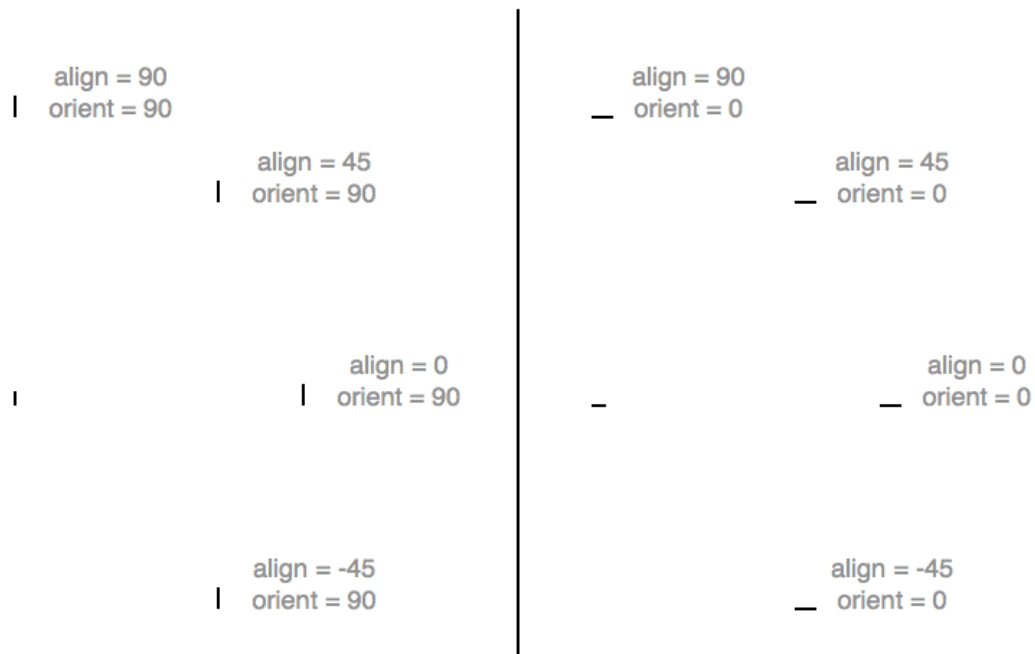


Figure 40. Alignment and Orientation Conditions for Experiment 3.

Each labeled line, paired with the unlabeled line, represents one of the combinations of alignment and orientation conditions. On the left, the vertical orientation conditions (orient = 90). On the right, the horizontal orientation conditions (orient = 0). The gray labels are for reference only.

Each block consisted of four practice trials followed by 48 test trials in a random order. In those trials, each of the two standard length lines was placed either in the left or right position (or top or bottom position, for vertical alignments; this is called “position of standard” in the table), and paired with each of its 6 comparison bars. Each of these combinations was shown 2 times. This made 48 test trials per block, for a total of 768 test trials over the course of the experiment.

Procedure

The procedure was almost exactly the same as that of Experiment 2. The only difference was the direction that the participant moved the left thumbstick. The instructions indicated only that the left thumbstick should be moved in the direction of the longer bar. For trials with vertical alignments, a response was registered when the left thumbstick was moved forward or backward. For trials with diagonal alignments, a response was registered when the thumbstick was moved forward, backward, left, or right. Because the left thumbstick moves omnidirectionally, the participant could move it diagonally, and did not have to choose one of these directions. Recall that response time was measured by when the 2 button was pressed, not when the thumbstick was moved, to minimize any differences caused by moving the thumbstick in different directions.

Results

As in Experiment 2, the comparability of the graphs is measured in both accuracy and time. The repeated measures ANOVAs have 5 factors: alignment (“align”), orientation (“orient”), the length of the standard line (“sLength”), the absolute difference between the standard and comparison lines (“absDif”), and which line was longer (“longerSC”). As in Experiment 2, absDif and longerSC are separated out from cDiff.

Accuracy

Table 15 shows the results of the repeated-measures ANOVA for Experiment 3, using accuracy (proportion correct) as the dependent variable. The conditions of the two blocks, two repetitions, and two standard positions were summed, for a total of 8 trials

per cell. Alignment was a significant predictor of response accuracy, but neither were orientation or the interaction of these factors. The horizontal alignment, with two lines to the left and right of one another, led to the most accurate responses. This effect was smaller than some significant effects related to the lengths of the lines. Figure 41 shows the effects of alignment and orientation on accuracy.

Effect	DFn	DFd	F	<i>p</i>	<i>p</i> <.05	Partial ω^2
align	3	123	4.24	.016 [GG]	*	0.0024
orient	1	41	0.72	.403		
sLength	1	41	76.32	<.001	*	0.0183
absDif	2	82	488.12	<.001 [GG]	*	0.1946
longerSC	1	41	12.47	.001	*	0.0028
align x orient	3	123	2.32	.090 [GG]		
align x sLength	3	123	0.95	.411 [GG]		
orient x sLength	1	41	0.12	.726		
align x absDif	6	246	2.58	.041 [GG]	*	0.0024
orient x absDif	2	82	1.07	.324 [GG]		
sLength x absDif	2	82	20.66	<.001 [GG]	*	0.0097
align x longerSC	3	123	2.46	.072 [GG]		
orient x longerSC	1	41	1.38	.248		
sLength x longerSC	1	41	3.29	.077		
absDif x longerSC	2	82	3.62	.048 [GG]	*	0.0013
align x orient x sLength	3	123	0.73	.524 [GG]		
align x orient x absDif	6	246	0.36	.828 [GG]		
align x sLength x absDif	6	246	1.81	.132 [GG]		
orient x sLength x absDif	2	82	0.27	.649 [GG]		
align x orient x longerSC	3	123	1.03	.378 [GG]		
align x sLength x longerSC	3	123	0.62	.597 [GG]		
orient x sLength x longerSC	1	41	3.08	.087		
align x absDif x longerSC	6	246	1.00	.400 [GG]		
orient x absDif x longerSC	2	82	0.81	.397 [GG]		
sLength x absDif x longerSC	2	82	0.15	.789 [GG]		
align x orient x sLength x absDif	6	246	0.31	.851 [GG]		
align x orient x sLength x longerSC	3	123	0.13	.938 [GG]		
align x orient x absDif x longerSC	6	246	0.79	.518 [GG]		
align x sLength x absDif x longerSC	6	246	1.26	.289 [GG]		
orient x sLength x absDif x longerSC	2	82	3.18	.065 [GG]		
align x orient x sLength x absDif x longerSC	6	246	1.34	.262 [GG]		

Table 15. Repeated-measures ANOVA table for accuracy (Experiment 3).

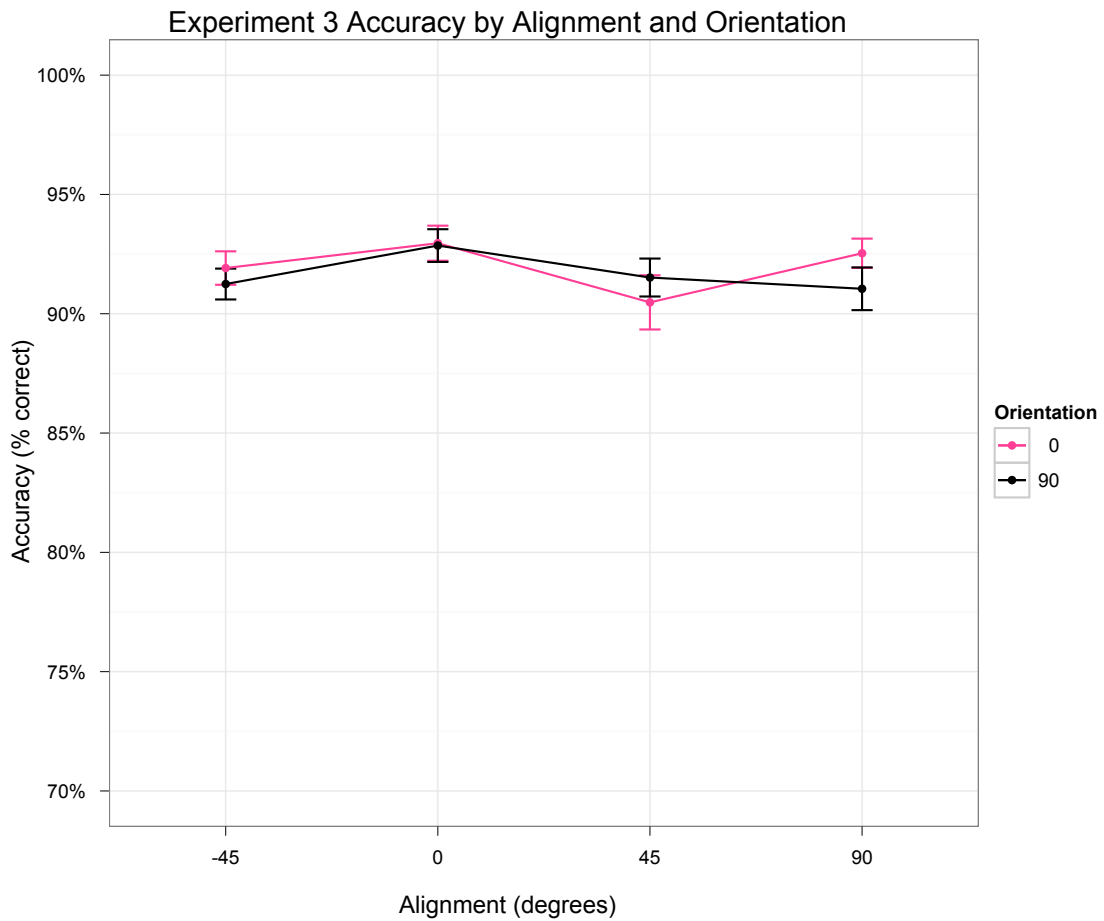


Figure 41. Accuracy by Alignment and Orientation (Experiment 3).

The black points represent vertical lines; the pink points represent horizontal lines. The alignment effect is significant; the orientation and interaction effects are not.

Figure 42 shows an interesting interaction of alignment and absDif. For the pairs of lines that were most similar in length ($\text{absDif} = 1$), responses were more accurate when the graphs were side by side (alignment = 0) than when they were arranged a different way. In other words, the alignment effect was stronger when the lines were more similar in length.

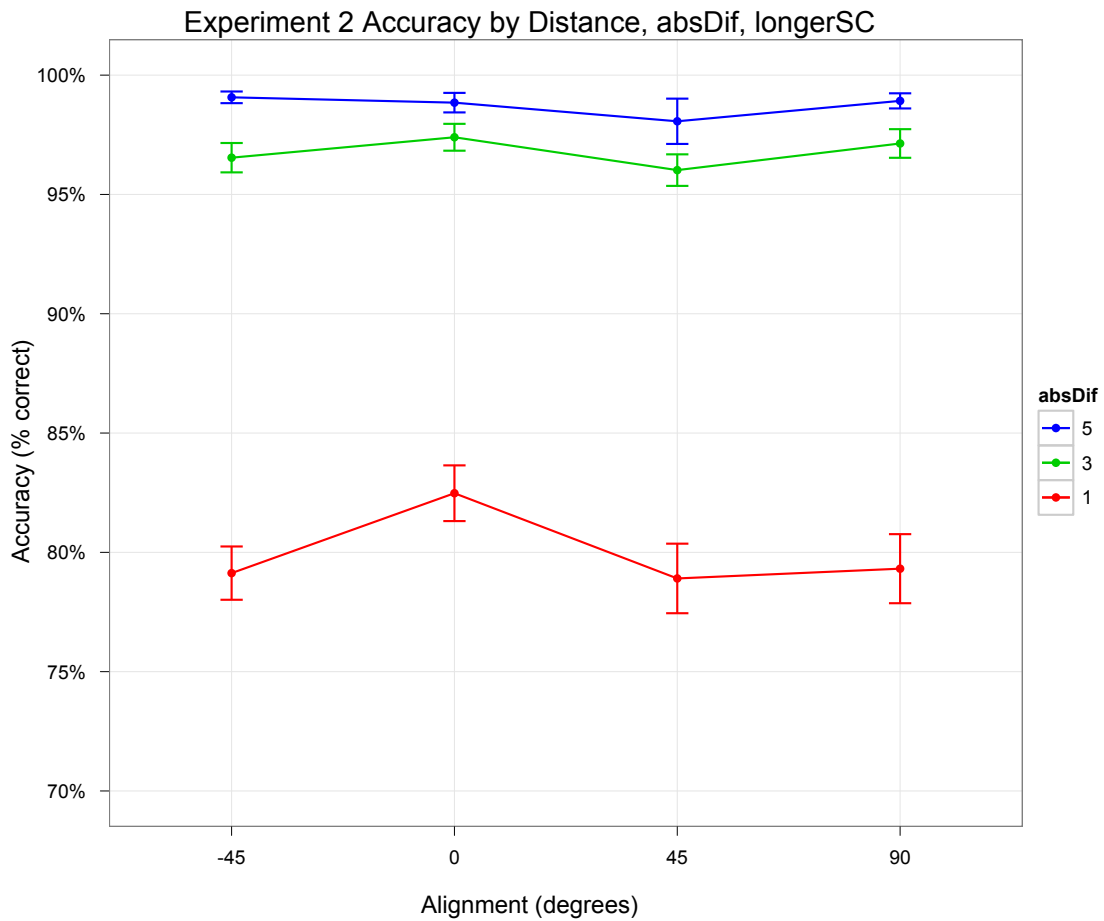


Figure 42. Accuracy by alignment and absDif (Experiment 3).

Response Time

Table 16 shows the results of the repeated-measures ANOVA for Experiment 3, using response time as the dependent variable. The conditions of the two blocks, two repetitions, and two standard positions were summed, for a total of 8 trials per cell. Orientation, but not alignment or the interaction of these factors, is a significant predictor of response time, with vertical lines being compared faster. Again, however, the factors related to the lengths of the lines had larger effects. Figure 43 shows the effects of alignment and orientation on response time.

Effect	DFn	DFd	F	p	p<.05	Partial ω^2
align	3	123	1.84	.155 [GG]		
orient	1	41	7.86	.008	*	0.0017
sLength	1	41	11.15	.002	*	0.0025
absDif	2	82	89.25	<.001 [GG]	*	0.0419
longerSC	1	41	11.27	.002	*	0.0025
align x orient	3	123	0.99	.396 [GG]		
align x sLength	3	123	2.14	.102 [GG]		
orient x sLength	1	41	0.07	.792		
align x absDif	6	246	1.75	.157 [GG]		
orient x absDif	2	82	1.55	.223 [GG]		
sLength x absDif	2	82	11.23	<.001 [GG]	*	0.0050
align x longerSC	3	123	1.20	.313 [GG]		
orient x longerSC	1	41	2.61	.114		
sLength x longerSC	1	41	0.69	.410		
absDif x longerSC	2	82	6.12	.010 [GG]	*	0.0025
align x orient x sLength	3	123	1.77	.159 [GG]		
align x orient x absDif	6	246	0.27	.869 [GG]		
align x sLength x absDif	6	246	0.68	.600 [GG]		
orient x sLength x absDif	2	82	0.60	.479 [GG]		
align x orient x longerSC	3	123	3.72	.018 [GG]	*	0.0020
align x sLength x longerSC	3	123	0.30	.802 [GG]		
orient x sLength x longerSC	1	41	0.01	.943		
align x absDif x longerSC	6	246	2.01	.109 [GG]		
orient x absDif x longerSC	2	82	0.12	.779 [GG]		
sLength x absDif x longerSC	2	82	1.45	.241 [GG]		
align x orient x sLength x absDif	6	246	1.02	.399 [GG]		
align x orient x sLength x longerSC	3	123	0.40	.732 [GG]		
align x orient x absDif x longerSC	6	246	2.70	.037 [GG]	*	0.0025
align x sLength x absDif x longerSC	6	246	0.67	.600 [GG]		
orient x sLength x absDif x longerSC	2	82	0.14	.792 [GG]		
align x orient x sLength x absDif x longerSC	6	246	0.80	.516 [GG]		

Table 16. Repeated-measures ANOVA table for response time (Experiment 3).

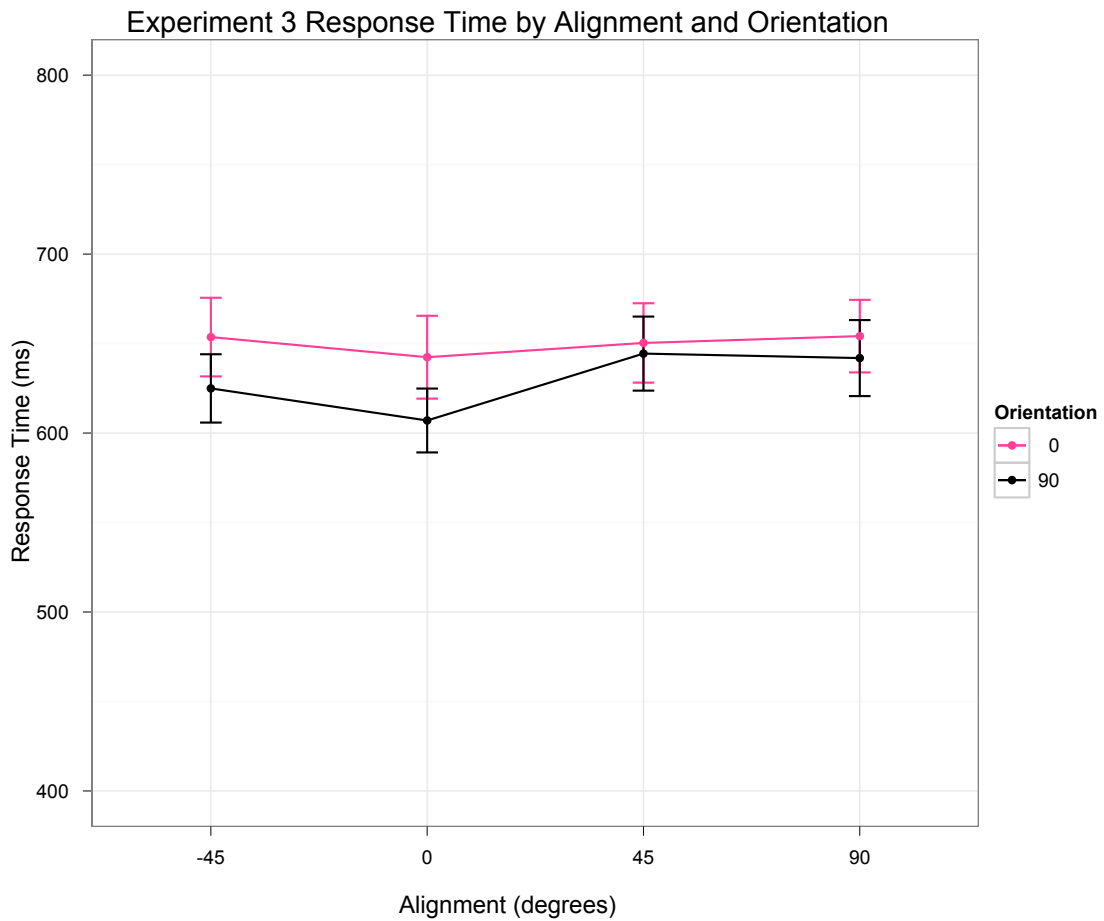


Figure 43. Response Time by Alignment and Orientation (Experiment 3).

The significant effect here is the faster response times for vertical bars, depicted as the black points, than the horizontal lines, depicted as pink points. Alignment and the interaction are not significant.

Absolute and relative differences

These data can be analyzed to test the psychophysical theory in the same way as Experiment 2. Figure 44 shows accuracy plotted by relative and absolute differences. The influence of relative difference is clear, although absolute difference makes a better prediction about the two cases in which the predictions are different.

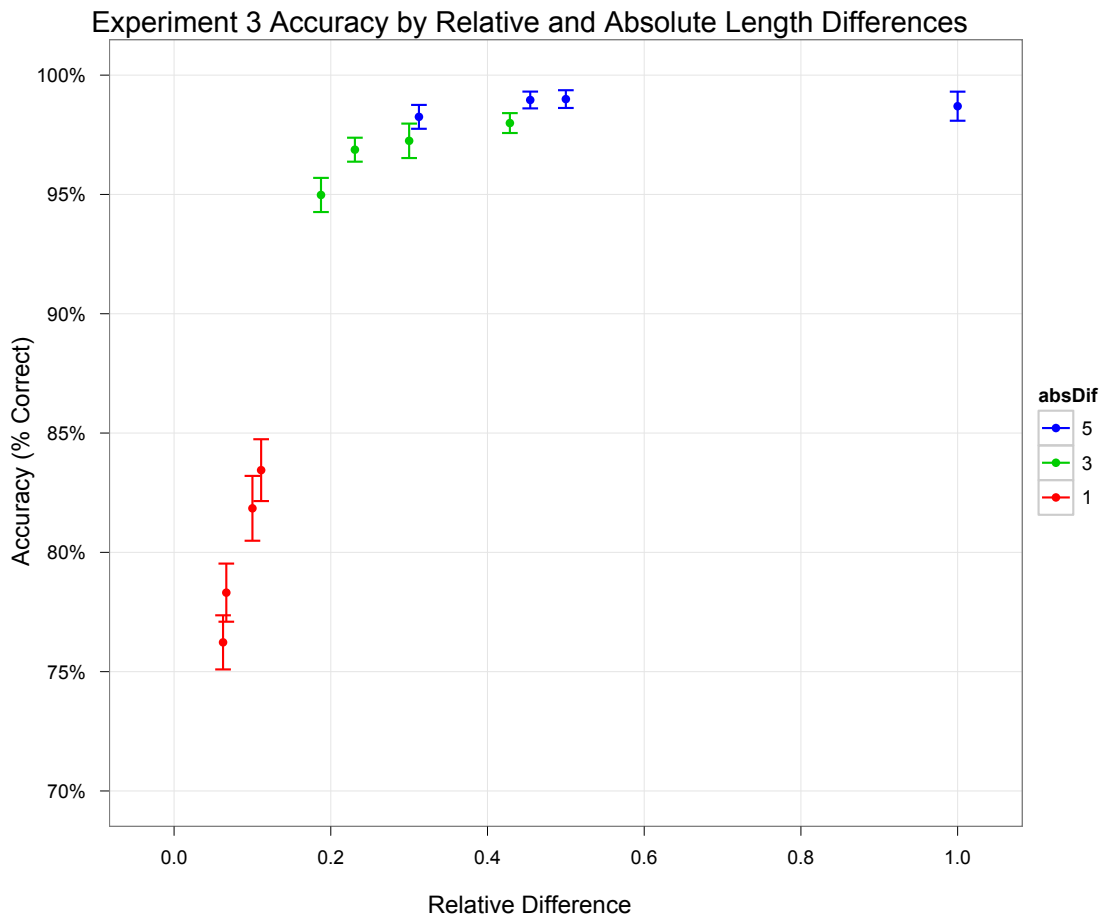


Figure 44. Accuracy by relative and absolute differences in length (Experiment 3).

Figure 45 shows the graph for response time. Again, there is a clear effect of relative differences. The two cases with different predictions have similar response times.

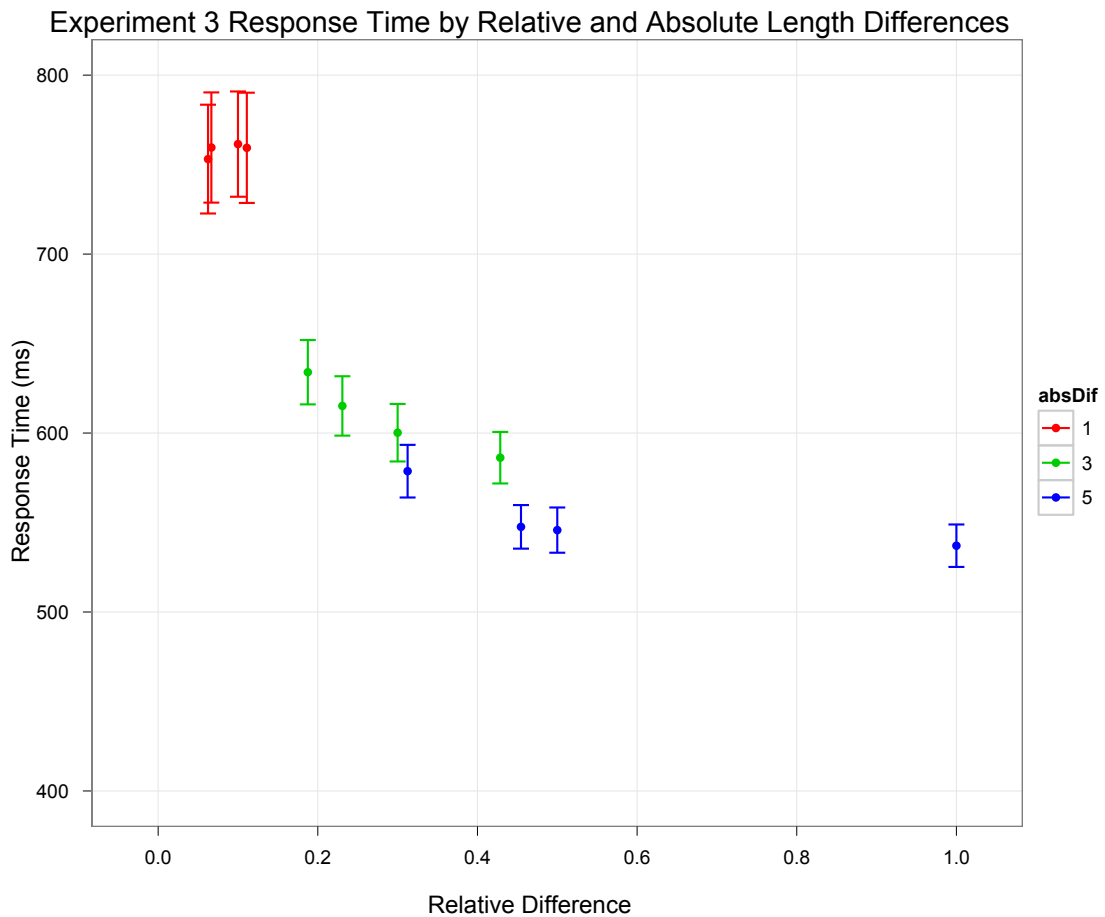


Figure 45. Response Time by relative and absolute differences in length (Experiment 3).

Discussion

Research Question 3-1. How does the alignment of lines affect their comparability?

The effects of alignment on comparability were small. Horizontal alignments led to more correct responses than other alignments, but not by much. There was no significant effect of alignment on response time, suggesting that the large effect seen in experiment 1 (Figure 24) was an artifact of the response procedure.

Research Question 3-2. How does the orientation of lines affect their comparability?

The orientation of the lines did not affect the accuracy of responses, but vertical lines were compared a bit faster than horizontal lines.

Research Question 3-3. How do alignment and orientation interact?

Surprisingly, alignment and orientation were not shown to interact significantly. This suggests that at the 200 pixel distance, for which viewers must move their eyes to compare graphs, the trapezoid effect described by Weber (1834/1978) and Ware (2004) does not play much of a role.

Research Question 3-4. How does the difference in length interact with these effects?

The most difficult comparisons were those of two lines that differed by only a single pixel in length, and these were performed most accurately, and in similar amounts of time, when the graphs are arranged side by side.

Based on these results, graph designers constructing visualizations of multiple bar graphs would likely best communicate information by making the bars vertical, and aligning them horizontally. Considering that the distance effects described in the previous chapter were generally larger than those described here, a compact, squarish arrangement would be generally preferred to a wider arrangement. A tall arrangement would be less preferred still. Of course, depending on the purpose of the graph, other alignments or orientations may be preferable.

Research Question 3-5. Is the absolute or relative difference in lengths a better predictor of comparability?

Both absolute and relative length differences affect comparability. Absolute difference is a slightly better predictor under the conditions measured here.

Chapter 5: Experiment 4

This chapter describes an experiment in which graphs of two lines each were placed at a distance from one another. The participant identified the graph for which the absolute length difference between the two lines was larger. In Experiment 4, the graphs were aligned horizontally, and all of the bars were vertical. The graphs were separated by one of four distances.

This experiment, and Experiment 5 described in the following chapter, were conducted at the same time, although by different participants. The goal was to see whether the patterns of results observed in Experiments 2 and 3 would carry over from comparing lines to comparing graphs. The graphs of two bars each are similar to the one-bar graphs of the previous experiments, and two of them make a simple small multiples graph. Such an arrangement of small graphs could represent a simple true spatiotemporal map representing two places at two times, for example. See Figure 6 for a chartmap with graphs of 3 bars each, and Figure 14 for a more complex example.

Each comparison screen showed four lines, each of which varied by trial as described in the design section below. Due to this larger number of varying lines, I had to reduce the number of levels of other conditions to keep the experiment at a reasonable length. I eliminated the 25 and 50 pixel distances for this reason. These shorter distances were eliminated to make it clear that the two graphs were distinct, as there were no axes or shading around each graph, and because I thought it would be more useful to know about the comparability of distant graphs than very close graphs.

Another simplification I made for this experiment, as well as the Experiment 5, was to have each graph consist of a shorter bar on the left and a longer bar on the right.

Of course this is only one of many possibilities of bar heights, but I wanted to make the task as straightforward as possible for the participants, and to collect many data points on the same conditions to increase the statistical power of the analysis.

Research Questions

Research Question 4-1. How does the distance between bar graphs affect their comparability?

I expected the response time to increase as the distance increased, but I did not know whether accuracy would be strongly affected by distance.

Research Question 4-2. How does the absolute difference in length increases in bar graphs interact with the distance effect?

The term absolute difference in length increases is explained in the design section below. It is roughly the equivalent of the absolute length difference in the previous experiments, the difference that the viewers are asked to judge. The question here is whether the more difficult comparison will be more severely affected by the larger differences than the easier comparison.

Research Question 4-3. What is the effect of the relative heights of lines on the comparability of bar graphs?

The task in this experiment is to compare two graphs and judge which has the larger increase from the left bar to the right bar. The left bar of each graph is of a different height. The question here is whether it is easier to compare graph that have similar base heights, as opposed to dissimilar base heights.

Method

Participants

Thirty-seven students in psychology classes participated for partial class credit as part of the psychology participant pool. There were 15 males and 22 females. The age range was 18 to 26, with a mean of 19.7. Participants were required to have normal or corrected-to-normal vision. No vision tests were performed on the participants, although none complained that they could not see the stimuli.

Apparatus and Materials

The apparatus and seating procedure were exactly the same as those used in Experiments 2 and 3. Each graph consisted of two lines instead of one. Lines were 2 pixels wide, and the two lines within a bar graph were separated by 2 pixels of white space. All lines were vertical, and the lines were aligned at the bottom.

Design

Table 17, below, summarizes the conditions of this experiment, excluding practice trials. The trials were grouped into blocks, with the positions of the two graphs always consistent within a block, and always changing from one block to the next. There was one practice block, followed by 8 test blocks. Excluding the practice block, there were 4 position conditions, each with the two graphs separated in the horizontal direction by a particular number of pixels, as measured from the center of the graph. The distances were 100, 200, 400, and 800 pixels (see Figure 46). In the practice block, the distance

was 300 pixels. There were two blocks of trials for each of the 4 positions. The order of the test blocks was randomized, with counterbalancing that ensured that the same block condition would not appear twice before another appeared once, and that no block condition would appear twice within a run of three blocks.

Each trial had two graphs: a standard graph and a comparison graph. Each graph had a shorter line on the left and a longer line on the right. There were no graphs with decreases or with equally sized lines. The left, base line of the standard graph (abbreviated *sBase*) was either 3 or 11 pixels long. The right line was equal to the base line plus either 8 or 14 pixels, for a total of 4 combinations. This value of 8 or 14 pixels is called the *standard increase*, abbreviated *sInc*. The base line of the comparison graph (*cBase*) was either 4 or 10 pixels long. The right line was equal to the base line plus one of four values (*comparison increase* or *cInc*) shown in Table 17, depending on the standard increase for that trial. The difference between the comparison increase and the standard increase (*csIncDif*) was what the participants were asked to judge; the values were always either -4, -2, 2, or 4. In the ANOVA tables and graphs, *csIncDif* is broken into *csIncAD* (the absolute difference in increase, either 2 or 4 pixels) and *csIncSign* (which increase was larger, equivalent to *longerSC*). The composition of the graphs by *sBase*, *sInc*, *cBase*, and *cInc* is depicted in Figure 47.

Each block consisted of eight practice trials followed by 64 test trials in a random order. The number of practice trials was increased from the previous experiments. This represented two pairings each of all of the combinations of standard and comparison graphs, one with the standard graph in the left position, and one with it in the right position. There were a total of 512 test trials in the experiment.

Between block conditions	Number of levels	Levels
Distance	4	100, 200, 400, 800
Block repetitions	2	
Total number of blocks	8	
Within block conditions	Number of levels	Levels
Standard base length	2	3, 11
Comparison base length	2	4, 10
Standard increase	2	8, 14
Comparison increase	4 per standard increase	8: 4, 6, 10, 12 14: 10, 12, 16, 18
Position of standard	2	left, right
Trial repetitions	1	
Number of trials per block	64	
Total number of trials	512	

Table 17. Conditions for Experiment 4.

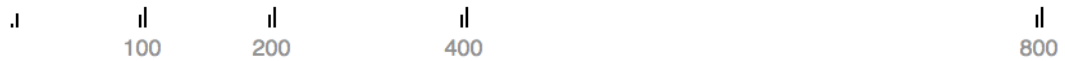


Figure 46. Distance Conditions for Experiment 4.

Each of the four labeled graphs, paired with the unlabeled graph, represents one of the four distance conditions. The lines and distances are to scale. The gray labels represent the distance in pixels between the graphs, measured from the center of each graph, and are for reference only.

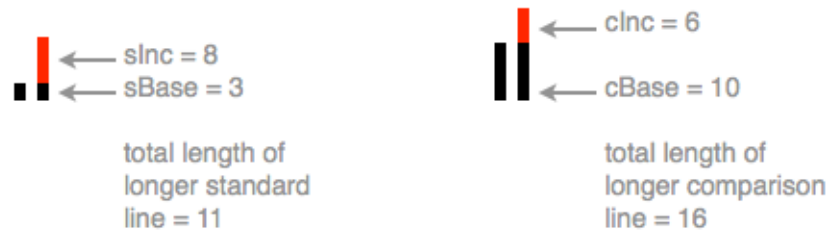


Figure 47. Length Conditions for Experiment 4.

These lines represent two graphs that were presented during a trial. In this example, the left pair is the standard graph, and the right pair, the comparison graph. The left line of each graph is the base, sBase and cBase, respectively, and the right line is longer than the base, by an amount indicated by sInc and cInc, respectively. The red color is shown here to highlight the increase amounts, but the lines in the experiment were all black. The gray labels indicate the number of pixels represented, and are for reference only. In this example, csIncDif is -2, csIncAD = 2, and csIncSign = -1, the standard increase being larger.

Procedure

The procedure was similar to Experiment 2. The participants moved the left thumbstick towards their choice of which graph had the larger increase from the left line to the right line.

Results

Accuracy

Table 18 below summarizes the results of the repeated-measures ANOVA for Experiment 4, using accuracy (proportion correct) as the dependent variable. Most 4-way, 5-way, and the 6-way interaction were not significant; these have been removed

from the table to save space. The conditions of the two blocks and two standard positions were summed, for a total of 4 trials per cell.

Distance was not a significant predictor of response accuracy. Figure 48 shows the effects of distance and csIncAD, which was the strongest predictor of accuracy.

Distance had no effect. The difference in increase between the lines of each graph had an effect of about 10 percentage points. For both conditions of csIncAD, accuracy was lower than in the 3 and 5 pixel conditions of Experiment 2 (Figure 33), but higher than the 1-pixel condition.

Effect	DFn	DFd	F	p	p<.05	Partial ω^2
dist	3	108	0.23	.859 [GG]		
sBase	1	36	0.02	.883		
cBase	1	36	8.13	.007	*	0.0015
sInc	1	36	66.99	<.001	*	0.0137
csIncAD	1	36	222.84	<.001	*	0.0447
csIncSign	1	36	3.1	.087		
dist x sBase	3	108	2.96	.043 [GG]	*	0.0012
dist x cBase	3	108	0.52	.635 [GG]		
sBase x cBase	1	36	45.98	<.001	*	0.0094
dist x sInc	3	108	0.94	.413 [GG]		
sBase x sInc	1	36	6.09	.018	*	0.0011
cBase x sInc	1	36	0.13	.721		
dist x csIncAD	3	108	0.28	.828 [GG]		
sBase x csIncAD	1	36	1.73	.197		
cBase x csIncAD	1	36	0.41	.527		
sInc x csIncAD	1	36	9.73	.004	*	0.0018
dist x csIncSign	3	108	0.20	.869 [GG]		
sBase x csIncSign	1	36	0.79	.380		
cBase x csIncSign	1	36	0.90	.348		
sInc x csIncSign	1	36	1.03	.316		
csIncAD x csIncSign	1	36	6.26	.017	*	0.0011
dist x sBase x cBase	3	108	1.70	.175 [GG]		
dist x sBase x sInc	3	108	0.51	.664 [GG]		
dist x cBase x sInc	3	108	0.14	.918 [GG]		
sBase x cBase x sInc	1	36	1.65	.207		
dist x sBase x csIncAD	3	108	0.91	.426 [GG]		
dist x cBase x csIncAD	3	108	1.37	.259 [GG]		
sBase x cBase x csIncAD	1	36	1.02	.318		
dist x sInc x csIncAD	3	108	0.81	.484 [GG]		
sBase x sInc x csIncAD	1	36	1.76	.193		
cBase x sInc x csIncAD	1	36	1.21	.279		
dist x sBase x csIncSign	3	108	1.96	.131 [GG]		
dist x cBase x csIncSign	3	108	1.89	.139 [GG]		
sBase x cBase x csIncSign	1	36	0.25	.619		
dist x sInc x csIncSign	3	108	0.55	.635 [GG]		
sBase x sInc x csIncSign	1	36	4.45	.042	*	0.0007
cBase x sInc x csIncSign	1	36	6.41	.016	*	0.0011
dist x csIncAD x csIncSign	3	108	0.12	.933 [GG]		
sBase x csIncAD x csIncSign	1	36	0.39	.539		
cBase x csIncAD x csIncSign	1	36	0.13	.720		
sInc x csIncAD x csIncSign	1	36	0.79	.380		
sBase x cBase x sInc x csIncAD	1	36	8.09	.007	*	0.0015
dist x cBase x sInc x csIncSign	3	108	3.88	.012 [GG]	*	0.0018

Table 18. Repeated-measures ANOVA table for accuracy (Experiment 4).

All non-significant 4-way, 5-way, and 6-way interactions have been omitted from this table to save space.

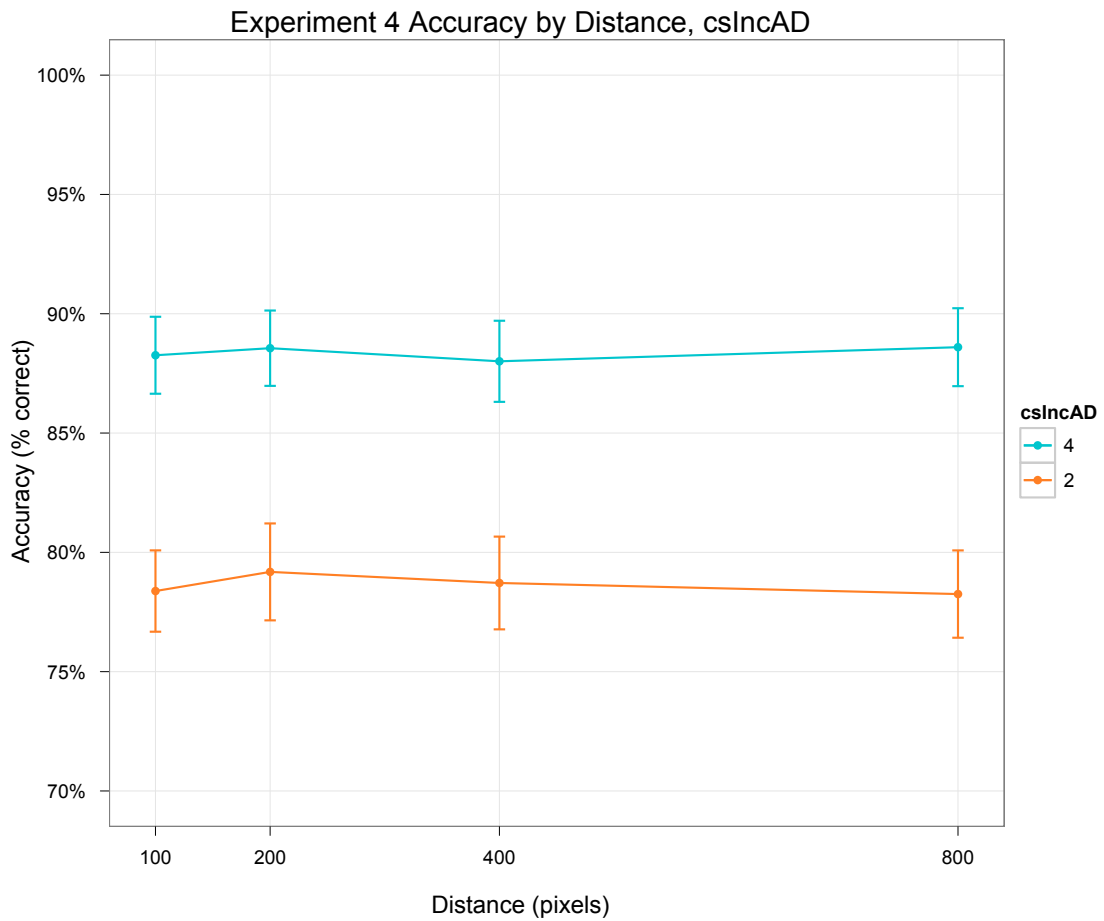


Figure 48. Accuracy by Distance and csIncAD (Experiment 4).

Response Time

Table 19 below summarizes the results of the repeated-measures ANOVA for Experiment 4, using response time as the dependent variable. Most 4-way, 5-way, and the 6-way interaction were not significant; the non-significant ones have been removed from the table to save space. The conditions of the two blocks and two standard positions were summed, for a total of 4 trials per cell.

Effect	DFn	DFd	F	p	p<.05	Partial ω^2
dist	3	108	4.64	.007 [GG]	*	0.0023
sBase	1	36	20.99	<.001	*	0.0042
cBase	1	36	19.82	<.001	*	0.0040
sInc	1	36	0.02	.899		
csIncAD	1	36	26.91	<.001	*	0.0054
csIncSign	1	36	0.07	.790		
dist x sBase	3	108	1.52	.218 [GG]		
dist x cBase	3	108	2.43	.090 [GG]		
sBase x cBase	1	36	0.48	.495		
dist x sInc	3	108	0.05	.981 [GG]		
sBase x sInc	1	36	0.03	.869		
cBase x sInc	1	36	1.37	.249		
dist x csIncAD	3	108	1.63	.204 [GG]		
sBase x csIncAD	1	36	0.04	.837		
cBase x csIncAD	1	36	0.07	.800		
sInc x csIncAD	1	36	0.56	.461		
dist x csIncSign	3	108	0.41	.724 [GG]		
sBase x csIncSign	1	36	0.82	.372		
cBase x csIncSign	1	36	1.46	.235		
sInc x csIncSign	1	36	0.00	.979		
csIncAD x csIncSign	1	36	0.03	.862		
dist x sBase x cBase	3	108	1.70	.187 [GG]		
dist x sBase x sInc	3	108	1.88	.145 [GG]		
dist x cBase x sInc	3	108	0.30	.798 [GG]		
sBase x cBase x sInc	1	36	0.19	.662		
dist x sBase x csIncAD	3	108	0.35	.759 [GG]		
dist x cBase x csIncAD	3	108	1.16	.326 [GG]		
sBase x cBase x csIncAD	1	36	5.04	.031	*	0.0009
dist x sInc x csIncAD	3	108	0.83	.460 [GG]		
sBase x sInc x csIncAD	1	36	0.41	.527		
cBase x sInc x csIncAD	1	36	0.15	.703		
dist x sBase x csIncSign	3	108	1.02	.379 [GG]		
dist x cBase x csIncSign	3	108	1.87	.148 [GG]		
sBase x cBase x csIncSign	1	36	0.00	.979		
dist x sInc x csIncSign	3	108	2.03	.124 [GG]		
sBase x sInc x csIncSign	1	36	1.78	.190		
cBase x sInc x csIncSign	1	36	0.94	.339		
dist x csIncAD x csIncSign	3	108	0.92	.416 [GG]		
dist x sBase x cBase x sInc x csIncSign	3	108	4.29	.010 [GG]	*	0.0021

Table 19. Repeated-measures ANOVA table for response time (Experiment 4).

All non-significant 4-way, 5-way, and 6-way interactions have been omitted from this table to save space.

Although distance was a significant predictor of response time, the effect was not monotonic. Figure 49 shows that a distance of 200 pixels produced the fastest responses. The response time difference between the two conditions of csIncAD was about 100 ms.

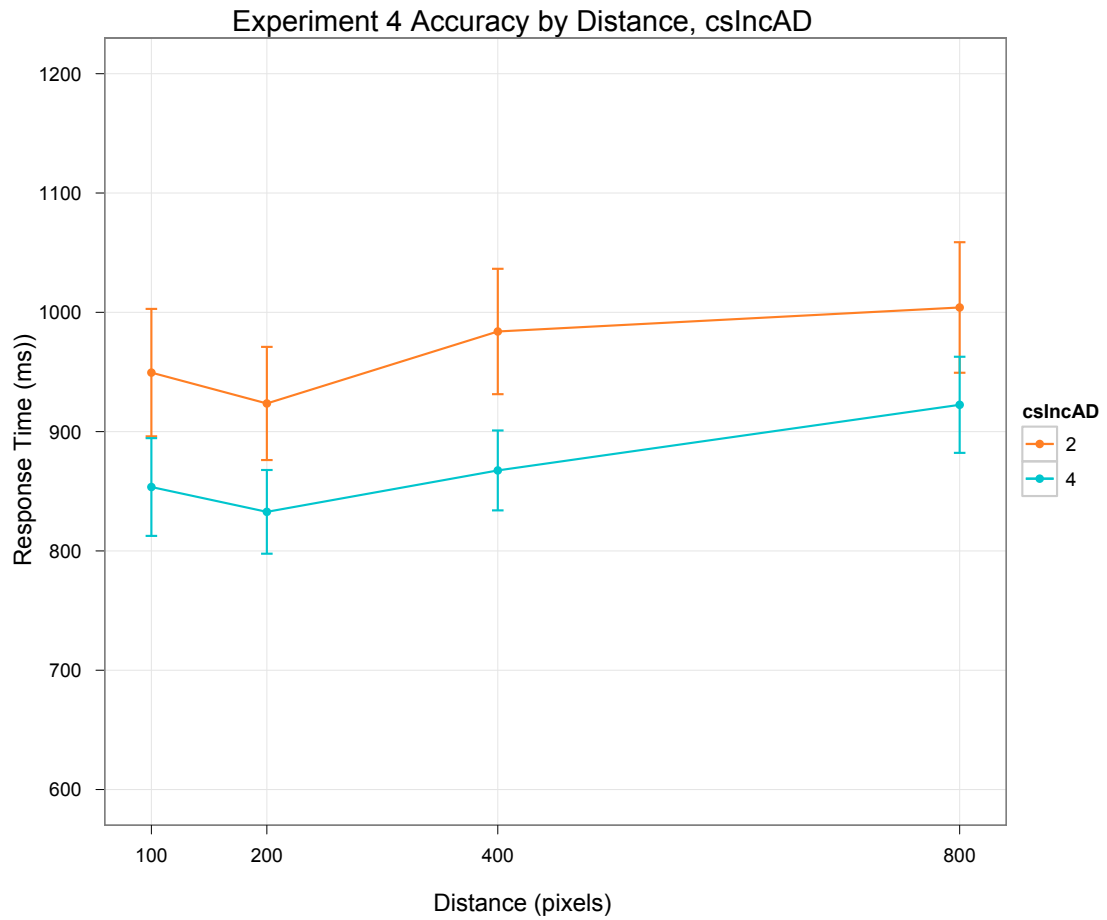


Figure 49. Response Time by Distance and csIncAD, (Experiment 4).

Baseline Heights

Participants in this experiment were asked to compare small graphs, not just individual lines. In some cases, the left line of each graph was of a similar height (differing by 1 pixel), and in some cases those lines differed by 7 pixels. It would be worthwhile to know how this difference affected the comparability of the graphs. Recall that we are assuming that all four pieces of information are necessary in this graph, and that reducing the graphs to show only a single bar depicting the change would cause a loss of otherwise necessary information contained in the baseline heights.

The variables we are interested in here are sBase and cBase, the lengths of the standard and comparison graph base lines. Using accuracy as the measure, the interaction of these effects is significant, with similar sBase and cBase values (4 and 3 or 11 and 10, respectively) leading to more accurate responses (Figure 50). Conditions with the smaller cBase value were also more accurate than those with a larger cBase value, although this was a smaller effect.

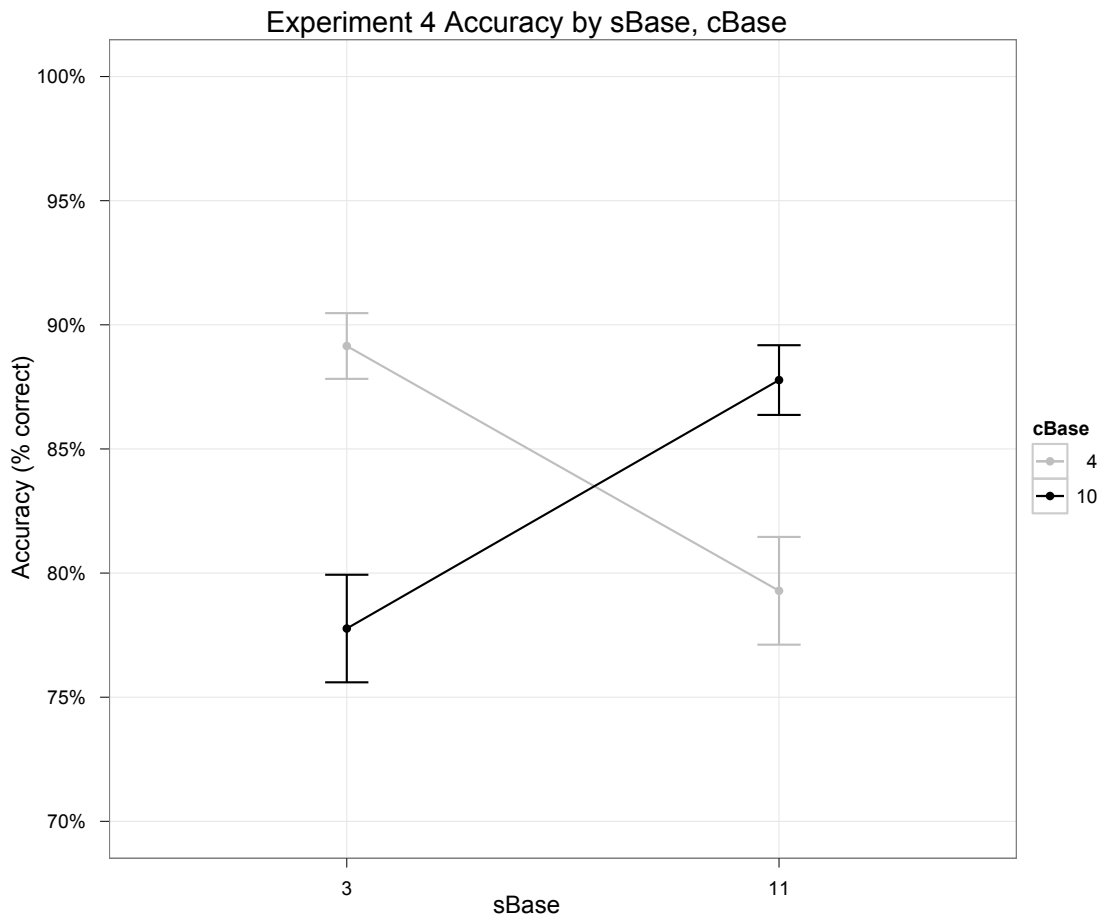


Figure 50. Accuracy by sBase, cBase (Experiment 4).

Using response time as the measure, there is no interaction between the sBase and cBase variables, but each variable is a significant predictor (Figure 51). In both cases, smaller baselines lead to faster responses.

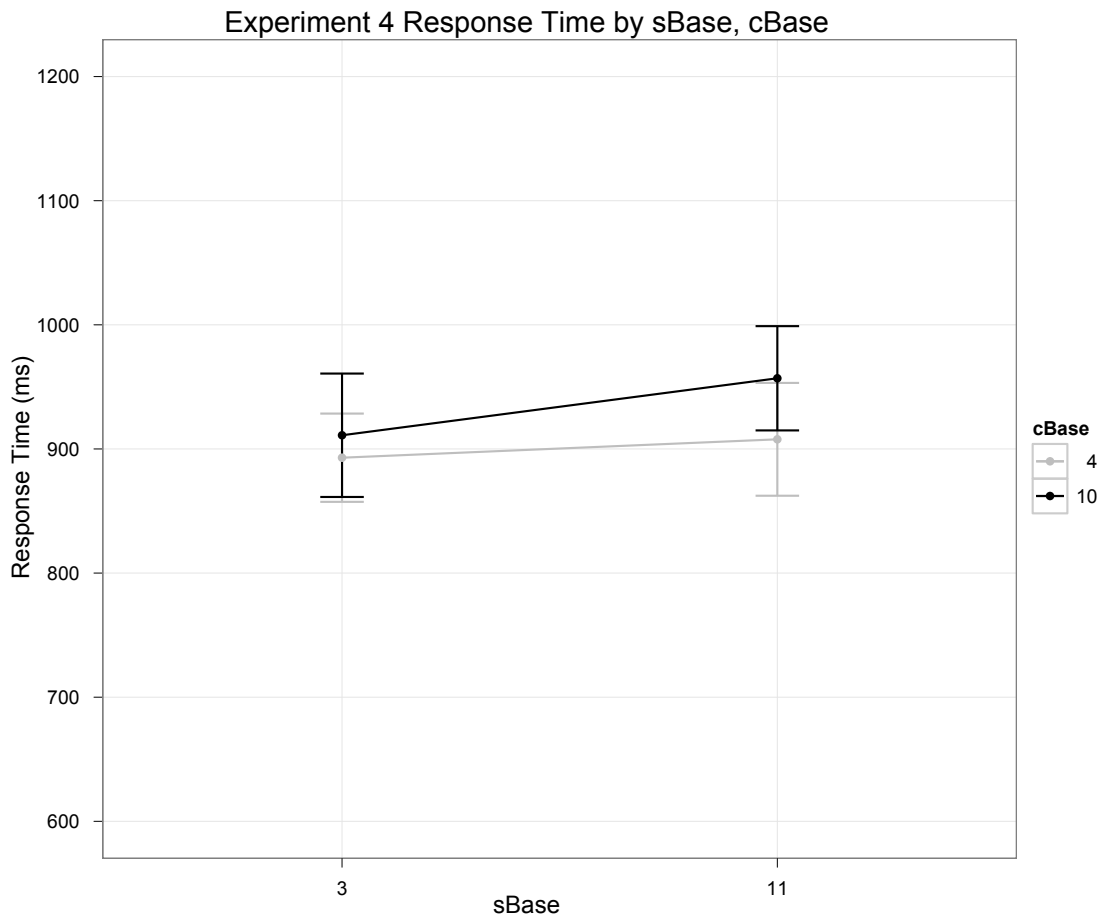


Figure 51. Response Time by sBase and cBase (Experiment 4).

Although the differences are small, response time increased significantly between the smaller and larger base line heights for each graph, with no significant interaction.

Discussion

Research Question 4-1. How does the distance between bar graphs affect their comparability?

Distance had little to no effect on the comparability of these bar graphs. It took slightly longer to respond when the graphs were 400 or 800 pixels apart than when they were 200 pixels apart or closer, but accuracy was not affected.

Research Question 4-2. How does the absolute difference in length increases in bar graphs interact with the distance effect?

As expected, the similarity of the graphs in terms of their increase, which was what was being compared, affected the accuracy and response time. More similar pairs of graphs were more difficult to compare.

Research Question 4-3. What is the effect of the relative heights of lines on the comparability of bar graphs?

Graphs with similar base line heights were compared with greater accuracy, by about 10 percentage points. There were also slight comparability advantages to smaller base line heights, suggesting effects of relative length differences.

These graphs were more difficult to compare than single lines, but this difficulty was not compounded by increases in distance. Provided the graph maker can ensure that the relevant information can be read and compared, he or she need not worry that two graphs will not be comparable because they are not near one another.

Chapter 6: Experiment 5

This chapter describes an experiment in which graphs of two lines each were placed at a distance to one another. The participant identified the graph for which the absolute length difference between the two lines was larger. In Experiment 5, the graphs were arranged in one of four alignments, the same as in Experiment 3. Unlike Experiment 3, all lines were oriented vertically.

Why study alignment and not orientation in this experiment, when Experiment 3 showed that both orientation and alignment have some effect on the comparability of lines? First, alignment is a more essential decision for the small multiple graph designer to make than orientation. Orientation is a decision about the small graph, but alignment is a decision about the larger graph. In other words, the knowledge gained about alignment is more likely to be broadly applicable to arrangements of small multiples, whereas the orientation knowledge is less likely to. Second, Experiment 3 established that vertically oriented lines were slightly more comparable than horizontally oriented bars across all alignments, as measured by response time, with no significant differences found for accuracy. Thus there is no compelling need to study small horizontally oriented bar graphs. Third, as in Experiment 4, the number of conditions had to be reduced for time considerations.

Research Questions

Research Question 5-1. How does the alignment of bar graphs affect their comparability?

In Experiment 3, lines were more comparable, as measured by accuracy, when they were side by side, as opposed to aligned diagonally or one over the other. It would be useful to know whether this effect persists for small graphs.

Research Question 5-2. How does the absolute difference in length increases in bar graphs interact with the alignment effect?

In Experiment 3, there was a significant interaction between the absolute difference between the lines and the alignment, as measured by accuracy. It would be useful to know whether this effect persists for small graphs.

Research Question 5-3. What is the effect of the relative heights of lines on the comparability of bar graphs?

This is the same as Research Question 4-3, and I therefore expect similar results.

Method

Participants

Thirty-eight students in psychology classes participated for partial class credit as part of the psychology participant pool. There were 16 males and 22 females. The age range was 18 to 23, with a mean of 19.7. Participants were required to have normal or corrected-to-normal vision. No vision tests were performed on the participants, although none complained that they could not make out the stimuli.

Apparatus and Materials

The apparatus and seating procedure were exactly the same as those used in Experiments 2, 3, and 4. The materials were similar to those used in Experiment 4.

Design

Table 20, below, summarizes the conditions for Experiment 5. The design was exactly the same as that of Experiment 4, except that there were four alignment conditions instead of four orientation conditions. The alignment conditions were the same as those in experiment 3, and are depicted in Figure 52. All pairs of graphs were separated by 200 pixels, as measured by the center of each graph.

Between block conditions	Number of levels	Levels
Alignment	4	-45, 0, 45, 90
Block repetitions	2	
Total number of blocks	8	
Within block conditions	Number of levels	Levels
Standard base length	2	3, 11
Comparison base length	2	4, 10
Standard increase	2	8, 14
Comparison increase	4 per standard increase	8: 4, 6, 10, 12 14: 10, 12, 16, 18
Position of standard	2	left, right
Trial repetitions	1	
Number of trials per block	64	
Total number of trials	512	

Table 20. Conditions for Experiment 5.



Figure 52. Alignment Conditions for Experiment 5.

Each of the four labeled graphs, paired with the unlabeled graph, represents one of the four alignment conditions. The lines and distances are to scale. The gray labels show the alignment angle, and are for reference only. There were no horizontal lines in Experiment 5.

Procedure

The procedure was similar to that in Experiments 3 and 4. The participants moved the left thumbstick in the direction of the graph with the larger increase.

Results

Accuracy

Table 21 below summarizes the results of the repeated-measures ANOVA for Experiment 4, using accuracy as the dependent variable. Most 4-way, 5-way, and the 6-way interaction were not significant; the non-significant ones have been removed from the table to save space. The conditions of the two blocks and two standard positions were summed, for a total of 4 trials per cell. The main effect of alignment was not significant, nor were any 2-way interactions including alignment. As in the experiment 4, the largest effect was that of csIncAD, the difference being compared. Figure 53 shows the effects of alignment and csIncAD.

Effect	DFn	DFd	F	p	p<.05	Partial ω^2
align	3	111	1.77	.169 [GG]		
sBase	1	37	0.41	.525		
cBase	1	37	16.91	<.001	*	0.0033
sInc	1	37	148.86	<.001	*	0.0295
csIncAD	1	37	451.05	<.001	*	0.0847
csIncSign	1	37	23.13	<.001	*	0.0045
align x sBase	3	111	0.91	.427 [GG]		
align x cBase	3	111	1.43	.240 [GG]		
sBase x cBase	1	37	34.43	<.001	*	0.0068
align x sInc	3	111	0.32	.792 [GG]		
sBase x sInc	1	37	4.77	.035	*	0.0008
cBase x sInc	1	37	0.22	.639		
align x csIncAD	3	111	0.41	.732 [GG]		
sBase x csIncAD	1	37	0.23	.636		
cBase x csIncAD	1	37	7.61	.009	*	0.0014
sInc x csIncAD	1	37	29.30	<.001	*	0.0058
align x csIncSign	3	111	0.23	.861 [GG]		
sBase x csIncSign	1	37	0.01	.918		
cBase x csIncSign	1	37	0.02	.881		
sInc x csIncSign	1	37	4.15	.049	*	0.0006
csIncAD x csIncSign	1	37	0.00	.969		
align x sBase x cBase	3	111	2.21	.096 [GG]		
align x sBase x sInc	3	111	1.37	.258 [GG]		
align x cBase x sInc	3	111	1.22	.306 [GG]		
sBase x cBase x sInc	1	37	0.42	.523		
align x sBase x csIncAD	3	111	0.99	.401 [GG]		
align x cBase x csIncAD	3	111	2.94	.041 [GG]	*	0.0012
sBase x cBase x csIncAD	1	37	4.48	.041	*	0.0007
align x sInc x csIncAD	3	111	1.64	.189 [GG]		
sBase x sInc x csIncAD	1	37	0.61	.442		
cBase x sInc x csIncAD	1	37	0.00	.999		
align x sBase x csIncSign	3	111	0.33	.789 [GG]		
align x cBase x csIncSign	3	111	0.73	.525 [GG]		
sBase x cBase x csIncSign	1	37	0.04	.845		
align x sInc x csIncSign	3	111	0.65	.578 [GG]		
sBase x sInc x csIncSign	1	37	4.79	.035	*	0.0008
cBase x sInc x csIncSign	1	37	4.12	.050	*	0.0006
align x csIncAD x csIncSign	3	111	0.45	.700 [GG]		
sBase x csIncAD x csIncSign	1	37	1.50	.228		
cBase x csIncAD x csIncSign	1	37	4.25	.046	*	0.0007
sInc x csIncAD x csIncSign	1	37	0.86	.359		
align x cBase x sInc x csIncAD x csIncSign	3	111	2.80	.049 [GG]	*	0.0011

Table 21. Repeated-measures ANOVA table for accuracy (Experiment 5).

All non-significant 4-way, 5-way, and 6-way interactions have been omitted from this table to save space.

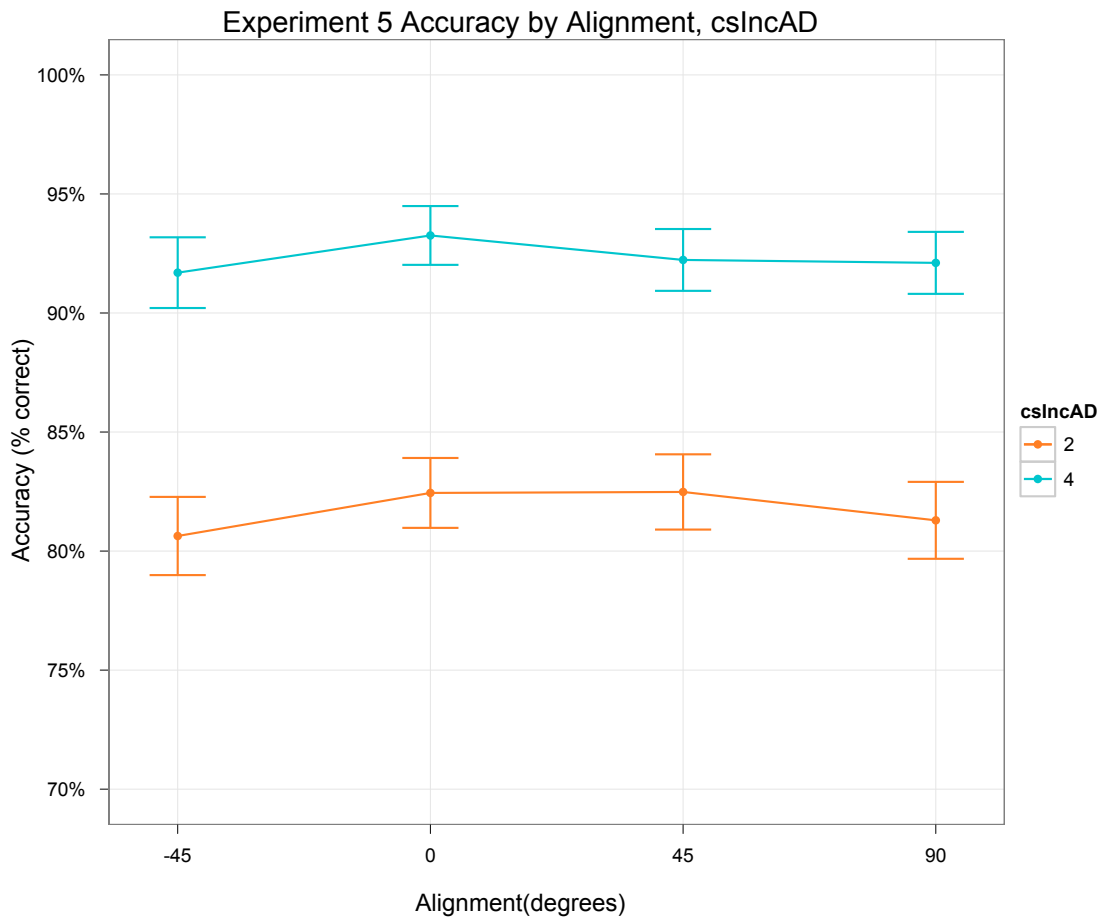


Figure 53. Accuracy by Alignment and csIncAD (Experiment 5).

Response Time

Table 22 below summarizes the results of the repeated-measures ANOVA for Experiment 4, using response time as the dependent variable. Most 4-way, 5-way, and the 6-way interaction were not significant; the non-significant ones have been removed from the table to save space. The conditions of the two blocks and two standard positions were summed, for a total of 4 trials per cell. Again alignment is not a significant effect, nor are its interactions. Figure 54 shows the effects of alignment and csIncAD.

Effect	DFn	DFd	F	p	p<.05	Partial ω^2
align	3	111	1.30	.279 [GG]		
sBase	1	37	16.09	<.001	*	0.0031
cBase	1	37	12.74	.001	*	0.0024
sInc	1	37	2.43	.128		
csIncAD	1	37	69.72	<.001	*	0.0139
csIncSign	1	37	2.57	.118		
align x sBase	3	111	1.81	.154 [GG]		
align x cBase	3	111	1.86	.155 [GG]		
sBase x cBase	1	37	2.26	.142		
align x sInc	3	111	2.69	.057 [GG]		
sBase x sInc	1	37	0.13	.725		
cBase x sInc	1	37	0.21	.653		
align x csIncAD	3	111	1.88	.142 [GG]		
sBase x csIncAD	1	37	0.00	.987		
cBase x csIncAD	1	37	0.01	.913		
sInc x csIncAD	1	37	2.29	.139		
align x csIncSign	3	111	0.49	.649 [GG]		
sBase x csIncSign	1	37	1.46	.235		
cBase x csIncSign	1	37	2.37	.132		
sInc x csIncSign	1	37	0.57	.455		
csIncAD x csIncSign	1	37	0.60	.444		
align x sBase x cBase	3	111	0.65	.546 [GG]		
align x sBase x sInc	3	111	1.25	.296 [GG]		
align x cBase x sInc	3	111	0.51	.613 [GG]		
sBase x cBase x sInc	1	37	13.14	.001	*	0.0025
align x sBase x csIncAD	3	111	0.32	.774 [GG]		
align x cBase x csIncAD	3	111	0.88	.448 [GG]		
sBase x cBase x csIncAD	1	37	9.13	.005	*	0.0017
align x sInc x csIncAD	3	111	1.71	.174 [GG]		
sBase x sInc x csIncAD	1	37	0.06	.811		
cBase x sInc x csIncAD	1	37	0.00	.995		
align x sBase x csIncSign	3	111	0.73	.508 [GG]		
align x cBase x csIncSign	3	111	0.15	.903 [GG]		
sBase x cBase x csIncSign	1	37	0.18	.670		
align x sInc x csIncSign	3	111	0.64	.583 [GG]		
sBase x sInc x csIncSign	1	37	0.45	.507		
cBase x sInc x csIncSign	1	37	0.15	.699		
align x csIncAD x csIncSign	3	111	2.08	.131 [GG]		
sBase x csIncAD x csIncSign	1	37	1.59	.216		
cBase x csIncAD x csIncSign	1	37	1.37	.250		
sInc x csIncAD x csIncSign	1	37	2.45	.126		
sBase x cBase x sInc x csIncSign	1	37	7.22	.011	*	0.0013
sBase x cBase x csIncAD x csIncSign	1	37	6.58	.014	*	0.0011

Table 22. Repeated-measures ANOVA for Response Time (Experiment 5).

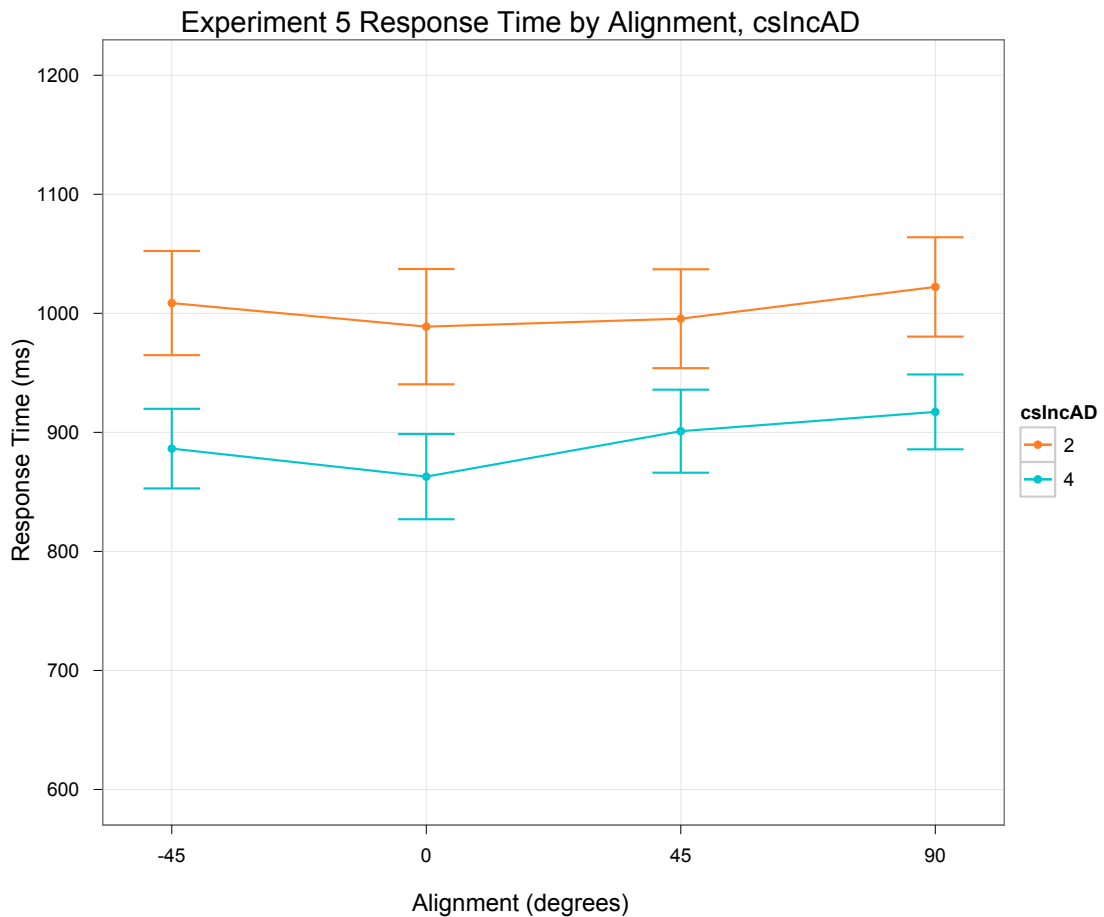


Figure 54. Response Time by Alignment and csIncAD (Experiment 5).

Discussion

Research Question 5-1. How does the alignment of bar graphs affect their comparability?

The alignment of small bar graphs did not significantly affect their comparability in any way, under the conditions of this experiment. All of the graphs were separated by 200 pixels; it is not possible to rule out the possibility that alignment would play some role in comparing graphs that were closer together.

Research Question 5-2. How does the absolute difference in length increases in bar graphs interact with the alignment effect?

There was no evidence of an interaction between alignment and any kind of length difference.

Research Question 5-3. What is the effect of the relative heights of lines on the comparability of bar graphs?

The effects are similar to those seen in Experiment 4. It was easier to compare graphs for which the critical length increase was greater, and easier to compare graphs with more similar base lengths.

Just as we saw that the effects of distance were more subtle when 2-line graphs were compared than when single lines were compared, the effects of alignment are more subtle, so much so that they are not significant here. The constructor of a graph containing multiple small bar graphs need not be concerned about how the alignment of those graphs will affect their comparability.

Chapter 7: General Discussion

This chapter summarizes the results of the five experiments and describes the limitations of the research. This is followed by suggestions of ways in which graph designers might use this knowledge to construct better graphs, and by discussion of how future research can answer remaining questions about the comparability of small graphs.

Summary of Findings

The general research questions are restated below, and the answers are grouped by topic in three sections: *Distance, Alignment and Orientation, and Complexity of Graphs*. The summaries highlight the results of the later experiments when these conflict with those of Experiment 1.

Research Question G-1. How is the comparability of lines affected by the distance between the lines? See *Distance*, below.

Research Question G-2. How is the comparability of lines affected by the alignment of the lines? See *Alignment and Orientation*, below.

Research Question G-3. How is the comparability of lines affected by the orientation of the lines? See *Alignment and Orientation*, below.

Research Question G-4. How is the comparability of lines affected by the interaction of the alignment and orientation of the lines? See *Alignment and Orientation*, below.

Research Question G-5. How is the comparability of small bar graphs affected by the factors described in questions 1, 2, 3, and 4? See *Distance, Alignment and Orientation, and Complexity of Graphs*, below.

Distance

Larger distances between lines decreased the comparability of those lines, measured by both the time taken to compare the lines and the accuracy of those comparisons. In terms of response time, the most substantial effect was the difference between bars at 50 pixels (about 0.72 degrees of visual angle) and 100 pixels (about 1.4 degrees of visual angle). Notably, for distances larger than 100 pixels, response time increased at a lower rate than did distance (see Figure 34). So we would not expect there to be some distance at which the information in a graph displayed on a computer screen could not be understood because it took too long for people to compare lines.

The accuracy of the comparisons was also affected by distance, although this effect was strongly influenced by the difference in length of the two lines. Very similar lines (those differing by 1 pixel, or .014 degrees of visual angle) could not be compared with an acceptable degree of accuracy when they were far enough apart that both could not be seen with central vision at the same time. Lines differing by 3 pixels or 5 pixels could be accurately compared, even at 800 pixels distance, with only mild effects of distance.

Neither of these effects were seen when 2-line graphs, rather than single lines, were compared. The distance between the graphs had no effect on the accuracy of the comparisons, and only slight, non-monotonic effects on the response time (see Figure 49). The fastest responses were made to graphs 200 pixels distant, not 100 pixels distant.

These results are consistent with the model of two visual memory systems, a fast, accurate iconic memory with a longer-term visual working memory that is prone to interference (Baddeley and Hitch, 1974, Phillips, 1974, Baddeley and Andrade, 2000).

When two lines are compared simultaneously (because they require no saccades or only tiny saccades) they can be compared accurately even if they are quite similar in length. This iconic memory fades as the saccades get longer, leading to increased interference and lower accuracy for comparisons of very similar lines. For less similar lines, the visual working memory is sufficient to maintain a representation of the two lines that allows accurate comparisons. Increased saccade time does not provide much further interference, so accuracy does not degrade sharply at longer distances.

For comparisons of graphs of two lines, the lines of each graph can always be compared simultaneously, but not the two graphs to each other, because they fall outside the 50-pixel range that facilitates these very fast comparisons. Thus accuracy is not much affected by distance for Experiment 4. In fact, response time was not much affected by distance either, suggesting that the slower time course in Experiment 2 is due to speed-accuracy trade off decisions made by the participants, not long saccade times. This would not be surprising given that saccades are very fast.

Alignment and Orientation

The effects of the alignment (how graphs were arranged relative to one another, side-by-side, one over the other, or diagonally) and orientation of bars (vertical or horizontal) at 200 pixels distance were quite subtle. People responded more accurately to bars aligned horizontally (graphs placed side-by-side) than to bars aligned vertically or diagonally (see Figure 41). This difference was most prominent for the most difficult comparisons, in which the lines differed by only 1 pixel. People responded a bit faster to

vertically oriented bars than horizontally oriented bars (see Figure 43), but with no significant difference in accuracy.

Alignment had no effect on the comparability of 2-line graphs by either measure, and orientation was not tested for these graphs.

Complexity of Graphs

What conclusion can we draw about how people compare small graphs, as opposed to comparing individual lines? It seems that something about the small graphs greatly reduces or eliminates the effects of distance and alignment. In that sense, small graphs might have a kind of protective effect on the information they depict, allowing it to be compared as part of a large or complex graph.

However, the overall accuracy and response times indicated that the small graphs were less comparable than corresponding single lines. For example, response accuracy for the 2-line graphs that differed in increase by 4 pixels was about 88%, worse than the response accuracy for single lines that differed in length by 3 pixels. The response times were also slower in experiments 4 and 5 than in 2 and 3 across the board.

It's worth remembering that these differences were expected. The 2-line graphs are more complex than the single-line graphs, and we are assuming that multiple lines are displayed, representing multiple data points per graph, because these data points, and not just the length increase between them, are potentially important information for the graph to convey. What is interesting here is how distance and alignment affect the different kinds of graphs differently.

Limitations

This study was conducted using a limited range of young student participants with normal or corrected-to-normal eyesight, doing repetitive tasks, devoid of context, in a small, windowless room for no pay. Only a small fraction of the potential variables were explored, and only a few conditions were tested among those variables. It would be foolish to assume that the specific outcomes will be applicable to all persons using all media in all situations. Graph designers should use common sense and test their designs on potential users. In the following section on practical advice, I have tried to frame the advice in general terms, not pixel-by-pixel instructions of how to draw any conceivable small multiple bar graph.

Practical Advice for Graph Makers

An aim of this project was that the results be directly relatable to the task of constructing small multiple graphs. Although the evidence gathered here comes only from bar graphs, it is likely that similar principles apply to other types of graphs. See later sections for a discussion of how future research can expand the graph difficulty principle for small multiples.

The first piece of advice people constructing small multiples graphs can take from this project is to be sure that if two individual elements must be compared to make a judgment, that those elements must be comparable with near perfect accuracy when they are positioned very close, or else their comparability will rapidly deteriorate when they are farther apart (see Figure 33). In Experiment 2, lines that differed by 3 pixels, about

.043 degrees of visual angle, were compared accurately over 90% of the time, even when they were 800 pixels apart.

Second, the most accurate comparisons are those made of elements that are very close to one another, within about half a degree of visual angle apart, (25 pixels, or .036 degrees of visual angle in Experiment 2). Even the tiniest differences were correctly seen nearly 90% of the time when they were this close. That said, if the differences are not tiny, distance has much less effect on accuracy, so there is no need for the designer to try to cram the elements close together. In general, it would be much better to make the elements distinct, and then include as many of them as necessary at whatever distances are appropriate to the situation.

Third, designers should remember that many small factors affect the comparability of graphs. It was easier to compare graphs for which the left line was of a similar height than those with left lines of different heights. Both the absolute and relative differences of the lengths of lines affect the comparability of the lines. The practical advice here is to make the graphs as similar as possible, except in ways that are relevant to the kinds of comparisons you expect people will do.

Fourth, the best way to arrange small bar graphs to make the bars as comparable as possible is to place them side by side, with the bars vertical. This advice is based on the single bars compared in Experiment 3. In Experiment 5, there was no significant effect of alignment on the comparability of 2-line graphs.

Directions for Future Research

The research described here was intended as a starting point, not an ending point, for determining a graph difficulty principle for small multiples. The small graphs, larger graphs, and tasks are all simplified as much as possible. There are a few main branches of future research that could prove beneficial to the development of a graph difficulty principle for small multiples:

Break the response time into its component parts

It would be possible to deconstruct the amount of time taken to compare distant graphs in the task presented here into its component parts of moving the gaze from one graph to another and possibly to other parts of the screen, and staring at each graph to extract information from it. From this it could be possible to determine why comparability decreases as it does under different circumstances. For example, it might be that the long eye movements interfere with the memory of one graph, or it might be that the time taken during that eye movement interferes with the memory, and clever experimental design could distinguish between these sources of error somewhat. This could also be done using an eye tracking system.

Maximize the comparability of graphs like the ones studied here

This is more promising than trying to reinvent vision research. Consider the problem of the comparability of 2-line graphs. These graphs were often less comparable than single lines under similar conditions of distance and alignment. In Experiment 1 (Figure 27, Figure 28), plain gray axes presented alongside small graphs did not seem to

interfere with the comparison of the graphs. Might informative gridlines or selective use of color, as suggested by Tufte (1983, 1990) for instance, help make small bar graphs more comparable? It might be the case that the relatively poor comparability of 2-line graphs can be alleviated, making small multiple bar graphs a more appealing practical option. Comparisons of different corrective measures could be quite practical.

Expand empirical studies to more complex graphs

I think a fruitful path for this project would be to increase our understanding of what makes small graphs readily comparable. This is where I would prefer to see this project go, along with some related maximizing studies. The world of complex data graphics is still growing, and it still needs guidance. Some suggested studies follow.

Comparisons of more than two graphs, and of complex graphs

The purpose of small multiples is to depict complex information, often representing different times, places, and other variables. With increased complexity of information comes increased complexity in integrating that information. The psychophysical methods used here may not work when more graphs are added because they rely on participants making simple, quick decisions. Already Experiments 4 and 5 pushed the limits of this simplicity with four bars. I doubt that they would work to effectively integrate even three graphs with three lines each, let alone something approaching the chartmap in Figure 14.

Comparisons of other types of graphs

This project was limited to bar graphs for simplicity. But many other types of graphs are popular. While I suspect that the basic lessons from this project would apply to other kinds of graphs, there might be some details about small multiple line graphs, scatterplots, or even pie charts that could be useful. For example, it can often be difficult to compare scatterplots. If the information in scatterplots could be drawn so as to make comparisons quick and accurate, then we would expect, based on the findings of the present research, that the large displays of scatterplots used to illustrate intercorrelations would be more interpretable.

Comparisons of Spatiotemporal Maps

The motivation for this project was spatiotemporal information, and how best to display it. The basic questions about how best to display this kind of information described by Andrienko and Andrienko (2006) still need to be answered. For instance, when a reader wants to understand the complex nature of a spatiotemporal data set, including the most complex questions of trends of distributions and distributions of trends, what kinds of small graphs (or maps) and larger graphs (or maps) best facilitate this kind of understanding? Do different types of questions require different types of visualizations, or might a very good chartmap or map array help a graph reader in unexpected ways? Of course, for information this complex, and interactive system can facilitate these queries, but what base visualizations will prove inviting to new users?

Final Thoughts

This research will not, by itself, relieve the tension between the desire for complex, data rich graphs, and the desire for simple, easy to comprehend graphs (Bertin 1967/1983; Pinker, 1990; Shneiderman, 1996; Tufte, 1983, 1990). I do believe that the results here can be interpreted as support for the idea, expressed and interpreted differently, but common to these theorists, that for graphs to be maximally useful, the information must be instantly perceptible, and that if they are, the mere size of a graph does not prevent it from being useful as a whole.

I also see it as part of a positive trend away from HCI research intended to prove that a particular complex system is better than another particular complex system, and towards a new kind of HCI research grounded in the tradition of basic research, trying to test prominent design principles and inform new ones.

References

- Andrienko, N. & Andrienko, G. (2006). *Exploratory Analysis of Spatial and Temporal Data*. Berlin: Springer.
- Andrienko, N., & Andrienko, G. (2004). Interactive visual tools to explore spatio-temporal variation. In *Proceedings of the working conference on Advanced visual interfaces. System papers: exploration and retrieval tools* (pp. 417-420). New York: ACM.
- Andrienko, G. & Andrienko, N. (1999). Interactive Maps for Visual Data Exploration. *International Journal of Geographic Information Science* 13 (4).
- Andrienko, N, Andrienko, G., & Gatalsky, P. (2003). *Exploratory spatio-temporal visualization: an analytical review*. *Journal of Visual Languages & Computing*.
- Baddeley, A. D., & Andrade, J. (2000). Working memory and the vividness of imagery. *Journal of Experimental Psychology: General* 129, (1) pp. 126-145.
- Baddeley, A. D., & Hitch, G. (1974). Working Memory. In Bower, G. H. (Ed.) *The Psychology of Learning and Motivation, Volume 8*. New York: Academic Press Inc.
- Bailey, J. B. (2002). Absolute threshold of vision. In Norton, T. T., Corliss, D. A., & Bailey, J. E. (Eds.) *The Psychophysical Measurement of Visual Function*. (pp. 35-54). Woburn, MA: Butterworth Heinemann.
- Bauer, D. T., Geurlain, S., & Brown, P. J. (2010). The design and evaluation of a graphical display for laboratory data. *Journal of the American Medical Informatics Association*. 17 (4) doi:10.1136/jamia.2009.000505

- Bedell, H. E. (2002) Spatial Acuity. In Norton, T. T., Corliss, D. A., & Bailey, J. E. (Eds.) *The Psychophysical Measurement of Visual Function*. (pp. 105-136). Woburn, MA: Butterworth Heinemann.
- Bederson, B., Shneiderman, B, & Wattenberg, M. (2002). Ordered and quantum treemaps: making effective use of 2D space to display hierarchies. *ACM Transactions on Graphics*, 21 (4).
- Bertin, J. (1983) *Semiology of Graphics*. (Wainer, H., Trans.). Madison, WI: The University of Wisconsin Press. (Original work published 1967).
- Billock, V. A., & Tsou, B. H. (2011). To Honor Fechner and Obey Stevens: Relationships Between Psychophysical and Neural Nonlinearities. *Psychological Bulletin* 137, (1). pp. 1-18.
- Boring, E. G. (1950). *A history of experimental psychology* (2nd ed.). New York: Appelton-Century-Crofts.
- Brewer, C. A. (2008). <http://www.colorbrewer.org> . accessed 4/15/2008.
- Card, S. K., Moran, T. P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Carr, D. B., Chen, J., Bell, B. S., Pickle, L., & Zhang, Y. (2002). Interactive linked micromap plots and dynamically conditioned choropleth maps. In *Proceedings of the 2002 annual national conference on Digital government research*. Los Angeles: ACM.
- Charles Joseph Minard. In *Wikipedia*. Retrieved from http://en.wikipedia.org/wiki/Charles_Joseph_Minard
- Chen, C., & Yu, Y. (2000). Empirical studies of information visualization: a meta-

- analysis. *International Journal of Human-Computer Studies*, 53 (5).
- Culbertson, H. M. & Powers, R. D. (1959). A study of graph comprehension difficulties. *Educational Technology Research and Development* 7 (3).
- Fechner, G. T. (1987). My own viewpoint on mental measurement. (Scheerer, E., Trans.) *Psychological Research*, 49, pp. 213-219. (Original work published 1887).
- Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology* 47 (6).
- Fitts, P. M., Jones, R. E., & Milton, J. L. (1950). Eye movements of aircraft pilots during instrument-landing approaches. *Aeronautical Engineering Review* 9 (2).
- Geschider, G. A. (1997). *Psychophysics: The Fundamentals*. (3rd Ed.) Mahwah, New Jersey: Lawrence Erlbaum Associates
- Goldberg, J. H., Kotval, X. P. (1999). Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics* 24. pp 631-645.
- Goldberg, M. E. (2000). The control of gaze. In Kandel, E. R., Schwartz, J. H., & Jessel, T. M. (Eds.), *Principles of Neural Science, Fourth Edition*. (pp 782-798). New York: McGraw-Hill.
- Griffin, A. L., MacEachren, A. M., Hardisty, F., Steiner, E., & Li, B.(2006). A Comparison of Animated Maps with Static Small-Multiple Maps for Visually Identifying Space-Time Clusters. *Annals of the Association of American Geographers*, 96 (4).
- Guo, D., Chen, J., MacEachren, A. M., & Lioa, K. (2006) A visualization system for space-time and multivariate patterns (VIS-STAMP). *IEEE Transactions on*

Visualization and Computer Graphics (12) 6.

- Hardisty, F., Myers, A., & Liao, K. (2007). GeoViz Toolkit [Computer Software]. Retrieved from <http://www.geovista.psu.edu/geoviztoolkit/>
- Henmon, V. A. C. (1906). The time of perception as a measure of differences in sensations. *Archives of Philosophy, Psychology, and Scientific Methods*, 8.
- Henmon, V. A. C. (1911). The relation of the time of a judgment to its accuracy. *Psychological Review* 18 (3), pp. 186-201.
- Hollands, J. G., & Spence, I. (1998). Judging Proportions with Graphs: The Summation Model. *Applied Cognitive Psychology* 12.
- Hollands, J. G., & Spence, I. (2001). The Discrimination of Graphical Elements. *Applied Cognitive Psychology* 15.
- Holway, A. H., & Boring, E. G., (1941). Determinants of Apparent Visual Size with Distance Variant. *The American Journal of Psychology*, 54 (1) pp 21-37.
- iPhone 3GS interface [Software]. (2009). Cupertino, CA: Apple, Inc.
- Jessa, M., & Burns, C. M. (2007) Visual sensitivities of dynamic graphical displays. *International Journal of Human-Computer Studies*, 65 (3).
- Konrad, C. M., Kramer, A. F., Watson, S. E., & Weber, T. A. (1996). A Comparison of Sequential and Spatial Displays in a Complex Monitoring Task. *Human Factors* 38 (3).
- Koussoulakou, A., & Kraak, M. J. (1992). Spatio-temporal maps and cartographic communication. *The Cartographic Journal* (29).

- Lee, M. D., Butavicius, M. A., & Reilly, R. E. (2003) Visualizations of binary data: A comparative evaluation. *International Journal of Human-Computer Studies* 59 (5).
- Link, S. W. (1971). Applying RT deadlines to discrimination reaction time. *Psychonomic Science*, 25 (6).
- Link, S. W., & Tindall, A. D. (1971). Speed and accuracy in comparative judgments of line length. *Perception & Psychophysics* 9 (3A).
- NASA (2003). Recent warming of Arctic may affect worldwide climate. Retrieved from <http://www.nasa.gov/centers/goddard/news/topstory/2003/1023esuite.html>
- Logitech (n. d.). Dual Action™ Gamepad. [Computer controller]. Photo retrieved from <http://www.logitech.com/en-us/gaming/controllers/devices/288>
- MacEachren, A. M. (1995). *How Maps Work: representation, visualization, and design*. New York: The Guilford Press.
- MacEachren, A. M., Dai, X., Hardisty, F., Guo, D., & Lengerich, G. (2003). Exploring High-D Spaces with Multiform Matrices and Small Multiples. *IEEE Symposium on Information Visualization*, Seattle, Washington.
- MacKenzie, I. S. (1992). Fitts' Law as a research and design tool in human-computer interaction. *Human-Computer Interaction* 7. pp 91-139.
- Martin, L. J., & Müller, G. E. (1899). Zur Analyse der Unterscheidsempfindlichkeit. Leipzig. Cited in Henmon, V. A. C. (1911). The relation of the time of a judgment to its accuracy. *Psychological Review* 18 (3), pp. 186-201.
- Mayhew, C. & Simmon, R. (2008). *Earth at night*. Retrieved from <http://apod.nasa.gov/apod/ap081005.html>

- Miniotas (2000). *Application of Fitts' Law to eye gaze interaction*. Poster presented at CHI, April 1-6, 200, The Hague, Netherlands.
- Monmonier, M. (1996). *How to Lie with Maps*. (2nd Edition). Chicago: The University of Chicago Press.
- Monmonier, M. (1990). Strategies for the visualization of geographic time-series data. *Cartographica* 27 (1).
- Morse, E., Lewis, M., & Olsen, K. A. (2000). Evaluating visualizations: using a taxonomic guide. *International Journal of Human-Computer Studies*, 53 (5).
- National Weather Service. (2011). *24 Hour Loop*. [Animated weather map]. Retrieved from http://www.weather.gov/sat_loop.php?image=ir&hours=24
- Parker, S., Schneider, B., & Kanow, G. (1975). Ratio Scale Measurement of the Perceived Lengths of Lines. *Journal of Experimental Psychology: Human Perception and Performance* 104 (2). pp. 195-204.
- Payne, D. G., & Lang, V. A. (1991). Visual Monitoring with Spatially versus Temporally Distributed Displays. *Human Factors* 33 (4).
- Peuquet, D. J. (2002). *Representations of Space and Time*. New York: The Guilford Press.
- Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics* 16 (2). pp. 283-290.
- Pinker, S. (1990). A Theory of Graph Comprehension. In Freedle (Ed.). *Artificial Intelligence and the Future of Testing*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Purchase, H. C. (2000). Effective information visualization: a study of graph drawing

- aesthetics and algorithms. *Interacting with Computers 13*.
- Schutz, H. G. (1961a). An evaluation of formats for graphic trend displays: Experiment II. *Human Factors 3* (2).
- Schutz, H. G. (1961b). An evaluation of methods for presentation of graphic multiple trends: Experiment III. *Human Factors 3* (2).
- Shneiderman, B., & Plaisant, C. (2005). *Designing the user interface: strategies for effective human-computer interaction, 4th edition*. Boston: Pearson Education.
- Shneiderman, B. (1996). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualization. In Bederson, B., & Shneiderman, B. (2003), *The Craft of Information Visualization*. Amsterdam: MorganKaufmann. (pp 364-371).
- Simcox (1984). A method for pragmatic communication in graphic displays. *Human Factors 26* (4).
- Simons, D. J. (2000). Current Approaches to Change Blindness. *Visual Cognition 7*.
- Smith, A. M. (2001). *Alhacen's Theory of Visual Perception*. Philadelphia: American Philosophical Society. (This work includes Smith's translation of Alhacen's Optics.)
- Spence, R, Witkowski, M., Fawcett, C., Craft, B., & deBruijn, O. (2002). Image Presentation in Space and Time: Errors, Preferences and Eye-gaze Activity. (Proceedings of) *Advanced Visual Interfaces International Working Conference Gallipoli*, Italy: ACM
- Stevens, S. S. (1975). *Psychophysics: Introduction to its Perceptual, Neural, and Social Prospects*. New York: John Wiley & Sons.

- Tessier-Lavigne, M. (2000). Visual processing by the retina. In Kandel, E. R., Schwartz, J. H., & Jessel, T. M. (Eds.), *Principles of Neural Science, Fourth Edition*. (pp 507-522). New York: McGraw-Hill.
- Tiger (2006). In *Wikipedia*. Retrieved from <http://en.wikipedia.org/wiki/Tiger>
- Titchener, E. B., & Pyle, W. H. (1907). The Effect of Imperceptible Shadows on the Judgment of Distance. *Proceedings of the American Philosophical Society* 46 (185). pp. 94-109.
- Tufte, E. R. (1983). *The Visual Display of Quantitative Information*. Cheshire, Connecticut: Graphics Press.
- Tufte, E. R. (1990). *Envisioning Information*. Cheshire, Connecticut: Graphics Press.
- Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: can it facilitate? *International Journal of Human-Computer Studies*, 57.
- Van den Berg, R., Cornelissen, F. W., & Roerdink, J. B. T. M. (2008). Perceptual Dependencies in Information Visualization Assessed by Complex Visual Search. *ACM Transactions on Applied Perception*, (4) 4.
- Wagenaar, W. A. (1975). Stevens vs Fechner: A plea for dismissal of the case. *Acta Psychologica* 39, pp. 225-235.
- Ware, C. (2004). *Information Visualization: Perception for Design*. San Francisco: Morgan Kauffman.
- Weber, E. H. (1978). *The Sense of Touch*. (Ross, H. E., Trans.) New York: Academic Press. (Original work published 1834).