8-2001

Graphical perception of nonlinear trends: discrimination and extrapolation

Lisa A. Best

GRAPHICAL PERCEPTION OF NONLINEAR TRENDS: DISCRIMINATION AND EXTRAPOLATION

By

Lisa A. Best

B.A. Glendon College, York University, 1992

M.A.P. University of Arkansas at Little Rock, 1994

A THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

(in Psychology)

The Graduate School

The University of Maine

August, 2001

Advisory Committee:

D. Alan Stubbs, Professor of Psychology, Co-Advisor

Laurence D. Smith, Associate Professor of Psychology, Co-Advisor

Roger B. Frey, Associate Emeritus Professor of Psychology

Michael A. Robbins, Research Associate Professor of Psychology

Laurence J. Latour, Associate Professor of Computer Science

GRAPHICAL PERCEPTION OF NONLINEAR TRENDS: DISCRIMINATION AND EXTRAPOLATION

By Lisa A. Best

Thesis Co-Advisors: Dr. D. Alan Stubbs and Dr. Laurence D. Smith

An Abstract of the Thesis Presented
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
(in Psychology)
August, 2001

This dissertation investigated several factors involved in the perception of nonlinear relationships in time series graphs. To model real-world data sets, the graphed data included different sample sizes and levels of variability, and represented different underlying trends. Graph format was also varied. The purpose of the experiments was to determine how these factors affect both trend discrimination and extrapolation accuracy, with the overall goal of determining what types of graphs are optimal in different situations. In Experiment 1, subjects viewed time series graphs on a computer screen and had to identify the type of trend that was present. Six trends (exponential increasing, asymptotic increasing, linear increasing, exponential decreasing, asymptotic decreasing, and linear decreasing) were presented on four graph types (histogram, line graph, scatterplot, and suspended bar graph). The same stimuli were presented in Experiment 2 and subjects

extrapolated future data points by adjusting the position of points on the screen. In Experiment 3, subjects were given feedback on their extrapolations in order to determine if this information would improve their forecasts. Experiment 4 examined discrimination and extrapolation accuracy with dynamic displays that included motion. In all experiments, accuracy was higher when variability was lower and sample size was higher. On discrimination tasks, choice accuracy was higher for nonlinear trends than for linear trends. On extrapolation tasks, in contrast, accuracy was lower when exponential trends were presented, due largely to subjects overestimating the rate of change. In regard to graph format, discrimination accuracy was highest when line graphs were used, but extrapolation accuracy was lowest with line graphs. Thus, the optimal graph format depends on the graphical perception task. Line graphs are optimal for discrimination but other graphical formats lead to higher extrapolation accuracy. Neither feedback nor dynamic displays improved accuracy.

ACKNOWLEDGEMENTS

While at the University of Maine I have been fortunate enough to work with several impressive researchers. I am deeply indebted to my advisors, Dr. Alan Stubbs and Dr. Laurence Smith. Without their support and guidance the completion of this project would have been virtually impossible. From the moment I entered the graduate program, Alan was encouraging and influenced how I approached my graduate career. He suggested taking several key multimedia courses. As a direct result of these courses, I was able to integrate technology into my classes and had the opportunity to teach a computer class outside of the psychology department. I have also worked closely on several research projects with Alan and I am very impressed by his research skills and ethic. Above all else, I feel fortunate to have known Alan on a personal basis and consider him a close friend.

During my second year at the University of Maine, I was part of a research group that was co-headed by Dr. Smith. As part of this group I worked on projects in graphical perception and the history of psychology. It was during this time that I was able to integrate two areas that I was very interested in--psychology and history. I am still involved in both research areas. Larry has proven to be a great source of knowledge and I appreciate his research skills and editing ability. The quality of this project would not

be as high had Larry not spent countless hours proofing each draft. I think that my writing is much improved as a direct result of his "red pen." I look forward to many years of collaboration and friendship (and conversations that eventually turn to food!).

Dr. Roger Frey played a key role in my graduate education. During my first year in Maine, Roger taught my graduate statistics course and his approach to statistics was novel and helped me appreciate how statistics really work. I continue to use notes from his class whenever I have statistical problems or questions. Roger's help extended far beyond the classroom. Without his computer programming knowledge (and his patience), the experimental programs would not yet be compete. Although my programming knowledge is still rudimentary, I am able to write simple programs. Without Roger, this would never had been possible.

Several others must be mentioned. Dr. Michael Robbins has been helpful in several ways...as a committee member, as a statistics teacher, and, for one summer, as an employer. Dr. Joel Gold's statistics class was a great learning experience and I hope to teach with a similar style. I would like to thank Dr. Laurence Latour for introducing me to the pleasures of teaching computer science classes. Teaching in a different department was a great experience and one of the things that helped in the process of finding a job. Joanne Hartt must also be commended. The patience and caring of a good teacher do not go unnoticed.

I would also like to thank my family. My husband, John Gallant, has loved and supported me throughout the dissertation process. He has been my friend, experimental subject, editor, and sounding board and I would like to thank him for the time he put into this project. My parents, Bill and Eleanor Best, have supported my education for almost thirty years. Thomas Edison said that success is 1% inspiration and 99% perspiration. My parents showed me the rewards of perspiration through their own work ethic. Last, but by no means least, I would like to thank my brother, David Best. David has always been a friend and someone I know I can count on (I guess I didn't really need a pony after all).

TABLE OF CONTENTS

List of Tables

List of Figures

Chapter 1

The Use of Data Graphs in Psychology

Previous research has shown that people have considerable difficulty perceiving nonlinear growth and change in the world, consistently underestimating rates of nonlinear change. This problem is important because many processes that bear on human life and welfare are nonlinear (e.g., the growth of pollutants, population growth, the spread of infectious diseases), and the misperception of nonlinear change may limit the effectiveness with which people can detect and respond to these types of phenomena. Although earlier research has posited mechanisms of bias by which the underestimation occurs, it remains unclear the extent to which people can discriminate linear from nonlinear change at all, even under favorable conditions.

Research in the area of graph perception has shown that in many situations graphs are a powerful means of inducing people to see phenomena that were previously undetected by them. In particular, time series graphs can make salient important changes in a natural process. Although some research indicates that people underestimate exponential growth even when shown graphs, the possibility that some types of graphs may be superior to others in revealing nonlinear change has not been investigated. One purpose of the research reported here is to examine the discriminability of nonlinear change using different graph formats. In order to make the discrimination tasks more like their real-world counterparts, the time series shown to subjects had different levels of

random variability; this also permitted an assessment of whether some graph types are more resistant than others to the effects of noise in the data. In addition to varying graph type, one of the studies described here explored the effects of motion on the detection of nonlinear change by using a form of dynamic graph.

Two types of tasks were used in this set of experiments. To assess the basic discriminability of linear and nonlinear trends, subjects performed an identification task in which they viewed hundreds of graphs and attempted to identify the time series in each graph as one of six linear or nonlinear trends. Detailed measures of the properties of each time series permitted psychophysical analyses of subjects' performance, as well as a determination of the graph type that produced optimal performance under different conditions of variability. Following previous research on the perception of nonlinear trends, these studies also used a second type of task—the extrapolation task—which required subjects to apply trend judgments to the prediction of future changes (as occurs in real-world prognostications of growth trends). In this task, subjects viewed graphs of linear and nonlinear trends and attempted to extrapolate future data points. Again, the graphs were of different types and showed data with different levels of variability. In order to assess the effects of feedback on extrapolation performance, in one study, subjects were given feedback on the accuracy of their responses.

Taken together, these experiments provide a clearer picture of the psychophysics of nonlinear trend perception and assessed various possible means for counteracting underestimation biases and promoting greater sensitivity to nonlinear trends embedded in noisy data.

The Importance of Graphs

For centuries, the question of how to best communicate complex data has concerned both philosophers and scientists (Wainer, 1997). The ancient Egyptians were the first to propose a revolutionary solution to this age-old problem—they transformed complex spatial information into a map and, in a sense, created the first graphical display. By Shakespearean times, it was recognized that one of the best ways to keep track of numbers was to use a "score and tally" system (Wainer & Thissen, 1993). This method of data organization represented an attempt to combine large sets of data coherently. Though this system was effective for simple counting tasks, it did not allow the user to examine any underlying data relationships. In 1801, William Playfair (see Wainer, 1997) published The Commercial and Political Atlas, which included examples of several graphs still in use today. In this one publication, Playfair brought the pie chart, the bar chart, the histogram, and the line graph into being.

It is difficult to convey complex information verbally (e.g., Wainer, 1997) or in tabular form (e.g., Chambers, Cleveland, Kleiner, & Tukey, 1983; Cleveland, 1993a, 1993b, 1994; Macdonald-Ross, 1977). As Playfair said in 1805:

I first drew the Chart in order to clear up my own ideas on the subject, finding it very troublesome to retain a distinct notion of the changes that had taken place. I found it answer[ed] the purpose beyond my expectation, by bringing into one view the result of details that are dispersed over a very wide and intricate field of universal history; facts sometimes connected with each other, sometimes not, and always requiring reflection each time they were referred to. I found the first rough draft g[a]ve me a better comprehension of the subject, than all that I had learnt from occasional reading, for half of my lifetime: and, on the supposition that what was of so much use to me, might be of some to others, I have given it with a tolerable degree of accuracy. (Playfair, 1805, quoted in Costigan-Eaves & Macdonald-Ross, 1990, 319-320.)

Thus, almost from their advent, graphs have been used to clarify patterns and make trends evident. These patterns may not have been evident in textual or tabular presentations, or may only become evident after a careful examination of the material (Tilling, 1975). In recent years there has been a surge of research on this topic and many researchers have concluded that graphical displays best communicate the patterns and idiosyncrasies inherent in any data set (e.g., Butler, 1993; Chambers, Cleveland, Kleiner, & Tukey, 1983; Cleveland, 1993a, 1993b, 1994; Cleveland & McGill, 1985; Henry, 1995; Kolata, 1984; Kosslyn, 1994; Kosslyn & Chambris, 1992; Lewandowsky & Spence, 1989; Macdonald-Ross, 1977; Meyer, 1997; Tufte, 1983, 1990; Tukey, 1990, 1993.).

A good graph can highlight important points that would otherwise be missed and, often, information presented graphically "can stop your mental flow in its tracks and make you think. A visual display can force you to notice what you never expected to see" (Tukey, 1990, p. 328). Effective data presentation is essential for both the researcher and the audience. When data are displayed properly, valid conclusions can be drawn quickly (Washburne, 1927) and novel hypotheses emerge. Because of the amount of information presented graphically, it is essential that a display convey the information in a clear, truthful manner (Tufte, 1983).

Though empirical research on the visual analysis of data has begun only recently, graphical analysis is one of the oldest methods of determining if there is a relationship between different sets of information (Parsonson & Baer, 1978). Parsonson and Baer outlined several advantages of using graphs to analyze relationships. This type of analysis is quick and easy for both the researcher and the audience. For the researcher, graphs are easy to make, requiring for rough drafts only a pencil, ruler, and grid paper. The audience is presented with information visually, which allows them to draw conclusions and form hypotheses. There are many different graphical formats available—many more than the statistics available to analyze the same data set. Furthermore, graphs allow people with a wide range of abilities to quickly see the message that is being conveyed — for example, whereas the interpretation of a graph is almost automatic (e.g., Cleveland & McGill, 1985; Tufte, 1983, 1990; Wainer, 1997), the

interpretation of the results of a statistical analysis requires formal training. When graphs are used, the data are not transformed (or are only minimally transformed), which allows the viewer to see data points that may otherwise be missed (e.g., outliers). Finally, the theoretical premises underlying graphical methods are simple and well-understood whereas the theoretical premises underlying statistical analysis are complex and might even be unknown to the user.

Although it has been argued that graphs may be optimal for data presentation, they are useful only if carefully constructed. Unfortunately, many graphs are poorly conceived and may convey messages that are incorrect. Wainer (1997) outlined several graphical mistakes, such as including irrelevant information or emphasizing trivial facts while ignoring the important findings. Many popular graphical displays fail to represent the data accurately or they hide the true meaning of the data with unnecessary elements, such as pictures, that fail to convey the intended message (Henry, 1995). Most researchers agree that it is important to keep in mind the nature of the conclusions that are to be conveyed and construct a display that allows a viewer to recognize these conclusions accurately and effortlessly (e.g., Cleveland, 1985; Kosslyn, 1994; Tufte, 1983, 1990; Wainer, 1997). Because the human visual system is very capable of recognizing and processing visual patterns (Goldstein, 1996, 1999), a well-constructed visual display can lead to better comprehension than a verbal description of the same phenomenon. The research on this topic has led to the general conclusion that graphs are

superior to tables (Butler, 1993; Legge, Gu, Luebker, 1989) and to text (Lewandowsky & Spence, 1989). Though it has been established that, in most cases, graphs are an optimal method to display data, the graph that best displays a particular data set depends on the nature of the data being displayed (Cleveland, 1993a, 1993b; Tufte, 1983, 1990). Although people are able to extract simple information from both graphs and tables, the detection and understanding of complex interactions is facilitated by the use of graphs (Meyer, Shinar, & Leiser, 1997).

Anscombe (1973) constructed one of the most famous illustrations of how graphs can make unclear relationships immediately obvious. Often, a good graph gives more information about a data set than even the most complicated statistics. Consider the four x-y data sets used by Anscombe. Each has several properties (n, mean of X's, mean of Y's, regression equation, correlation coefficient, etc.) that are the same, in terms of linear regression (see Table 1). In spite of these identical properties very different patterns emerge when the four data sets are presented graphically (see Figure 1). These differences might not be apparent from the tabular data, but are immediately obvious from the graphs. This is precisely the point that Tufte (1983, 1990) makes — a good graph should force us to see what we did not expect to see.

Table 1

Four Data Sets with Identical Linear Properties

SOURCE: Anscombe, F. J. (1973). Graphs in statistical analysis. American Statistician, 27, 17-21.

X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.00	9.14	10.0	7.46	8.00	6.58
8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
13.0	7.58	13.0	8.74	13.00	12.74	8.00	7.71
9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
11.0	8.33	11.0	9.26	11.00	7.81	8.00	8.47
14.0	9.96	14.0	8.10	14.00	8.84	8.00	7.04
6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25
4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.5
12.0	10.84	12.0	9.13	12.0	8.15	8.00	5.56
7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91
5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89

Linear Properties of the Data Sets

N=11

Mean of X's = 9.0

Mean of Y's = 7.5

Equation of regression line: Y=3+0.5X

Standard error of estimate of slope = 0.118

t = 4.24

Sum of squares X = 110.0

Sum of squares Y = 13.75

Residual sum of squares of Y = 13.75

Correlation coefficient = .82

r^2=.67

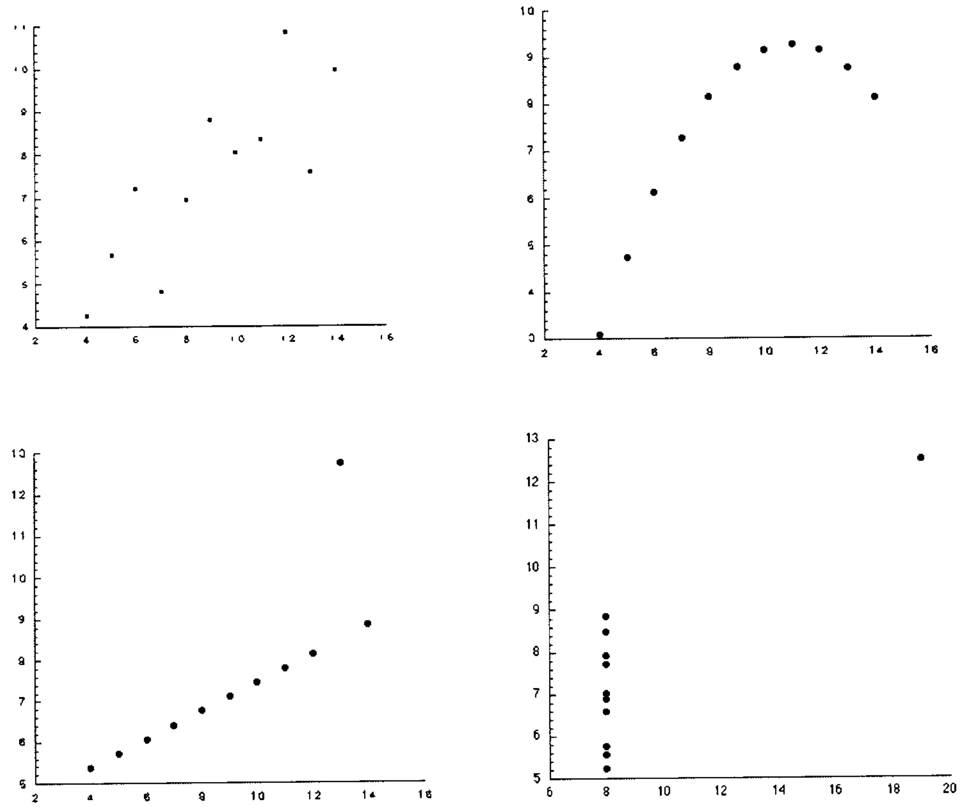


Figure 1. Scatterplots of four data sets with identical linear properties.

SOURCE: Anscombe, F. J. (1973). Graphs in statistical analysis. American Statistician, 27, 17-21.

A good graph can be used to enhance or, in some cases, even replace traditional hypothesis testing. A well-constructed graph can convey important research findings without highlighting unimportant (but perhaps statistically significant) results. Some researchers have suggested that there is no statistic as important as a well-chosen graph (Chambers et al., 1983; Tukey, 1990). Visual displays often lead viewers to notice relationships that are totally unexpected and often lead to intuitive analyses that are supported by subsequent statistical analyses. Empirical findings have led to the conclusion that humans are often quite accurate "intuitive statisticians" (e.g., Best, Smith, Stubbs, & Frey, 1998, 1999; Nisbett & Ross, 1980; Smith, Boynton, Stubbs, & Doble, 1992). The use of more "intuitive" statistical methods, such as graphs, may allow researchers to uncover important findings and, at the same time, remove the focus from some of the less important components of the data.

An important distinction must be drawn between graphs used for data analysis (the so-called analytic or exploratory graphics) and graphs used to present research findings (the so-called presentation graphics). When analyzing data, researchers focus on examining a data set and are interested in the nuances present in the data. Though data analysts are interested in uncovering relationships in a data set, they are also interested in finer details, such as the presence of outliers and the shape of the underlying sample distribution (Chambers et al., 1983; Cleveland, McGill, & McGill, 1988). During this

stage of research, the focus is on the data and not on the audience. Once the researchers have drawn conclusions, they can make decisions about which points to highlight and the type of presentation that accurately conveys the intended message (Wainer, 1997).

Even though it is generally accepted that graphs are useful to convey large amounts of information, it is essential that individual data sets be carefully examined prior to the decision about which specific display technique is to be used (Cochran, Albrecht, & Green, 1989). Integrating both data analytic and presentation graphics into the research process allows researchers to become aware of their data in a way that would be difficult or impossible if only descriptive and inferential statistics were used to describe the data properties. When graphs are used at an exploratory level, decisions concerning which presentation methods to use will be focused on the data themselves and not solely on a p-value associated with an inferential test.

Chapter 2

Graphical Perception

The field of graphical perception focuses on visualization and involves presenting the same information using different display methods in order to determine which methods of data presentation (e.g., text, tables, types of graphs) lead to better interpretation. Information on a graph is encoded in terms of geometry, texture, and color, and when an observer examines the display, this information is visually decoded. This decoding process has been labeled "graphical perception" (Cleveland, 1985).

In order to provide an overview of the current research trends in the field, four different research perspectives will be discussed. Cleveland (e.g., 1993a; 1993b; 1994) approached graphical perception from a psychological perspective and focused on the ability of subjects to visually decode the information encoded in a graph. Tufte (1983, 1990) theorized about different graphical methods and advocated avoiding graphical elements that do not add a significant amount of information to the graphic. Tufte argued that this "chart junk" should be avoided at all costs — the result is a graph that includes only the elements that are absolutely necessary. Pinker (1990, 1997) constructed a theory of graphical comprehension that is more cognitive in nature and described how graphical elements are decoded and understood by the brain. Other theorists are interested in other aspects of graphical perception, such as how to construct graphs that are effective at conveying the intended message. Kosslyn (1994), like Pinker, focused on the perceptual

and cognitive abilities of humans and suggested that the effectiveness of graphical displays could be optimized by taking perceptual abilities into account during the design process. Some of the basic research findings of these researchers will be examined.

The Perceptual Theories of Graphical Perception

William Cleveland

There has been much research on graphical perception and the graphical elements that make it easy (or difficult) to decode the information embedded in a graph. One task of graphical perception research is to determine the optimal display methods for different types of data. Although one method may be suited to one data set, it is not necessarily appropriate for all types of data. In today's data-laden society, it is important that guidelines concerning graph construction be understood and available to researchers.

Although graphs provide an efficient way to display a data set, the question still remains as to which type of graph is optimal. Cleveland (1993a, 1994 b; Cleveland, Diaconis, & McGill, 1982; Cleveland & McGill, 1986) examined different types of data and outlined different graphical displays that enhance human perception and data interpretation. Cleveland and his associates suggested that the purpose of a graph is twofold—a good graph allows a viewer to quickly see the information and data relationships and allows him or her to communicate this information to an audience.

Following the basic tenets of visual perception, Cleveland (1985) suggested that when a graph is observed, many mental and visual tasks are carried out before quantitative information can be extracted from the display. In order to understand a graph, it is necessary to first understand what variables are being presented. Once an observer has this knowledge, the basic quantitative information can be extracted. It is the extraction of quantitative information that is the primary task of graphical perception and is achieved as an observer scans the symbols used, the lines that connect the data points, and the axis scales. At this point in the perception process, the observer has some idea of what variables are being presented, the general relationships that exist, and some of the dips and rises in the data series. These early visual tasks are carried out by the visual system almost instantly and effortlessly.

After the observer has some basic knowledge about what is being presented, he or she carefully examines the display in order to discover some of the finer, but still informative points. Tick marks and scale lines make it possible to determine exact data points and points at which shifts in the data occurs. Cleveland (1985) argues that there are 10 basic graphical perception tasks that are performed when information from graphs is decoded (see Figure 2).

In a series of experiments, Cleveland and his colleagues (Cleveland & McGill, 1984, 1985) found that observer accuracy was best for judgments of position along a common scale and along identical, nonaligned scales (A & B in Figure 2). They

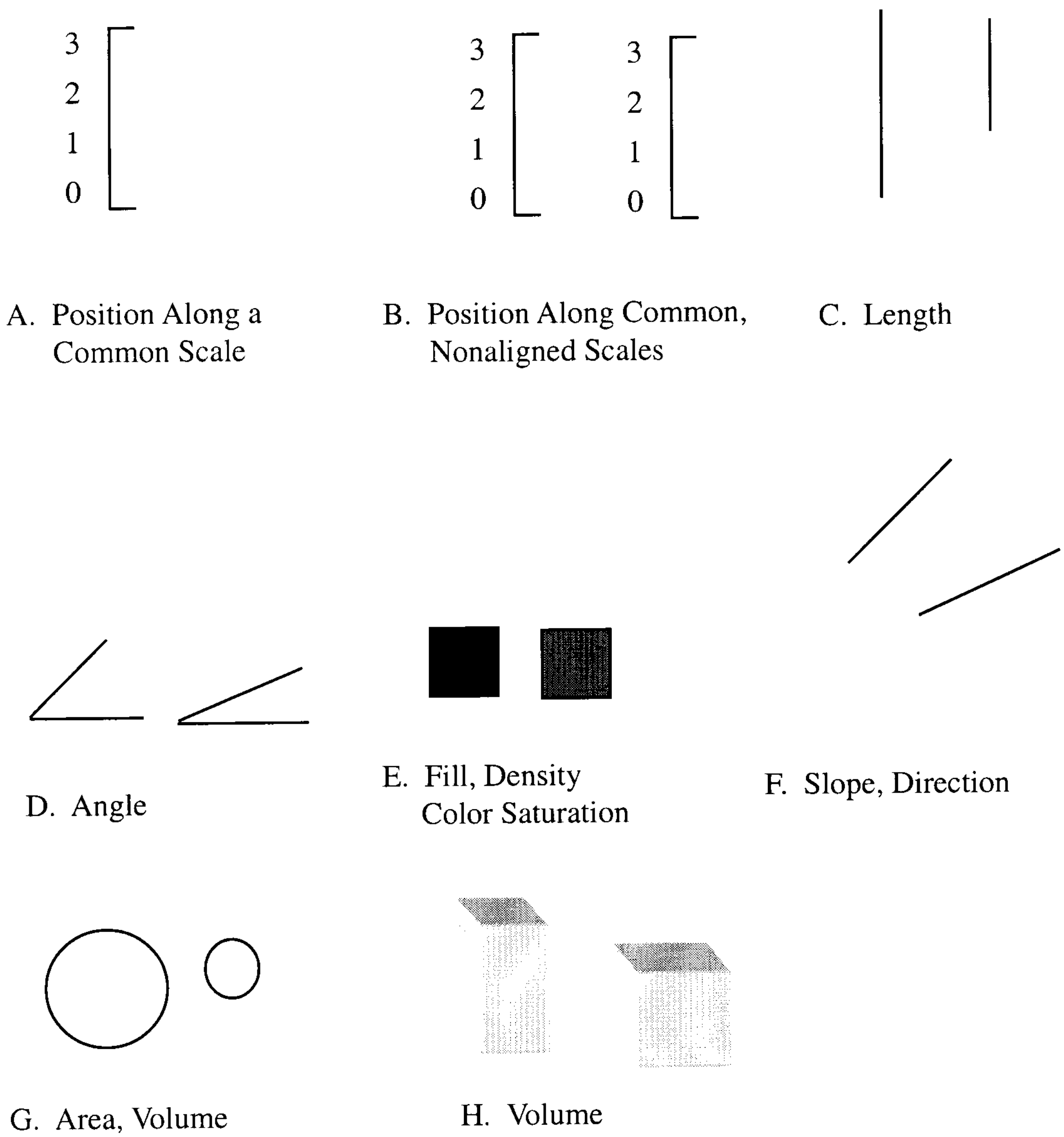


Figure 2. Graphical perception tasks, ordered from the most accurate to the least accurate.

SOURCE: Created from information provided in Cleveland (1985).

recommended the use of a common scale simply because the statistical error of visual estimates was lower when this method was used. Thus, it is easier for subjects to draw conclusions if different sets of data points are presented on a single scale. Subjects were accurate if they were asked to determine which of two lines (or bars on a bar graph) was longer (C in Figure 2) but were less able to judge differences if they had to differentiate between two data points using angle, slope, area, volume, density, color saturation, or color hue differences (D through H in Figure 2).

In a review of the literature, Carswell (1992) examined 39 graphical perception experiments that had been published in scholarly journals. In order to assess Cleveland's ranking scheme, she compared how subjects responded to pairs of graphs and predicted which should be more efficient according to Cleveland's ranking of graphical specifiers. She found general support for the task ordering model. In this meta-analysis, the differences between position, length, and angle were minimal, though performance on these tasks was consistently better than for area or volume judgments. Carswell concluded that Cleveland's ordering scheme is most successful at predicting performance in comparison and point-reading tasks but is less effective when more complex information must be extracted from a graph. Thus, it appears that some perceptual tasks are easier than others and result in more accurate judgments. It follows that graphs

should be constructed in order to optimize the decoding ability to humans. When constructing a graph, it is important to take into account the effectiveness of a particular display and the type of information to be conveyed.

Often, the purpose of a graph is to display a large amount of information in a small space. Working within Cleveland's ordering schema, Cochran et al. (1989) suggested that line graphs should be used to show trends in a data set but should be avoided if an observer is expected to detect exact amounts. Line graphs are a good method of presentation if the data have successive data points, cover a long period of time, or are presented in cases where judgments or data extrapolations are necessary. Bar graphs are optimal if comparisons of magnitude or size are necessary and can be used to convey differences in time or between items. Pie charts have a limited usage and can be used for part-to-whole judgments. If pie charts are used, they should not contain more than five segments. The selection of a graph type should be driven by the type of data being presented and the limitations of the human visual system when performing certain perceptual tasks.

Edward Tufte

A good graphic should "induce the viewer to think about the substance rather than about methodology, graphic design, the technology of graphical production, or something else" (Tufte, 1983, p. 91). In this single statement, Tufte emphasized the importance of having viewers focus their attention on the data rather than on how it was collected or analyzed (Henry, 1990). Hence, it is essential for both the researcher and the intended audience to focus on the data and the message that they convey.

The primary purpose of constructing a graph is to communicate a complex idea with clarity, precision, and efficiency (Tufte, 1983, 1990). Graphs should be constructed so that a lot of data are displayed in a coherent manner that does not lead to visual distortions. The result of these attempts should force the viewer to compare different parts of the data and make conclusions about the data at different levels of detail. A good graph should not replace statistical and verbal descriptions of a data set but rather integrate these into an easily comprehended picture of the data under scrutiny.

Tufte (1983) made several recommendations concerning the proper display of data. One of his primary suggestions is to remove as much "non-data" ink as possible. Non-data ink is defined as any graphical component that does not convey information — even if this results in the elimination of symmetry, closure, and other organizational

properties. Tufte labeled this extra ink as "chartjunk" and argued that the best way to make an effective and easily readable graphic is to maximize the data-ink ratio (the amount of ink used to convey data compared to the total ink in the graph).

Although the extra ink in graphics does often hide the intended message, it is possible that maximizing the data-ink ratio can render a graphic less appealing, and hence decrease readability (Spence, 1990). Spence argued that extra ink is often helpful because it can lead to a graph with fewer graphical elements that must be decoded (for example, extra ink may clarify an important point that would otherwise have to be considered). Furthermore, Carswell (1992) examined 39 graphical perception experiments in order to compare the assertions of Cleveland and Tufte. She concluded that Cleveland's graphical perception system was superior. In the vast majority of experiments, Cleveland's model had greater predictive power than Tufte's data-ink principle.

Information Processing Models of Graphical Perception

Models of graphical perception that center on information processing go beyond examining the components of an effective graphical display. Information processing models of graphical perception focus on how the brain organizes, stores, and processes information (Lohse, 1993).

When a graph is first presented to a reader, the early perceptual processes detect visual primitives, such as shape, position, color, and length (Julesz, 1980; Lohse, 1993). The visual description is constructed from these primitives. This memory system is limited and therefore, only a certain amount of information can be organized and interpreted initially. After initial processing, the information triggers a memory trace (a graph schema, in the case of graphical perception). In general, the memory schema contains standard procedures for interpreting information.

Theorists who focus on graphical cognition are interested in determining how graphical information is processed by the cognitive system. Although these theories are not incompatible with those of graphical perception, they do focus on different aspects of graph comprehensibility.

Steven Pinker

Visual (and graphical) cognition focuses on how the brain creates a symbolic representation of a scene or visual display (Pinker, 1990). This representation can include information about the individual elements of the display, their shape, size, location, color, and texture. When a graph is examined, two types of mental representations are made: a visual description and a graph schema (Pinker, 1990, 1997).

The visual description is the structural description that represents a graph and, therefore, contains the information that allows an observer to decode the underlying meaning of the graph. Pinker (1990) suggested that the visual description is constrained

in a way that makes graphical comprehension easy for the observer. In order to decode a visual display, the brain has to decode the spatial information that is contained in the display. First, the variables must be recognized. Once recognition has occurred, the other perceptual dimensions must be decoded. For example, in a bar graph, the individual bars give some information about what information is being conveyed, but additional information is necessary for an accurate interpretation of the graph. When observing a bar graph, one has to be aware of the variables on the graph, but it is also necessary to examine other attributes, such as the horizontal and vertical positions of the patterns.

The higher-order analysis of a graphic allows a viewer to determine the meaning conveyed and the conclusions that can be made based on this meaning. In decoding a display correctly, one must group the components according to how each component is related to the other components. The Gestalt laws of grouping provide an explanation of how the visual system accomplishes this task (Lohse, 1993; Pinker, 1990). Several of these laws are useful in understanding how graphs are interpreted. Perceptual elements that are either close to one another, similar to each other, continuations of one another, or parallel to each other are perceived as belonging to a single group (Goldstein, 1996, 1999). Thus, the Gestalt principles may help observers determine how the different graphical components are related and thus aid in interpretation.

After the basic elements of a graph have been deciphered, it is necessary for the cognitive system to translate the information in the visual description, specify how to access the relevant components of this description, and determine what type of graph is being represented (Pinker, 1990). Graph schemas function to allow a person to recognize the type of graphic that is currently being viewed. After the correct schema is chosen, it is necessary to translate the visual information into conceptual information. To accomplish this goal, the graph schema is searched for relevant messages and when these messages are encountered, they are added to the list of information that the reader already has about the data.

It has been proposed that we have different graph schemas for different types of graphics (Pinker, 1990, 1997). For example, if we are presented with a bar graph, the properties necessary for decoding bar graphs make up the "bar graph schema". Examples of these properties may include "higher bars mean more of variable X," or "look at the legend to see what color represents variable X." Also contained in the schema is information concerning how to answer questions about the information contained in the graph. When we are presented with a specific question, we first locate the legend to determine which element to focus on, then we ascertain the necessary location of this element on the graph, and finally, we are able to determine an answer to the question (Lohse, 1993). As with other cognitive tasks, the time required to answer a specific question depends on the difficulty level of the question. Thus, it is essential to take into

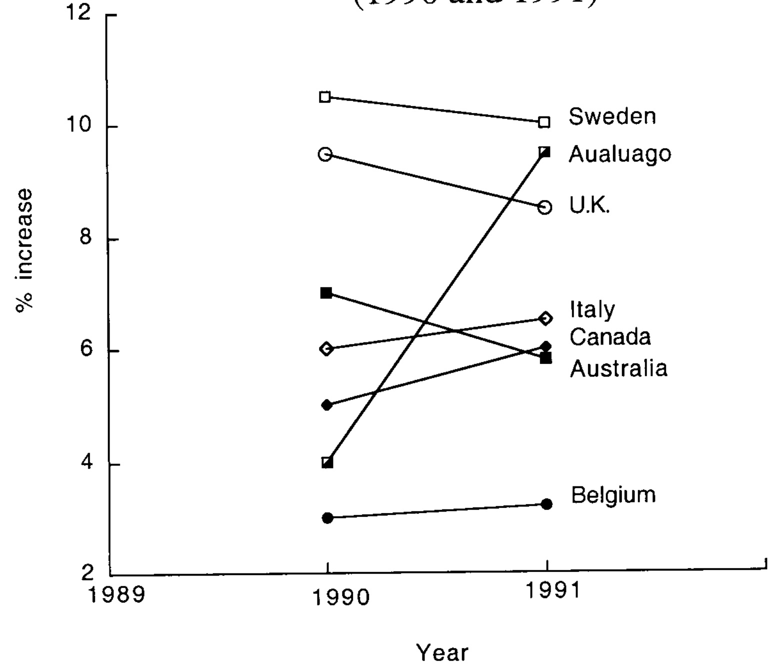
account both the processing capacity of the mind and the types of graphs that maximize this capacity. If done correctly, graphs can be constructed that allow for high levels of comprehension with little data distortion.

Stephen Kosslyn

As a cognitive scientist, Kosslyn (1994) was interested in how the brain processes and stores information. He emphasized that, "A picture can be worth a thousand words—but only if you can decipher it" (Kosslyn, 1994, p. 1). Kosslyn (1994; Kosslyn & Chabris, 1992) formulated a theory of graphical perception that included many cognitive elements. His theory emphasized the cognitive organization of the brain and how this organization affects what types of graphical displays are most effective. Given this theoretical orientation, the focus of Kosslyn's theory is on the graphical attributes that allow a reader to detect and correctly interpret the message underlying the graphical display.

Kosslyn (1994) stressed three "psychological maxims" that must be considered before a theory of graph comprehension can be formulated. First, it is important to realize that the mind is not a simple receiving system, comparable to a camera. Humans do not register information exactly as it is presented, but use several mechanisms to integrate and interpret information as it is presented. Psychologists have formulated laws of perceptual organization (Goldstein, 1996, 1999) to explain how the mind organizes information into meaningful groups. For example, in Figure 3, the lines are registered

Forecast for Consumer Prices
(1990 and 1991)



Real GDP% Increase in a Year

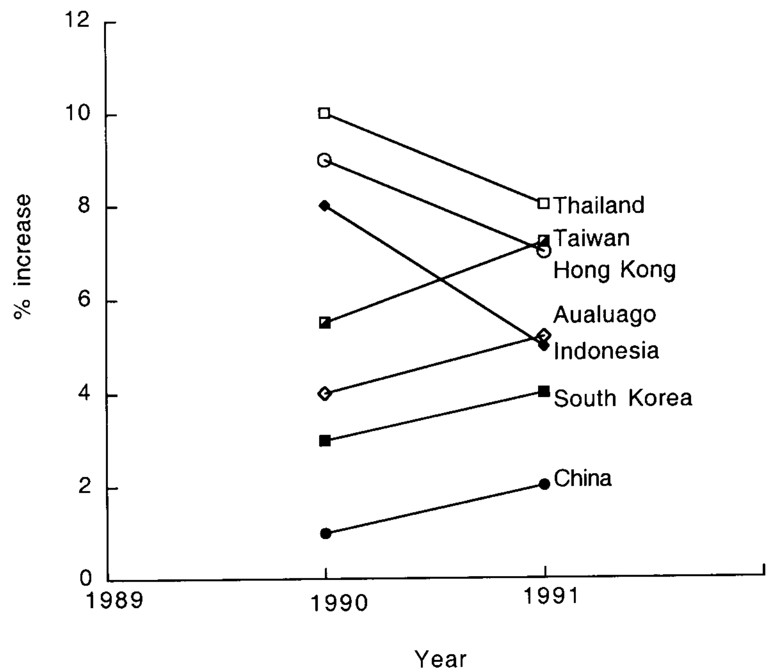


Figure 3. The effects of perceptual organization.

SOURCE: Created from information provided in Kosslyn (1994).

easily when they are parallel (they are registered as a single group) but the display is more difficult to read when each line is organized separately (there are more perceptual groups in the left panel).

The second maxim focused on the fact that we tend to make direct comparisons between graphical properties and the properties of the entities symbolized by the graphical pattern (Kosslyn, 1994). For example, when we are presented with a line graph, the rise and fall of the line is interpreted as change in the variable being represented. If a noncontinuous variable, such as gender, is represented on a line graph, this conclusion is misleading. Thus, when constructing a graph, it is important to consider how the variables will be interpreted—"'more' of something in a display should correspond to 'more' of a substance" (Kosslyn, 1994, p. 8).

Finally, when constructing a display, it is important to recognize that the capacity of the brain is limited (Kosslyn, 1994). Because the brain can only process a certain amount of information at any one time, a graph should not require readers to process more groups than this (which would be an impossible request). Graphs are useful because they often contain a large amount of information that is presented it in a form that does not overload our information processing system.

Based on these maxims, Kosslyn (1994; Kosslyn & Chabris, 1992) formulated a number of rules of graphical construction. Before any information is graphed, it is important to decide what major point(s) should be communicated and what questions

readers should be able to answer after examining the display. It is also important to determine what graphical formats are familiar to the reader—expert graph readers will be able to interpret certain complicated formats but inexperienced readers will require more familiar formats. These decisions should drive construction. For example, graphs should be used to illustrate relative quantities but if it is important to communicate specific values, a table better communicates this information.

Kosslyn and his colleagues (1994; Kosslyn & Chabris, 1992) were interested in examining the relationship between information processing processes and the interpretation of graphical information. Based on this research, a number of rules of graphical construction that focused on determining an optimal display method for different types of information were formulated. For example, Kosslyn and his colleagues argued that a line graphs are optimal when displaying the interaction between two variables but bar graphs better convey relative point values. In general, many of these rules are similar to those of other researchers. Kosslyn's conclusions about graphical construction may not have been original but his theory of graphical comprehension integrated several key elements of perception and cognition.

Thus, models of graphical cognition focus on the brain processes that occur as observers examine and decode graphical displays. These theories complement those of graphical perception and attempt to examine "how" graphs are interpreted. Hence,

whereas graphical perception theorists focus on how to construct a graph efficiently, researchers in graphical cognition focus on the processes that allow one to interpret and make conclusions about the graph in question.

Graphical perception researchers are interested in how humans read, understand, and form conclusions based on graphical information. The trend towards graph use began with Playfair and is now recognized as a data analytic method that can be used to supplement the results of inferential statistics. Research in the area has led to important advances and to several guidelines on graph construction. Although researchers in graphical perception focus on diverse issues, all are ultimately interested in the factors that aid the readability and comprehensibility of graphical displays.

Chapter 3

The Graphical Perception of Time Series Data

Time series graphs are commonly used to convey change in a variable over time. These displays are used for many purposes—for example, by clinical psychologists to chart patient progress or by stock market analysts to convey rises and dips in the market. In order for such graphs to be effective, it is important to consider carefully what information has to be conveyed and the best way to convey this information (Cleveland, 1985, 1993; Tufte, 1983; Wallgren, Wallgren, Persson, Jorner, & Haaland, 1996).

Time series analysis is the study of temporal variation in some discrete or continuous dimension (see McDowall, McCleary, Meidinger, & Hay, 1980). Typically, time series data are presented graphically as a line graph, with time on the horizontal axis and the "amount" of the variable in question on the vertical axis (Kosslyn, 1994). An example of a time series analysis concerns the question of whether different environments have differential effects on behavior. The effects of different environments on a behavior can be tested by alternating exposure of the two environments (ABAB design). The data can be presented in two ways: tabular (see Table 2) or graphical (see Figure 4).

Table 2

Hypothetical Data Gathered Under Two Conditions

SOURCE: Created from information in Parsonson & Baer (1978).

Environment 1	Environment 2	Environment 1	Environment 2
1	5	4	11
1	6	2	10
2	6	0	11
0	8	2	9
1	7	1	9
2	9	1	8
1	8	2	7
0	11	0	7
1	10	0	9
1	10	1	8
	11	2	7
	9	1	6
	8		7
	8		5
	9		5
	8		
	10		

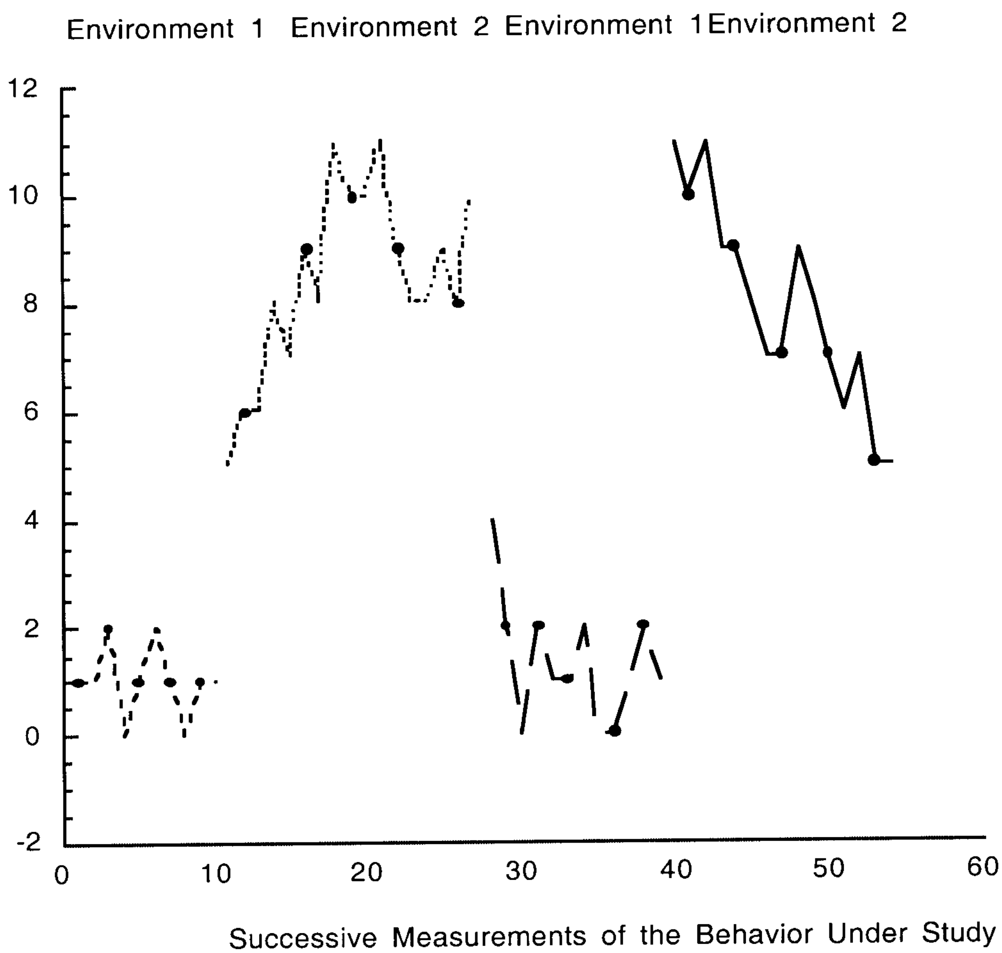


Figure 4. A graphic representation of the hypothetical data of an ABAB single-subject experimental design.

SOURCE: Created from information in Parsonson & Baer (1978).

Although the conclusion that Environment 2 leads to "more" of the behavior than Environment 1 is evident from both the table and the graph, the effects are more quickly and easily seen when the series is presented in graphical format. In a single glance, it is obvious that Environment 2 produces more of the behavior, and that the effects of the environment on the behavior increase gradually during the first exposure to Environment 2. Also, during the second exposure to Environment 2, the effects of the environment decrease over time. The graph also conveys that Environment 2 always produces higher levels of the behavior than Environment 1. Though the same information can be extracted from Table 2, extraction is much easier when the graph is used (Parsonson & Baer, 1978).

One major focus of time series analysis is whether there is a change in the mean level of some variable. For example, in charting client progress, clinical psychologists are interested in determining whether or not their clients are improving. One way to determine this is to use a statistical model to determine if there is a significant change from the baseline level. Visual inspection techniques represent an alternative to the inferential statistic approach. These techniques focus on whether or not an observer detects a change in time series data presented graphically (Kazdin, 1982; Kratochwill, 1978). Proponents of visual analysis argue that these techniques are optimal because, when using them, one is less likely to detect small (but possibly significant) effects. One researcher, quoted in Gottmann and Glass (1978) stated,

Applied behavior analysis doesn't need sensitive statistics to salvage small effects. The field needs basic effects, that is, effects that are large, reliable, powerful, clear, and durable. Such effects will give us strong principles. The field doesn't need dubious, small, and sometimes but statistically significant results, because they would result in the development of more and weaker principles. (pp. 197-198)

As with any data analytic technique, there are advantages and disadvantages associated with this method of analysis (Gibson & Ottenbacher, 1988). The graphical presentation of data provides a comprehensive and compact account of the data. Whereas traditional statistical methods give summary data, visual inspection techniques present individual raw data points. Typically, researchers using visual techniques are searching for the "big picture" underlying any data set and are concerned with questions such as whether a trend is present, whether there are any increases or decreases that occur over time, and how much variability is present.. Researchers using visual analytic techniques are not as concerned with relatively minor trends and effects. Effects that are not visually apparent will not be reported and, thus, ineffective independent variables will not be included in future research or clinical interventions.

Loftus (1993) has gone so far as to argue that traditional hypothesis testing is an overused and useless way to analyze data:

The main argument that I have tried to make in this article is that hypothesis testing is a wave of the past (and never should have been a wave at all). Characterizing

conclusions in hypothesis-testing terms requires reducing the complex, multidimensional information that generally emerges from an experiment into one or more binary decisions that are almost always logically predetermined to begin with. (p. 255)

Although most researchers do not take a stance this strong, many do recognize the importance of graphical data analysis.

In spite of their strengths, graphical methods have been criticized on various grounds (see Ottenbacher, 1990). For example, interrater reliability is sometimes low when visual inspection techniques are employed. Raters are more likely to agree when a time series plot does do not include a statistically significant temporal change and, therefore, interrater reliability is higher for graphs that do not contain significant trends. Given that visual inspection techniques are commonly used, by both professionals and laypersons, it is important to systematically examine the effectiveness of these techniques in order to improve the accuracy of analysts as they draw conclusions based on time series data.

Before visual inspection techniques are employed, it is important to consider the amount of data being analyzed. Under optimal conditions, human performance may rival that of statistical tests (Best et al., 1998, 1999). However, if not enough data are presented (Park, Marascuilo, & Gaylord-Ross, 1990), if the level of variability is high (DeProspero & Cohen, 1979; Furlong & Wampold, 1982; Krathochwill, 1978;

Ottenbacher, 1986, 1990), or if the display method is inadequate (Cleveland, 1983; Tufte, 1990), accuracy can decline. Thus, it is important that these issues be examined and the effects of these variables determined.

Munger, Snell, and Lloyd (1989) examined how accurate teachers were when assessing student progress by graphical means. When continuous variables presented on a graph that had an upward trend (representing an improvement), the judgments were consistent but when the trends did not change or decreased, the consistency of the judgments was lower and depended upon the frequency of data collection (in other words, the sample size). This finding is not consistent with previous research that suggested that visual inspection techniques are unreliable, regardless of whether the judges were teachers (Munger, Snell, & Lloyd, 1989) or skilled behavior analysts (Jones, Weinrott, & Vaught, 1978). At least in the case of student progress, it appears that ascending trends are discriminable from both negative and flat trends. Furthermore, other research has suggested that including a trend line that runs through the data improved performance (Bailey, 1984; Rojahn & Schultze, 1985).

Visual inspection techniques can be analyzed through the comparison of visual judgments and statistical tests (Jones et al., 1978; Park et al., 1990). Jones et al. asked subjects to determine whether a set of graphs that had been published in the Journal of Applied Behavior Analysis (JABA) included a change in level or not (they could also reply that they were unsure if a change occurred). The correlation between the human

judgments and statistical analyses were low, suggesting that there was little agreement between the judges and the statistic. Further analyses revealed that the human judgment was affected by the level of serial dependency in the data and the different levels of significance of the effects. Serial dependency, or autocorrelation, refers to the relationship that can exist between successive data points in a series (Kratochwill, 1978). For example, in a time series of daily temperatures, two successive days (e.g., July 1st and July 2nd) are likely to have similar temperatures. If many adjacent points in a series are similar, the autocorrelation will be high. According to Jones et al. (1978) agreement between visual judgments and statistical methods was highest when the data sets contained large temporal changes and the serial dependency was low. The level of agreement fell monotonically as the level of serial dependency increased. Interestingly, in contrast to Munger et al. (1989), it was found that agreement was highest if there was no significant increase in the series.

Park et al. (1990) had expert judges determine whether graphs (previously published in JABA) represented a nonsignificant, significant, or unclear level change. These researchers found that the agreement between judges was only slightly above chance. Agreement did improve if the researchers only had to decide if the trend was significant or nonsignificant. In support of Jones et al. (1978), there was more agreement when the time series had a nonsignificant level change. Unfortunately, the levels of interrater agreement were below the established, acceptable levels, though agreement was

higher if the trend was nonsignificant. In addition, time series statistics were more conservative than the visual inspection techniques, which is interesting given that one of the proposed strengths of visual techniques is they tend to be more conservative, preventing the detection of unimportant (but statistically significant) effects.

Generally, the higher the variability in a data series, the more difficult it is for subjects to determine the strength of the underlying relationship between the variables being represented (e.g., Bobko & Karren, 1979; Cleveland et al., 1982; Erlick & Mills, 1967; Lane, Anderson, & Kellam, 1985). It appears that visual analysts do not systematically compare the size of the mean change to the level of variability in the data (Furlong & Wampold, 1982). In this sample of experts, only 20% took the variation of different phases into account when assessing whether there was an increase or decrease in a time series graph. If this is the case, it is necessary to examine the effects of variation carefully and devise graphing techniques that take the consequences of these effects into account.

DeProspero and Cohen (1979) examined the reliability of visual inspection in series that included a consistent mean increase. They found that, if the series had a large mean shift, low variability, and no other discernible underlying trends, there was more agreement between the raters. The raters appeared to use each of these factors in order to come to a decision about the series, which suggests that each must be taken into account when displaying time series data. In fact, the statistical orientation of the analyst affected

the judgment of time series analysts. Analysts with a background in multivariate analysis of variance tended to focus on the change in level in relation to the variability of the time series whereas behavior analysts focused on the mean change, regardless of changes in variability (Wampold & Furlong, 1981). Thus, when plotting time series data, it is important to consider the characteristics of the data set and the background of the audience. The effectiveness of a time series graph is affected by both of these factors.

A simple comparison of the performance of visual analysts and the corresponding statistical techniques does not address the different types of errors that may occur (Matyas & Greenwood, 1990; Swets, 1995). In signal detection theory, two types of errors, misses and false alarms, are possible. A miss occurs when one fails to detect a genuine change in the time series. A miss corresponds to a Type II error in statistical decision making. A false alarm is an error of reporting a change that is nonexistent. A Type I error is the statistical counterpart to a false alarm.

Matyas and Greenwood (1990) argued that it may be incorrect to assume that an error occurs when a judge declares an effect that a statistical test does not find. In both statistical and visual analyses, there is a chance that mistakes will occur. When testing hypotheses, researchers accept that a certain amount of error is unavoidable. Matyas and Greenwood argued that, if the results of statistical and visual analyses are incongruent, it

may be wrong to assume that the human observer made a mistake (either a miss or a false alarm). An alternate, and equally plausible explanation, is that an effect actually exists but the statistical test is not sensitive enough to detect the difference.

In order to examine these issues, it is possible to examine the performance of visual analysts when the effects are known to be significant or nonsignificant (Matyas & Greenwood, 1990). These researchers presented subjects with time series displays that varied in terms of effect size, variability, and serial dependency and asked them to determine if the plot represented an intervention effect in terms of level, trend, or another systematic changes. As variability and serial dependency increased, the rate of false alarms also increased while the miss rates remained relatively stable.

Typically, time-series data are presented on a standard line graph with axes representing the actual data points. Although this method is certainly common, it is not necessarily the optimal way to present this type of information. To address the efficiency of different graphing techniques, Knapp (1983) examined the assessment of change when different display methods are utilized. Knapp presented subjects with different plots of time-series data and asked them to determine if a change occurred in the series. He used three different plotting methods—cumulative plot, semi-log paper, and frequency polygon. Performance was similar regardless of the type of display used. Though these

results suggest that performance is comparable regardless of the display, further research is needed to determine if different display methods (e.g., bar graphs or dot plots) produce different judgments.

Thus, although research on time series data has often shown visual analysis to be less than optimal, much of the research on graphical perception suggests that observers are able to detect data trends (Gibson & Ottenbacher, 1988; Munger, Snell, & Lloyd, 1989; Wampold & Furlong, 1981) and mean changes (e.g., Bailey, 1984; DeProspero & Cohen, 1979; Knapp, 1983). The detection of both mean and trend shifts depends on other data attributes, such as variability (DeProspero & Cohen, 1979; Furlong & Wampold, 1982; Wampold & Furlong, 1981), serial dependency (Jones, Weinrott, & Vaught, 1978), and the presence of trend lines (Bailey, 1984).

Chapter 4

Graphical Perception of Exponential Relationships

It is nearly impossible to turn on the television or radio or read a newspaper without being exposed to some issue involving growth or decline. We are faced daily with population growth, economic growth, and stock market dips and rises (Carroll, 1997; DeSanctis, 1984). Although humans are fairly accurate at predicting changes when presented with linear trends, accuracy is much lower when exponential trends are presented (e.g., Timmers & Wagenaar, 1977; Wagenaar & Sagaria, 1975). This finding is unfortunate because many growth functions represent some type of exponential growth or decline. Thus, it is important to examine exponential relations systematically in order to determine why linear and nonlinear relationships might be perceived differently.

As Wagenaar and Timmers (1979) point out, phenomena exhibiting exponential growth can have an important bearing on human affairs. For example, population growth and changes in pollution levels often follow exponential patterns, and failures to comprehend such changes will likely prevent adequate responses to them on the part of society. Yet exponential functions have some dramatic, and probably unintuitive, features that may make them difficult to perceive accurately. At the most basic level, exponential functions describe processes in which a variable changes in a multiplicative fashion for each linear increment in time. The result of repeated multiplications is the familiar phenomenon of compounding, as in the case of interest compounding over time in a

savings account. Graphically speaking, this means that the exponential function remains fairly level for a period of time but then shows steep, positively accelerated growth. One way to convey the difficulty that exponential relationships can pose for perceivers is to consider some examples. For the example of compounding interest, a sum of money earning interest at 7% will double every ten years. Perhaps surprisingly, this means that a person who saves for retirement over a period of five decades will see their savings grow as much in the last ten years as they did in the previous forty. As Price (1963) has shown, the number of scientists in the world has followed an exponential trajectory since the 17th century, doubling every 15 years. This remarkable rate of growth is hard to comprehend in various ways. One is that such exponential growth entails the striking conclusion that about 88% of all scientists who ever lived are alive today. Perhaps equally striking is the fact that this has always been true for any other period since the inception of science.

Although many experiments have examined the factors that affect the perception of linear relationships, little has been done to determine if observers can accurately perceive exponential relationships between variables. Research on the human perception of linear relationships between correlated variables has been partly focused on determining the function (or "rule") that subjects use when asked to determine the degree of association. Similarly, much of the research on exponential relationships has focused

on determining the polynomial or exponential function that subjects intuitively use when they examine these types of curves (e.g., Jones, 1977, 1979, 1985; Timmers & Wagenaar, 1977; Wagenaar & Timmers, 1978, 1979).

Subjects consistently make systematic errors when extrapolating information from exponential curves. When presented with a series of numbers that were distributed exponentially and asked to extrapolate the function $y=a^x$, subjects took into account only a small proportion of the exponential curve and underestimated future values (Wagenaar & Sagaria, 1975). Thus, when extrapolating future data points from an exponential curve, subjects used only a small proportion of the information available to them to make judgments concerning the underlying functions. In addition, their forecasts were systematic underestimations of the actual future point.

Jones (1977, 1979, 1985) conducted a series of empirical studies designed to determine how subjects perceive exponential curves. He suggested that subjects are unable to extrapolate exponential functions because they are intuitively fitting a polynomial curve to the data, not an exponential one. Jones (1977) reexamined Wagenaar and Sagaria's (1975) data and concluded that subjects were extrapolating using a quartic polynomial. Jones calculated the quartic polynomials for Wagenaar and Sagaria's data points and found that the generated polynomial points led to extrapolations that were closer to the actual data points than did the exponential extrapolations.

One practical application of these studies focuses on the determination of whether humans can accurately forecast future events. For example, when a clinician examines a time series graph of client progress, the detection of past trends may be less relevant than the forecasting of future status. In extrapolation tasks, several factors are important. These may affect the ability of humans to forecast future events and should be taken into account before the conclusion that humans are inefficient at extrapolation tasks can be drawn.

Extrapolation of Time Series Data

A common purpose of plotting time series data is so that future points can be forecasted from past points. Forecasting is often a component of decision making processes (Ashton, 1994; Winklhofer, Diamantopoulos, & Witt, 1996), and the consequences of mistakes can be large. In the business world, for example, forecasts based on human judgments are common (Flores & Olson, 1992; Goodwin & Wright, 1994) and forecasting performance can result in the gain or loss of millions of dollars. It is possible that human judgments are better than statistical models in recognizing unique or changing circumstances (Loftus, 1993; Tufte, 1990) . If so, it is extremely important that the data available to the forecaster are presented clearly and in a way that allows quick and accurate conclusions to be drawn.

Given a set of points, can people make accurate judgments as to what the next point in the series will be? How accurate are they if they are asked for points farther in the future? Using a stock market example, when examining a time series graph, people are reasonably accurate if they are asked to forecast tomorrow's value, which should be quite similar to the current value (MacKinnon & Wearing, 1991), but if the following week, month, or year must be forecasted, the accuracy of the forecast decreases. It is possible that errors that occur during forecasting tasks become compounded over time. To illustrate, if asked to estimate how far a yard (3 feet) is, using a pace would yield a fairly accurate estimate but using the same method to estimate how far a mile (5280 feet) would lead to a much less accurate estimate (though Weber's Law suggests that the error would be proportional to the overall distance).

In time series extrapolation, the forecaster typically examines a graph or table of past values and projects future values. During the extrapolation task, different heuristics and biases can affect the forecast that is made (Bolger & Harvey, 1993). Tversky and Kahneman (1974) argued that people use different general rules, or heuristics, when predicting the future from the past. These heuristics lower the cognitive demands by reducing the amount of data that must be remembered, and the complex computations that would otherwise have to be performed mentally become easier. Though these practices lead to quicker estimates, they also lead to biases that may reduce accuracy in the estimation task.

The "anchor and adjust" heuristic is relevant in the extrapolation task (Tversky & Kahneman, 1974). When performing this task, subjects often formulate their response after they make a mental adjustment based on some predetermined value (the anchor). The anchor and adjustment heuristic can take different forms. Lawrence and O'Connor (1992) concluded that people use an estimate of the long-term mean of the series as their anchor and adjust their estimate by mentally calculating the weighted average of the anchor and the value of the last available data point. This is the averaging method. The adding method is a variation and involves using the last observation as the anchor and making adjustments by adding an estimate of trend based on the last change in the series (Harvey, Bolgar, & McClelland, 1994).

Generally, people do use anchor and adjust heuristics when faced with an extrapolation task. Bolger and Harvey (1993) examined the adding and averaging versions of the adjust and anchor heuristic to determine if their use depended on the context. If is no trend is present in the data set, the averaging method is used. If a trend is present, the type of heuristic that is employed depends upon the amount of serial dependency — with low levels, the adding method is used but if the level of serial dependence is high, subjects use a combination of the averaging and adding methods when producing estimates.

It has been suggested that subjects extrapolate data from a time series in a two-step process (Wagenaar & Timmers, 1978). First, the properties of the time series are identified (e.g., whether the series is stationary or exhibits an increase or a decrease, or accelerates or decelerates, etc.). After these properties are identified, subjects use the rules they have discovered to extrapolate the further elements of the series. Subjects make errors in both of these stages and, typically, extrapolations are lower than the actual series value (e.g., Jones, 1977; Keren, 1983; Timmers & Wagenaar, 1977). However, as will be discussed later, this underestimation can be minimized if subjects are given feedback as they make their extrapolations (MacKinnon & Wearing, 1991).

Lawrence and Makridakis (1989) presented subjects with time series graphs that had either an upward (increasing) linear trend, a downward (decreasing) linear trend, or a flat (no) trend. When asked to make forecasts concerning the increasing trend, the judgments of the subjects fell below the least squares regression line (underestimation). Forecasts for the decreasing series were higher than the regression line (overestimation). There were only slight overestimation errors when subjects forecasted the flat trend. Lawrence and Makridakis concluded that subjects "dampened" the slope for all types of trend. In essence, the rate of change (the slope of the time series) was underestimated for both increasing and decreasing series.

The tendency of subjects to make judgments that underestimate the slope of time series data has been found in many modifications of the forecasting paradigm (Mullett & Cheminat, 1995). Typical experiments present successive values and then ask subjects some variant of the question, "what will the value be if the growth continues as it is?" This rate of change underestimation occurs in a number of research paradigms and does not appear to be affected by the shape of the underlying trend (Jones, 1978; Lawrence & Makridakis, 1989), the phrasing used to explain the task, the presentation methods used to display the data (Lawrence, Edmundson, & O'Connor, 1985, 1986; Lawrence & Makridakis, 1989), an awareness of the tendency of the bias (Andreassen & Kraus, 1990), or the level of expertise of the judges (Sanders & Ritzman, 1992; Wagenaar & Sagaria, 1975).

The underestimation bias that occurs when subjects are asked to extrapolate future points in an exponential series could be due to the difficulty of the task. When subjects are asked to make a judgment about a linear time series, they have to determine the line that best describes the data. When judging exponential trends, they have to determine which curve best describes the time series. Thus, with exponential time series, the subject must first decide the shape of the function and then make a judgment about subsequent data points.

The underestimation associated with exponential functions occurs regardless of whether the data are presented as a numerical series or on a graphical display. Timmers and Wagenaar (1977) focused on decreasing trends and found that subjects were more accurate when the functions were presented graphically. Although there was an underestimation bias regardless of whether data was presented in a tabular or graphical format, the bias was greatest when a tabular format was used.

Based on the fact that there is an underestimation of exponential growth curves, it is important to determine if corrective factors can be implemented to minimize this bias. The research on exponential relationships and forecasting has focused primarily on the comparison of responses under graphical and tabular conditions. Although these are certainly the most popular methods to display time series data, there is no research examining whether they are the optimal methods. As Cleveland (1983, 1993, 1994) has pointed out, choosing an appropriate graphical display should depend upon the task that an observer is expected to carry out. Because there is no universal "best" display method, it is worthwhile to examine different displays to determine which is most appropriate for a particular decision making task. As of yet, very little research has examined the effectiveness of other graphical presentation methods (for exceptions, see Knapp, 1983 and Legge, Gu, & Luebker, 1989, who used luminance displays and found them less effective than scatterplots).

The Effects of Feedback on Extrapolation

There are two types of feedback that can be provided in a decision making task (Goodwin & Wright, 1993). Outcome feedback provides an indication of whether a judgment was correct or incorrect and cognitive feedback about task properties gives the forecaster information about the relationship between his or her estimate and the actual data point (Balzer, Doherty, & O'Connor, 1989). Thus, when subjects receive outcome feedback, they are informed whether their response was correct but when they receive cognitive feedback, they learn how their judgment differed with respect to the actual data point. In general, these researchers have found that subjects who receive cognitive feedback are more successful than those who receive outcome feedback.

The question of whether feedback can lead to more accurate judgments is important and should be addressed. Though little research has been done in this area, MacKinnon and Wearing (1991) examined the effects of feedback on the ability of subjects to forecast exponential change. These researchers recognized that the problems people have with interpreting exponential relationships correctly could be due to the fact that they do not usually receive feedback during the extrapolation task.

In typical extrapolation tasks, subjects are shown a set of data points, presented either graphically or in tabular form, and are asked to forecast one or several points in the future or to determine when the curve will reach a certain point. MacKinnon and Wearing (1991) pointed out that, in everyday situations, people do not extrapolate in the

absence of feedback. For example, when examining a time series display of the stock market, the information concerning the previous extrapolation is generally known before it is necessary to forecast to the next point. Even when it is necessary to forecast farther into the future, feedback is eventually available and can aid in the decision making process.

The data sets used by MacKinnon and Wearing (1991) were functions with 100 data points and a growth rate of 6%. In this case, subjects were presented with a single data point and asked to extrapolate to the value of the series in the next time period. Two types of feedback were provided for each extrapolation: feedback concerning the actual value of the next series point and how far their estimate was from this value. After subjects made their last forecasting decision, they were required to identify the function they had seen (they picked out the function from 30 possibilities). The results showed that subjects were quite accurate at forecasting but there was a slight tendency towards underestimation. However, subjects were not able to identify the series that they had seen (only 25% of the subjects were able to pick out the series that they had seen). When feedback was delayed (by having subjects forecast beyond the next data point before receiving feedback), performance declined. These researchers concluded that forecasting performance is optimal if feedback is immediate but the ability to identify the trend that

was forecasted is extremely limited for subjects. Given that this study is the only investigation to systematically examine the effects of feedback on performance, it is obvious that the effects of feedback should be examined further.

Although MacKinnon and Wearing (1991) examined how feedback affects subject extrapolation accuracy, there are several problems associated with their methodology. The design of the study was ecologically invalid; subjects were presented with a single numerical value and asked to make a judgment concerning the next data point in the series. Furthermore, the data presented to subjects did not include random variability. The between-subjects design was inefficient as subjects were only presented with a single trend type. Information is not presented in this way in the "real-world." Time series predictions are typically based on a graphical display of several, noisy data points. The underlying trend of the series often varies and the reader must determine which trend is present before decisions concerning the value of future points can be made.

Graph Type and the Detection of Exponential Trends

As noted earlier, several questions arise when one considers how humans best decode the information contained in a visual display. One pertinent question involves the display method used to present the data. Do some types of graphs lead to greater accuracy in trend discrimination?

The decision as to which display method to use when presenting data should be driven by what is expected of the reader. Bar graphs are particularly useful if observers are required to estimate specific quantities (Culbertson & Powers, 1959) but line graphs may be optimal if subjects are required to identify patterns in the data set or predict future trends (Kosslyn, 1994). Hence, it appears that the optimal graphic depends upon the task that is required of the observer.

Wallgren et al. (1996) suggested that bar graphs are only appropriate for time series data if a small number of points are displayed, but line graphs are acceptable regardless of the number of points that must be displayed. Although this point of view is generally accepted, other researchers have found that the type of trend present in the data should drive the selection of graph type—whereas line graphs best convey linear trends, bar graphs and suspended bar graphs were superior if the underlying data trends were followed step-increase or sine-wave patterns (Best, Smith, Frey, & Stubbs, 1998). Currently, no research has examined whether line graphs are optimal to convey exponential relationships. It is possible that the degree of subject underestimation is affected by the type of display that is used to present nonlinear data. Although line graphs may be optimal to display linear time series, alternate display methods may best convey nonlinear trends.

Using Motion Cues to Improve Extrapolation Ability

Another issue that could be addressed when determining an optimal display method is the manner in which each data point is presented. Typically, when one is presented with a graph, all of the data points are laid out on a grid. This presentation may not allow an observer to fully appreciate the underlying trends. Is a static graph the best way to present this information? The research on exponential growth, extrapolation, and forecasting has typically presented a series of points simultaneously and asked subjects to make decisions based on this presentation. It is possible that this presentation method hampers the ability of subjects to recognize the trend inherent in the time series.

Perceived motion is an important component of everyday life and affects how observers perceive different situations (Lappin, 1995). Two things must be considered if graphical perception is to be understood—the correspondence between graphical components and the resulting visual image and the visibility of the scene. Typical experiments on the visual perception of graphical displays only take into account the correspondence between the properties of the display and the image that is formed by those properties. Lappin (1995) argued that visual psychophysicists must also consider the visibility of a scene. Visibility refers to the visual sensitivity to the specific image properties that give information about an entire scene. Most importantly, the visual system uses motion as a source of information and, thus, graphical displays that take this into account may prove more effective.

The path that the eyes follow when scanning an image is referred to graphical movement (Tannenbaum, 1998). This movement is not random—in our culture, most people tend to scan from left to right and top to bottom, in a manner similar to the way that most written works are displayed (Goldstein, 1996, 1999). Thus, it is possible to create displays that direct vision along a certain path. Typical time series graphs present all data points simultaneously, but changing the presentation of the points by presenting the data points in succession could improve both the detection of trends and forecasting accuracy.

One possible variant of the static time series display is a display that simulates motion. If the points in a time series were presented one at a time so as to simulate motion the trend in the data may be easier to decipher. This added dimension may give subjects information that could enhance their ability to make more accurate extrapolations and forecasts. Lyon and Waag (1995) examined the ability of subjects to predict the future position of an object relative to an observer. Because this "visual extrapolation" is used in everyday situations, subjects may be more adept these tasks. This presentation may cause the subject to perceive motion and make it easier to detect the degree of trend present in the data.

Lyon and Waag (1995) postulated that subjects use motion to make forecasting judgments using a two step process. During the first step, the target is tracked while it is visible and, in the second step, the tracking continues even after the target has

disappeared. As has been found with time series extrapolation, subjects are more accurate if they are asked to extrapolate points that are in close proximity to the point at which the motion disappeared (e.g., Lyon & Waag, 1995; MacKinnon & Wearing, 1991). Lyon and Waag concluded that subjects are able to make accurate forecasts when tracking a moving object. Based on these results, it is possible that a more dynamic presentation may lead to more accurate judgments.

Wagenaar and Timmers (1979) raised a valid point concerning the graphical and tabular display of growth functions—these displays accurately present the data but omit the real-time component associated with the growth. Exponential functions (both increasing and decreasing) have an initial relatively stable period followed by a rapid period of growth or decline. In contrast, asymptotic growth and decline functions begin with an initial, rapid increase or decrease, followed by a stable period. The static presentation of these functions may impair the ability of subjects to perceive the rapid increases or decreases and thus, lead to the tendency towards underestimation.

Wagenaar and Timmers (1979) employed an interesting display method that better approximates how exponential growth is perceived as observers experience real-time change in their environment. They presented subjects with a empty large square representing a "pond". The square began to fill with points ("duckweeds"), which increased in number at an exponential rate (this simulating a natural growth process). When the duckweeds stopped appearing, subjects had to indicate how much longer it

would take for the entire pond to be filled. Wagenaar and Timmers found that subjects vastly overestimated the time required to fill the pond, indicating that they were <u>under</u>estimating the rate of duckweed growth. Thus, the underestimation bias that occurs when subjects are presented with graphical and tabular formats to predict future data points was also found when subjects were presented with this paradigm. It appears that when extrapolating future data points of a data set with an exponential trend, there is a tendency towards underestimation of growth rates and this bias is present even if a variety of presentation methods are used.

Given that subjects are generally not accurate when asked to extrapolate future data points in a series of exponential data, it is important to examine some alternate display methods. Wagenaar and Timmers (1979) examined an important issue (real-time growth) in their study of the perception of exponential growth, although the real-time display method they used failed to improve subjects' judgments. Perhaps, these results were due to the unfamiliarity of the display method. Presumably, the display method was novel to subjects and this could have affected their judgments (Kosslyn, 1992; Lohse, 1993; Pinker, 1990). It is possible that judgment accuracy may improve if subjects are presented with familiar displays that present the data in a more dynamic manner. This may represent a compromise between typical, static displays and the real-time displays used by Wagenaar and Timmers.

The Effects of Random Variability

When noise is introduced into a data set, it becomes more difficult to detect underlying data trends. Variability can mask the true trend and lead to the identification of false trends (Goodwin & Wright, 1994). Typically, the performance of human observers decreases as the level of variability in the data increases, in comparison with the performance of statistical tests (Sanders, 1992). Perhaps observers mistake the underlying noise for systematic changes and this leads to extrapolation errors. Other researchers have found that subjects were able to "see" through noisy data and could detect the presence of trends in spite of random variability (Best et al., 1998). Although performance was optimal if the levels of random variability were low, subjects were able to detect trends in noisy data. Given that extrapolation is difficult even if the data sets contains no variability (the trend is a smooth function), an extrapolation task becomes even more difficult when subjects are presented with a time series display of noisy data (O'Connor, Remus, & Grigg, 1993).

The information that is encountered in everyday situations always contains some degree of variability. Given this inherent trait, it is important to examine how observers perform with different levels of variability in a data set. The addition of random variability enhances the difficulty of perceiving real-world trends and this makes it important to discover effective ways of presenting these types of trends. If it is true that trend discrimination and extrapolation become more difficult as random variability

increases, research investigating these effects may lead to ways to minimize these problems. To date, no research has systematically examined the effects that variability has on the ability of subjects to discriminate among different trend types and extrapolate future data points.

The Effects of Sample Size

It is well accepted that having more information available enhances the ability of subjects to determine the trend that is present in a graphical display (Best, Smith, & Stubbs, 1998; Best et al., 1998, 1999; Lewandowsky & Spence, 1989). Research examining the effect of including more points on a scatterplot (or more bars in a bar graph or suspended bar graph) suggests that the addition of data points increases accuracy and improves the ability to detect trends and mean differences between two samples of data. The power of both intuitive and statistical tests generally increases as a function of sample size. As sample size increases, more information becomes available and this enhances the ability to detect changes. Thus, at least in the case of graph perception of trends, "more is better".

These findings have not been extended to the perception of exponential trends and judgment tasks (Goodwin & Wright, 1994). Wagenaar and Timmers (1979) found that the underestimation bias was greater if more data were presented to subjects. In a classic study, Wagenaar and Timmers (1978) examined the biases associated with exponential growth. These researchers presented subjects with a number of data points (n=3, 5, or 7)

and required subjects to determine the value that would be reached after another k years if the process went on uninhibited. The judgments of subjects were greater (and hence more accurate) when fewer data points were presented, but they were not more sensitive in identifying the growth function that was presented. Wagenaar and Timmers argued that series with fewer data points impart the message that the next value should be large. The results of this study are consistent with Wagenaar and Timmers (1979)—extrapolations improve if fewer data points are presented.

Andreassen and Kraus (1990) proposed the salience hypothesis to explain why there is less underestimation when fewer data points are presented. This hypothesis can be illustrated using an example from Wagenaar and Timmers (1978). When five data points (3, 7, 20, 55, 148) were presented there was more underestimation than when three data points (3, 20, 148) were presented. The salience hypothesis explains the greater degree of underestimation in terms of the larger differences between the individual data points when fewer points exist. In the case of exponential growth, the steep increase is more salient if fewer points are present because there is a larger difference between the individual data points. These large differences lead subjects to interpret the series as growing more quickly. When presented with a series that contains more data points, the steep exponential increase is not as noticeable because the magnitude of the differences between the individual data points is not as great. As an example, consider a plot of population growth. If individual days (or even months were plotted), it would be obvious

that the population was increasing but the rate of growth would be partially obscured by the number of days plotted. If, instead, the plot included yearly (or decade) averages, the differences between the individual points would be larger and this could lead to a more accurate perception of population growth. If this explanation is true, increases in the number of data points presented should not affect the ability of subjects to detect a trend but may affect the perception of the rate of underlying growth.

Thus, in the case of graph perception, two processes are at work. When more data points are presented, the efficiency of both human estimates and corresponding statistical analyses increase and it is easier to detect data trends and differences between two data sets. This seems incongruent with the findings of Wagenaar and Timmers (1978, 1979). These researchers found that increasing the number of data points actually increased the tendency of subjects to underestimate the rate of growth. Thus it seems that the perceptual processes underlying trend discrimination are different from those that operate during forecasting tasks. Given this possibility, it is important to examine the effects of sample size systematically in order to determine if the optimal number of data points is dependent upon the underlying task.

Different Types of Exponential Functions

Research on linear relationships has generally led to the conclusion that estimates are more accurate if the correlation between two variables is positive than if it is negative (Erlick & Mills, 1967). When presented with a scatterplot having a negative linear trend,

the ability of subjects to estimate the degree of covariation between the variables decreases. Much of the research on exponential relationships appears to contradict this finding (Goodwin & Wright, 1994), as researchers have concluded that subjects are more accurate if the exponential series is decreasing. Although the research appears convincing, it is possible that there is one simple explanation—ceiling and floor effects.

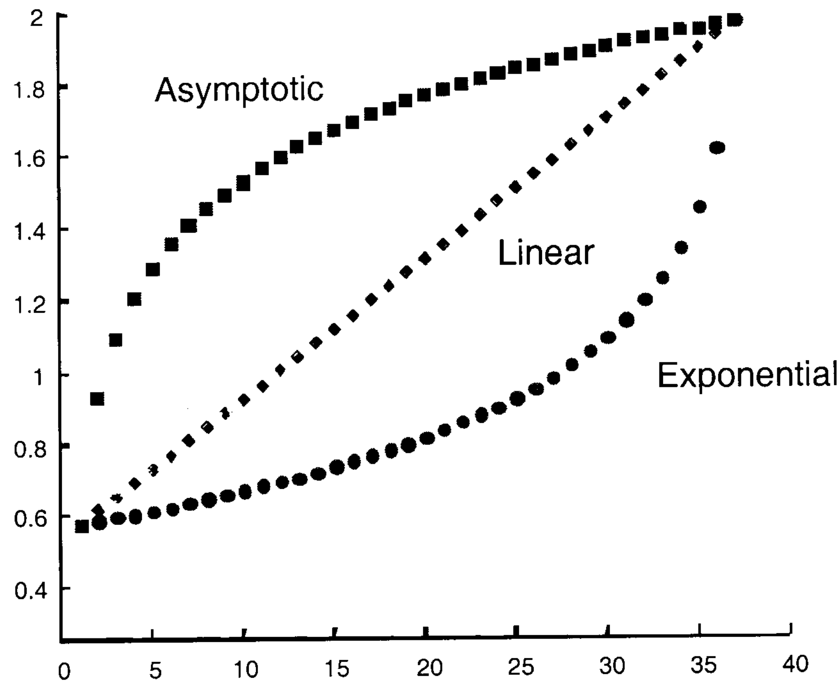
When subjects examine a time series display of data points representing an exponential relationship and are asked to extrapolate to future points, their judgment may be an underestimation simply because there is no absolute maximum point in the data series. On the other hand, a function that is declining exponentially does have a minimum point—zero. Thus, the decrease in the underestimation bias that occurs when subjects extrapolate points in a data set that decreases exponentially may simply be due to the fact that decreasing series have a concrete minimum point and this absolute minimum creates a floor effect in the extrapolations.

Simple exponential functions can be represented by one of four possible functions (MacKinnon & Wearing, 1991). Exponential growth functions begin at some minimum point, initially increase slowly, and, at some point, rapidly begin an upward increase. Exponential declining functions begin at some maximum point, followed by a period of slow decrease, and at some point begin to fall rapidly. Asymptotic growth functions have

an initial, rapid growth period, followed by a stage of relative stability, where little growth occurs. Asymptotic declining functions initially surge downward and level off at some minimum value (see Figure 5).

In order to determine whether the increased accuracy that occurs with decreasing functions is simply due to a floor effect, it is possible to vary the placement of the data points on the graphical display of curve so that subjects will have a wider range of extrapolation options.

MacKinnon and Wearing (1991) examined the ability to extrapolate future data points for each of these trend types but used an ineffective design. Each subject only extrapolated points for one type of trend. Other researchers have not systematically examined whether discrimination and extrapolation accuracy is affected by the underlying trend. One purpose of the current experiments was to systematically examine how people respond to different type of trends.



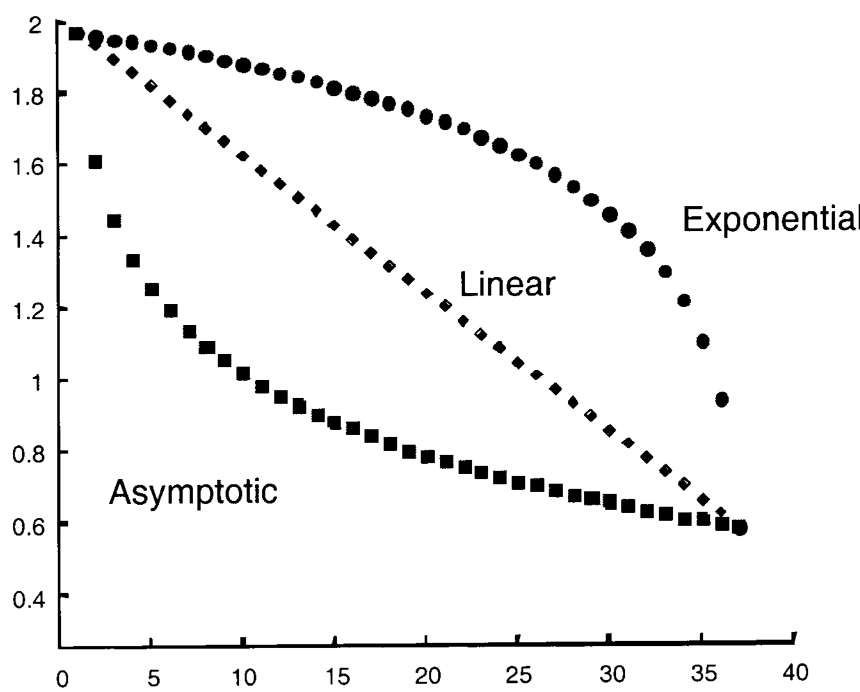


Figure 5. Four types of nonlinear change.

Statement of the Problem

The present experiments were designed to investigate several of the issues involved in the perception of nonlinear relationships. Real-world data sets are not perfect but include variability, have different sample sizes, and represent different underlying trends. Given this, it is important to realize that, when data sets are plotted, the resulting graphs will be inherently different. It is possible that the optimal display method differs according to the characteristics of a specific set of data. The purpose of these experiments was to integrate several independent variables in order to determine how each factor affects extrapolation accuracy and discrimination ability.

The amount of variability in a data set affects the ability of subjects to make inferences and conclusions about the intended message. The effects of variability were examined systematically to determine how different levels of variability affect the accuracy of subject judgments. The number of data points displayed on a graph also affects extrapolation ability. Some researchers have found that subjects are more accurate at making judgments when more data points are presented, but others have found that presenting more data points does not enhance extrapolation accuracy. Thus, another purpose of these experiments was to examine the effects of sample size on both trend discrimination and extrapolation ability.

When graphics are used for presentation purposes, the type of trend conveyed in a time series display should drive which type of graph is used to represent the data. Although line graphs are often used to convey trend information, researchers should not assume that they are equally effective in all situations. Bar and suspended bar graphs have been found to better convey step and sine trends (Best et al., 1998), which suggests that different data characteristics should be taken into account when making decisions about which graphical methods are optimal. MacKinnon and Wearing (1991) examined the ability of subjects to detect the presence of four different trend types (exponential increase, exponential decrease, asymptotic increase, and asymptotic decrease) and found that subjects underestimated rates of change. Therefore, another purpose of the present experiments was to examine whether the type of trend affects the ability of subjects to make judgments about data presented graphically.

When subjects receive feedback, they can use this information to adjust their subsequent extrapolations. Feedback was manipulated in Experiment 3 to determine if subjects are able to use this information to adjust their forecasts and increase their accuracy. In order to examine the effects of feedback under a variety of conditions, it was included as a variable crossed factorially with other variables. In addition, it is possible that optimal display methods have dynamic components, such as motion. Thus, Experiments 4a and 4b examined discrimination and extrapolation accuracy with displays that included motion.

This set of experiments represents an attempt to determine how people perceive linear and nonlinear relationships. Every day we are forced to determine the processes that underlie many diverse situations and we often make predictions based on the information that we encounter. This information is not perfect. When people make predictions about naturally occurring phenomena, they must make decisions about the type of trend that is present, regardless of how noisy the data is or how many data points are presented. Once they determine trend type, they can extrapolate future points. The type of graph used to display data should be data-driven and, therefore, an overall goal of these studies was to determine if different graphs are optimal in different situations.

Chapter 5

Experiment 1: The Discrimination of Trends

As noted earlier, much of the research on growth and decline functions has focused on how accurately subjects can extrapolate future data points. Although this research is important and interesting, it is also important to assess subjects' ability to discriminate between different types of curves. Even in an extrapolation task, the first step is presumably to detect the presence of an underlying trend. After an individual makes a decision concerning the trend, it becomes possible to make judgments concerning future points. At this point, different strategies can be used to make those judgments.

The purpose of this experiment was to determine how well subjects can discriminate between six different trends (linear increase, linear decrease, exponential increase, exponential decrease, asymptotic increase, asymptotic decrease). Examples of the nonlinear trends were shown in Figure 5. Exponential trends begin with relatively little change followed by a sharp increase or decrease. Asymptotic trends are the mirror images of the complementary exponential trends and begin with sharp growth or decay followed by a period of leveling off.

To date, no research has systematically examined the ability of subjects to discriminate between different exponential, asymptotic, and linear trends. In the present research, the different trends were presented in four different kinds of graph—line graph,

histogram, scatterplot, and suspended bar graph. In order to determine discrimination performance under different conditions, sample size and variability were manipulated. Each trend type was presented with 9, 17, or 33 data points. Variability was determined by calculating a standard normal deviate from a population with a mean of 0 and a standard deviation of 1.and transforming that to a deviation value to be added to or subtracted from the underlying curve point. For low variability, the standard normal deviate was multiplied by .22, for moderate variability the standard normal deviate was multiplied by .44, and for high variability, the standard normal deviate was multiplied by .66. The result was a 3 (sample size) x 3 (level of variability) x 4 (graph type) x 6 (type of trend) repeated measures design.

Subjects were expected to be able to discriminate between the linear, asymptotifc, and exponential trends. In addition, subjects were expected to discriminate the direction of the trend (increasing and decreasing). For all curves, discrimination was expected to be highest when the sample size was high and the variability was low. Based on previous results, it was hypothesized that histograms and line graphs would lead to more accurate discriminations than suspended bar graphs or scatterplots.

Subjects

Six subjects participated in the experiment. As is typical with psychophysical experiments, the number of subjects was small, with each subject making a large number of perceptual judgments.

Apparatus

A PC computer (with a screen resolution of 640 by 480 pixels) was used to present the stimuli to the subjects, provide feedback, record subject responses, and compute necessary statistical information.

Method

Subjects were presented with a graphical display on a computer screen and asked to identify the underlying trend in the series. Because subjects had to discriminate between different types of trend, it was necessary to familiarize them with the different functions. Before the experiment, subjects were shown examples of the underlying curves (with standard deviation = 0) and samples of the curves with added variability that they were to make judgments about (see Appendix 1 and 2).

This sort of task is one that signal detection theorists refer to as identification with "signals known exactly." The use of signals-known-exactly arguably limits the ecological validity of the task, but real-world situations often involve an a priori restricted range of trend types to be detected (as when certain cases of population growth are known in advance to be either exponential or logistic). Also, despite the fact that overall levels of performance may be enhanced by prior familiarity with potential trends, it was expected that this familiarity would not qualitatively alter the effects of such variables as sample size and variability on performance. Furthermore, the incorporation

of variability in the curves introduced considerable difficulty into the task even when the curve types were known ahead of time. This can be seen in the examples of graphs shown in Appendix 2, where different graph types and levels of variability are illustrated.

To prevent response biases, subjects were instructed that their responses should be distributed equally among the six different trends. In an effort to familiarize subjects with the procedure and the proper allocation of responses, they were given repeated practice sessions until they were able to distribute their responses nearly equally. At the end of each practice session, subjects recorded the number of times they used each response type based on information displated on the screen.

At the beginning of each session of 54 trials, subjects were prompted to enter the session number, their initials, the sample size, the level of variability, and the type of graph to be used in the session. Within each session, the type of trend varied. The order of the sessions was partially randomized so that graph type and sample size varied within blocks of sessions at each level of variability. Each subject received a different ordering of conditions. This partial randomization was done to control for possible carry-over effects between the conditions. On each trial, subjects identified the underlying trend by using a mouse to select from buttons representing the six different types of trend. Each display remained on the screen until a response was made. The computer recorded reaction time data. Following each response, subjects received auditory feedback (a brief tone for a correct response and no tone for an error) and automatically moved on to the

next trial. At the end of each session, the computer presented subjects with the percentage of correct discriminations and their distribution of responses. After completing six sessions, they summed their response types and were instructed to alter their distribution of responses if the sums were not approximately equal. Each subject completed 36 sessions.

Results and Discussion

Overall Results

The purpose of this experiment was to examine the effects of variability, sample size, graph type, and trend type on discrimination performance. In order to assess the discrimination accuracy of subjects, a 3 x 3 x 4 x 6 repeated measures analysis of variance was conducted. There were significant main effects for variability, $F(2, 10)=109.30$, $p<.0001$, sample size, $F(2, 10)=4.31$, $p<.05$, graph type, $F(3, 15)=9.15$, $p<.001$, and trend type, $F(5, 25)=8.31$, $p<.0001$. There were also significant interactions between sample size and graph type, $F(6, 30)=2.49$, $p<.05$, sample size and trend type, $F(10, 50)=5.36$, $p<.05$, graph type and trend type, $F(15, 75)=1.98$, $p<.05$ and variability and graph type, $F(6, 30)=4.33$, $p<.05$.

The type of graph used to display the data affected the accuracy of subjects (see Figure 6). Accuracy was highest when subjects were presented with line graphs (accuracy=66%) or scatterplots (65.5%) and was lower when subjects were presented with histograms (59%) or suspended bar graphs (57.5%). Post hoc tests revealed that

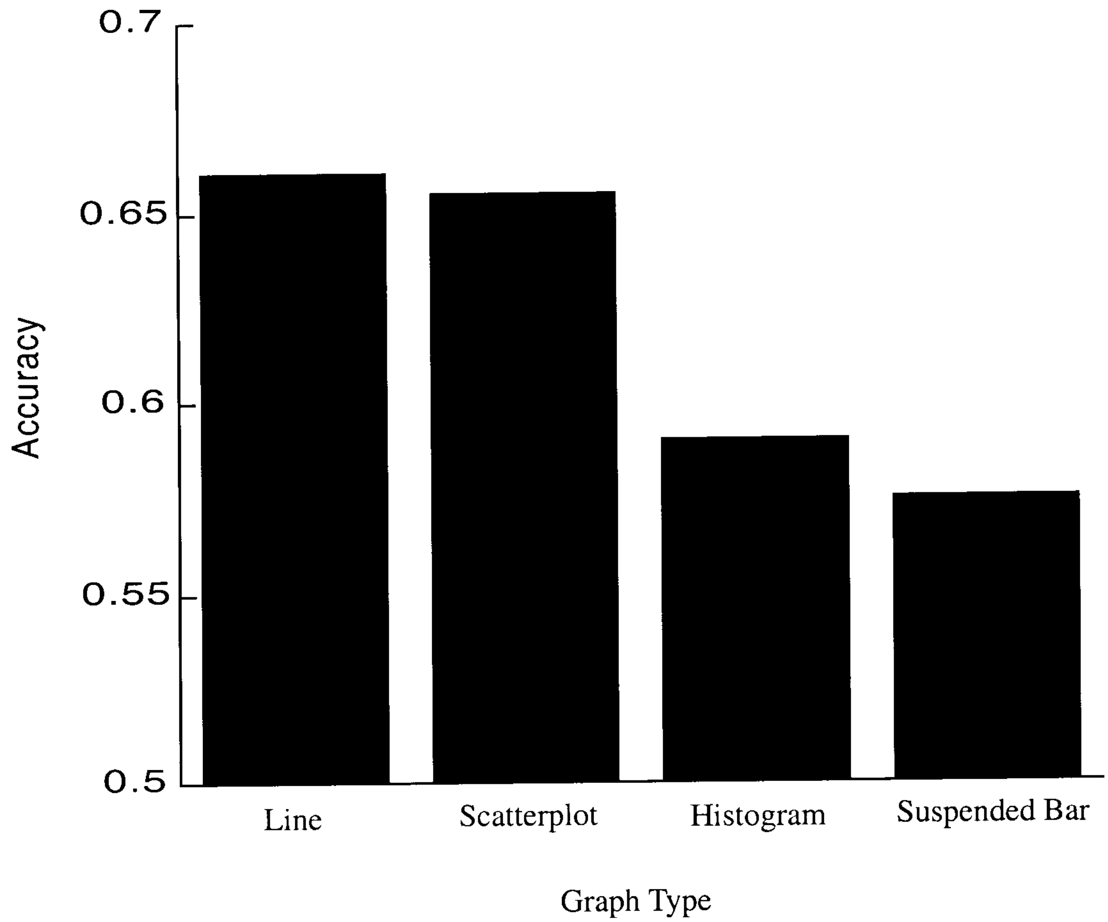


Figure 6. The effects of graph type on accuracy.

line graphs and scatterplots led to higher accuracy than histograms and suspended bar graphs (the results of all post hoc tests are presented in Appendix 3). There were no significant differences between line graphs and scatterplots or histograms and suspended bar graphs. Thus, the hypothesis that line graphs and histograms would lead to higher accuracy was only partially confirmed.

When variability was low, subjects were correct on 76% of discrimination trials but accuracy dropped to 48% on trials when variability was high. Post hoc tests revealed that there were significant differences between all levels of variability. Accuracy was

highest when variability was low, intermediate when the variability was moderate, and lowest when the variability was high. Thus, as variability increased, the ability of subjects to discern trends in the data dropped and they made more discrimination errors.

The number of data points displayed on a graph also affected subject accuracy. Accuracy was highest when sample size was high and lowest when sample size was small. Post hoc analyses revealed that subjects were more accurate when the graphs displayed 33 data points and performance dropped significantly when the sample size was 9. There were no significant differences between either the low and intermediate or high and intermediate sample sizes. Overall, performance was best when variability was low and the graph contained 33 data points.

Although it was expected that larger sample sizes would lead to higher accuracy when the variability was high, increasing sample sizes were not expected to lead to increases in accuracy when the variability was low. However, there was no significant interaction between these variables, $\underline{F}(4, 20)=1.23$, $\underline{p}>.05$. Figure 7 shows the relationship between variability and sample size. As can be seen from the figure, the effects of sample size were fairly constant across all levels of variability. When the variability was low, accuracy was highest when sample size was high and decreased as sample size decreased. At moderate and high levels of variability, the differences between the three sample sizes were smaller.

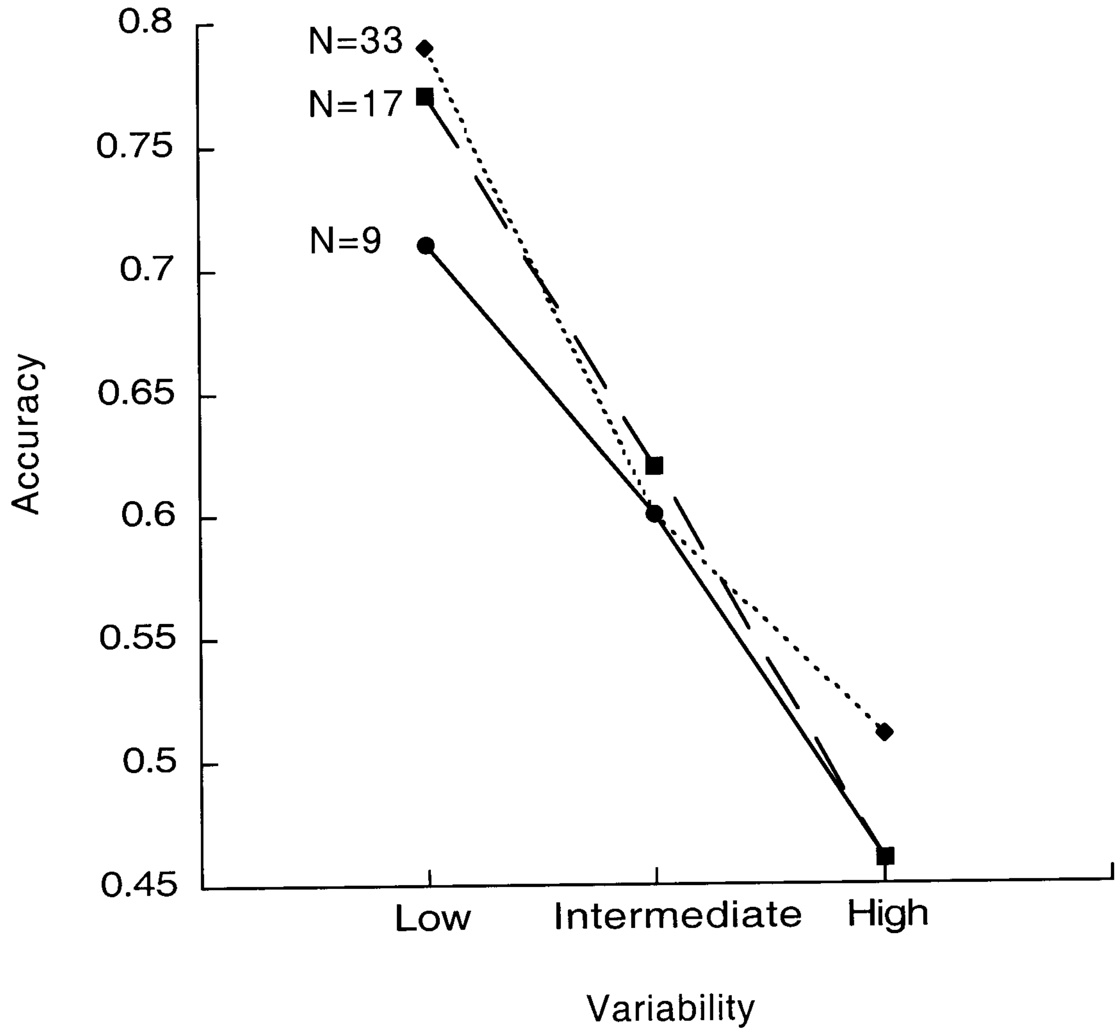


Figure 7. The effects of sample size at each level of variability.

The type of trend affected discrimination accuracy. Accuracy was higher when subjects were presented with nonlinear trends (asymptotic or exponential) and fell when subjects were presented with linear trends. Accuracy means and standard deviations are shown in Table 3. Although accuracy was slightly higher when asymptotic trends were presented, post hoc tests revealed no significant differences between the four nonlinear graph types. Discrimination accuracy was lower when linear trends were presented, and post hoc tests revealed significant differences between each linear and nonlinear trend. Accuracy was similar for each exponential–asymptotic trend pair. In addition, there were no differences in accuracy that could be accounted for by the direction of the trend. For each curve performance was similar, regardless of whether the trend was increasing or decreasing. Figure 8 shows the effects of trend type on accuracy. As can be seen from the figure, the direction of the trend did not affect discrimination accuracy.

Table 3

Accuracy for the Six Types of Trend

Trend	Mean	Standard Deviation
Decreasing Exponential	.672	.055
Increasing Exponential	.673	.088
Decreasing Asymptotic	.637	.142
Increasing Asymptotic	.670	.128
Decreasing Linear	.553	.083
Increasing Linear	.521	.088

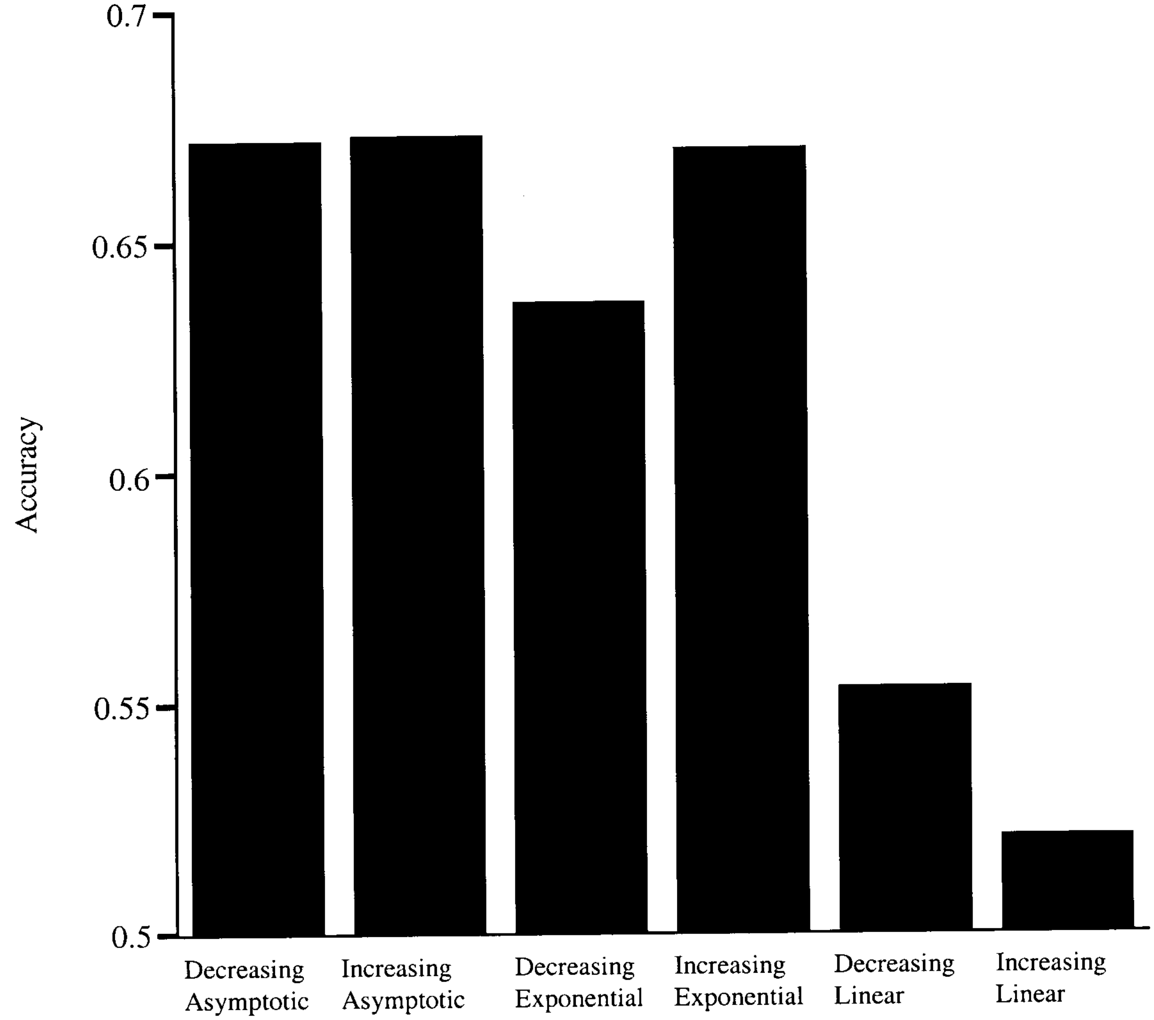


Figure 8. Subject accuracy for the six trend types.

Figure 9 depicts the interaction between sample size and graph type. Regardless of the sample size, performance was higher for line graphs and scatterplots and lower for histograms and suspended bar graphs. As can be seen in Figure 9, with histograms, line graphs, and suspended bar graphs, performance was fairly similar regardless of the number of data points displayed on the graph. Post hoc tests revealed that, for these types of graphs, sample size had no significant effect on discrimination accuracy.

When scatterplots were used, discrimination accuracy increased as sample size increased. Accuracy was significantly lower for small data sets and was higher when the sample size was high or intermediate. Accuracy was similar for intermediate and large

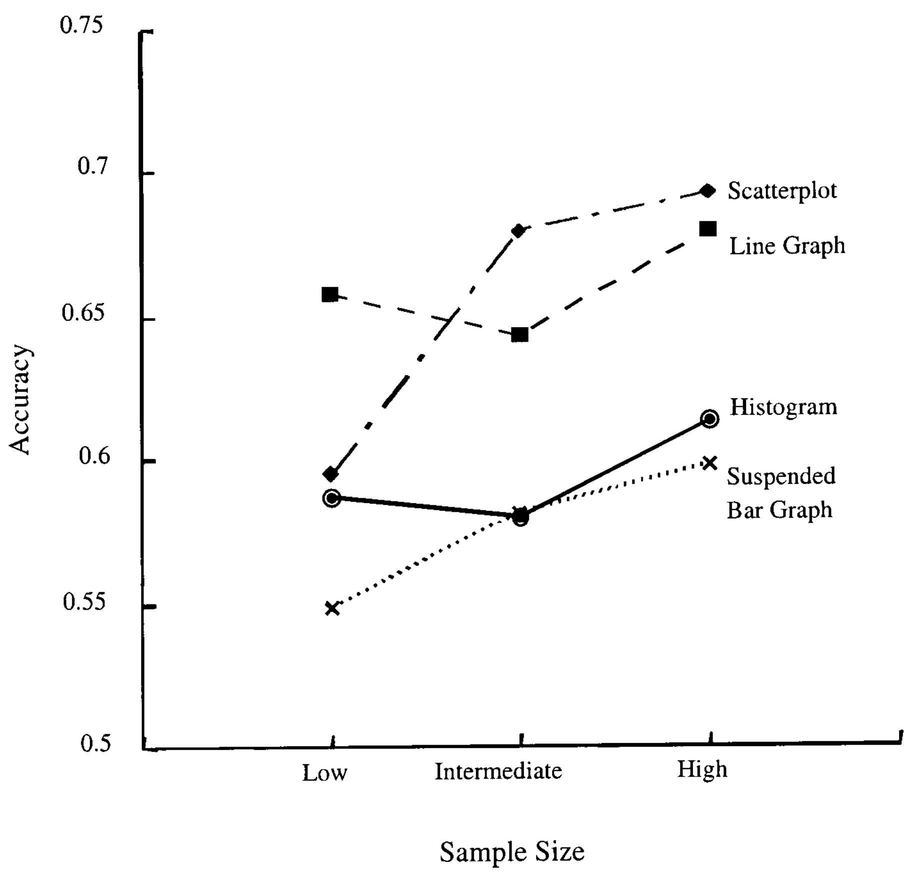


Figure 9. Accuracy as a function of sample size and graph type.

data sets. Thus, although increases in sample size did not lead to increases in performance when histograms, line graphs, and suspended bar graphs were used, larger data sets did lead to higher accuracy than smaller data sets in the case of scatterplots.

There was also a significant interaction between sample size and trend type (see Figure 10). For all four nonlinear trend types, sample size differences had no significant effect on discrimination accuracy. For linear trends, performance was best if the sample size was high. When these trends were presented, accuracy was lowest if sample size was low and increased systematically as sample size increased. With increasing linear trends, there were significant differences across all levels of sample size—low sample sizes led to poorer performance than intermediate sample sizes and intermediate sample sizes led to poorer performance than high sample sizes. With decreasing linear trends, accuracy was higher when the sample size was high and was significantly lower when the sample size was intermediate or low. For this trend, there were no significant accuracy differences between the low and intermediate sample sizes (see Figure 10a). Regardless of trend direction, sample size did not affect performance for nonlinear trends. Figure 10b shows the accuracy of each trend type as a function of sample size, regardless of whether it was increasing or decreasing. Thus, it appears that with linear trends, accuracy tends to improve as sample size increases.

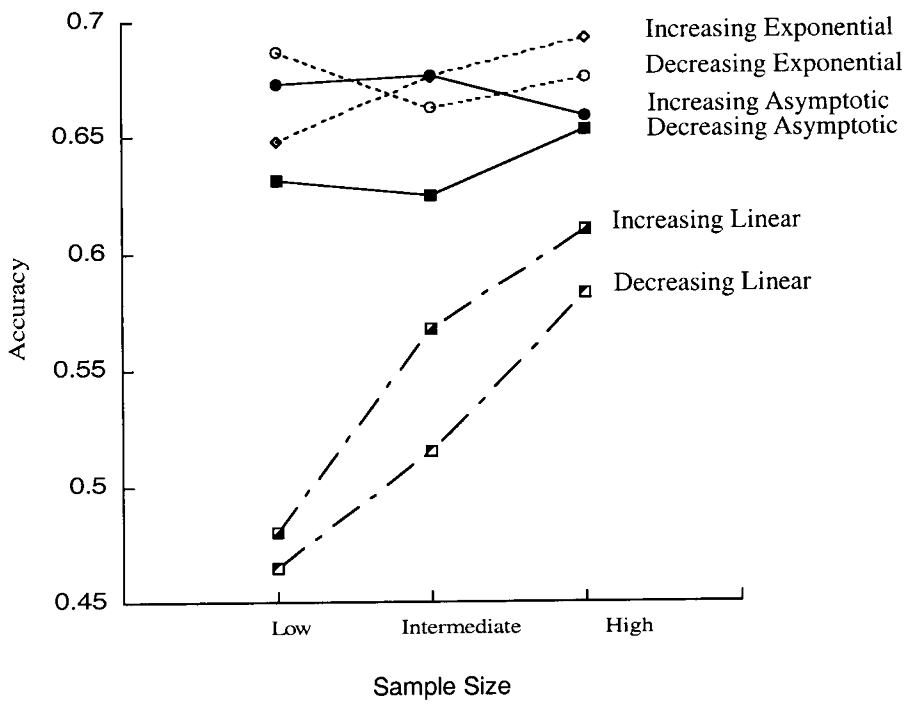


Figure 10a. Accuracy as a function of sample size and trend type.

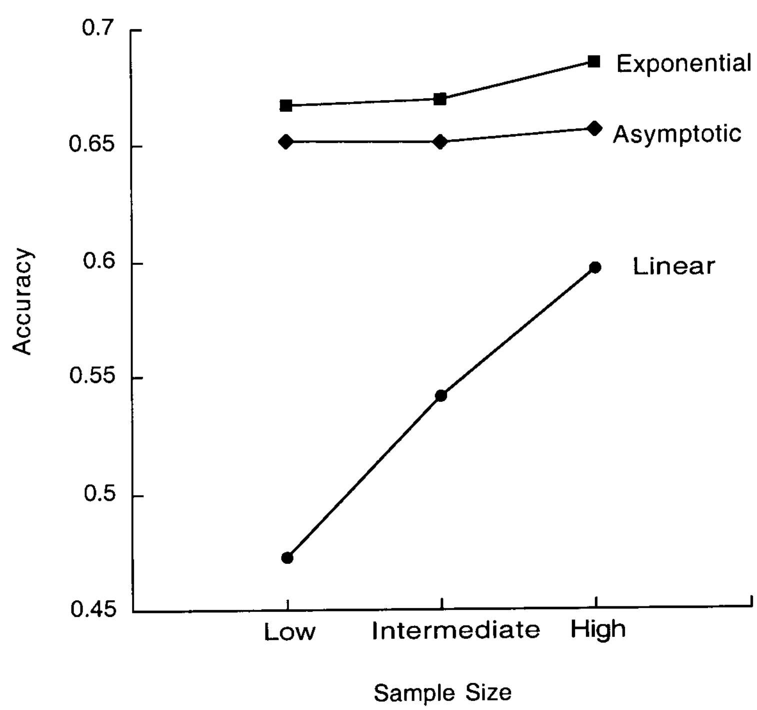


Figure 10b. Accuracy for asymptotic, exponential, and linear trends, as a function of sample size.

There was also a significant interaction between trend type and graph type. As can be seen in Figure 11a, when subjects were presented with decreasing asymptotic trends, accuracy was significantly higher when line graphs and scatterplots were presented and was lower when histograms or suspended bar graphs were used. When subjects were presented with increasing asymptotic trends, accuracy did not depend on graph type. Overall, for asymptotic trends, accuracy was highest when line graphs and scatterplots were used to present the data (see Figure 11b). When decreasing exponential trends were presented, accuracy was similar for each of the graph types. When increasing exponential trends were presented, accuracy was significantly higher with line graphs. Overall, for exponential trends, accuracy was similar for the different types of graphs. Because performance was significantly better when line graphs were used to present increasing exponential trends, they may be the best way to present exponential trends (see Figure 11b). When increasing linear trends were presented, accuracy was highest with line graphs or scatterplots and was significantly lower with histograms or suspended bar graphs. For decreasing linear trends, performance was significantly higher when line graphs were used to present that data. Thus, for linear trends, performance was significantly better with line graphs than when other graphical formats are used (see Figure 11b).

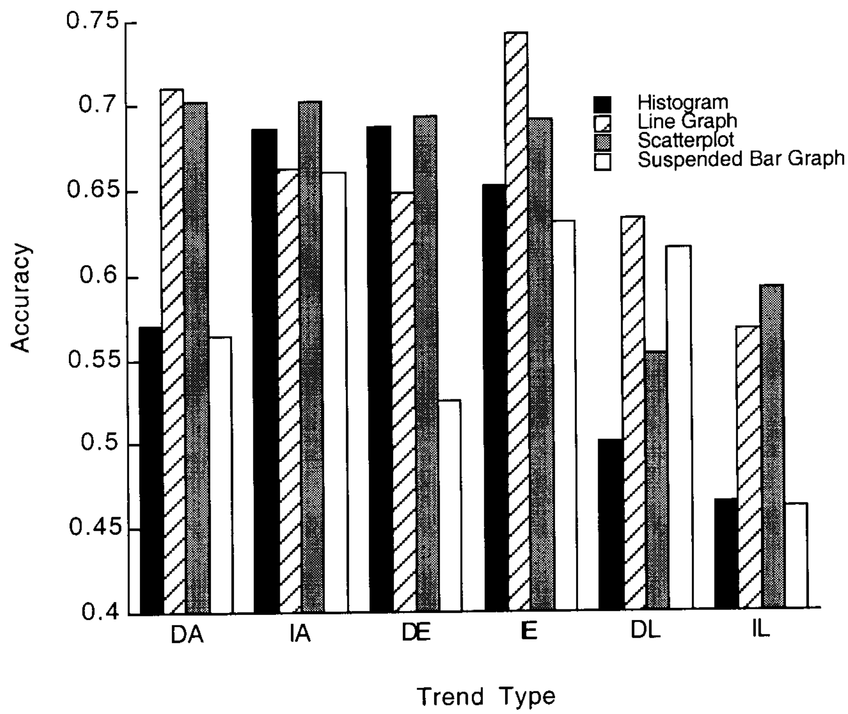


Figure 11a. Accuracy of the six trend types as a function of graph type.

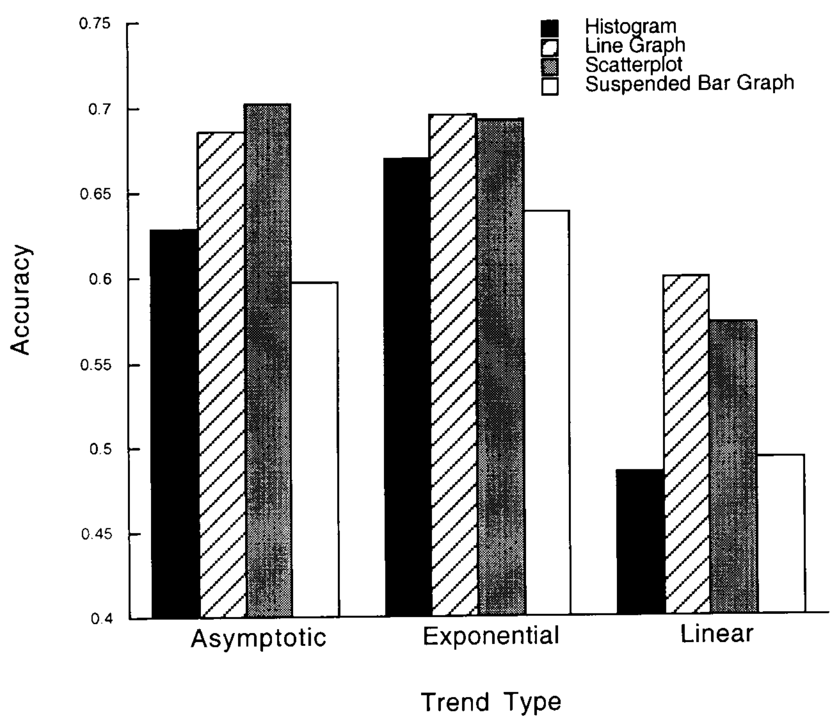


Figure 11b. Accuracy of the three trend types as a function of graph type.

To further examine the effects of the experimental conditions on discrimination accuracy, an ideal observer analysis was conducted, using all 11,664 of the trials that were presented to subjects. The ideal observer is a theoretical observer that makes optimal use of the data (Legge et al., 1989). In this case, the ideal observer response is defined as the underlying curve on each trial that best fit the displayed curve points (the curve with the least squared residual).

As can be see in Figure 12, the performance of the ideal observer decreased as variability increased and sample size decreased. On trials with low variability, performance was equally high regardless of the sample size (perhaps because of a ceiling effect), but as variability increased, the effects of sample size became apparent. On trials with high variability, the effects of sample size were more pronounced and it was clear that the ideal observer's performance (like that of subjects') was lowest on trials with high variability and low sample size. Comparison of Figure 12 with Figure 7 (on page xx) shows the subjects' performance fell well below the ideal observer's in all nine of the conditions shown. This deficit in performance is reflected in the fact that subjects' responses matched those of the ideal observer only 63% of the time. Agreement between the subjects and the ideal observer did depend on the level of variability, with agreeement rates at 77% when variability was low, 63% when variability was moderate, and 49% when variability was high.

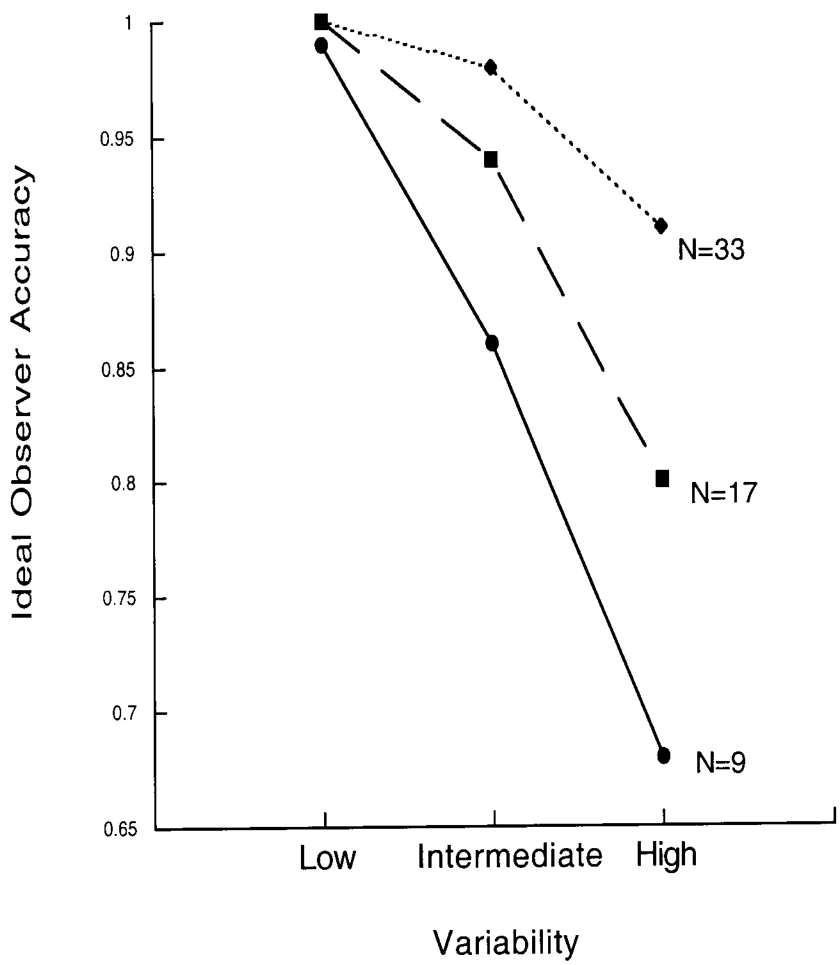


Figure 12. Performance of the ideal observer.

Signal Detection Analyses

Traditional hypothesis testing statistics are used to determine if population means are the same or different. In perceptual experiments, Signal Detection Theory parallels these methods and is often used to assess the fine perceptual discriminations of a human observer. In signal detection theory, the performance index, d', measures subject sensitivity to a particular stimulus and provides a measure of how well a signal is detected against a background of noise. Unlike traditional methods, d' takes into account both how often subjects make a correct response (hit) and how often they falsely report that a signal is present (false alarm). In this way, d' is able to provide a measure of discrimination performance that is, unlike the percent correct measure, independent of response biases (response biases in the present experiment will be discussed later).

Another advantage of using d' as an accuracy measure is that it, again unlike percent correct, is metrically suitable for computations of relative efficiency. Relative efficiency represents the performance of a human subject relative to the performance of an ideal observer. Relative efficiency is thus a measure of how effectively information is used by a human observer. Both d' and relative efficiency can be used to assess the ability of subjects to discriminate between different signals based on samples of data. In order to determine how sensitive subjects were to differences between linear and nonlinear curves, d' was calculated for the following pairs of trends: increasing linear/ increasing asymptotic, increasing linear/increasing exponential, decreasing linear/

decreasing asymptotic, and decreasing linear/decreasing exponential. These comparisons give an indication of how well subjects were able to discriminate between these linear and nonlinear trends (Because subjects were able to discriminate quite accurately between exponential and asymptotic trends and between increasing and decreasing trends, these pair combinations were not included in the analysis.) In order to assess sensitivity, the d' was calculated for each session by extracting hit and false alarm rates from confusion matrices (see Macmillan & Creelman, 1991). Overall sensitivity was calculated by averaging the d' values from each of the four pairings.

A 3 (variability) x 3 (sample size) x 4 (graph type) repeated measures analysis of variance was conducted on the d' measure. Subject sensitivity was affected by variability, $F(2,10)=36.33$, $p<.05$, sample size, $F(2,10)=9.07$, $p<.05$, and graph type, $F(3,15)=7.24$, $p<.05$. When the level of variability was low, the average sensitivity was 2.64. Sensitivity was 1.71 when the variability was moderate and dropped to 1.01 when the variability was high. Post hoc tests revealed significant differences between the three levels of variability. There was a similar pattern for sample size. When the sample size was high, the average sensitivity was 2.04, when sample size was intermediate the average sensitivity was 1.79, and when sample size was low, the average sensitivity fell to 1.52. There were significant sensitivity differences between large and small samples. Subjects were most sensitive when line graphs (d'=2.04) and scatterplots (d'=1.92) were used to present the data. Sensitivity was lower when histograms (d'=1.66) and suspended

bar graphs ($\underline{d}'$=1.52) were used to present the data. Post hoc tests revealed that line graphs lead to significantly higher levels of sensitivity than suspended bar graphs. There were no additional significant effects for graph type.

The interaction between variability and sample size was significant, $\underline{F}(4,20)=3.0$, $\underline{p}<.05$. As can be seen in Figure 13, sensitivity was highest when the variability was low and sample size was large. When the sample size was small, there were significant decreases in sensitivity with each increase in variability. When sample size was intermediate or large, sensitivity was significantly higher when the variability was low. Thus, for small sample sizes, the effects of variability are most evident and improvements in sensitivity occurred as variability decreased. When sample size was intermediate or large, sensitivity was highest when variability was low but there were no significant differences between the moderate and high variability samples.

The interaction between variability and graph type was also significant, $\underline{F}(6,30)=4.32$, $\underline{p}<.05$. In all instances, line graphs and scatterplots led to higher sensitivity than histograms and suspended bar graphs. These sensitivity differences were most pronounced when the variability was low and they became less evident as the variability increased. When the variability was low, sensitivity ranged from 2.27 for histograms to 3.03 for scatterplots but when the variability was high, the range of sensitivity decreased to a low of .78 for suspended bar graphs and a high of 1.22 for line graphs.

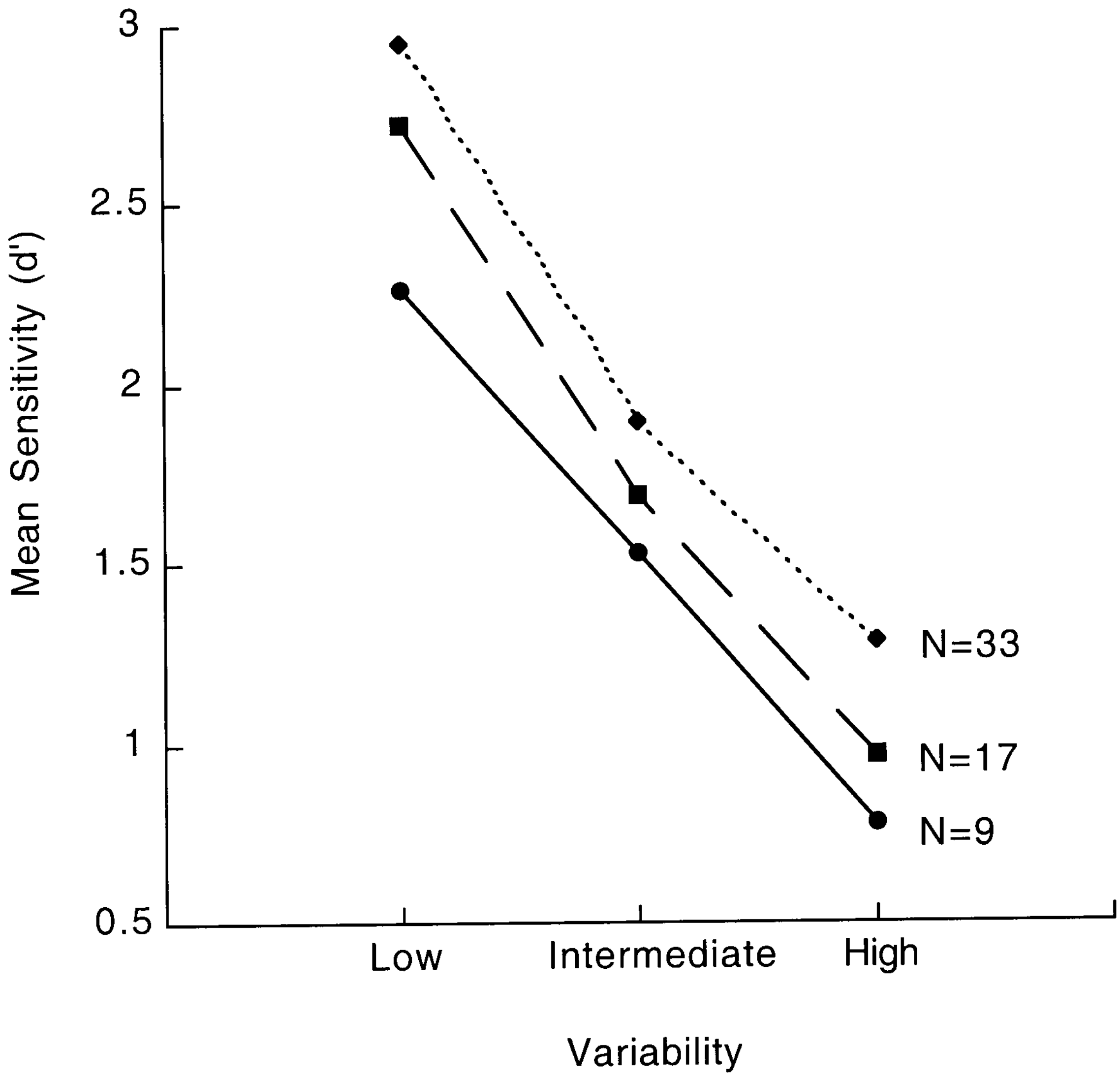


Figure 13. Mean sensitivity as a function of sample size and variability.

There was an additional interaction between sample size and graph type, $F(6,30)=3.35$, $p<.05$. As can be seen from Figure 14, among the four graph types, high sample sizes led to better sensitivity than low sample sizes. This difference was most pronounced for scatterplots and suspended bar graphs. When sample size was low these types of graphs led to low sensitivity but when sample size was intermediate or high, scatterplots led to high sensitivity. Although overall sensitivity was lowest when suspended bar graphs were used to present the data, it was particularly low with the small sample size.

The measure of relative efficiency provides a comparison of the human observer and the ideal observer. Relative efficiency allows an examination of how effectively information is used by subjects compared to optimal numerical methods. It can be defined as the ratio of the subject sensitivity to the sensitivity of the ideal observer. Past research on graph perception has reported relative efficiency scores of approximately .75 (Legge et al., 1989) and .70 (Best et al., 1999). The relative efficiency scores in the present experiment were lower than previous scores, possibly because the present task was more difficult than past tasks (which involved two-sample comparisons analogous to t tests). In the present experiment, the overall relative efficiency was .44. There were significant main effects for variability, $F(2, 10)=4.66$, $p<.05$ and graph type, $F(3,15)=8.30$, $p<.05$. When variability was low, relative efficiency was highest (RE=.51). Relative efficiency dropped to .43 when variability was moderate and to .39 when

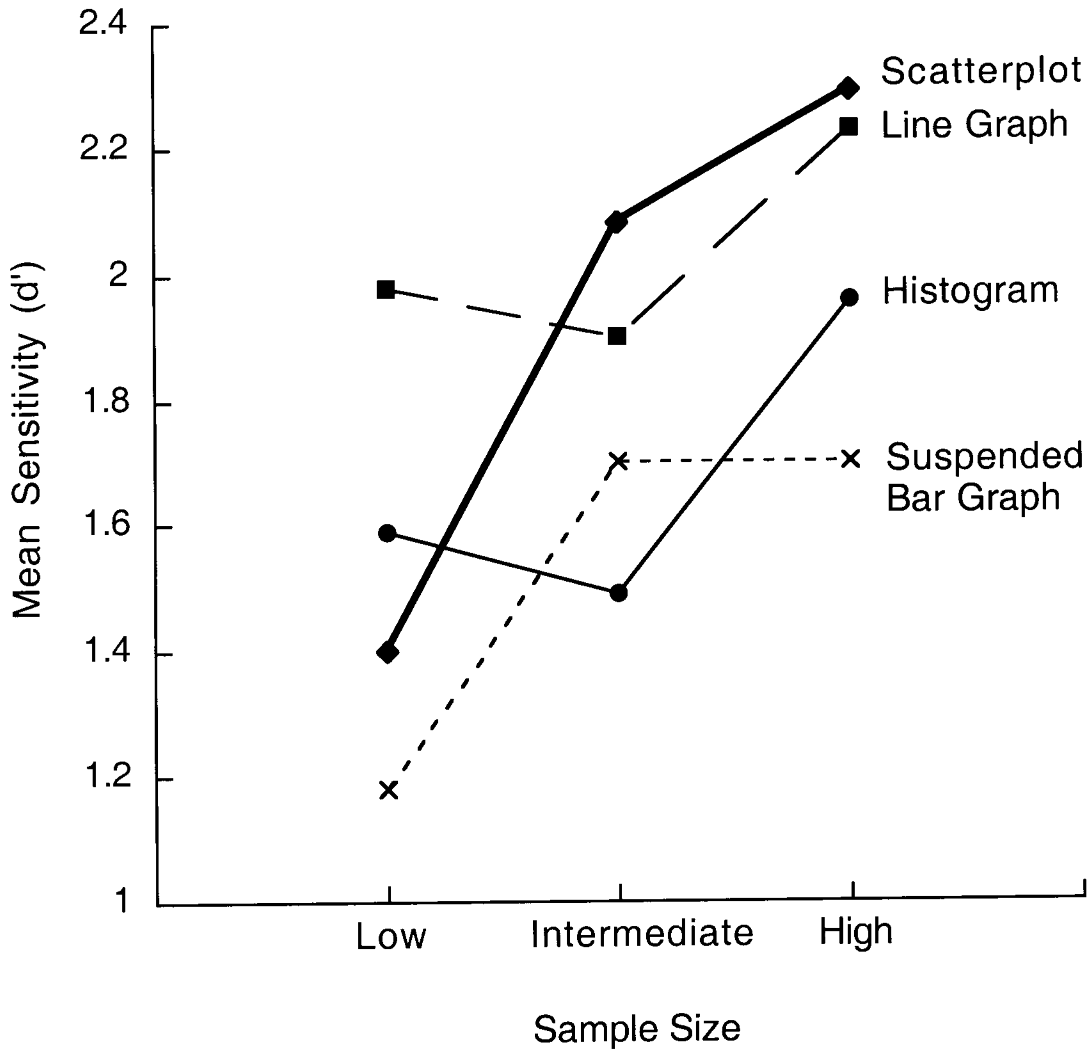


Figure 14. Mean sensitivity as a function on graph type and sample size.

variability was high. Relative efficiency was highest when line graphs (RE=.55) and scatterplots (RE=.48) were used to present the data. Efficiency dropped when histograms (RE=.39) and suspended bar graphs (RE=.35) were used to present the data.

There was a significant interaction between variability and sample size, $F(4, 20)=4.05$, $p<.05$. As can be seen in Figure 15, when sample size was low, relative efficiency was similar across the levels of variability. Post hoc tests revealed that variability had no significant effect on relative efficiency at this sample size. This result suggests that, when sample size is low, the effects of variability are comparable for the human subjects and the ideal observer. In contrast, when sample size was intermediate or high, relative efficiency was impaired by high levels of variability. Thus, when variability was high but sample size was low, subjects were able to see through the noise in order to determine which trend was present. As sample size increased, the relative efficiency on high-variability trials decreased, which suggests that, although increasing the number of data points on a graph made it easier for subjects to discriminate between different trends, the ideal observer used the additional information more effectively and was better able to integrate information from larger data sets in combatting the effects of variability.

The interaction between sample size and graph type was also statistically significant, $F(6, 30)=2.97$, $p<.05$. The patterns of relative efficiency for scatterplots and suspended bar graphs were similar across the different sample sizes (see Figure 16). The

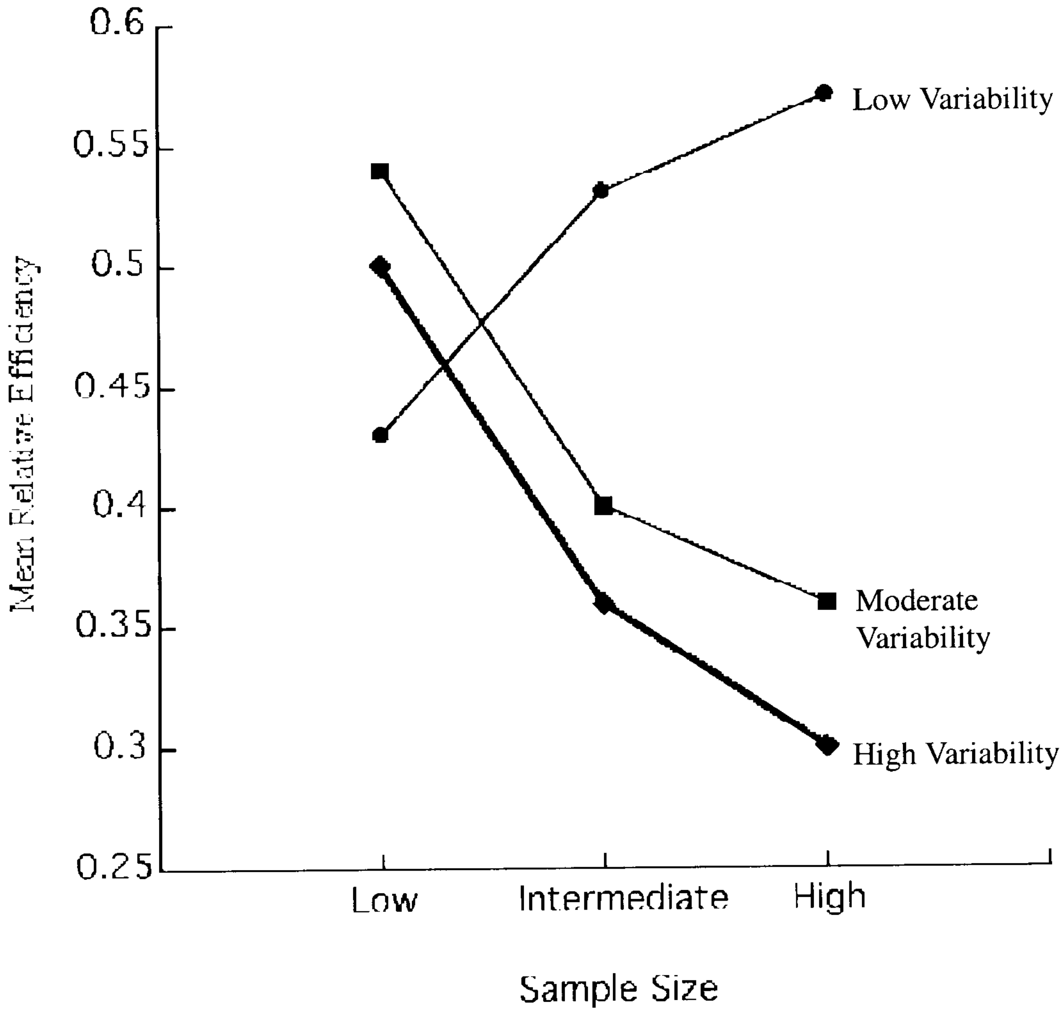


Figure 15. Relative efficiency as a function of sample size and variability.

relative efficiency for histograms and line graphs was higher when the sample size was low and decreased as sample size increased. It appears that line graphs and histograms are superior when sample size is low but as sample size increases relative efficiency is less dependent upon graph type.

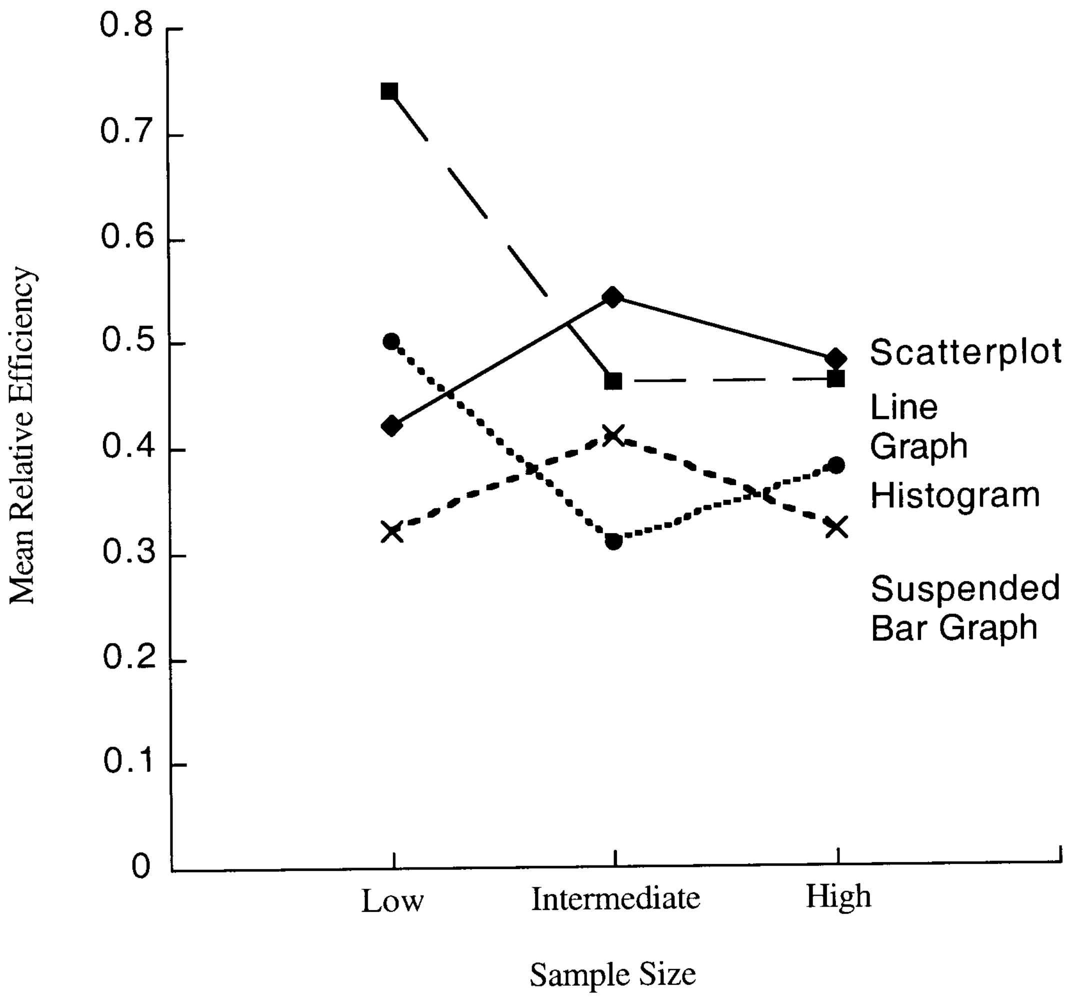


Figure 16. Average relative efficiency as a function of sample size and graph type.

To conclude, the purpose of Experiment 1 was to determine how well subjects could discriminate between linear and nonlinear trends. Three measures of discrimination accuracy were examined--percent correct, sensitivity, and relative efficiency. Table 4 summarizes and compares the results of ANOVA's for the three measures, with check marks indicating significant effects. As can be from the Table, the results of these analyses were fairly consistent. In all instances, performance was best when data with a low level of variability were presented on line graphs or scatterplots. In addition, the similarity of results for percent correct and the bias-free sensitivity index d' provides evidence that the results for the percent correct measure are not attributable to any confounding effects of bias.

Table 4
Summary of the Results from Experiment 1

Main Effect or Interaction	Percent Correct	Sensitivity	Relative Efficiency
Variability	sig	sig	sig
Sample Size	sig	sig	ns
Graph Type	sig	sig	sig
Trend Type	sig	N/A	N/A
Var x Graph	sig	sig	ns
N x Graph	sig	sig	sig
N x Trend	sig	N/A	N/A
Trend x Graph	sig	N/A	N/A
Var x N	ns	sig	ns

Response Time

A 3 x 3 x 4 x 6 repeated measures analysis of variance was used to assess if response time varied across the different levels of the independent variables. There was a significant main effect of graph type, $F(3, 15)=13.38$, $p<.05$. The average response time for suspended bar graphs (3.26 seconds) was higher than for histograms (3.01 seconds), line graphs (2.75 seconds), and scatterplots (2.77 seconds). Post hoc tests revealed that the time it took to make discriminations was longer when suspended bar graphs were used.

There was also a significant interaction between sample size and graph type, $F(6,30)=4.68$, $p<.05$. When line graphs were used to present the data, the response times for the different sample sizes were not significantly different. When histograms, suspended bar graphs, and scatterplots were used to present the data, response times were significantly slower when fewer points were presented.

Tests of the Validity of the Experimental Design

Several factors were examined in order to examine the validity of the experimental paradigm. First, given the nature of the experimental task, it was possible that subject performance would improve with experience. If this were the case, it would be difficult to ascertain whether differences in discrimination accuracy were due to the effects of the independent variables or simply due to experience. In order to examine this possibility, mean accuracy was calculated for the first 18 and second 18 sessions.

Accuracy means are presented in Table 5. There were no significant differences across time, $t(5)=-1.63$, $p>.05$. Thus, the overall results of the experiment were not affected by shifts in accuracy due to practice.

Table 5

Mean Percent Correct on Early and Later Sessions

Subject	Sessions 1 - 18	Sessions 19 - 36
1	68.8	68.5
2	64.3	64.5
3	65.7	70.9
4	62.2	62.2
5	60.7	63.6
6	45.3	46.0

Second, the computer calculated the fit between the displayed curve on each trial and each of the six underlying curves. In order to assess the overall difficulty of each trial, a measure of difficulty was calculated by taking the difference in fit (squared residuals) between the two underlying curves with the lowest squared residuals. For example, if an exponential increasing curve was presented, the computer calculated the squared residual between the displayed trend and each of the six underlying curves, and the absolute difference in fit between the two curves with the best fits was calculated. Large differences would thus indicate that the trend was best represented by a single underlying curve and, in these cases, discrimination would presumably be easy. When the difference was small, the curve could be represented equally well by (at least) two

different underlying curves and therefore the difficulty of discrimination would be greater. When subjects made a correct response, the mean level of difficulty by this measure was 0.056 and when an incorrect response was given, the mean difficulty was 0.039. A t-test indicated that this difference was significant, $t(5)=3.22$, $p<.05$. Hence, trials that had a higher difficulty rating were indeed more difficult for subjects and discrimination accuracy on those trials was lower.

A third check of the experimental design involved the basis on which subjects were discriminating between trends. A likely candidate for the perceptual cue being used by subjects is the amount and direction of curvature in the displayed trends. Exponential curves are accelerating and have positive curvature values, asymptotic curves are decelerating and have negative curvature values, and linear curves do not accelerate and have values approximating zero. Thus, it is possible that subjects focus on curvature when discriminating these three trends. If so, then the curvilinear trends used in this experiment should be more discriminable on trials where the degree of curvature is higher than on trials where the degree of curvature is less, making them appear more like linear trends. In fact, the addition of random variability to the underlying curves means that it is possible for a trial with a particular underlying curve to have a curvaturein the displayed data that is characteristic of a different curve. For example, a linear underlying underlying curve with unusually high random deviates toward the end of the curve would, by chance alone, have the positive curvature characteristic of an exponential

curve. This issue will be discussed further in the General Discussion.

To assess whether discrimination performance is related to amounts of curvature in this fashion, a measure of curvature was recorded by the computer for each trial. The measure was the mean of the middle third of the time series subtracted from the mean of the data points in the first and last thirds of the series. Thus, when the curve is negatively accelerated, as with the asymptotic trends, the mean of the middle third will be greater than the mean of the remainder of the curve and produce a negative value for the curvature measure. Conversely, when the curve is positively accelerated, as with the exponential trends, the mean of the middle third will be less than the mean of the rest of the curve and the curvature index will be positive. Linear trends would be expected to have curvature values near zero. (For decreasing curves the curvature index values were multiplied by -1 in order to reverse their signs, so that accelerating or positive curvature would always be reflected as a positive value in the index and decelerating or negative curvature would always have a negative value.) In the case of the two curvilinear trend types, the curvature values will become more extreme (i.e., further from zero) as the curves become more curvilinear.

As can be seen in Table 6, the mean absolute index of curvature on trials where the correct response was made was higher in value than on trials where an incorrect response was made. Thus, as expected, it appears that the degree to which a curve accelerates or decelerates affected discrimination accuracy. The index of curvature will be explored later in the General Discussion in connection with a theoretical analysis of the perceptual basis of trend discrimination.

Table 6

Mean Absolute Index of Curvature Values as a Function of Accuracy

Curve type	Curvature on Incorrect Trials	Curvature on Correct Trials
Decreasing Exponential	0.130	0.160
Decreasing Asymptotic	0.094	0.135
Decreasing Linear	0.010	0.016
Increasing Exponential	0.145	0.177
Increasing Asymptotic	0.095	0.150
Increasing Linear	0.003	0.001

A fourth check of the experimental design concerned the issue of the subjects' response allocations. It will be recalled that subjects were instructed to allocate responses as evenly as possible to the six response types. The reason for this was that the percent correct measure of accuracy is not an unbiased measure of sensitivity unless the allocations are approximately equal, that is, the pattern of responses is unbiased. On discrimination tasks, bias exists when subjects tend to respond in a particular manner regardless of the displayed stimuli. To take an extreme example, a subject who

responded "linear increasing" on all trials regardless of which stimulus was presented (and thus had a strong bias) would have an accuracy of 100% on linear increasing trials and an accuracy of 0% on all others. Clearly, the 100% accuracy does not reflect any sensitivity in this case, but rather reflects the bias. As a general rule, the bias (i.e., response allocations) and percentage of correct discriminations are not independent, and percent correct becomes a pure measure of sensitivity only when responding is unbiased (i.e., response allocations are equal).

To determine whether subjects responding was biased (despite the instructions to allocate responses equally) and whether any such biased responding might have influenced the results of the experiment, several analyses were conducted. To begin with, the overall allocations across all subjects were tabulated to see if there were any general biases to respond in a particular way (see Table 7).

Table 7

Overall Response Allocations

Curve Type	Percentage of Responses
Decreasing Exponential	15.08
Decreasing Asymptotic	17.84
Decreasing Linear	16.91
Increasing Exponential	17.05
Increasing Asymptotic	17.45
Increasing Linear	15.67

Although these allocations are reasonably well balanced, it is possible that they were influenced by some of the factors that were varied in the experiment. For example, in view of the fact that cues for curvilinearity might be expected to be masked by high levels of noise in the data, it is possible that high variability would bias subjects to use the "linear" responses. As a second example, it is possible that certain graph types tend to make time series look inherently less linear. To check on these possibilities, a 3 (variability) x 4 (graph type) x 6 (response type) repeated measures analysis of variance was conducted, using frequencies of responses allocated to categories as a dependent measure. There were no significant main effects or interactions, which indicated that the response allocations of subjects were not affected by these variables.

To further examine the effects of bias, the response allocations of each subject were examined individually. Some of the subjects (e.g., Subject 1) allocated their responses quite equally but others (e.g., Subject 6) were much less effective at doing so. Whether such deviations from equal allocations were independent of, or related to, accuracy was assessed for each subject by computing correlation coefficients between the percent correct for each trend type and the number of responses allotted to that trend type (see Table 8). The overall correlation between response allocation and percent correct was $r=.43$. The individual correlations were low for Subject 1 ($r=.08$) and Subject 3 ($r=.07$). There were moderate (but nonsignificant) correlations for Subject 2 ($r=.60$), Subject 4 ($r=.43$) and Subject 5 ($r=.67$). For Subject 6, there was a significant

Table 8

Response Allocations and Discrimination Accuracy for Each Subject. Percent correct is given in parenthesis below the response allocations

	Subject					
Trend	1	2	3	4	5	6
Inc. Exp.	15.5%	18%	17.1%	17.5%	20.7%	15.8%
	(65)	(74)	(72)	(75)	(74)	(42)
Dec. Exp.	14.2%	13.9%	17%	16.4%	17.4%	11.6%
	(63)	(62)	(77)	(73)	(70)	(37)
Inc. Asy.	16.8%	18.%	16.5%	15.5%	15.1%	20.4%
	(68)	(75)	(76)	(69)	(64)	(52)
Dec. Asy.	18.8%	17.3%	15.1%	17.5%	14.1%	24.1%
	(71)	(70)	(66)	(74)	(60)	(62)
Inc. Lin.	17.7%	14.1%	16.5%	17.3%	14.3%	14%
	(54)	(50)	(60)	(63)	(47)	(39)
Dec. Lin.	16.8%	18.6%	18%	15.7%	18.5%	14%
	(53)	(58)	(64)	(59)	(58)	(40)

relationship between response allocations and accuracy (r=.99). To determine whether the overall findings presented above depended critically on this subject's aberrant data, Subject 6's data were removed from the data set and the analyses were run again on the remaining five subjects The results for the remaining five subjects were the same as those for the six subjects (all significant main effects and interactions remained significant) and, therefore, Subject 6's data were included in the overall results, as reported earlier.

Chapter 6

Experiment 2: Extrapolation of Trends

Whereas the first experiment focused on how well subjects could discriminate between data sets with different underlying trends, this experiment focused on the ability to extrapolate future points in a data series. Much of the previous research on the ability to extrapolate has not taken into account several key factors. This experiment focused on answering several questions: 1. How does the level of variability affect subject extrapolations? 2. Do different types of graphs affect the ability to extrapolate? 3. Does extrapolation accuracy depend on the type of trend? and 4. Does the number of data points on the curve affect extrapolation ability? As in Experiment 1, graph type, variability, sample size, and trend type were manipulated. In this case, the dependent variable was extrapolation accuracy. Thus, the result was a 3 (sample size) x 3 (level of variability) x 4 (graph type) x 6 (type of trend) repeated measures design.

It was expected that accuracy would be highest when the sample size was large and the level of variability was low. In addition, based on the existing literature, subjects were expected to make more accurate extrapolations when making judgments concerning linear and asymptotic trends than for exponential trends. Furthermore, it was tentatively hypothesized that as variability increased and the visual cues for the presence for curvilinearity became lost in the noise, subjects would increasingly treat all trend types in terms of the linear component of the data. As a result it was expected that subjects'

extrapolations would become increasingly similar to predictions made by linear regression as levels of variability increased. It may be that, with increasing noise, simple linear regression provides increasingly good estimates of the next point in a series relative to estimates projected by nonlinear curve fits. To investigate this possibility, data were recorded for each trial on the predictions of linear fits compared to both the subjects' forecasts and the location of the next point on the underlying curve. Although extrapolations based on simple linear regression may never surpass the accuracy of forecasts that take curvilinearity into account, they may come close in accuracy and would presumably be efficient to intuitively perform in terms of the cognitive load required.

Subjects

The same subjects participated in this experiment as in Experiment 1.

Apparatus

A PC computer (with a screen resolution of 640 by 480 pixels) was used to present the stimuli to the subjects, provide feedback, record subject responses, and compute necessary statistical information.

Method

Subjects viewed a graphical display on a computer screen and were asked to extrapolate the next data point in the series. For information on the generation of the six underlying curves, see Appendix 1. Each subject was presented with 36 sessions of 54

trials, resulting in 1944 trials per subject. Given this design, each unique combination of the independent variables was presented to subjects nine times.

At the beginning of each session, subjects were prompted to enter the session number, their initials, the sample size, the level of variability, and the type of graph to be used in the session. On each trial, subjects were required to predict where the next data point in the series would fall. A data point appeared on the screen and subjects adjusted its vertical position until they were satisfied that its placement represented their estimate of the next data point in the series. (The initial height of the data point to be adjusted was randomized in order to control for overestimation and underestimation biases based solely on initial location.) In essence, subjects manipulated the graph itself. For example, on a trial that presented the data in a histogram, the next data point was represented as a bar and subjects used the up-arrow and down-arrow keys to adjust the height of the bar until they were satisfied with their forecast. In the case of line graphs, the line connecting the adjustable point to the preceding point was continuously redrawn in real time so as to follow the moving point. After subjects made their selection, they clicked "OK" to move on to the next trial. Although the trials were self-paced, reaction time data were collected to determine if response latencies associated with the different conditions varied.

Results and Discussion

Extrapolation Accuracy

In order to assess how well subjects were able to extrapolate future points on the graphs, subject accuracy was calculated. Extrapolation error was defined as the absolute difference between the point predicted by the subjects and the next point on the underlying curves. This provided a measure of how close the extrapolations were to the underlying curve points, with smaller deviations indicating lower errors and higher extrapolation accuracy.

There was a significant main effect for graph type, $F(3, 15)=5.01$, $p<.05$. When histograms or suspended bar graphs were used to present the data, the mean error was .50. The corresponding error was .56 when scatterplots were used to present the data. When line graphs were used, error was highest at .73. Post hoc tests showed that accuracy was significantly lower for line graphs than for scatterplots or suspended bar graphs. There were no significant differences between line graphs and histograms. Thus, at least for extrapolation tasks, it appears that scatterplots and suspended bar graphs lead to higher accuracy.

There was also a main effect for variability, $F(2, 10)=33.24$, $p<.05$. When variability was low, the mean error was .41. Error rose to .60 when the variability was moderate and to .70 when the variability was high. Post hoc tests revealed significant differences between each level of variability. Thus, as expected, the ability to accurately

extrapolate future points on a time series graph became worse as variability increased.

The number of data points on a graph had a significant effect on extrapolation accuracy, $\underline{F}(2, 10)=5.5$, $\underline{p}<.05$. When the number of data points was large, extrapolation error was .53. The average error increased to .58 when the sample size was intermediate and to .60 when the sample size was small. Post hoc tests showed that small sample sizes led to significantly more error than either intermediate or high sample sizes. There were no significant decreases in error when sample size was increased from 17 points (intermediate) to 33 points (large). Thus, when intermediate or large samples are presented, judgments about the next curve points are more accurate.

The type of data trend presented significantly affected extrapolation accuracy, $\underline{F}(5, 25)=5.3$, $\underline{p}<.05$. Overall, error was lowest for asymptotic curves (for both increasing and decreasing trends, the average error was .52). For linear trends, subject error was .51 for decreasing trends and .58 for increasing trends. As expected, accuracy was lowest for exponential trends. The average error was .67 for increasing exponential trends and .62 for decreasing exponential trends. Thus, the ability to extrapolate future data points was best if asymptotic or linear trends were presented. Although error was higher when exponential trends were presented, increasing exponential curves led to more error than decreasing exponential curves.

There was also a significant interaction between variability and trend type, $\underline{F}(10, 50)=2.51$, $\underline{p}<.05$). As can be seen in Figure 17a and Figure 17b, regardless of the level of variability, error was highest when exponential trends (increasing or decreasing) were presented. When variability was low, the overall error was lower for asymptotic trends than for linear trends. Post hoc tests revealed a significant difference between the increasing linear and increasing asymptotic curves. There were no other significant differences between the linear and asymptotic curves. When variability was moderate, there were no significant differences among the two linear and two asymptotic curves. When variability was high, decreasing linear curves led to significantly lower error than increasing and decreasing exponential curves. Post hoc tests revealed no further differences among the curves. Thus, the variability in a data set affected how well subjects extrapolated future data points but these effects were dependent upon the type of trend present. Overall, exponential curves led to significantly higher error but there were few differences between the asymptotic and linear curves.

There was an additional interaction between sample size and trend type, $\underline{F}(10, 50)=2.11$, $\underline{p}<.05$. For exponential trends (increasing or decreasing), error was similar across all levels of sample size (see Figure 18a and Figure 18b). In contrast, when linear and asymptotic trends were presented, error was significantly lower when more data

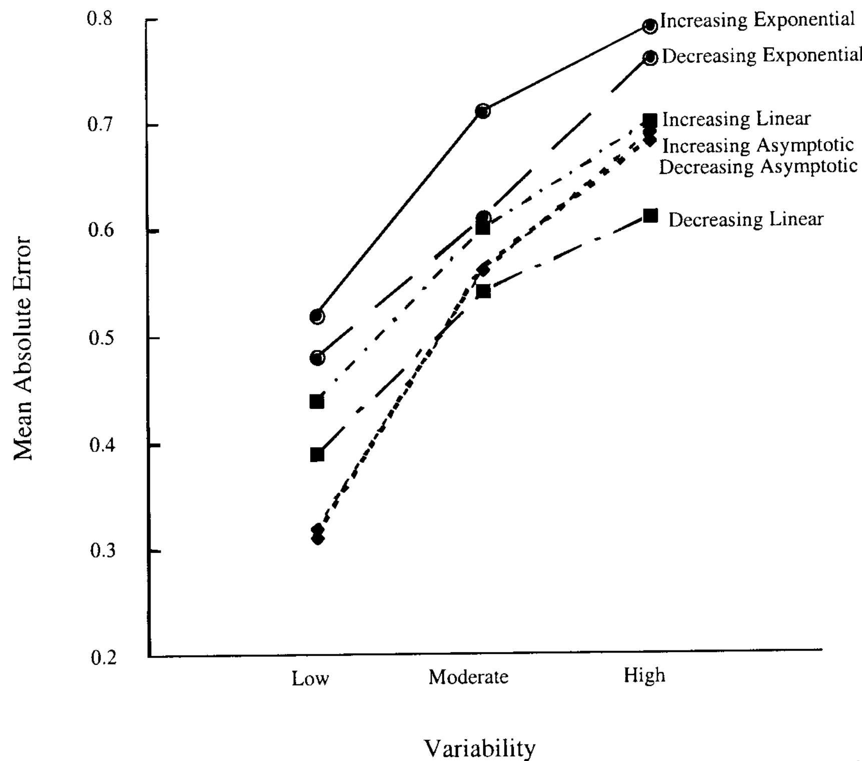


Figure 17a. Extrapolation error as a function of variability and trend type.

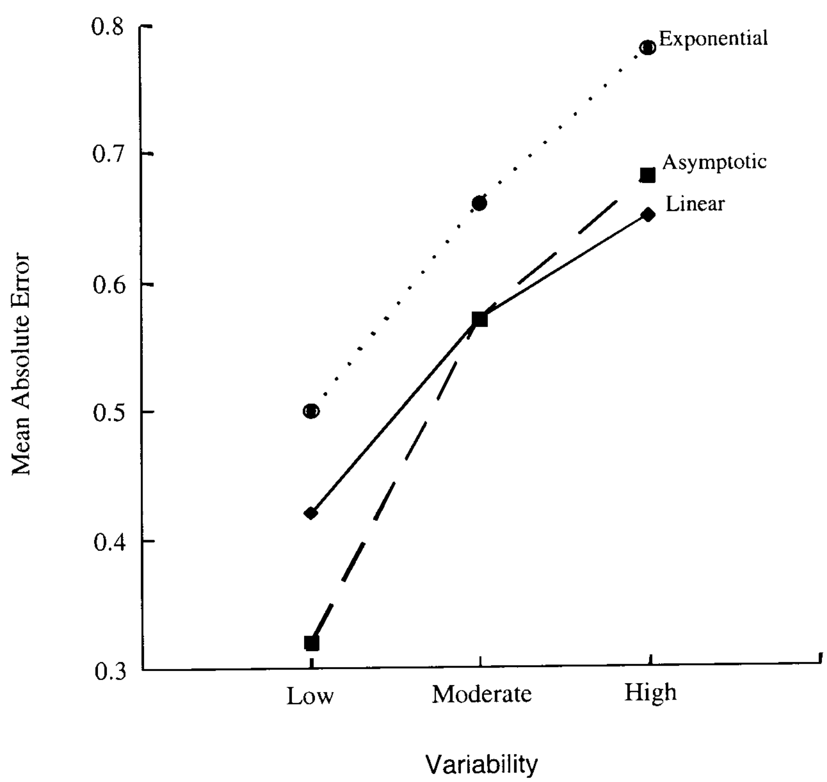


Figure 17b. Extrapolation error for exponential, asymptotic, and linear trends as a function of variability.

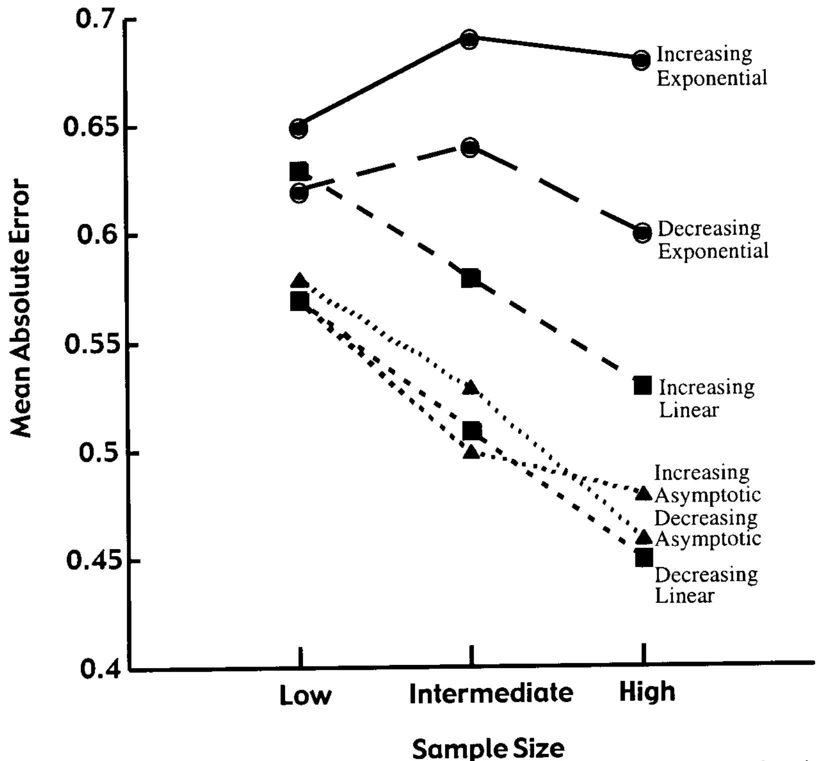


Figure 18a. Extrapolation error as a function of trend type and sample size.

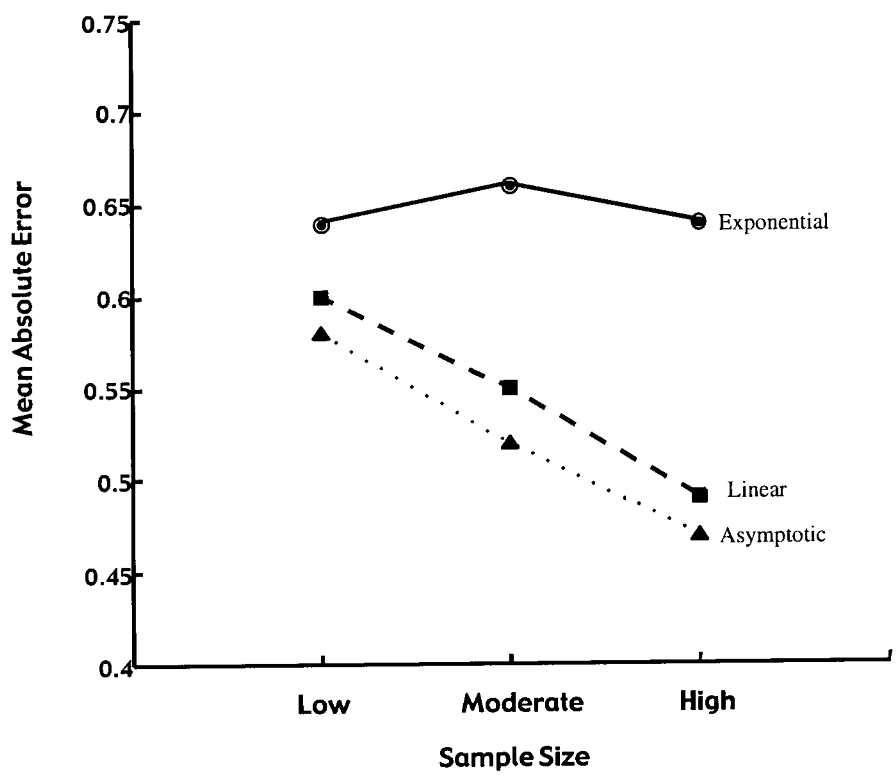


Figure 18b. Extrapolation error for exponential, asymptotic, and linear trends as a function of sample size.

points were presented. Thus, it appears that increasing the number of data points on a graph had little effect on the extrapolation of exponential trends but improved extrapolation accuracy with linear and asymptotic trends.

Over- and Underestimation Errors

An examination of the mean of the signed extrapolation errors for each trend type revealed that subjects made systematic over- and underestimation errors (see Figure 19). When the trend type was increasing, subjects tended to overestimate the next data point, predicting points higher than the next underlying curve point. When the trend type was decreasing, subjects underestimated the next data point and their extrapolations were lower than the underlying curve point. In all cases, the change in the trend was judged to be greater than the actual change in the underlying curve, indicating that the rate of change was being overestimated. As can be seen in Figure 19, the degree of this overestimation was greater when the trend was increasing and was closer to zero for curves with decreasing trends. As expected, the degree of overestimation of change was smallest when asymptotic curves were presented.

When linear and exponential trends were presented, errors involving overestimation of the rate of change occurred regardless of the level of variability in the data set (see Figure 20). When asymptotic trends were presented, these overestimation errors were slightly more pronounced when variability was moderate or high and were virtually nonexistent when variability was low. This indicates that, although subjects

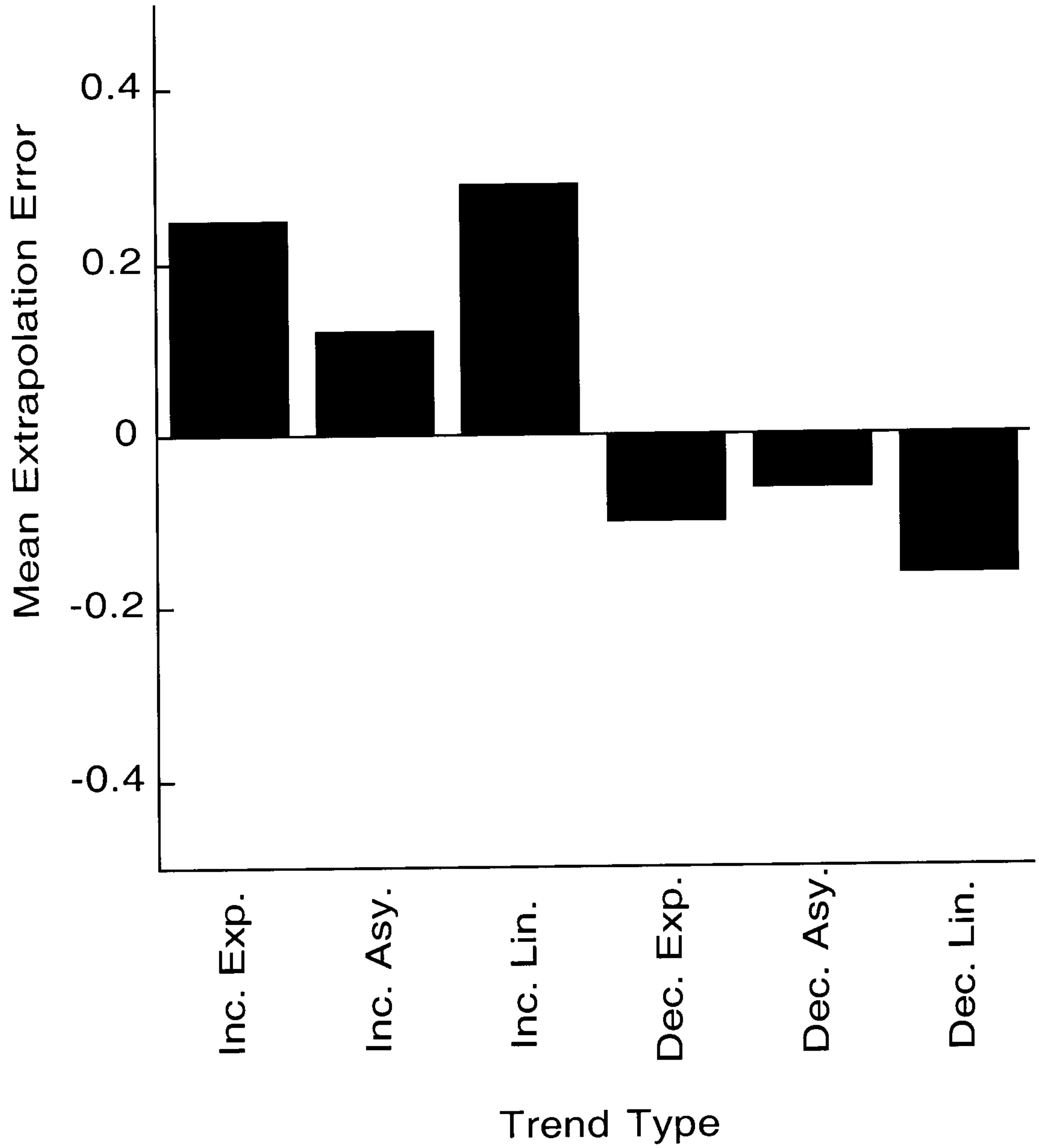


Figure 19. The estimation errors for each trend type.

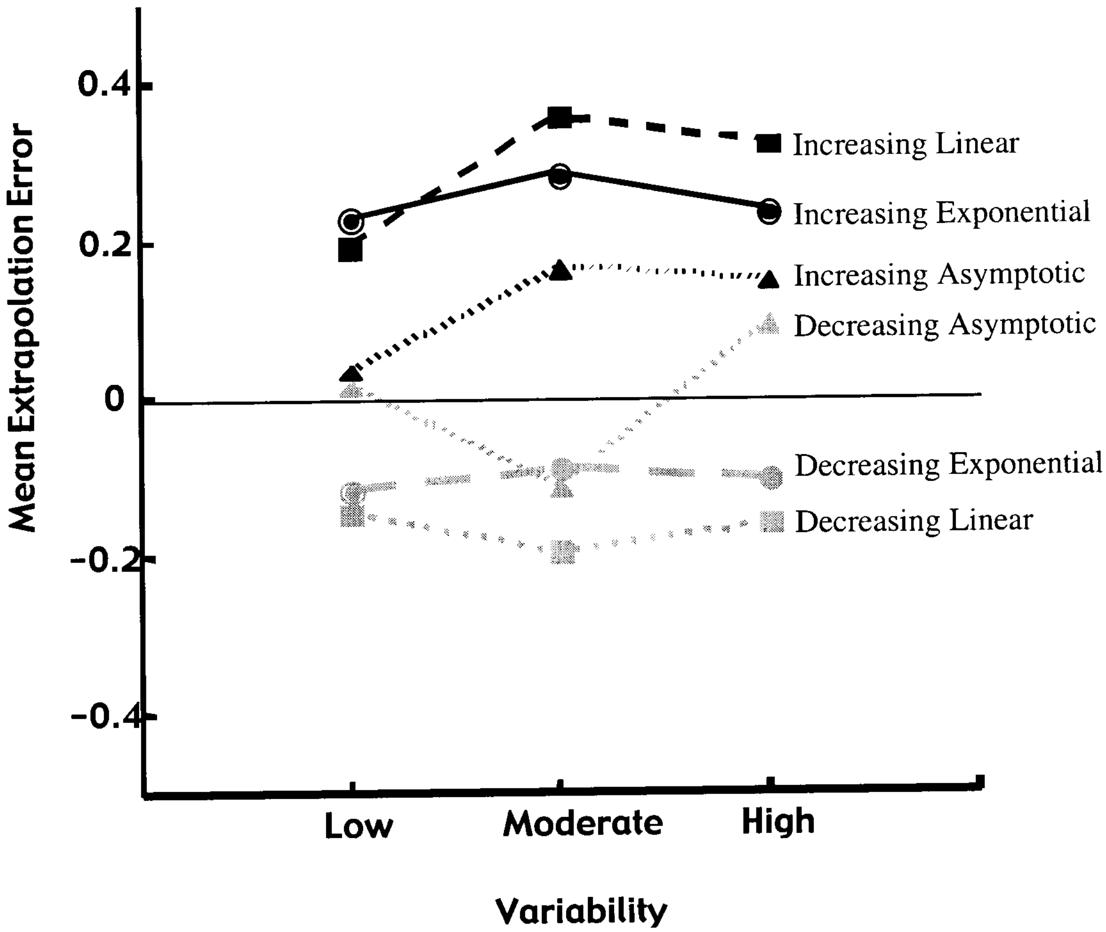


Figure 20. The estimation errors for each trend as a function of variability.

were fairly accurate when judging future points on asymptotic curves with low variability, their accuracy tended to drop with increases in variability. However, the variability x trend type interaction fell short of significance, $F(10, 50)=1.31$, $p>.05$, in part because variability had no effect on errors for the other two trend types. For these trend types, there were equivalent overestimation errors across all levels of variability.

There was a significant interaction between sample size and trend type, $F(10, 50)=35.7$, $p<.05$. For asymptotic and linear trends, sample size had very little effect on the over- and underestimation extrapolation errors (see Figure 21a). Regardless of the number of data points presented, the subject extrapolations were too high for increasing curves and too low for decreasing curves, indicating a tendency towards overestimation of change in the trends. This pattern did not hold for exponential trends (see Figure 21b). For both increasing and decreasing exponential trends, there was a tendency to overestimate the rate of change when the sample size was high or intermediate. However, when the sample size was low, this pattern reversed and the subjects underestimated the change. Thus, for exponential curves, smaller sample sizes led to underestimation and larger sample sizes led to overestimation.

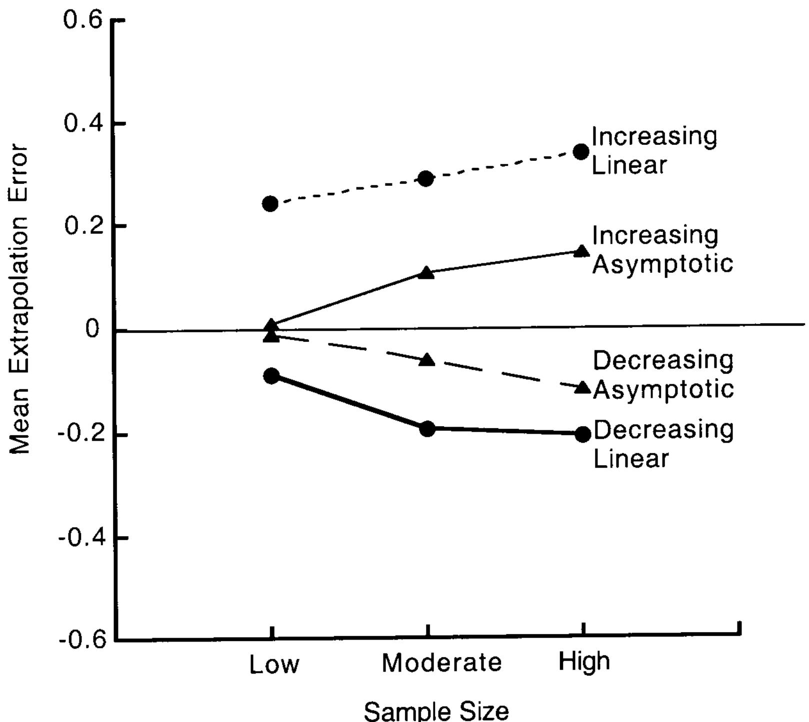


Figure 21a. The estimation biases for linear and asymptotic trends as a function of sample size.

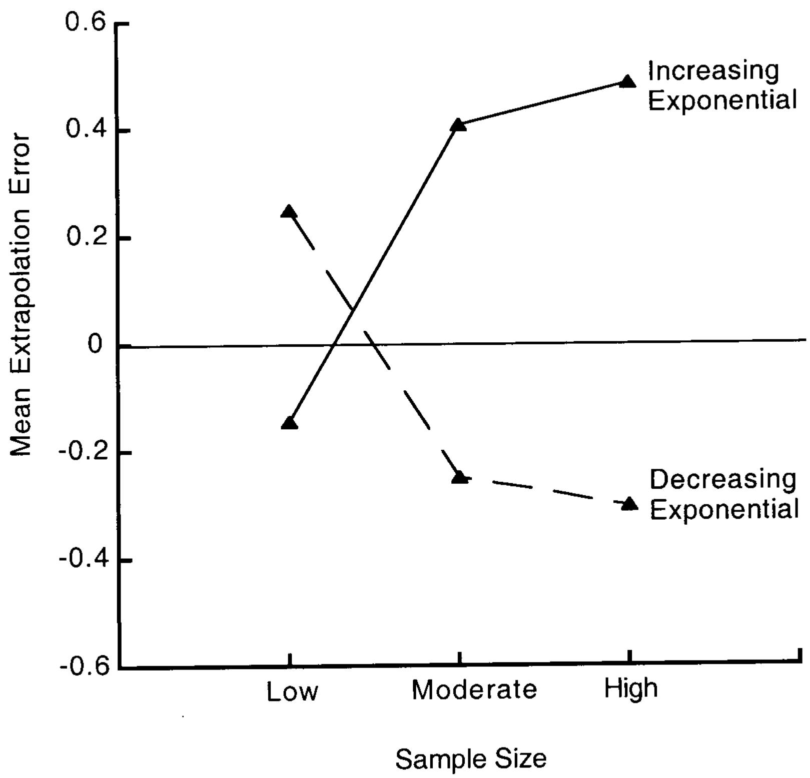


Figure 21b. The estimation biases for exponential trends as a function of sample size.

Extrapolation Heuristics

Because a time series graph contains several kinds of information that are relevant to predicting trends, there is a range of methods that could potentially be used to make extrapolations. Each of these methods, or heuristics, takes into account different amounts and types of information and therefore imposes different cognitive demands on those who use them. For example, if the subject takes each curve point into account when predicting (e.g., while performing some sort of intuitive curve-fitting), the cognitive load would be high. On the other hand, if the subject based his or her extrapolation on only the last few curve points, the cognitive load would be low. In general, it would be expected that simpler methods that use less of the available information would produce less accurate extrapolations, whereas more complicated methods that take into account more information would produce more accurate forecasts.

In this section, subjects' extrapolations will be compared with the extrapolations generated by two heuristics, both in terms of similarity between those extrapolations and in terms of their relative accuracy. Two extrapolation heuristics will be discussed--the use of linear components of the curve through linear regression and the use of the most recent data points as a basis for extrapolation. Examination of the heuristics involving the most recent data points was suggested by the literature on anchor and adjust heuristics (Tversky & Kahneman, 1974), in which subjects anchor their estimates to, say, the mean of the last three points before making final adjustments based on other features of the

data; the cognitive load for such a method should be quite low given that the anchor is based only on a few recent points. Examination of the linear regression heuristic was suggested by the fact that in the present experiment all three underlying curves began and ended at approximately the same points; examination of this heuristic was also suggested by the tentative hypothesis that subjects would shift toward using linear components of the data when variability becomes so high as to mask the cues for curvilinearity. The use of linear regression would be expected to impose a heavier cognitive load than the use of the mean of the last three points because it takes into account all of the data in the series. For this same reason it would also be expected to be more accurate than the mean of the last three (except perhaps for the case of asymptotic curves).

In order to determine the similarity of the subjects' methods to these two heuristics, absolute difference scores were calculated between the subject extrapolations and the points generated by each extrapolation heuristic (this was done for each of the experimental trials). Low deviations indicate that the subject extrapolation was similar to the extrapolation based on the heuristic in question.

Table 9 presents the mean deviations of the subjects' extrapolations from the two extrapolation heuristics. As can be seen from the table, subjects' extrapolations did not approximate either heuristic very closely. Also, at all levels of variability, the extrapolations were closer to the point predicted by linear regression than to the mean of the last three points. Thus, it appeared that subjects came closer to basing their

extrapolations on the linear characteristics of the curves than to using anchor and adjust heuristics. In addition, Table 9 shows that subjects' extrapolations deviated more from the linear regression predictions as variability increased. Thus, the hypothesis that, at high levels of variability, subjects would shift toward using linear regression was not confirmed.

Table 9

Deviation of Subject Extrapolations from Two Heuristics

	Level of Variability		
	Low	Moderate	High
Point predicted by linear regression	.50	.63	.71
Mean of last three data points	.52	.69	.76

The second issue to be addressed here is the relative accuracy of subject extrapolations and heuristic extrapolations. Table 10 presents the mean absolute error of predictions from linear regression, the mean of the last three points, and the subject extrapolation. When variability was low or moderate, the points predicted by linear regression and the mean of the last three points were both quite accurate. At higher levels of variability, the point predicted by linear regression was slightly more accurate predictor than the mean of the last three points. This suggests that linear regression is, as expected, a more accurate method than the mean of the last three points. However, the

differences are surprisingly small, suggesting that the mean of the last three points is viable as a simple rule of thumb.

Table 10

Extrapolation Error of Subjects Compared to that of the Extrapolation Heuristics

	Level of Variability		
	Low	Moderate	High
Linear Regression	.24	.30	.37
Mean of the Last Three Points	.25	.32	.41
Subject Extrapolation	.41	.60	.70

As can be seen from the third row of Table 10, the subject extrapolations were less accurate than the points predicted by either of the heuristics. This finding, which is surprising given that subjects could use cues for curvilinearity that were not being used by either heuristic, will be discussed in the General Discussion. It is worth noting here, however that the most effective way to make predictions about future curve points in the present task is to base the prediction on the underlying curve. This would presumably be done by intuitively "smoothing" the data and then projecting the smoothed curve onto subsequent points to be predicted. If something like this is done, extrapolations will be quite accurate, and, over a number of trials, this method of prediction would lead to the highest level of accuracy.

It may be instructive to compare the subjects' extrapolations not only to the underlying curve but to the subsequent point on the displayed curve (i.e., the curve that includes the random variability). When the variability was low, the extrapolations were slightly closer to the underlying curve point (error=.41) than to the displayed curve point (error=.46). At high levels of variability, the extrapolations were substantially closer to the underlying curve point (error=.70) than the displayed curve point (error = .89). Thus, despite subjects' relatively poor performance compared to the two heuristics, they were able, at least in part, to see through the noise at all levels of variability, including the high levels.

Response Time

An analysis of response time indicated no significant main effects or interactions. Therefore, at least for extrapolation tasks, there were no differences in the amount of time spent on each trial as a function of the experimental conditions. It is possible that there were no significant effects because of the nature of the task. On average, subjects spent 8.07 seconds on extrapolation trials and 2.9 seconds on discrimination trials. This increase in time spent on extrapolation trials may have eliminated or masked any possible differences arising from the experimental conditions. Given this finding, response time will not be analyzed in the remaining experiments.

Chapter 7

Experiment 3: The Effects of Feedback on Extrapolation Accuracy

Although it is apparent that feedback is an important component of any decision making task, it may be especially important in extrapolation tasks that require multiple judgments. In view of MacKinnon and Wearing's (1991) conclusion that feedback is an essential component of successful extrapolation, Experiment 3 examined the extrapolation accuracy of subjects with and without feedback. Feedback concerning the accuracy of previous extrapolations was expected to improve accuracy of subsequent extrapolations (MacKinnon & Wearing, 1991). The purpose of this experiment was to determine the degree to which feedback affects performance and how feedback might interact with variability.

If subjects are presented with a graphical display and asked to estimate several points beyond the available data set, their accuracy should be affected by whether or not they receive feedback after each estimate is made. In addition, the effects of feedback should be cumulative. For example, if subjects are asked to extrapolate the next five points in a series, the differences between the performance of subjects receiving feedback and those who do not receive feedback should be most evident after several judgments have been made.

When no feedback is given, subjects could potentially use their previous estimate as an anchor for their next estimate (Goodwin & Wright, 1994). Because the accuracy of any estimation depends on many factors (such as level of variability, sample size, and type of trend), using an estimate as an anchor for future estimations may be problematic. On the other hand, if feedback is given, subjects have the opportunity to use an actual data point as an anchor for their next estimate and, thus, future estimates, although still not perfect, may be more accurate. The incorporation of feedback gives subjects additional information and may increase extrapolation accuracy, especially in cases where the underlying variability is low. This prediction follows from the fact that feedback given under high variability conditions is more likely to be misleading in that the feedback point is more likely to represent random variability and less likely to represent the underlying trend.

In order to reduce the number of sessions required, and because the issue of sample size has been addressed in Experiments 1 and 2, sample size was not manipulated in this experiment. Rather sample size was set at the intermediate value (17 points) used in the previous studies. The result was a 3 (level of variability) x 4 (graph type) x 6 (type of trend) x 2 (feedback) repeated measures design.

Subjects

The same six subjects participated in this experiment as participated in Experiments 1 and 2.

Apparatus

An PC computer (with a screen resolution of 640 by 480 pixels) was used to present the stimuli to the subjects, provide feedback, record subject responses, and compute necessary statistical information.

Method

Subjects viewed time series graphs on a computer screen and were asked to extrapolate the next three data points in the series. Information concerning the generation of the curve points is given in Appendix 1. Each subject completed 24 sessions of 36 trials, resulting in 864 trials per subject. With this design, each unique combination of the independent variables was presented to subjects six times. The order of the blocks was randomized so that subjects received the conditions in different orders. This was done to neutralize any carry-over effects from condition to condition.

The method of presentation was similar to that of the previous experiments. Before each session, subjects were prompted to enter the session number, their initials, the level of variability, type of trend, and whether feedback was to be given. They then viewed the display for each trial and vertically adjusted the last point of the series by using the up- and down- arrow keys until they were satisfied that the adjustment represented their predictions (as in Experiment 2, the initial location of the adjustable point was randomized). When they were satisfied with their estimate they pressed the end key and another point was presented at the end of the data series. In the no-feedback

conditions, subjects then made second and third extrapolations. After making the third estimate, they pressed the end key to move on to the next trial. In the feedback conditions, subjects were given feedback after the first estimate was made. The feedback consisted of an open circle representing the next point on the displayed curve (i.e., the curve to which random variability had been added). Both the feedback and the subject estimate remained on the computer screen. This gave the subjects two pieces of information—their estimate and the next point in the series--that could potentially be used in making subsequent extrapolations. After the subjects received the feedback, they were prompted to make a second extrapolation and were given feedback concerning its accuracy. Finally, they made a third estimation and received feedback. After the third estimate, subjects pressed the end key to move on to the next trial.

Results and Discussion

A 3 (variability) x 6 (trend type) x 4 (graph type) x 2 (feedback) repeated measures analysis of variance was used to determine if the accuracy (measured in terms of error scores) was higher when subjects received feedback. The error scores of the second and third extrapolations were averaged to give an overall error score (the first extrapolation was omitted because in both the feedback and no-feedback conditions, subjects made the first extrapolation without the benefit of feedback--i.e., even in the feedback condition, feedback was given only after the first extrapolation was made).

As can be seen in Table 11, subjects tended to be more accurate when they received feedback concerning their extrapolations, with error scores being lower at all three levels of variability. However, contrary to expectation, the effects of feedback were not significant, $F(1,5)=3.16$, $p>.05$. Thus, accuracy was essentially similar in the feedback and no-feedback conditions and performance was not reliably higher when subjects received feedback. The effect of variability itself was significant, as in previous experiments. However, there was no significant interaction between variability and feedback condition, suggesting that the effects of variability were, contrary to expectation, the same across the feedback and no-feedback conditions.

Table 11

The Effects of Feedback on Extrapolation Accuracy

	Variability			
	Low	Moderate	High	Mean
No Feedback	.61	.71	.80	.71
Feedback	.53	.61	.72	.62

The feedback point represented the next point in the series and, depending on the amount of variability in the data, it could have provided a basis for subsequent extrapolations. It was hypothesized that, with increases in variability, the usefulness of the feedback point in providing information would decrease. Given this, subjects could have based their extrapolations on the feedback point or on their own previous extrapolation, using it as an anchor.

Past research has shown that subjects are able to see through the noise as they make decisions about data presented graphically (Best et al., 1999; see also Experiment 2 above). In order to determine if subjects were basing their extrapolations on the feedback point or on their own previous predictions (which may come closer to representing the underlying curves), the deviations of each extrapolation from the previous feedback point and the previous subject extrapolation were calculated. When the variability was low, the extrapolations were closer to the feedback point (deviation=.70) than to the previous extrapolation (deviation=.78). Thus, with low variability, subjects came closer to matching the feedback point than the previous extrapolation. In contrast, as variability increased, subjects began to use their previous extrapolations as an anchor. When the variability was moderate, the average deviation between the extrapolation and the previous extrapolation was .77 and the deviation between the extrapolation and the feedback point was .79. When the variability was high, the average deviation between the subject extrapolation and the previous extrapolation was .80, as compared with a .92

deviation between the extrapolation and the previous feedback point. Thus, as variability increased, subjects were more likely to use their own predictions as a basis for future predictions.

Although the effects of feedback were not significant, the fact that the feedback did appear to increase extrapolation accuracy suggests the need for further study. A power analysis indicated that, given the same difference between means of the feedback and no-feedback conditions, the use of 8 subjects rather than 6 would have would have produced statistical significance. Thus, feedback may aid in extrapolation tasks but future testing is needed.

Chapter 8

Experiment 4: Dynamic Presentation of Trends

Because motion often allows us to see patterns in things, it is possible that extrapolation accuracy would improve if a more active display were used. Active presentation of data points may allow subjects to detect trends in the data that are not perceivable when motion is absent. Typically, graphical displays are static—graphs are laid out on paper or on a computer screen. Dynamic graphs, although not as common, use more unconventional methods in an attempt to enhance certain aspects of the data. One way that data sets can be enhanced is motion.

When presenting data on a computer screen, there are two possible ways to present the data point—simultaneously and sequentially. Simultaneous presentation is most common and simply displays all of the points at one time. In this case, subjects view the data as a single display made up of a number of points. Sequential presentation of data points is rarely used. In this presentation method, data points are presented sequentially and with proper timing of the presentation of points, the perception of motion can be induced. Given the vast amount of literature on motion and the structure that motion adds to different detection situations, simulated motion in graphical displays might be expected to enhance both discrimination and extrapolation accuracy. Pilot testing of sequential motion displays revealed that exponential and asymptotic trends produced the appearance of accelerating and decelerating motion, respectively, whereas

linear trends produced apparent motion occurring at a uniform rate. (The appearance of acceleration and deceleration was due to changes in the average vertical distance between consecutive points in the two curvilinear trends, given equal time intervals between points and equal distances along the x-axis.) It was hypothesized that these changes in the rate of apparent motion, which were phenomenally quite salient, would enhance subjects' performance by adding cues unavailable in the original static displays. In addition, it was hypothesized that the benefits of motion would be greatest at low levels of variability and decrease as variability increases. High levels of variability were expected to break up the apparent motion and thus eliminate the acceleration and deceleration cues. Because this research was preliminary, the sample size was held constant (at 17 points) and the type of graph presented to subjects was limited to the scatterplot and histogram formats. A 3 (variability) x 2 (graph type) x 6 (type of trend) x 2 (presentation method) repeated measures design was employed.

In sum, the purpose of this experiment was to examine the effectiveness of sequential presentation of data points. It was expected that the simulated motion arising from the sequential presentation would enhance both discrimination ability and extrapolation accuracy.

Subjects

The same six subjects participated in the experiment as in the previous three experiments.

Apparatus

An PC computer (with a screen resolution of 640 by 480 pixels) was used to present the stimuli to the subjects, provide feedback, record subject responses, and compute necessary statistical information.

Method 4a: Trend Discrimination

Essentially, the procedure of the present experiment mirrored that of Experiment 1. Subjects were presented with graphical displays on a computer screen and were prompted to identify the underlying trend. Each subject completed 12 sessions of 48 trials, resulting in 576 trials per subject. There were thus results at eight trials at each combination of the independent variables. As before, sessions were grouped into blocks, which were presented to subjects in different orders.

Subjects began each session by entering the session number, their initials, the level of variability, the graph type, and whether the points would be presented sequentially or simultaneously. On each trial, subjects were asked to identify the underlying trend of the data series, as in Experiment 1. In the case of the static display, the graph was presented on the computer screen and subjects were able to examine it until they were satisfied with their judgment. When the graphs were presented dynamically, each data point appeared individually at the rate of about 0.03, so as to simulate motion. On each trial, the sequence was repeated three times (the three repetitions resulted in a total viewing time of 2.25 seconds which approximated the mean

time spent viewing the static displays in Experiment 1). As in Experiment 1, after a response was made, the computer presented a tone as feedback for a correct discrimination. Following the feedback, the next trial began.

Results and Discussion

In order to determine if motion affected discrimination ability, a 3 (variability) x 2 (graph type) x 2 (motion) x 6 (trend type) repeated measures analysis of variance was conducted. There was no main effect for motion, $F(1, 4) = 1.67$, $p > .05$. Overall, subjects were correct on 71% of trials that did not include motion and on 69% of trials that included motion.

Although it was expected that the effects of motion would be dependent upon the amount of noise in the data, there was no significant interaction between motion and variability. When variability was low, subjects were correct on 83% of the static trials, as compared with 77% of trials that contained motion. When variability was moderate, subjects were accurate on 64% of the no-motion trials and on 67% of the motion trials. When the variability was high, accuracy was 54% for both motion and no-motion trials. Thus, at all levels of variability, accuracy was similar regardless of the method of presentation.

Motion did not differentially affect discrimination accuracy for the different trend types. There was no significant interaction between trend type and motion. In particular, it was possible that motion would make exponential trends more salient and selectively

improve discrimination ability for these types of trends, but the means for the motion and no-motion trials were not significantly different. Mean accuracy for exponential increasing curves was 65% for no motion trials and 73% for motion trials. Accuracy for exponential decreasing curves was 74% for motion trials and 66% for no motion trials. These mean differences are not significant and, therefore, it does not appear that the inclusion of motion made exponential trends easier to discriminate.

In addition, there was no interaction between graph type and motion. When histograms were presented, subjects were correct on 65% of no-motion trials versus 66% of motion trials. When scatterplots were presented, subjects were correct on 69% of no-motion trials and on 66% of motion trials. Therefore, it does not appear that presenting data sequentially on scatterplots and histograms differentially affects the ability to discriminate between different types of trends.

As can be seen in Table 12, these nonsignificant findings are not likely to be due to any biases resulting from unequal response allocations. Subjects were able to allocate their responses fairly equally among the six trend types and there was no clear relationship between response allocations and percentage of correct discriminations. The overall correlation between response allocation and accuracy was $r=.42$, $p>.05$.

Table 12

Response Allocations and Discrimination Accuracy for the Six Trend Types. Percent correct are presented in parentheses

	Subject					
Trend	1	2	3	4	5	6
Inc. Asy.	16.8% (65)	16.0% (77)	17.2% (80)	14.8% (70)	15.8% (72)	17.7% (48)
Dec. Asy.	15.8% (59)	15.5% (68)	17.5% (78)	16.8% (73)	18.9% (78)	23.4% (65)
Inc. Exp.	15.5% (69)	20.7% (80)	18.2% (77)	18.9% (84)	20.0% (81)	12.5% (50)
Dec. Exp.	18.2% (75)	15.3% (78)	13.9% (71)	17.5% (77)	14.8% (73)	8.2% (29)
Inc. Lin.	15.3% (54)	13.0% (51)	13.9% (51)	16.6% (66)	13.5% (52)	21.9% (64)
Dec. Lin.	18.4% (64)	19.6% (76)	18.5% (68)	15.3% (63)	17.1% (65)	16.3% (45)

Method 4b: Extrapolation Accuracy

In essential respects, Experiment 4b mirrored Experiment 2, but with the addition of a motion condition. Following the method of Experiment 4a, variability, graph type, trend type, and presentation method were manipulated. Each subject completed 576 extrapolations, representing eight trials at each unique combination of the independent variables.

At the beginning of each session, subjects were prompted by the computer to enter the session number, their initials, the level of variability, the graph type, and presentation method. When the presentation method was static, subjects were simply presented with a standard scatterplot or histogram. As in Experiment 2, they adjusted the last point by using the up-arrow and down-arrow keys until they were satisfied that the placement represented their estimate of the next point in the series. They hit the end key to terminate the trial and move on to the next one. In the dynamic condition, subjects were presented with sequential data points and prompted to adjust the last data point until they were satisfied with their extrapolation. Subjects could hit the end key to repeat the sequence as many times as they wished. After the first presentation, the data point being adjusted was included in the dynamic presentation. After subjects were satisfied with their judgments, they hit the delete key to move on to the next trial.

Results and Discussion

A 3 (variability) x 2 (graph type) x 2 (motion) x 6 (trend type) repeated measures analysis of variance was conducted in order to determine if motion affected the accuracy of extrapolations. As in Experiment 2, error scores were obtained by taking the absolute deviation of each extrapolation from the corresponding point on the underlying trend. There was a main effect of motion, $F(1, 10)=85.26$, $p<.05$. On motion trials, mean absolute error was 1.40 as compared with .44 on no-motion trials. These effects of motion were not expected and thus the hypothesis that motion would improve extrapolation accuracy was not confirmed.

There was a significant interaction between motion and trend type, $F(5, 25)=3.16$, $p<.05$. As can be seen in Table 13, error was lower on no-motion trials, for all six trend types. However, the difference between the errors for motion and no-motion trials was larger when decreasing trends were presented. Thus, the negative effects of motion on accuracy were greater for trends with negative slopes.

Table 13

Absolute Error Scores for Motion and No-Motion Trials

Trend Type	Motion Trials	No-Motion Trials	Difference
Exponential Increasing	1.36	0.57	0.79
Exponential Decreasing	1.50	0.48	1.02
Asymptotic Increasing	1.27	0.42	0.85
Asymptotic Decreasing	1.50	0.42	1.16
Linear Increasing	1.25	0.41	0.84
Linear Decreasing	1.44	0.33	1.11

Over- and Underestimation Errors

Table 14 shows the over- and underestimation errors for each trend type. As can be seen from the table, the extrapolations were too high for increasing trends and too low for decreasing trends. In both cases, these biases represent overestimation of the amount of change in the trends as was found in Experiment 2. On motion trials, these biases tended to be even more pronounced and the overestimation errors even larger.

Table 14

Bias for the Six Different Trend Types

Trend Type	Overall Bias	Motion Trials	No Motion Trials
Exponential Increasing	0.43	0.51	0.34
Exponential Decreasing	-0.54	-0.81	-0.26
Asymptotic Increasing	0.17	0.34	0.01
Asymptotic Decreasing	-0.53	-1.00	-0.06
Linear Increasing	0.44	0.20	0.67
Linear Decreasing	-0.56	-1.00	-0.12

In the past, researchers have found that subjects underestimate the degree of growth or decay in exponential trends. This finding was not confirmed in the current extrapolation studies. For all extrapolation tasks, the predicted points were too high or too low, which indicates that subjects predicted more change than was represented in the trends. It was hypothesized that motion would make the rate of growth or decline more salient and lead to more accurate predictions. Although motion did not lead to more accurate extrapolations, it appears that the inclusion of motion did lead to even more over- and underestimation. If this motion-induced over- and underestimation is in fact due to enhanced salience of growth and decline in the motion condition, it is unclear why motion failed to enhance discrimination performance in Experiment 4a.

Chapter 9

General Discussion

One of the goals of the studies reported here was to systematically examine how variability, sample size, and graph type affect discrimination and extrapolation accuracy. Overall, graphs with a high degree of variability and low sample size were difficult and accuracy on all tasks was lower in these conditions. The effects of graph type and trend type were dependent upon the required task. For discrimination tasks, line graphs and scatterplots led to higher accuracy than histograms and suspended bar graphs. However, on extrapolation tasks, line graphs led to lower accuracy than histograms, scatterplots, and suspended bar graphs. Although discrimination accuracy was highest when nonlinear trends were presented, exponential trends led to lower extrapolation accuracy than asymptotic and linear trends. These findings will be further examined and analyzed in the following sections.

The Discrimination of Trends

Experiment 1 showed that variability and sample size affected the ability of subjects to discriminate between linear and nonlinear trend types. Overall, accuracy was highest when variability was low and sample size was high. This finding confirms previous results (e.g., Goodwin & Wright, 1994; Sanders, 1992) and suggests that humans are readily able to discriminate between different trends when the level of variability is low. Furthermore, even when the level of variability was high,

discrimination accuracy averaged 48%, which is well above the chance level of 33% (assuming a correct discrimination between increasing and decreasing trends). Thus, subjects were often able to see through the noise even in high-variability conditions. In addition, the effect of sample size suggests that including more data points in a time series graph results in improvements in discrimination ability. Finally, subjects were more accurate when nonlinear trends were present. Accuracy was similar for exponential and asymptotic curves but was lower when linear curves were presented.

In terms of graph type, accuracy was highest when line graphs and scatterplots were presented and dropped when bar graphs and suspended bar graphs were used. This confirms the previous claims (Cochran et al., 1989; Wallgren et al., 1996) that line graphs accurately convey information about trend. Overall, discrimination accuracy was best when nonlinear trends with a low level of variability and a large sample size were presented on line graphs or scatterplots.

In Experiment 1, signal detection analyses showed that the sensitivity of the subjects was lower when small sample sizes were displayed. The measure of relative efficiency compared how well the subjects did relative to an ideal observer. At lower levels of variability, relative efficiency was independent of sample size but when variability was high, larger sample sizes led to lower efficiency. This means that, compared to the ideal observer, subjects were less able to extract useful information from additional data when the data points were noisy.

Previous studies have found relative efficiencies of approximately .72 (Best et al., 1999; Legge et al., 1989). This previous research examined the ability of subjects to determine if two samples were drawn from the same or different populations. The average relative efficiency in Experiment 1 was .44, which is considerably lower than previous results. Because the current task involved discriminating between subtly different trends, it is possible that it was more difficult than tasks used in previous studies—discriminating between these trends involves using subtle cues such as curvature information, whereas deciding if two samples were drawn from the same or different populations involves using information about the relative position of two samples on a graph. In addition, average relative efficiency, as computed here, was based on only four trend pairs—increasing asymptotic/increasing linear, increasing exponential/increasing linear, decreasing asymptotic/decreasing linear, and decreasing exponential/decreasing linear. The relatively easy discriminations between increasing and decreasing trends and exponential and asymptotic trends were not included in the analysis of sensitivity and relative efficiency. Given this, if all possible trend pairs had been examined, average relative efficiency would have been considerably higher.

Experiment 4 manipulated the way that the data points were presented. This experiment examined the discrimination and extrapolation ability of subjects in a condition that simulated motion. In the motion condition, the data points were presented individually in a sequential fashion. Discrimination ability was not affected by the

inclusion of motion—subjects were equally accurate in the motion and no motion conditions. Thus, it appears that, at least for discrimination, the inclusion of motion neither helps nor hinders the ability to discriminate between different linear and nonlinear trend types.

The Use of Curvature in Discrimination

When discriminating between trends of the sort used in these experiments, subjects must analyze two crucial kinds of information. First, they must use slope information to determine whether the trend is increasing or decreasing. Although previous studies (e.g., Jones et al., 1978) have reported that visual analysts are often unable to detect differences between increasing and decreasing trends, in these experiments subjects were able to determine the direction of the trend. On 98% of trials, they were correct in judging whether the displayed trend was increasing or decreasing. Second, the amount of curvature in the trend may be used to determine how much acceleration or deceleration occurs in a set of data and thus discriminate between various linear and nonlinear curves. Exponential curves are accelerating and have positive curvature values, asymptotic curves are decelerating and have negative curvature values, and linear curves show a steady rate of growth and have curvature values near zero.

Figure 22 presents frequency distributions of the responses for the three response types as a function of the index of curvature of the displayed data points. As can be seen in the figure, when the curvature was positive, subjects were more likely to respond that

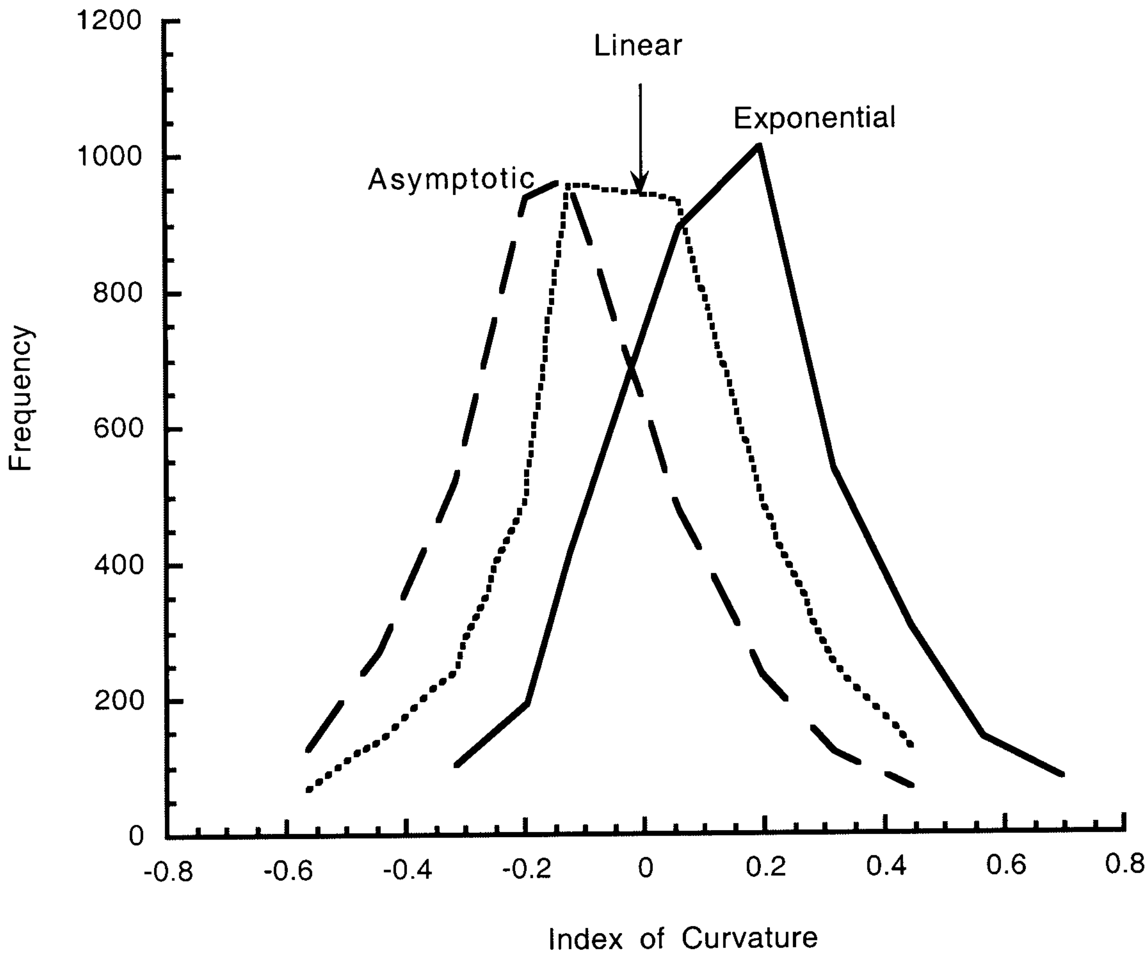


Figure 22. Number of responses of each type as a function of the index of curvature.

an exponential trend was present. When the curvature was negative, subjects were more likely to respond that an asymptotic trend was present. When the curvature was close to zero, subjects tended to respond that a linear trend was present. This finding suggests that subjects were using information about curvature when they made decisions about what type of trend was present.

On trials where a correct discrimination was made, there was a clear separation between the distributions of response types as a function of curvature values (see Figure 23a). On these trials, subjects were able to use curvature information to determine which trend type was most likely present. Subjects made more correct responses when they were presented with exponential and asymptotic curves. When subjects made incorrect discriminations (see Figure 23b), the frequency distributions were flatter and, for all three trend types, the mean curvature was closer to zero. These distributions showed a greater degree of overlap between the response types which suggests that on some trials subjects were not using curvature information when they made discriminations. On these trials, the curvature information was ambiguous and, therefore, it was more difficult for subjects to determine which trend was present.

The curvature distributions of incorrect responses (see Figure 23b) reveal another interesting aspect of the data. When subjects responded that the trend type was exponential, the curvature value tended to be positive, when they responded that the trend was asymptotic, the curvature was typically negative, and when they responded that the

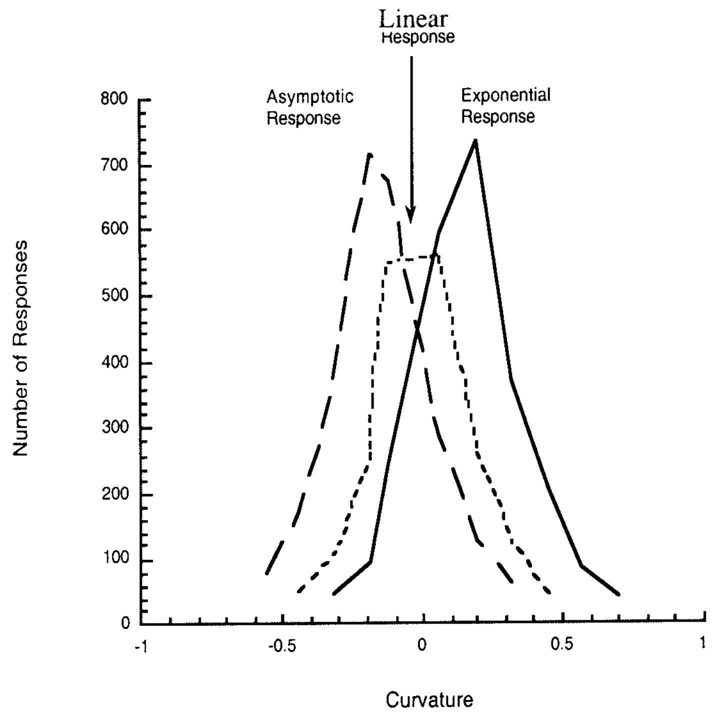


Figure 23a. Response type as a function of the index of curvature on trials where a correct discrimination was made.

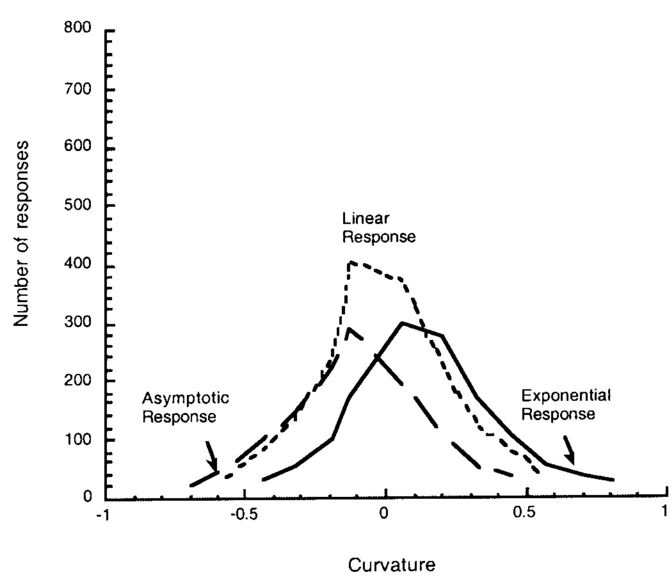


Figure 23b. Response type as a function of the index of curvature on trials where a incorrect discrimination was made.

trend was linear, the curvature was generally near zero. Thus, it appears that subjects were making some use of curvature information even as they made incorrect discriminations. On incorrect trials, the curvature distributions were overlapping, which indicates that these trials were difficult. Although subjects made an incorrect response, on many trials they were using curvature information appropriately, given the information presented to them For example, in Figure 23b, the responses falling the right-hand tail of the "exponential" distribution were appropriate ones; they turned out to be wrong only because random variability misleadingly caused a linear or asymptotic underlying trend to have positive curvature in the displayed data.

One unexpected (and counterintuitive) finding from Experiment 1 was that subjects were least accurate at identifying linear trends. Accuracy was 67% for exponential trends, 65% for asymptotic trends, and 54% for linear trends (note that in Figure 23 the peak of the "linear" response distributions was higher than the other two distributions when an incorrect response was made and lower when a correct response was made). These results are consistent with previous research on the intermediate size problem, in which subjects are required to discriminate between three stimuli with overlapping distributions and show low accuracy for the middle stimulus. This phenomenon, which has been documented in both animals (Laverty, Werboff, & Frey, 1969) and humans (Baker & Ulness, 1969), could offer an explanation as to why discrimination accuracy was lower when linear trends were presented.

Figure 24 shows the distributions of trend types as a function of the index of curvature for all of the trials that occurred in Experiment 1. Because the three distributions overlap, there was some ambiguity on most trials regardless of the trend type. As can be seen in the figure, the degree of overlap is greatest for linear trends, because the linear distribution overlaps with the exponential distribution on one side and the asymptotic distribution on the other. In contrast, although both the exponential and asymptotic distributions overlap with the linear distribution, the degree of overlap is smaller for these trend types, which have relatively little overlap with one another. This means that discriminations for these trends will be less difficult. For example, although the exponential distribution overlaps with both the asymptotic and linear distributions, when the curvature is high there is little or no overlap and on these trials it would be relatively clear that an exponential trend was present. On the other hand, when linear trends were presented, some of the linear trends would appear exponential and other linear trends would appear asymptotic. Therefore, for linear trends, discrimination was more difficult because there were relatively more trials that contained ambiguous curvature information. Even when the index of curvature was 0 (the strongest possible evidence for a linear underlying trend), there was a good chance that the trend was actually asymptotic or exponential.

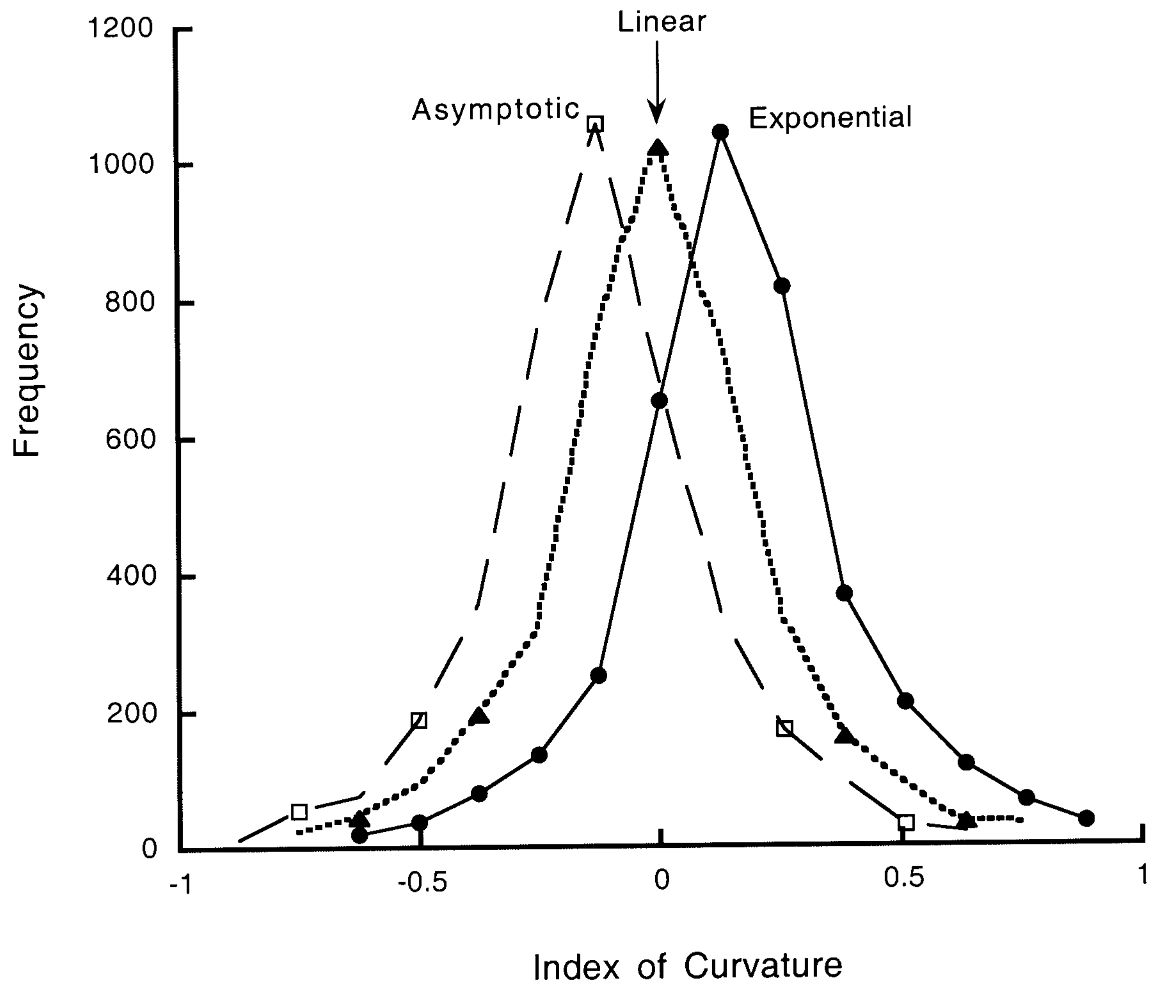


Figure 24. The index of curvature for the three trend types.

In order to further examine the intermediate size problem, an analysis was conducted of the performance of an ideal observer who responds solely on the basis of the index of curvature, but makes optimal use of that information. To simulate the requirement that subjects allocate responses equally across the categories, response criteria were set for the ideal observer along the curvature continuum in such a way that the three trend responses were made equally often. Based on these cutoffs, the ideal observer was correct on 59% of exponential trials, 57% of asymptotic trials, and 42% of linear trials. Thus, the intermediate size problem occurs even when information from the curvature index is used optimally. This result suggests, in line with the discussion above, that the lower accuracy on linear trials is a consequence of the way that curvature information is structured in the task rather than due to a perceptual inability to detect linear trends.

To conclude, when discriminating between different trends, subjects can use information about slope and curvature to help them make a decision about whether an exponential, asymptotic, or linear trend is present. Both the human subjects and the ideal observer were affected by the intermediate size problem and had lower accuracy when linear trends are presented. Although subjects did not respond optimally on all trials, they do appear to have used curvature and slope information as they made decisions about what trend was present.

The Extrapolation of Future Points on Time Series Graphs

Experiment 2 examined the ability of subjects to predict future points on a time series graph. As expected, extrapolation accuracy was best when variability was low and sample size was high. Extrapolation ability was lowest for exponential curves (which begin with an initial stable period, followed by a period of sharp growth or decay) and was higher when asymptotic trends were presented. Past research (Jones, 1977, 1979, 1985; Wagenaar & Sagaria, 1975) has shown that the prediction of this sharp growth or decay is difficult. The present results confirmed these findings and extended them to the case of graphically displayed trends, which were expected to provide perceptually favorable conditions for making such judgments.

Extrapolation accuracy was highest when the curves were presented on bar graphs, scatterplots, and suspended bar graphs and was lower when line graphs were used. This confirms previous claims (Wallgren et al., 1996) that although line graphs are more effective for discrimination of patterns or trends, other graphical formats are better suited for extrapolation tasks involving specific data values. In the extrapolation task subjects had to predict an exact point and it is possible that the increased accuracy for scatterplots, histograms, and suspended bar graphs was due to the fact that these graph formats emphasize individual points, which may have made the points more salient and thus easier for subjects to use in predicting future data points.

In Experiment 3, subjects predicted three future points on a time series graph. After each extrapolation, subjects were presented with feedback in the form of the next point (with variability added) in the series. Although previous research has shown that subjects are more accurate when they receive feedback, the results of this study did not find that feedback significantly improved the ability to predict future points. Although, extrapolation accuracy was not significantly different in the feedback and no-feedback conditions, subjects were slightly more accurate when they received feedback. A power analysis suggested that including two additional subjects would have strengthened the power of the test to the point of statistical significance. The role of feedback in enhancing the ability to predict future points on time series graphs clearly merits further study.

In Experiment 5, subjects were presented with graphs that simulated motion and had to predict the next point in the series. The sequential presentation of data points did not improve extrapolation ability. On motion trials, the over- and underestimation errors were larger than on the no motion trials. Thus, the inclusion of motion did increase the appearance of growth or decay, but in these experiments, led to more error. Given the possibility that prior knowledge of the trend types led to the extrapolation overestimations, future studies could examine the effects of motion when subjects have no prior knowledge of trend types being presented.

Extrapolation Biases

Although much of the previous research (e.g., Mullett & Cheminat, 1995) found that, on extrapolation tasks, subjects underestimated future points on exponential trends, the current studies produced an opposite finding. Overall, subjects overestimated the degree of change in both increasing and decreasing trends (see Figure 25). As can be seen in the figure, the extrapolations were shifted to the right of zero for increasing trends and to the left of zero for decreasing trends. The overestimation errors were highest for exponential and linear trends and were closer to zero for asymptotic trends. It is possible that prior knowledge about the six trend types affected the predictions about future data points. Before data collection began, subjects were made aware of the six trend types that would be presented and they had considerable experience with detecting the trends in Experiment 1. This prior knowledge could have led to predictions that were higher (or lower for decreasing trends) than those that would have been made by naïve subjects. Thus, the use of the "signals known exactly" procedure may have yielded knowledge about the different signals that affected extrapolation accuracy.

Extrapolation involves two different perceptual tasks. First, subjects have to decide what type of trend is present and second, they have to predict future data points. In many real-world situations, deciding on the type of trend is difficult because there is, in principle, an unlimited number of possibilities. When subjects made extrapolations, they were familiar with the underlying trends and were aware that one of six possible

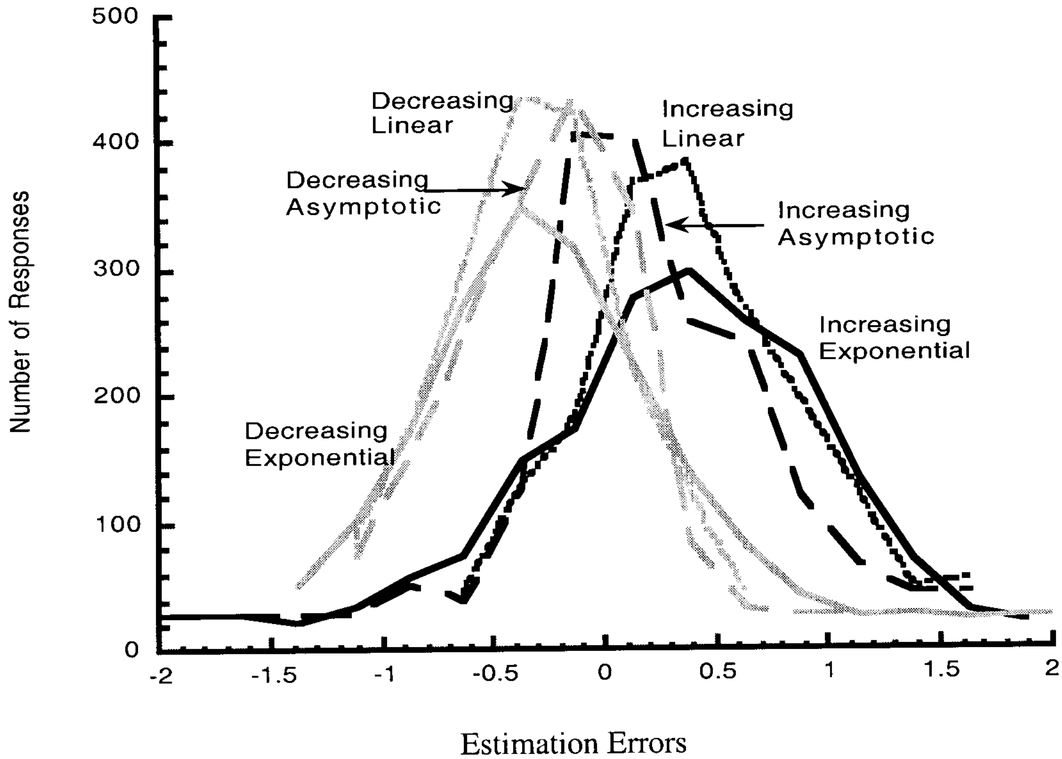


Figure 25. The estimation errors of the subjects for each trend type.

trends was present. This knowledge could have led subjects to overestimate future data points. For example, subjects were aware that exponential trends would be presented on one third of the trials. This awareness, paired with the knowledge that the last few data points on exponential trends represent a sharp increase or decrease, could have led subjects to make judgements that were too high for increasing trends and too low for decreasing trends. The exact mechanism by which such knowledge could have produced overestimations of change remains unclear, but it is possible that change-enhancing memory distortions play a role (Silka, 1989).

There was one interesting exception to the finding of overestimation bias. When exponential trends with a small sample size were presented, subjects underestimated the amount of change in future curve points (see Figure 23b). This finding contrasts with previous research on exponential trends (Wagenaar & Timmers, 1978, 1979) in which small sample sizes reduced the underestimation found with larger sample sizes, thus producing more perception of change. In an effort to explain this previous result, Andreassen and Kraus (1990) proposed the salience hypothesis, according to which smaller sample sizes result in larger point-to-point changes and hence make sharp increases more salient. However, it is apparent that this hypothesis cannot account for the bias with low sample sizes in the present research, given that the tendency in this case was to report less change rather than more. How best to account for these discrepancies is unclear, but it is possible that they stem from differences in the methods used to present

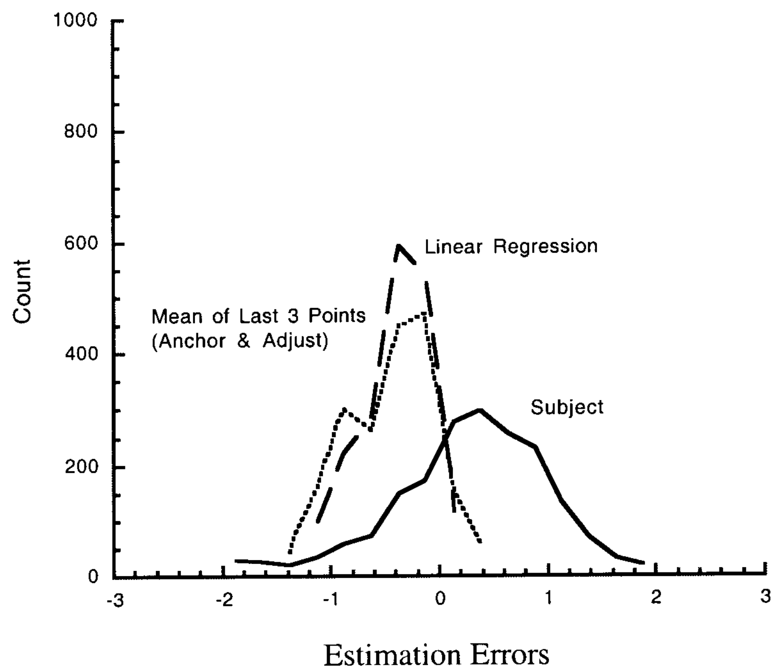
the time series data to subjects. In the previous research subjects typically viewed lists of numbers and gave their extrapolations as numerical values. In the present experiments subjects viewed graphs and made extrapolations by adjusting the last point on the graph.

<u>Extrapolation Heuristics</u>

The subjects' extrapolation, the point predicted by linear regression, and the mean of the last three points were compared in order to determine which method of extrapolation was most accurate. Although it was expected that subject extrapolations would be more accurate than the two heuristic methods (especially when variability was low), this was not the case. At all levels of variability, the points predicted by linear regression and the mean of the last three points were more accurate than the extrapolations. As can be seen in Figure 26, the distributions of errors for the subjects were more spread out than the error distributions for the points predicted by linear regression and the mean of the last three points. Thus, one reason that subjects were less accurate than the other methods was the fact that they had more overestimation errors, as represented by the long tails of these distributions. In addition, as can be seen from the figure, the means of the subject error distributions tended to be shifted away from 0 toward the direction of change in the trend.

Previous research on extrapolation (e.g., Bolgar & Harvey, 1993) suggests that subjects use the last few displayed points as an anchor and adjust their extrapolations relative to these points. The present research has suggested that subjects are not using

Increasing Exponential



Decreasing Exponential

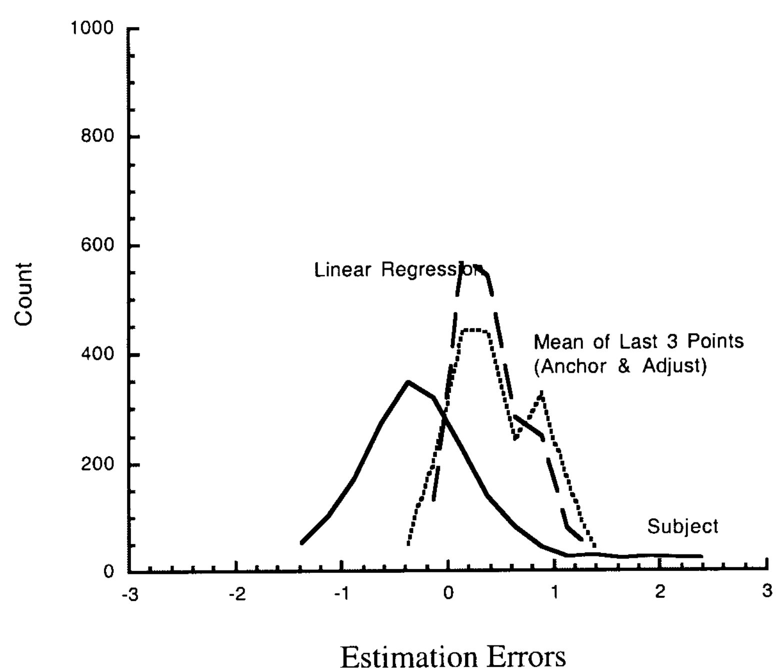
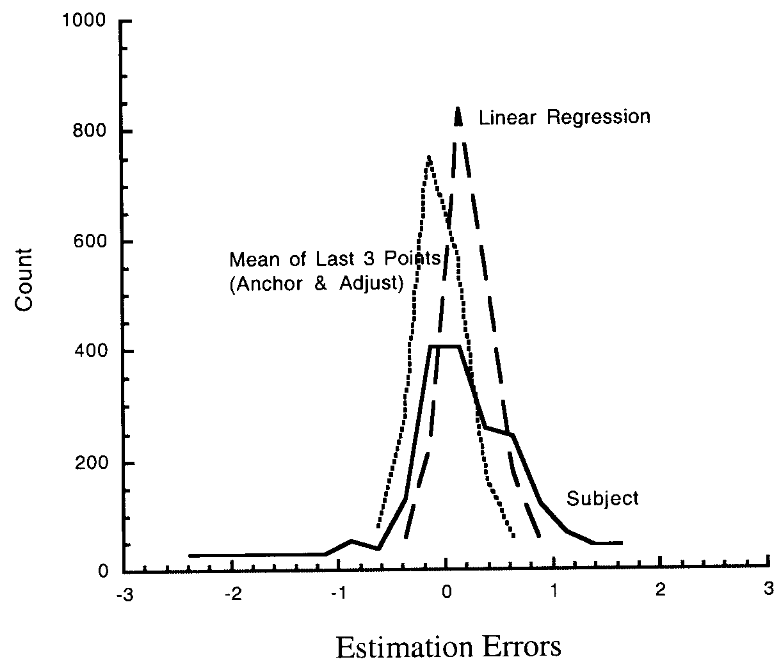


Figure 26a. The estimation errors of the subjects, the point predicted by linear regression, and the mean of the last three data points on exponential curves.

Increasing Asymptotic



Decreasing Asymptotic

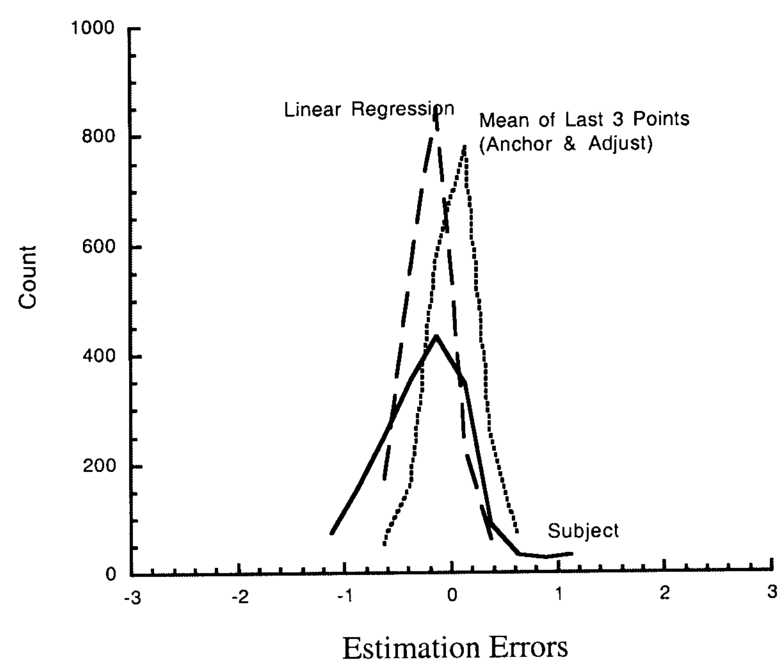
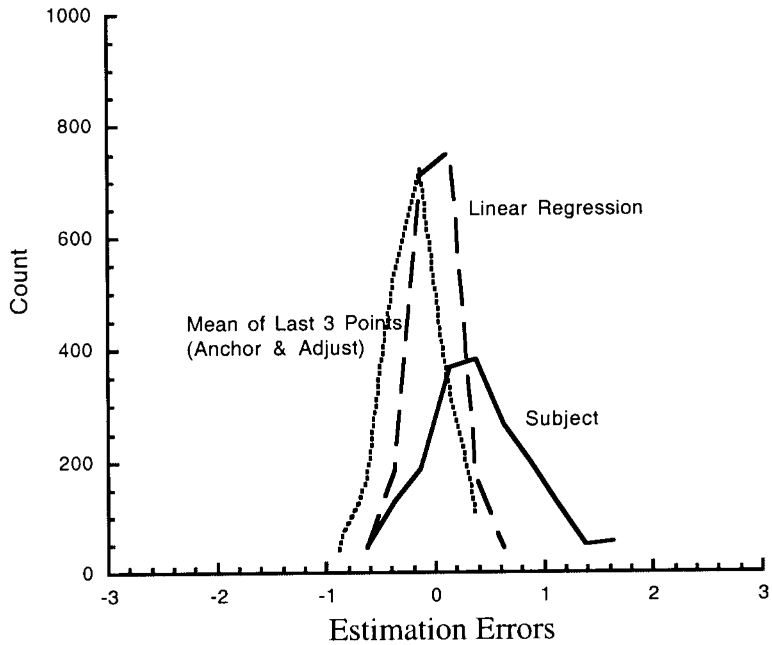


Figure 26b. The estimation errors of the subjects, the point predicted by linear regression, and the mean of the last three data points on asymptotic curves.

Increasing Linear



Decreasing Linear

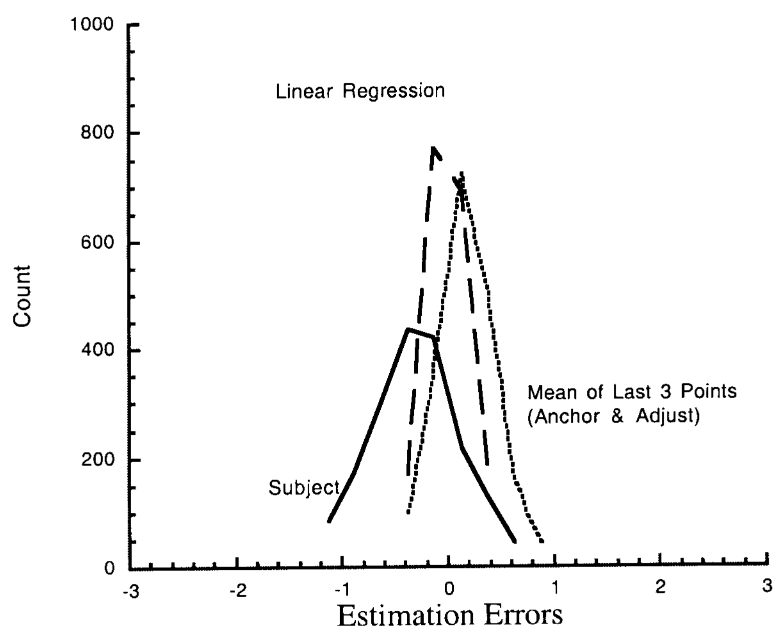


Figure 26c. The estimation errors of the subjects, the point predicted by linear regression, and the mean of the last three data points on linear curves.

anchor and adjust heuristics when they make predictions. In almost all cases, the predictions were closer to the points predicted by linear regression, which suggests that the linear components of trends have a greater influence on the predictions than the last few curve points.

Given the interest in the ability to predict future points on a time series graph, this finding is significant. It was not expected that the points predicted by linear regression would be accurate predictors of future points on nonlinear time series. Much of the previous research on extrapolation heuristics has focused on variations of the anchor and adjust method of extrapolation. The current findings suggest two conclusions—first, subjects are not using anchor and adjust heuristics or linear regression when they make predictions and, second, the predictions of the subjects would have been more accurate on average if they had used either of these heuristics.

Conclusions

There has been much debate about which graph types best convey numerical information. The results of the present experiments suggest that the optimal graph type is dependent upon the task that is required. When subjects had to discriminate between different types of trends, line graphs and scatterplots led to higher levels of accuracy, but when they had to predict future points on a time series graph, line graphs led to much lower levels of accuracy than the other graph types. It appears that line graphs effectively convey trend information but make it difficult to predict exact data points. When

histograms, scatterplots, and suspended bar graphs are used to present data, each individual point is emphasized and this may make it easier for subjects to predict future points. This confirms the finding that bar graphs effectively convey specific quantitative information and line graphs best convey information about the underlying trend (Cochran et al., 1989).

Previous research has shown that increasing the number of data points improves the ability to detect linear relationships (e.g., Lewandowsky & Spence, 1989), yet similar increases reduce the ability to detect exponential trends (Wagenaar & Timmers, 1978, 1979). The current studies confirmed that the effects of sample size were dependent upon the type of curve that was present. On discrimination tasks, accuracy for exponential and asymptotic trends was similar regardless of the number of data points present. When linear trends were presented, discrimination ability increased as the number of data points increased.

When asymptotic and linear trends were presented, extrapolation accuracy increased systematically as sample size increased. For these curves, accuracy was highest when sample size was high, intermediate when sample size was moderate, and lowest when sample size was low. However, when exponential trends were presented, absolute accuracy was similar regardless of the number of data points presented. In these experiments, there was an overall tendency toward overestimation of change for all trend types. For linear and asymptotic trends this tendency was greater when sample size was

small and decreased with rises in sample size. For exponential trends, there was a tendency toward overestimation when sample size was small or intermediate but there was a tendency toward underestimation when sample size was high. Thus, as with the previous studies, there seems to be an underestimation bias with larger sample sizes.

Given the current interest in nonlinear growth (e.g., in the stock market, in pollution levels, or in human populations), it is important to consider the implications of the current findings for strategies that could be applied to the real-world problem of detecting and extrapolating from patterns of change. Contrary to previous research, it seems that extrapolation accuracy could be improved if people were instructed to use the linear regression line as an extrapolation guide and to adjust their predictions relative to this line. Thus, it may be that a modified anchor and adjust heuristic based on linear components of trends is advisable. Rather than using the last few curve points as a basis for prediction, predictions may be more accurate if the linear regression line were used as the anchor and predictions were adjusted from this point. Therefore, if perceivers were presented with an exponential curve, they could visually fit a linear regression line through the data and adjust the slope of this line until they were satisfied that it represented the rise or fall in the data. For asymptotic trends, they could decrease the slope of the line until they were satisfied with the fit of the line and the data. For linear trends, the slope of the line would presumably not have to be adjusted and the visually fitted line could be used as a basis for prediction. The possibility that some such strategy

combined with the use of well-chosen graphical displays, would help improve the handling of problems associated with nonlinear growth merits further research in applied settings.

References

Anderson, J. R., & Bower, G. H. (1973). Human associative memory. Washington, DC: Winston & Sons.

Andreassen, P. B., & Kraus, S. J. (1990). Judgmental extrapolation and the salience of change. Journal of Forecasting, 9, 347-372.

Anscombe, F. J. (1973). Graphs in statistical analysis. American Statistician, 27, 17-21.

Ashton, A. H. (1984). A field test of implications of laboratory studies of decision making. Accounting Review, 59, 361-375.

Baker, K. E., & Ulness, J. (1968). The dependency of intermediate size problem responses upon sets in human Ss. Psychonomic Science, 13, 75-76.

Balzer, W. K., Doherty, M. E., & O'Connor, R., Jr. (1989). Effects of cognitive feedback on performance. Psychological Bulletin, 106, 410-433.

Best, L. A., Smith, L. D., & Stubbs, D. A. (1998). Perception of graphs: Discriminability and variability. Paper presented at the Maine Biological and Medical Sciences Symposium, Biddeford, ME.

Best, L. A., Smith, L. D., Frey, R. B., & Stubbs, D. A. (1998). Discrimination of trends in time series graphics. Poster presented at the annual meeting of the Eastern Psychological Association, Boston, MA.

Best, L. A., Smith, L. D., Stubbs, D. A., & Frey, R. B. (1999). Sensitivity and bias in graphical perception. Poster presented at the annual meeting of the Eastern Psychological Association, Providence, RI.

Bobko, P., & Karren, R. (1979). The perception of Pearson product moment correlations from bivariate scatterplots. Personnel Psychology, 32, 313-325.

Bolger, F., & Harvey, N. (1990). Context-sensitive heuristics in statistical reasoning. Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 46A, 779-811.

Butler, D. L. (1993). Graphs in psychology: Pictures, data, and especially concepts. Behavior Research Methods, Instruments, & Computers, 25, 81-92.

Carroll, J. M. (1997). Human-computer interaction: Psychology as a science of decision. Annual Review of Psychology, 48, 61-83.

Carswell, C. M. (1992). Choosing specifiers: An evaluation of the basic tasks model of graphical perception. Human Factors, 34, 535-554.

Chambers, J. M., Cleveland, W. S., Kleiner, B., & Tukey, P. A. (1983). Graphical methods for data analysis. Pacific Grove, CA: Wadsworth & Brooks/Cole.

Cleveland, W. S. (1993a). A model for studying display methods of statistical graphics. Journal of Computational and Graphical Statistics, 2, 323-343.

Cleveland, W. S. (1993b). Visualizing data. Summit, NJ: Hobart Press.

Cleveland, W. S. (1994). The elements of graphing data. Summit, NJ: Hobart Press.

Cleveland, W. S., Diaconis, P., & McGill, R. (1982). Variables on scatterplots look more highly correlated when scales are increased. Science, 216, 1138-1141.

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

Cleveland, W. S., & McGill, R. (1986). An experiment in graphical perception. International Journal of Man-Machine Studies, 25, 491-500.

Cleveland, W. S., McGill, M. E., & McGill, R. (1988). The shape parameter of a two-variable graph. Journal of the American Statistical Association, 83, 289-300.

Cochran, J. K., Albrecht, S. A., & Green, Y. A. (1989). Guidelines for evaluating graphical designs: A framework based on human perception skills. Technical Communication, 36, 25-32.

Costigan-Eaves, P., & Macdonald-Ross, M. (1990). William Playfair (1759-1823). Statistical Science, 5, 318-326.

Culbertson, H. M., & Powers, R. D. (1959). A study of graph comprehension difficulties. Audio-Visual Communication Review, 7, 97-100.

DeProspero, A., & Cohen, S. (1979). Inconsistent visual analyses of intrasubject data. Journal of Applied Behavior Analysis, 12, 573-579.

DeSanctis, G. (1984). Computer graphics as decision aids: Directions for research. Decision Sciences, 15, 463-487

Doble, T. M., Boynton, D. M., Smith, L. D., & Stubbs, D. A. (1992). Judging statistical significance from graphs: Effects of mean lines and scaling. Paper presented at the meeting of the American Psychological Association, Washington, DC.

Erlick, D., & Mills, R. J. (1967). Perceptual quantification of conditional dependency. Journal of Experimental Psychology, 73, 9-14.

Flores, B. E., & Olson, D. L. (1992). Judgmental adjustment of forecasts: A comparison of methods. International Journal of Forecasting, 7, 421-433.

Furlong, M. J., & Wampold, B. E. (1982). Intervention effects and relative variation as dimensions in experts' use of visual inference. Journal of Applied Behavior Analysis, 15, 415-421.

Gibson, G., & Ottenbacher, K. (1988). Characteristics influencing the visual analysis of single-subject data: An empirical analysis. Journal of Applied Behavioral Science, 24, 298-314.

Goldstein, E. B. (1999). Sensation and perception (4th Ed.). Pacific Grove, CA: Brooks/Cole.

Goldstein, E. B. (1996). Sensation and perception (5th Ed.). Pacific Grove, CA: Brooks/Cole.

Goodwin, P., & Wright, G. (1994). Heuristics, biases and improvement strategies in judgmental time series. Omega: The International Journal of Management Science, 22, 553-568.

Gottmann, J. M., & Glass, G. V. (1978). Analysis of interrupted time series experiments. In T. R. Kratochwill (Ed.), Single subject research: Strategies for evaluating change (pp. 197-235). New York: Academic Press.

Harvey, N., Bolgar, F., & McClelland, A. (1994). On the nature of expectations. British Journal of Psychology, 85, 203-229.

Henry, G. T. (1990). Practical sampling. Newbury Park, CA: Sage.

Henry, G. T. (1995). Graphing data: Techniques for display and analysis. Thousand Oaks, CA: Sage.

Jones, G. V. (1977). Polynomial perception of exponential growth. Perception and Psychophysics, 21, 197-200.

Jones, G. V. (1979). A generalized polynomial model for perception of exponential series. Perception and Psychophysics, 25, 232-234.

Jones, G. V. (1985). Perception of inflation: Polynomial not exponential. Perception and Psychophysics, 36, 485-487.

Jones, R. R., Weinrott, M. R., & Vaught, R. S. (1978). Effects of serial dependency on the agreement between visual and statistical inference. Journal of Applied Behavior Analysis, 11, 277-283.

Julesz, B. (1984). A brief outline of the texton theory of human vision. Trends in Neuroscience, 7, 41-45.

Kazdin, A. E. (1982). Single case research designs: Methods for clinical and applied settings. New York: Oxford University Press.

Keren, G. (1983). Cultural differences in the misperception of exponential growth. Perception and Psychophysics, 34, 289-293.

Keren, G. (1991). Calibration and probability judgments: Conceptual and methodological issues. Acta Psychologia, 77, 217-273.

Knapp, T. J. (1983). Behavior analysts' visual appraisal of behavior change in graphic display. Behavior Assessment, 5, 155-164.

Kolata, G. (1984). The proper display of data. Science, 226, 156-157.

Kosslyn, S. M. (1992). Cognitive neuroscience and the human self. In A. Harrington (Ed.), So human a brain: Knowledge and values in the neurosciences. (pp. 37-56). Boston: Birkhaeuser.

Kosslyn, S. M. (1994). Elements of graph design. New York: W. H. Freeman.

Kosslyn, S. M. & Chabris, C. F. (1992). Minding information graphics. Folio, 70, 69-71.

Kratochwill, T. R. (1978). Foundations of time series research. In T. R. Kratochwill (Ed.), Single subject research: Strategies for evaluating change (pp. 1-90). New York: Academic Press.

Lane, D. M., Anderson, C. A., & Kellam, K. L. (1985). Judging the relatedness of variables: The psychophysics of covariation detection. Journal of Experimental Psychology: Human Perception and Performance, 5, 640-649.

Lappin, J. S. (1995). Visible information about structure from motion. In W. Epstein & S. Rogers (Eds.), Perception of space and motion. Handbook of perception and cognition (2nd ed.) (pp. 165-199). San Diego, CA: Academic Press.

Laverty, J. J., Werboff, J., & Frey, R. B. (1969). Solution of intermediate size problem and discrimination transfer by rats. Journal of Comparative and Physiological Psychology, 68, 262-267.

Lawrence, M. J., Edmundson, R. H., & O'Connor, M. J. (1986). The accuracy of combining judgmental and statistical forecasts. Management Science, 32, 1521-1532.

Lawrence, M. J,. Edmundson, R. H., & O'Connor, M. J. (1985). An examination of the accuracy of judgmental extrapolation of time series. International Journal of Forecasting, 1, 25-35.

Lawrence, M., & Makridakis, S. (1989). Factors affecting judgmental forecasts and confidence intervals. Organizational Behavior and Human Decision Processes, 42, 172-187.

Lawrence, M., & O'Connor, M. (1992). Exploring judgmental forecasting. International Journal of Forecasting, 8, 15-26.

Legge, G. E., Gu, Y., & Luebker, A. (1989). Efficiency of graphical perception. Perception and Psychophysics, 46, 365-374.

Lewandowsky, S., & Spence, I. (1989). The perception of statistical graphs. Sociological Methods and Research, 18, 200-242.

Loftus, G. R. (1993). A picture is worth a thousand p values: On the irrelevance of hypothesis testing in the microcomputer age. Behavior Research Methods, Instruments, & Computers, 25, 250-256.

Lohse, G. L. (1993). A cognitive model for understanding graphical perception. Human-Computer Interaction, 8, 353-388.

Lyon, D. R., & Waag, W. L. (1995). Time course of visual extrapolation accuracy. Acta Psychologia, 89, 239-260.

Macdonald-Ross, M. (1977). How numbers are shown: A review of research on the presentation of quantitative data in texts. Audio-Visual Communication Review, 25, 359-409.

MacKinnon, A. J., & Wearing, A. J. (1991). Feedback and the forecasting of exponential change. Acta Psychologia, 76, 177-191.

Macmillan, N. A., & Creelman, C. D. (1991). Detection theory: A user's guide. New York: Cambridge University Press.

Matyas, T. A., & Greenwood, K. M. (1990). Visual analysis of single-case time series: Effects of variability, serial dependence, and magnitude of intervention effects. Journal of Applied Behavior Analysis, 23, 341-351.

Mayer, R. E. (1993). Comprehension of graphics in texts: An overview. Learning and Instruction, 3, 239-245.

McDowall, D., McCleary, R., Meidinger, E. E., & Hay, R. A., Jr. (1980). Interrupted time series analysis. Beverly Hills, CA: Sage.

Meyer, J. (1997). A new look at an old study on information display: Washburne (1927) reconsidered. Human Factors, 39, 333-340.

Meyer, J., Shinar, D., & Leiser, D. (1997). Multiple factors that determine performance with tables and graphs. Human Factors, 39, 268-286.

Mullet, E., & Cheminat, Y. (1995). Estimation of exponential expressions by high school students. Contemporary Educational Psychology, 20, 451-456.

Munger, G. F., Snell, M. E., & Lloyd, B. H. (1989). A study of the effects of frequency of probe data collection and graph characteristics on teachers' visual analysis. Research in Developmental Disabilities, 10, 109-127.

Nisbett, R. E., & Ross, L. (1980). Human inference: Strategies and shortcomings on social judgment. Englewood Cliffs, NJ: Prentice-Hall.

O'Connor, M., Remus, W., & Grigg, K. (1993). Judgmental forecasting in times of change. International Journal of Forecasting, 9, 163-172.

Ottenbacher, K. J. (1986). Reliability and accuracy of visually analyzing graphed data from single-subject designs. American Journal of Occupational Therapy, 40, 464-469.

Ottenbacher, K. J. (1990a). Visual inspection of single-subject data: An empirical analysis. Mental Retardation, 28, 283-290.

Ottenbacher, K. J. (1990b). When is a picture worth a thousand *p* values? A comparison of visual and quantitative methods to analyze single subject data. Journal of Special Education, 23, 436-449.

Park, H. S., Marascuilo, L., & Gaylord-Ross, R. (1990). Visual inspection and statistical analysis in single-case designs. Journal of Experimental Education, 58, 311-320.

Parsonson, B. S. & Baer, D. M. (1978). The analysis and presentation of graphic data. In T. R. Kratochwill (Ed.), Single Subject Research: Strategies for Evaluating Change (pp. 101-167). New York: Academic Press.

Pinker, S. (1990). A theory of graph comprehension. In R. Freedle (Ed.), Artificial intelligence and the future of testing (pp. 72-126). Hillsdale, NJ: Lawrence Erlbaum.

Pinker, S. (1997). How the mind works. New York: Norton.

Price, D. J. (1963). Little science, big science. New York: Columbia University Press.

Rojahn, J., & Schultze, H. (1985). The linear regression line as a judgment aid in visual analysis of serially dependent A-B time series data. Journal of Psychopathology and Behavior Assessment, 7, 191-206.

Sanders, N. R. (1992). Accuracy of judgmental forecasts: A comparison. Omega: The International Journal of Management Science, 20, 353-364.

Sanders, N. R., & Ritzman, L. P. (1992). The need for contextual and technical knowledge in judgmental forecasting. Journal of Behavioral Decision making, 5, 39-52.

Silka, L. (1989). Intuitive judgments of change. New York: Springer-Verlag.

Spence, I. (1990). The visual psychophysics of graphical elements. Journal of Experimental Psychology: Human Perception and Performance, 16, 683-692.

Swets, J. A. (1996). Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers. Mahwah, NJ: Lawrence Erlbaum.

Tannenbaum, R. S. (1998). Theoretical foundations of multimedia. New York: Computer Science Press.

Tilling, L. (1975). Early experimental graphs. British Journal for the History of Science, 8, 193-213.

Timmers, H., & Wagenaar, W. A. (1977). Inverse statistics and misperception of exponential growth. Perception and Psychophysics, 21, 558-562.

Tufte, E. R. (1983). The visual display of quantitative information. Cheshire, CT: Graphics Press.

Tufte, E. R. (1990). Envisioning information. Cheshire, CT: Graphics Press.

Tukey, J. W. (1990). Data-based graphics: Visual display in the decades to come. Statistical Science, 5, 327-339.

Tukey, J. W. (1993). Graphics comparisons of several linked aspects: Alternatives and suggested principles. Journal of Computational and Graphical Statistics, 2, 1-49.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1127-1131.

Wagenaar, W. A., & Sagaria, S. D. (1975). Misperception of exponential growth. Perception and Psychophysics, 18, 416-422.

Wagenaar, W. A., & Timmers, H. (1978). Extrapolation of exponential time series is not enhanced by having more data points. Perception and Psychophysics, 24, 182-184.

Wagenaar, W. A., & Timmers, H. (1979). The pond-and-duckweed problem: Three experiments on the misperception of exponential growth. Acta Psychologia, 43, 239-251.

Wainer, H. (1992). Understanding graphs and tables. Educational Researcher, 21, 14-23.

Wainer, H. (1997). Some multivariate displays for NAEP results. Psychological Methods, 2, 34-63.

Wainer, H., & Thissen, D. (1981). Graphical data analysis. Annual Review of Psychology, 32, 191-241.

Wainer, H., & Thissen, D. (1993). Graphical data analysis. In G. Keren & C. Lewis (Eds.), A Handbook for data analysis in the behavioral sciences: Statistical issues (pp. 391-457). Hillsdale, NJ: Lawrence Erlbaum.

Wallgren, A., Wallgren, B., Persson, R., Jorner, U., & Haaland, J. (1996). Graphing statistics & data: Creating better charts. London: Sage.

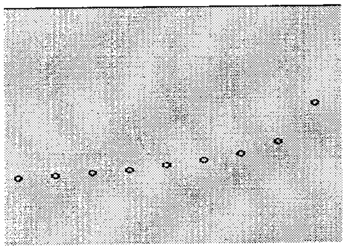
Wampold, B., & Furlong, M. (1981). The heuristics of visual inference. Behavior Assessment, 3, 79-92.

Washburne, J. N. (1927). An experimental study of various graphic, tabular and textual methods of presenting quantitative material. Journal of Educational Psychology, 18, 361-376.

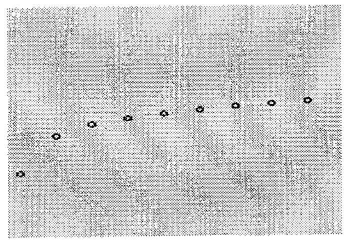
Winklhofer, H., Diamantopoulos, A., & Witt, S. F. (1996). Forecasting practice: A review of the empirical literature and an agenda for future research. International Journal of Forecasting, 12, 193-221.

Appendix A

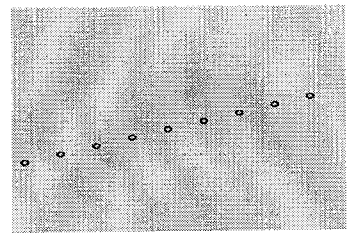
Examples of the Different Curves with No Variability



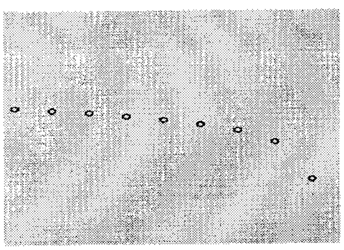
Exponential Increase
Variability = 0
Sample Size = 9



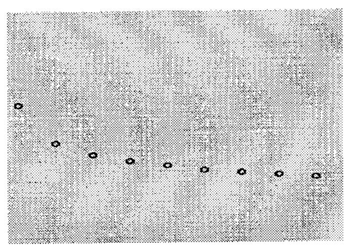
Asymptotic Increase
Variability = 0
Sample Size = 9



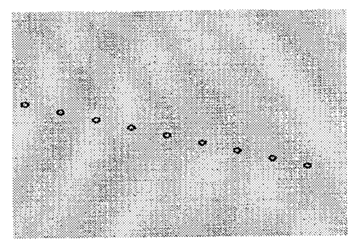
Linear Increase
Variability = 0
Sample Size = 9



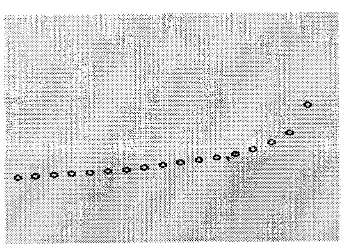
Exponential Decrease
Variability = 0
Sample Size = 9



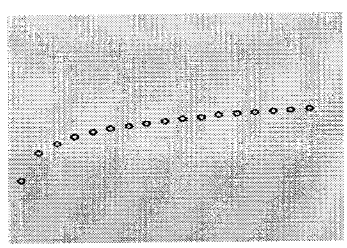
Asymptotic Decrease
Variability = 0
Sample Size = 9



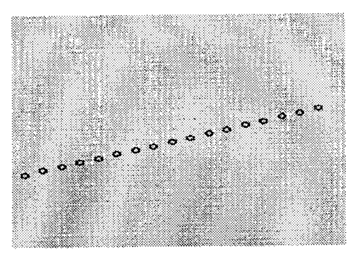
Linear Decrease
Variability = 0
Sample Size = 9



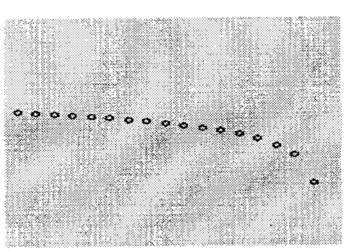
Exponential Increase
Variability = 0
Sample Size = 17



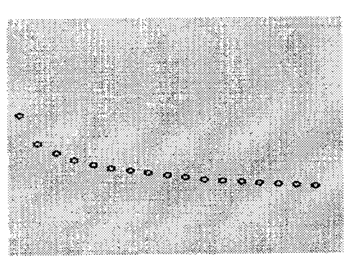
Asymptotic Increase
Variability = 0
Sample Size = 17



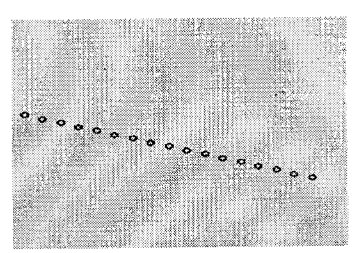
Linear Increase
Variability = 0
Sample Size = 17



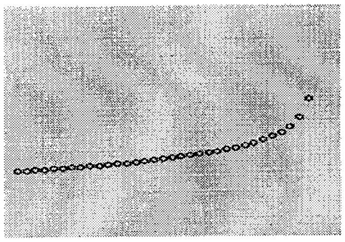
Exponential Decrease
Variability = 0
Sample Size = 17



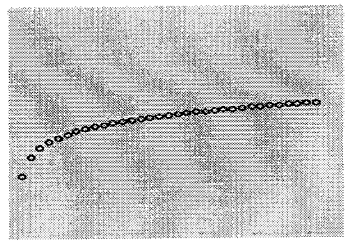
Asymptotic Decrease
Variability = 0
Sample Size = 17



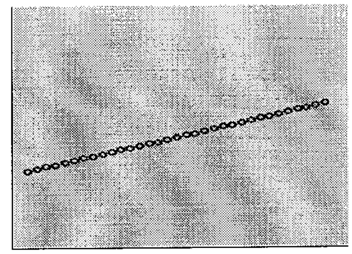
Linear Decrease
Variability = 0
Sample Size = 17



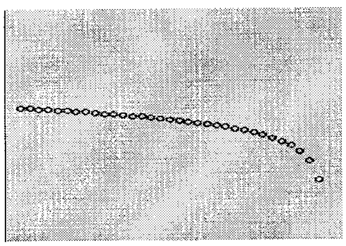
Exponential Increase
Variability = 0
Sample Size = 33



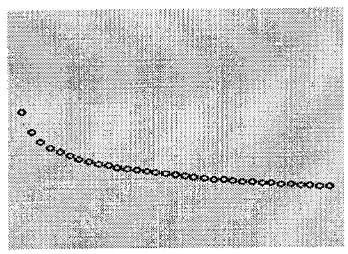
Asymptotic Increase
Variability = 0
Sample Size = 33



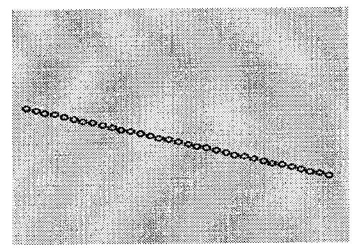
Linear Increase
Variability = 0
Sample Size = 33



Exponential Decrease
Variability = 0
Sample Size = 33



Asymptotic Decrease
Variability = 0
Sample Size = 33



Linear Decrease
Variability = 0
Sample Size = 33

The Generation of the Curve Points

Discrimination Experiments (Experiment 1 and Experiment 4a)

The curve points for the discrimination tasks are presented in Table 15. These points were generated using the following formulae:

Decreasing Asymptotic:	Y = 1.75 - 0.75 (LOG (X - .5) / LOG(10))
Decreasing Exponential:	Y = 0.804 + .75 (LOG (37.5 - (X+4)) /LOG(10))
Decreasing Linear:	Y = 2.02 - .041 (X)
Increasing Asymptotic:	Y = 0.804 + .75 (LOG (X - 0.5) / LOG(10))
Increasing Exponential:	Y = 1.75 - 0.75 (LOG (37.5 - (X+4)) / LOG(10))
Increasing Linear:	Y = .54 + .04 (X+4)

Table A1
Underlying Data Points (Y) for the Six Trend Types.

DA	DE	DL	IE	IA	IL
1.976	1.938	1.976	0.616	0.578	0.748
1.618	1.928	1.935	0.626	0.938	0.790
1.452	1.917	1.894	0.637	1.102	0.831
1.342	1.906	1.852	0.648	1.212	0.872
1.260	1.895	1.811	0.659	1.294	0.913
1.195	1.883	1.770	0.671	1.359	0.954
1.140	1.871	1.729	0.683	1.414	0.996
1.094	1.859	1.688	0.695	1.460	1.037
1.053	1.846	1.646	0.708	1.501	1.078
1.017	1.832	1.605	0.722	1.537	1.119
0.984	1.818	1.564	0.736	1.570	1.160
0.954	1.803	1.523	0.751	1.600	1.202
0.927	1.788	1.481	0.766	1.627	1.243
0.902	1.772	1.440	0.782	1.652	1.284
0.879	1.754	1.399	0.800	1.675	1.325
0.857	1.736	1.358	0.818	1.697	1.366
0.837	1.717	1.317	0.837	1.717	1.408
0.818	1.697	1.275	0.857	1.736	1.449
0.800	1.675	1.234	0.879	1.754	1.490
0.782	1.652	1.193	0.902	1.772	1.531
0.766	1.627	1.152	0.927	1.788	1.573
0.751	1.600	1.111	0.954	1.803	1.614
0.736	1.570	1.069	0.984	1.818	1.655
0.722	1.537	1.028	1.017	1.832	1.696
0.708	1.501	0.987	1.053	1.846	1.737
0.695	1.460	0.946	1.094	1.859	1.779
0.683	1.414	0.904	1.140	1.871	1.820
0.671	1.359	0.863	1.195	1.883	1.861
0.659	1.294	0.822	1.260	1.895	1.902
0.648	1.212	0.781	1.342	1.906	1.943
0.637	1.102	0.740	1.452	1.917	1.980
0.626	0.936	0.698	1.618	1.928	2.026
0.616	0.578	0.657	1.976	1.938	2.067

The Following curves were generated based on the previous formulae:

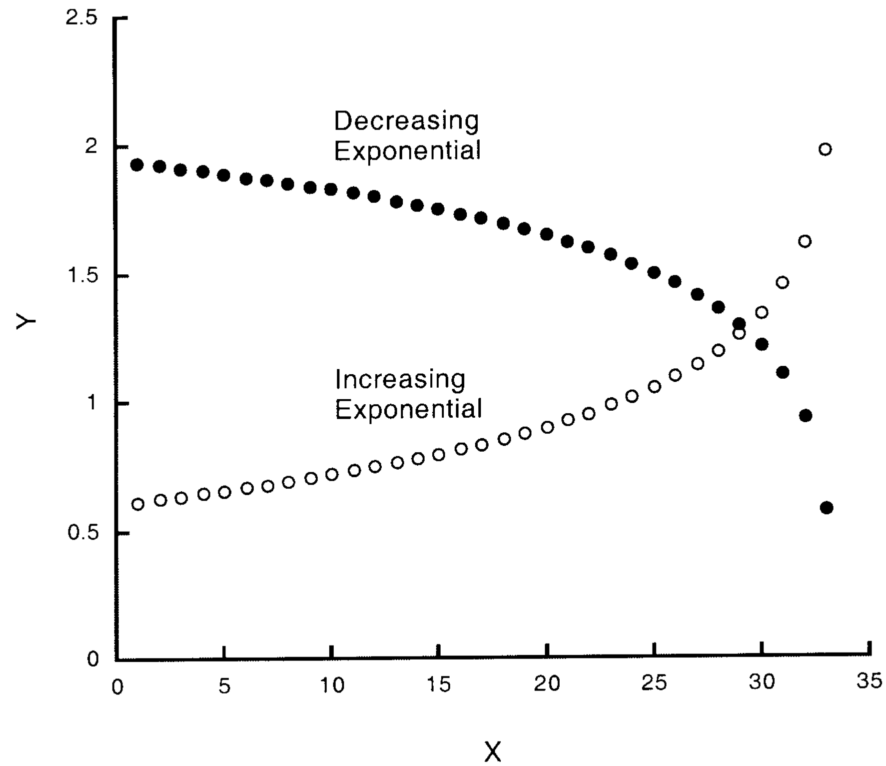


Figure A1. Exponential curves with 33 data points.

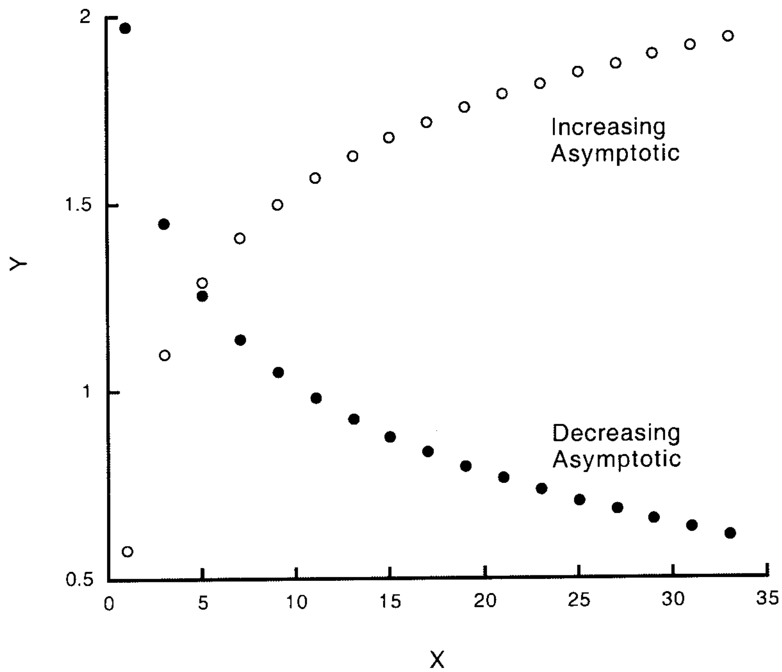


Figure A2. Asymptotic curves with 17 data points. The curves were generated by plotting every second curve point on Table 15.
Note: In Experiment 4a (Discrimination with Motion), all stimuli contained 17 points.

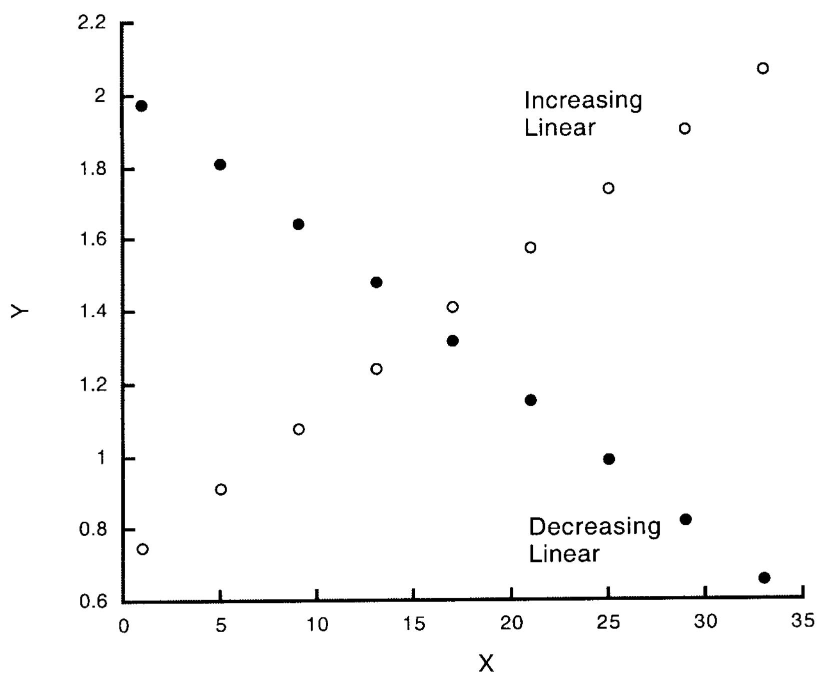


Figure A3. Linear curves with 9 data points.
The curves were generated by plotting every fourth curve point on Table 15.

Extrapolation Experiments (Experiment 2 and Experiment 4b)

The curve points were generated using the following formulae:

Decreasing Asymptotic:	Y = 1.75 - 0.75 (LOG (X - .5) / LOG(10))
Decreasing Exponential:	Y = 0.804 + .75 (LOG (37.5 - (X)) /LOG(10))
Decreasing Linear:	Y = 2.01 - .04 (X)
Increasing Asymptotic:	Y = 0.804 + .75 (LOG (X - 0.5) / LOG(10))
Increasing Exponential:	Y = 1.75 - 0.75 (LOG (37.5 - X) / LOG(10))
Increasing Linear:	Y = .54 + .04 (X)

In these experiments, 37 curve points were generated (X= 1 to 37).

The Following curves were generated based on the previous formulae:

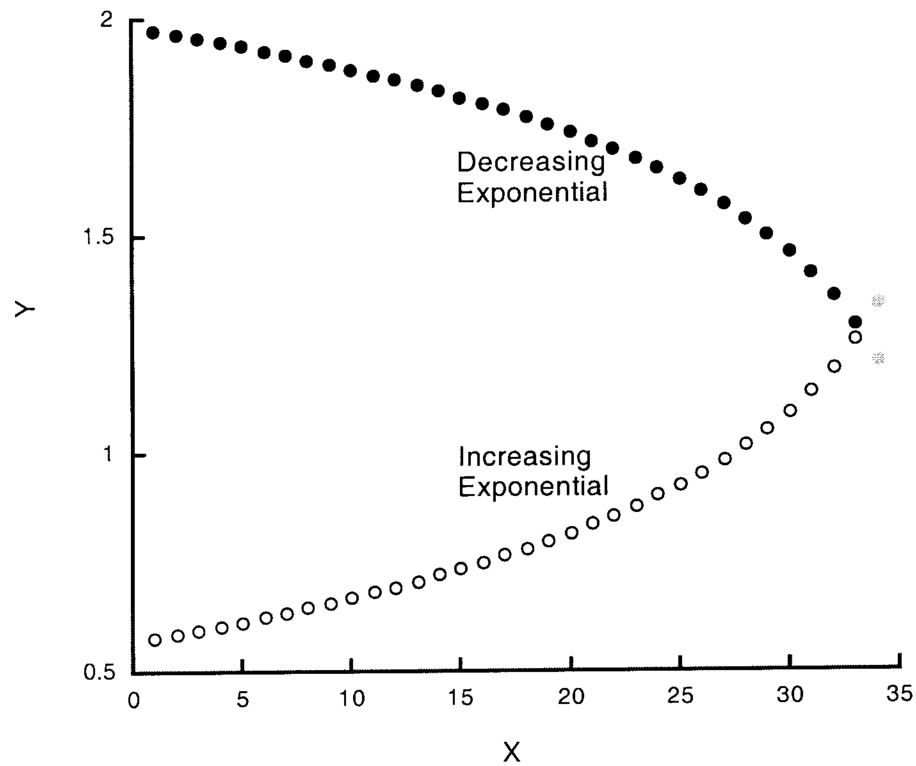


Figure A4. Exponential curves with 33 data points (used in Experiments 2 and 4b). Subjects extrapolated Underlying Curve Point 34 (represented by a gray circle).

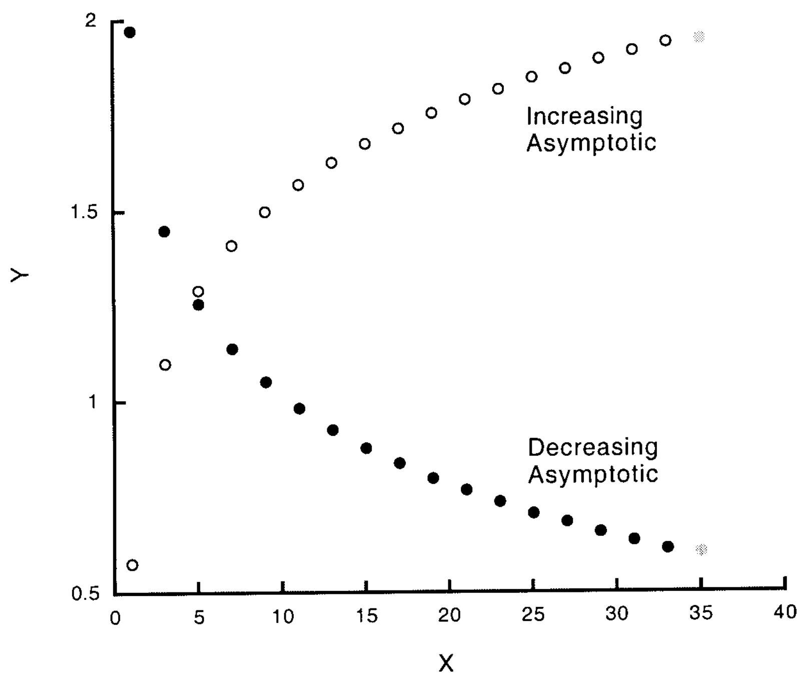


Figure A5. Asymptotic curves with 17 data points (used in Experiments 2 and 4b). Subjects extrapolated Underlying Curve Point 35 (represented by a gray circle).

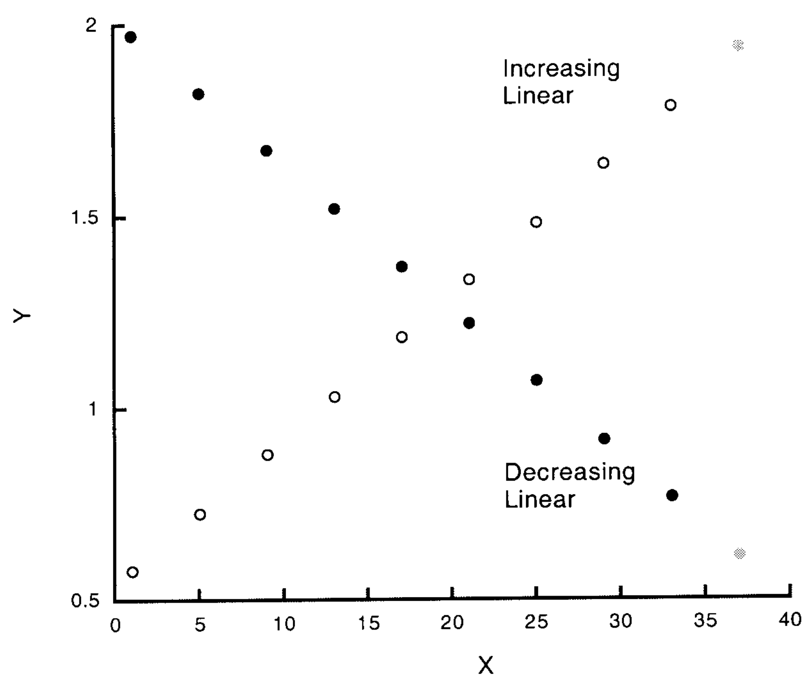


Figure A6. Linear curves with 9 data points (used in Experiments 2 and 4b). Subjects extrapolated Underlying Curve Point 37 (represented by a gray circle).

Feedback Experiment (Experiment 3)

The curve points were generated using the following formulae:

Decreasing Asymptotic:	Y = 1.75 - 0.75 (LOG (X - .5) / LOG(10))
Decreasing Exponential:	Y = 0.787+ .75 (LOG (39.5 - (X)) /LOG(10))
Decreasing Linear:	Y = 2.01 - .33 (X)
Increasing Asymptotic:	Y = 0.787+ .75 (LOG (X - 0.5) / LOG(10))
Increasing Exponential:	Y = 1.75 - 0.75 (LOG (39.5 - X) / LOG(10))
Increasing Linear:	Y = .52 + .04 (X)

In these experiments, 39 curve points were generated (X = 1 to 39). Sample size was not manipulated in this experiment. In all conditions, the sample size was 17.

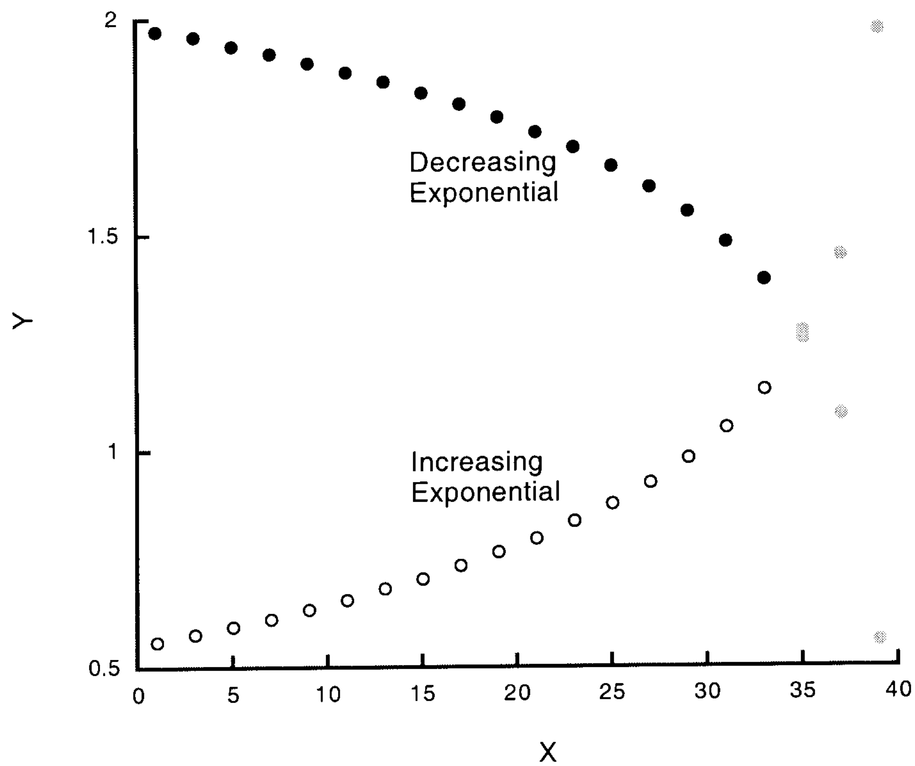


Figure A7. Exponential curves with 17 data points (used in Experiment 3). Subjects extrapolated Underlying Curve Point 35, 37, and 39 (represented by a gray circles).

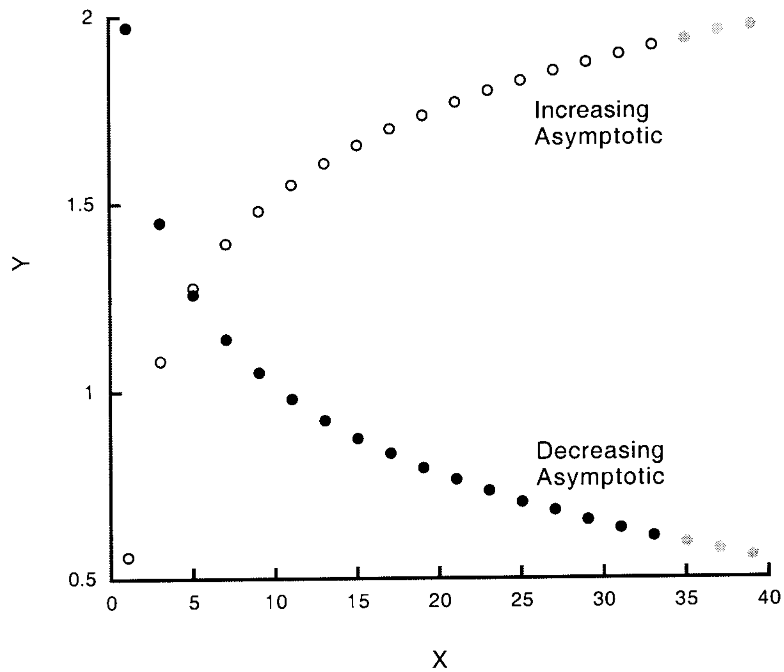


Figure A8. Asymptotic curves with 17 data points (used in Experiment 3). Subjects extrapolated Underlying Curve Point 35, 37, and 39 (represented by a gray circles).

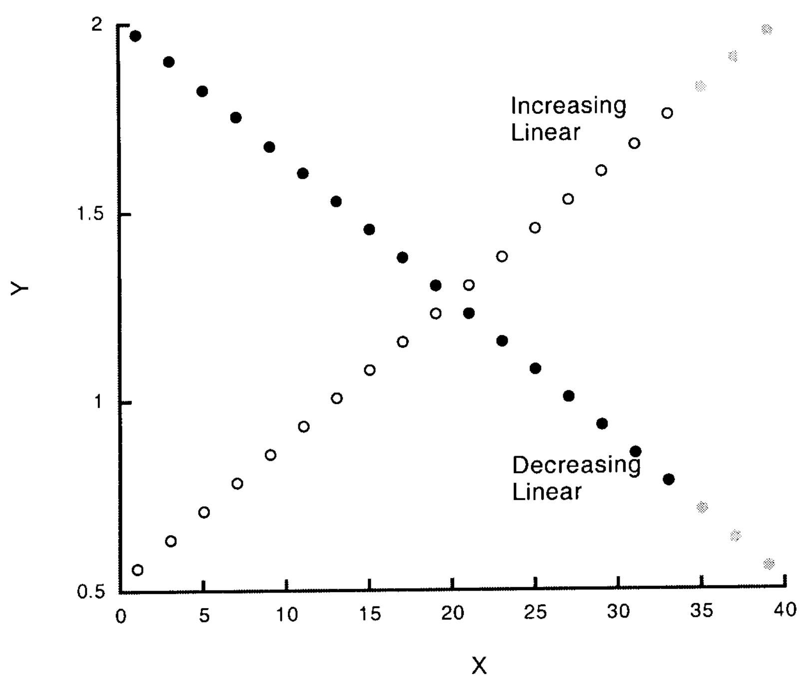
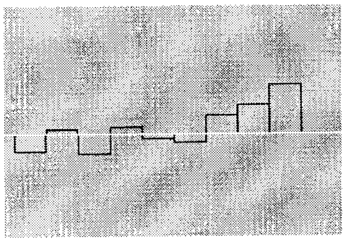


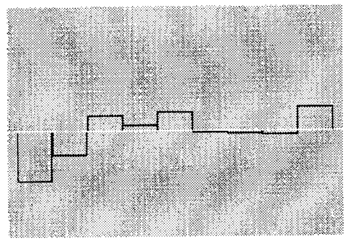
Figure A9. Linear curves with 17 data points (used in Experiment 3). Subjects extrapolated Underlying Curve Point 35, 37, and 39 (represented by a gray circles).

Appendix B

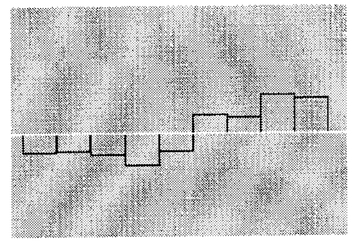
Samples of the Stimuli that were used in the Experiments



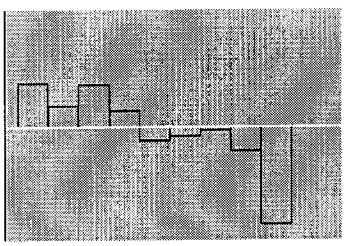
Exponential Increase
Variability = 1
Sample Size = 9



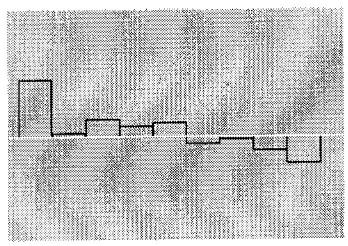
Asymptotic Increase
Variability = 1
Sample Size = 9



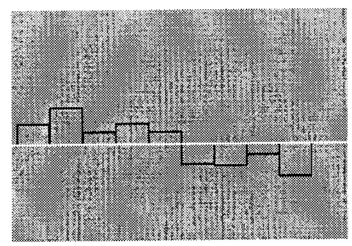
Linear Increase
Variability = 1
Sample Size = 9



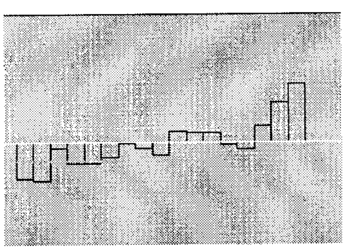
Exponential Decrease
Variability = 1
Sample Size = 9



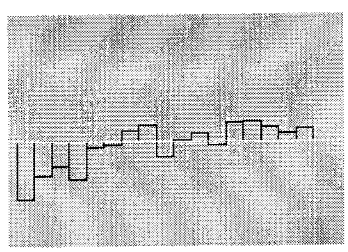
Asymptotic Decrease
Variability = 1
Sample Size = 9



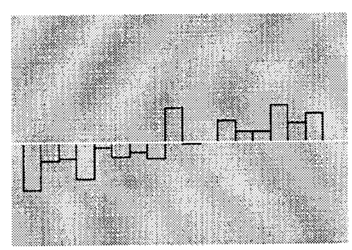
Linear Decrease
Variability = 1
Sample Size = 9



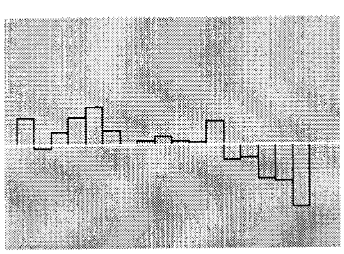
Exponential Increase
Variability = 1
Sample Size = 17



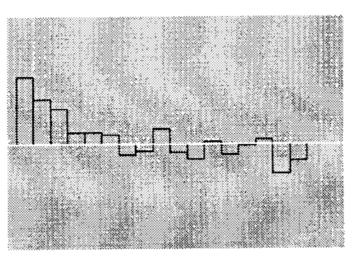
Asymptotic Increase
Variability = 1
Sample Size = 17



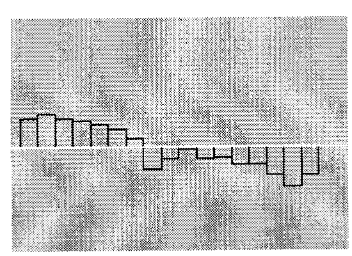
Linear Increase
Variability = 1
Sample Size = 17



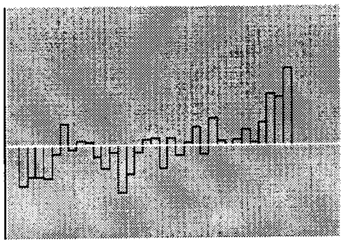
Exponential Decrease
Variability = 1
Sample Size = 17



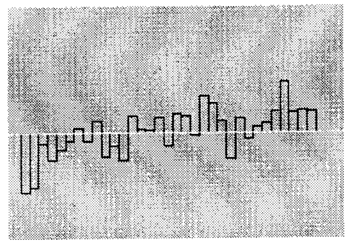
Asymptotic Decrease
Variability = 1
Sample Size = 17



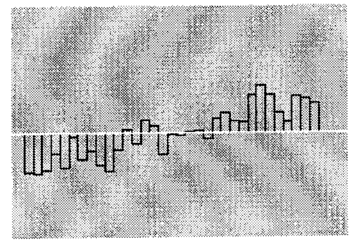
Linear Decrease
Variability = 1
Sample Size = 17



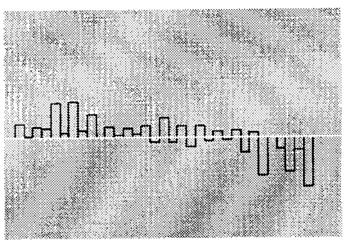
Exponential Increase
Variability = 1
Sample Size = 33



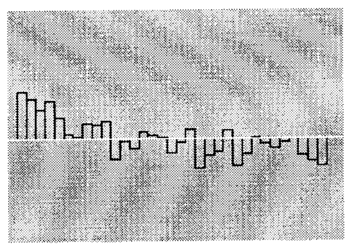
Asymptotic Increase
Variability = 1
Sample Size = 33



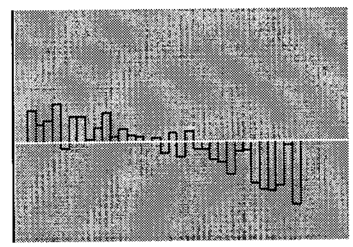
Linear Increase
Variability = 1
Sample Size = 33



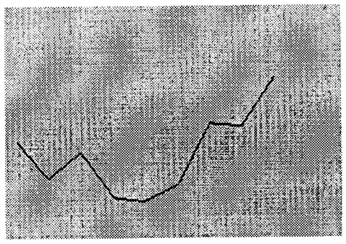
Exponential Decrease
Variability = 1
Sample Size = 33



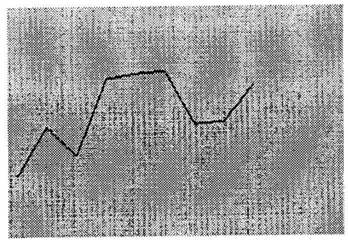
Asymptotic Decrease
Variability = 1
Sample Size = 33



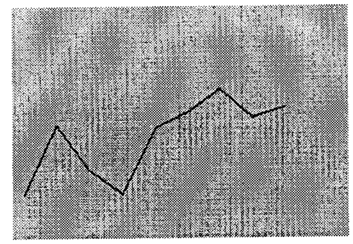
Linear Decrease
Variability = 1
Sample Size = 33



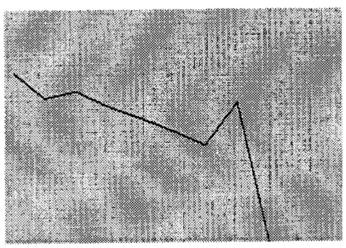
Exponential Increase
Variability = 2
Sample Size = 9



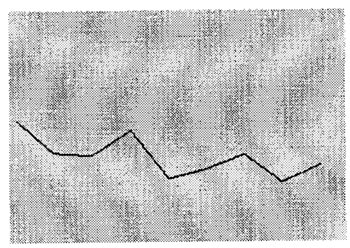
Asymptotic Increase
Variability = 2
Sample Size = 9



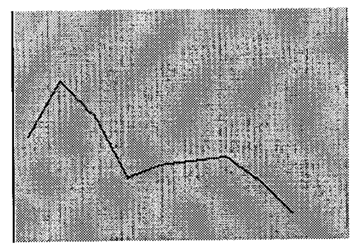
Linear Increase
Variability = 2
Sample Size = 9



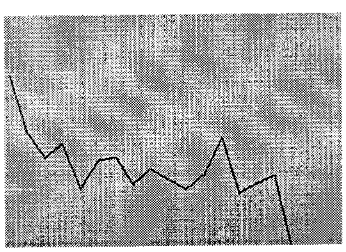
Exponential Decrease
Variability = 2
Sample Size = 9



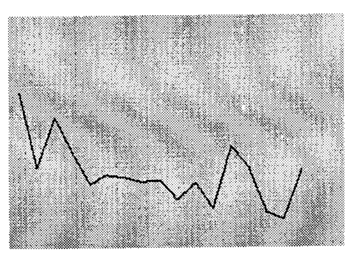
Asymptotic Decrease
Variability = 2
Sample Size = 9



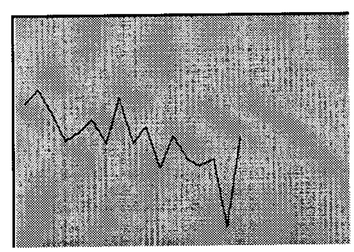
Linear Decrease
Variability = 2
Sample Size = 9



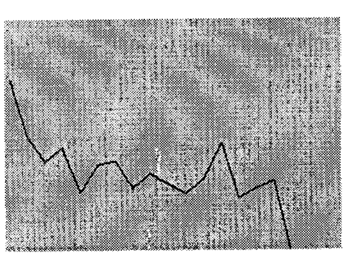
Exponential Increase
Variability = 2
Sample Size = 17



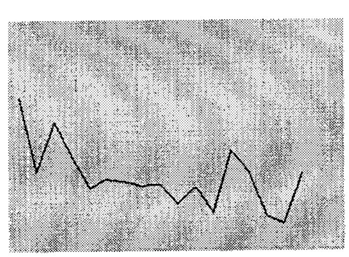
Asymptotic Increase
Variability = 2
Sample Size = 17



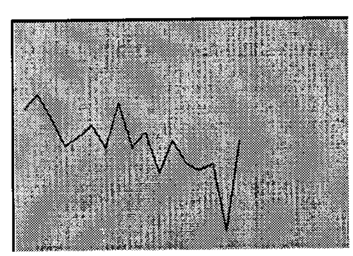
Linear Increase
Variability = 2
Sample Size = 17



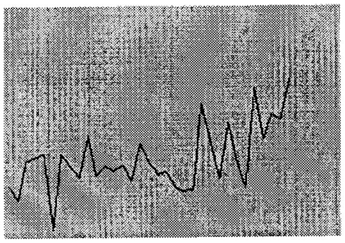
Exponential Decrease
Variability = 2
Sample Size = 17



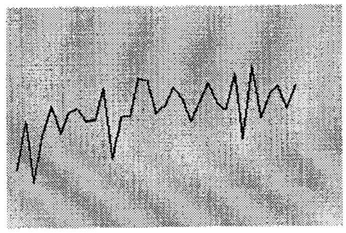
Asymptotic Decrease
Variability = 2
Sample Size = 17



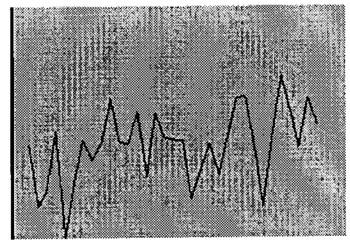
Linear Decrease
Variability = 2
Sample Size = 17



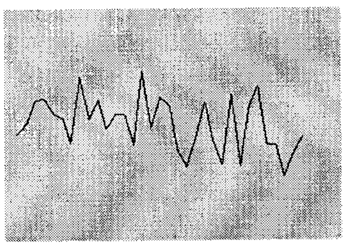
Exponential Increase
Variability = 2
Sample Size = 33



Asymptotic Increase
Variability = 2
Sample Size = 33

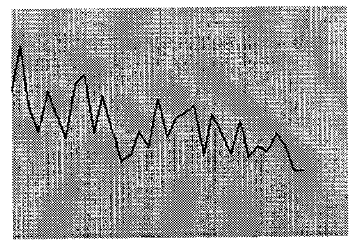


Linear Increase
Variability = 2
Sample Size = 33

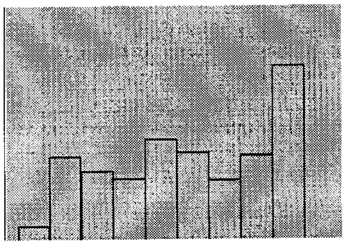


Exponential Decrease
Variability = 2
Sample Size = 33

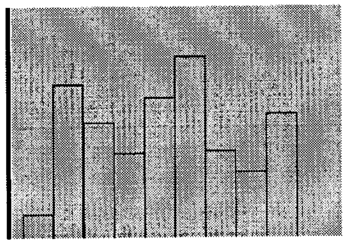
Asymptotic Decrease
Variability = 2
Sample Size = 33



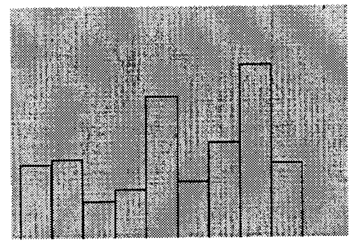
Linear Decrease
Variability = 2
Sample Size = 33



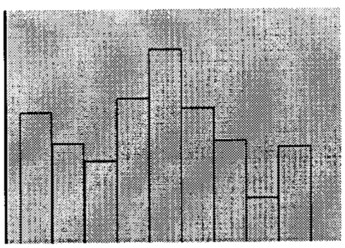
Exponential Increase
Variability = 3
Sample Size = 9



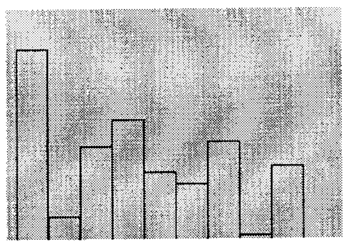
Asymptotic Increase
Variability = 3
Sample Size = 9



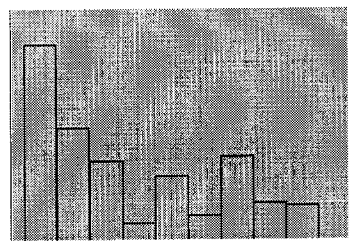
Linear Increase
Variability = 3
Sample Size = 9



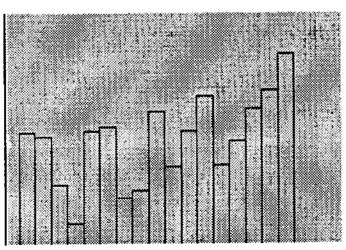
Exponential Decrease
Variability = 3
Sample Size = 9



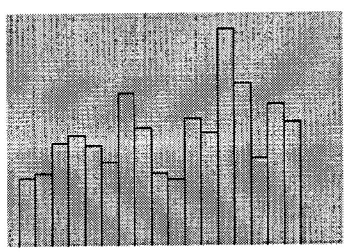
Asymptotic Decrease
Variability = 3
Sample Size = 9



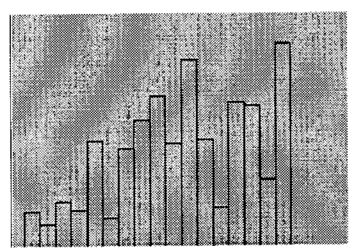
Linear Decrease
Variability = 3
Sample Size = 9



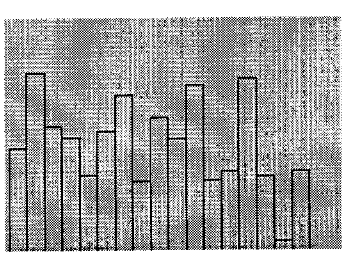
Exponential Increase
Variability = 3
Sample Size = 17



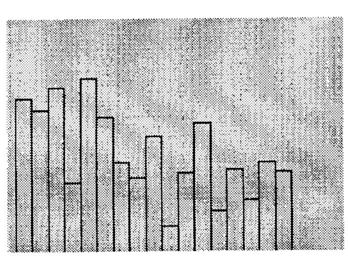
Asymptotic Increase
Variability = 3
Sample Size = 17



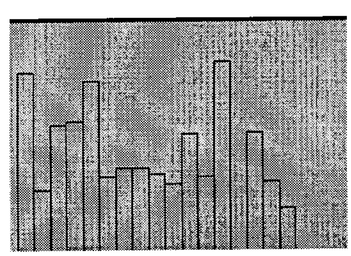
Linear Increase
Variability = 3
Sample Size = 17



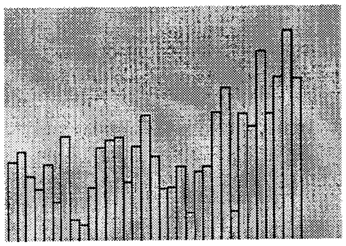
Exponential Decrease
Variability = 3188188
Sample Size = 17



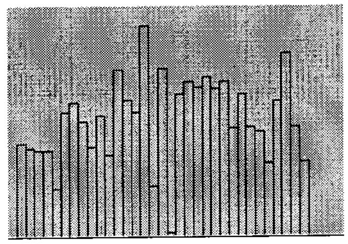
Asymptotic Decrease
Variability = 3
Sample Size = 17



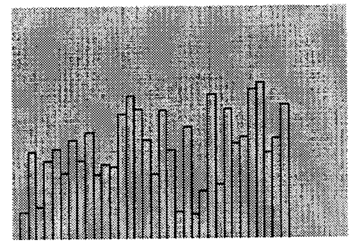
Linear Decrease
Variability = 3
Sample Size = 17



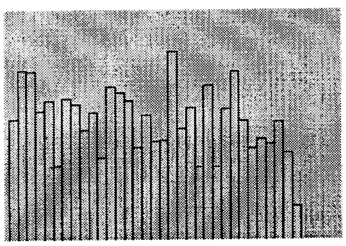
Exponential Increase
Variability = 3
Sample Size = 33



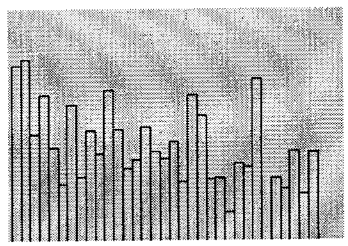
Asymptotic Increase
Variability = 3
Sample Size = 33



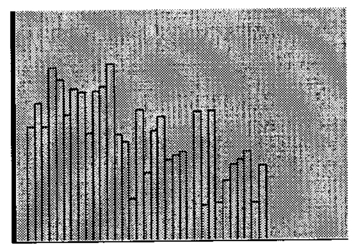
Linear Increase
Variability = 3
Sample Size = 33



Exponential Decrease
Variability = 3
Sample Size = 33



Asymptotic Decrease
Variability = 3
Sample Size = 33



Linear Decrease
Variability = 3
Sample Size = 33

Appendix C

The Results of the Post Hoc Comparisons

Table C1

Experiment 1 Significant Main Effects

Variable	Simple Comparison	Mean Difference
Variability		
	Low-High	.28****
	Low-Intermediate	.15***
	Intermediate-High	.14****
Sample Size		
	Low-High	-.05*
	Low-Intermediate	-.02*
	Intermediate-High	-.02*
Graph Type		
	Histogram-Line Graph	-.07*
	Histogram-Scatterplot	-.07*
	Histogram-Suspended Bar	.01
	Line Graph-Scatterplot	.0006
	Line Graph-Suspended Bar	.08*
	Scatterplot-Suspended Bar	.08*
Trend Type		
	Decreasing Asymptotic-Decreasing Exponential	-.04
	Decreasing Asymptotic-Decreasing Linear	.08*
	Decreasing Asymptotic-Increasing Exponential	-.04
	Decreasing Asymptotic-Increasing Asymptotic	-.04
	Decreasing Asymptotic-Increasing Linear	.11*
	Decreasing Exponential-Decreasing Linear	.12*
	Decreasing Exponential-Increasing Exponential	-.0013
	Decreasing Exponential-Increasing Asymptotic	-.002
	Decreasing Exponential-Increasing Linear	.15**
	Decreasing Linear-Increasing Exponential	-.11***
	Decreasing Linear-Increasing Asymptotic	-.11**
	Decreasing Linear-Increasing Linear	.04
	Increasing Exponential-Increasing Asymptotic	.0007
	Increasing Exponential-Increasing Linear	.15**
	Increasing Asymptotic-Increasing Linear	.15**

+ p<.06 * P<.05 ** P<.01 *** P<.001 **** p<.0001

Experiment 1 Significant Interactions

Table C2

Experiment 1: Variability by Graph Type Interaction

Level of Variability	Comparison	Mean Difference
Low	Histogram-Line Graph	-.09*
	Histogram-Scatterplot	-.13**
	Histogram-Suspended Bar	-.04
	Line Graph-Scatterplot	-.04+
	Line Graph-Suspended Bar	.05
	Scatterplot-Suspended Bar	.09+
Moderate	Histogram-Line Graph	-.06*
	Histogram-Scatterplot	-.01
	Histogram-Suspended Bar	.07*
	Line Graph-Scatterplot	.05+
	Line Graph-Suspended Bar	.13**
	Scatterplot-Suspended Bar	.09*
High	Histogram-Line Graph	-.05
	Histogram-Scatterplot	-.06
	Histogram-Suspended Bar	.02
	Line Graph-Scatterplot	.01
	Line Graph-Suspended Bar	.07**
	Scatterplot-Suspended Bar	.08*

+ p<.1 * P<.05 ** P<.01 *** P<.001 **** p<.0001

Table C3

Experiment 1: Sample Size by Graph Type Interaction

Sample Size	Comparison	Mean Difference
Low Sample Size	Histogram-Line Graph	-.07*
	Histogram-Scatterplot	-.02
	Histogram-Suspended Bar	-.04+
	Line Graph-Scatterplot	.06*
	Line Graph-Suspended Bar	.11*
	Scatterplot-Suspended Bar	.05
Moderate Sample Size	Histogram-Line Graph	-.06*
	Histogram-Scatterplot	-.11*
	Histogram-Suspended Bar	.001
	Line Graph-Scatterplot	-.04
	Line Graph-Suspended Bar	.06*
	Scatterplot-Suspended Bar	.11*
High Sample Size	Histogram-Line Graph	-.07
	Histogram-Scatterplot	-.08+
	Histogram-Suspended Bar	-.02
	Line Graph-Scatterplot	-.02
	Line Graph-Suspended Bar	.08*
	Scatterplot-Suspended Bar	.10*

+ p<.1 * P<.05 ** P<.01 *** P<.001 **** p<.0001

Table C4

Experiment 1: Sample Size by Trend Type Interaction

Sample Size	Trend Comparison	Mean Difference
Low	Decreasing Asymptotic-Decreasing Exponential	-.02
	Decreasing Asymptotic-Decreasing Linear	.14*
	Decreasing Asymptotic-Increasing Exponential	-.05
	Decreasing Asymptotic-Increasing Asymptotic	.05
	Decreasing Asymptotic-Increasing Linear	.16**
	Decreasing Exponential-Decreasing Linear	.16**
	Decreasing Exponential-Increasing Exponential	-.03
	Decreasing Exponential-Increasing Asymptotic	.03
	Decreasing Exponential-Increasing Linear	.19**
	Decreasing Linear-Increasing Exponential	- .19****
	Decreasing Linear-Increasing Asymptotic	-.19**
	Decreasing Linear-Increasing Linear	-.02
	Increasing Exponential-Increasing Asymptotic	.0004
	Increasing Exponential-Increasing Linear	.21****
	Increasing Asymptotic-Increasing Linear	-.21**
Moderate	Decreasing Asymptotic-Decreasing Exponential	-.05
	Decreasing Asymptotic-Decreasing Linear	.05+
	Decreasing Asymptotic-Increasing Exponential	-.04
	Decreasing Asymptotic-Increasing Asymptotic	-.05+
	Decreasing Asymptotic-Increasing Linear	.11*
	Decreasing Exponential-Decreasing Linear	.10+
	Decreasing Exponential-Increasing Exponential	-.10
	Decreasing Exponential-Increasing Asymptotic	.002
	Decreasing Exponential-Increasing Linear	.16**
	Decreasing Linear-Increasing Exponential	-.09+
	Decreasing Linear-Increasing Asymptotic	-.10**
	Decreasing Linear-Increasing Linear	-.06+
	Increasing Exponential-Increasing Asymptotic	.01
	Increasing Exponential-Increasing Linear	.15**
	Increasing Asymptotic-Increasing Linear	.16*
High	Decreasing Asymptotic-Decreasing Exponential	-.04
	Decreasing Asymptotic-Decreasing Linear	.04
	Decreasing Asymptotic-Increasing Exponential	-.02
	Decreasing Asymptotic-Increasing Asymptotic	-.07
	Decreasing Asymptotic-Increasing Linear	.07
	Decreasing Exponential-Decreasing Linear	.08+
	Decreasing Exponential-Increasing Exponential	-.02
	Decreasing Exponential-Increasing Asymptotic	.03
	Decreasing Exponential-Increasing Linear	.11**
	Decreasing Linear-Increasing Exponential	-.07+
	Decreasing Linear-Increasing Asymptotic	-.05
	Decreasing Linear-Increasing Linear	-.03
	Increasing Exponential-Increasing Asymptotic	.02
	Increasing Exponential-Increasing Linear	.09**
	Increasing Asymptotic-Increasing Linear	.07+

+ p<.1 * P<.05 ** P<.01 *** P<.001 **** p<.0001

Table 5

Experiment 1: Trend Type by Graph Type Interaction

Trend Type	Graph Type Comparison	Mean Difference
Decreasing Asymptotic	Histogram - Line Graph	-.13*
	Histogram - Scatterplot	-.13*
	Histogram - Suspended Bar	.01
	Line Graph - Scatterplot	.09
	Line Graph - Suspended Bar	.15+
	Scatterplot - Suspended Bar	.14+
Decreasing Exponential	Histogram - Line Graph	.04
	Histogram - Scatterplot	-.01
	Histogram - Suspended Bar	.03
	Line Graph - Scatterplot	-.05+
	Line Graph - Suspended Bar	.12
	Scatterplot - Suspended Bar	.03
Decreasing Linear	Histogram - Line Graph	-.13+
	Histogram - Scatterplot	-.08
	Histogram - Suspended Bar	-.02
	Line Graph - Scatterplot	.05
	Line Graph - Suspended Bar	.11
	Scatterplot - Suspended Bar	-.05
Increasing Asymptotic	Histogram - Line Graph	.02
	Histogram - Scatterplot	-.02
	Histogram - Suspended Bar	.06
	Line Graph - Scatterplot	-.04
	Line Graph - Suspended Bar	.03
	Scatterplot - Suspended Bar	.08
Increasing Exponential	Histogram - Line Graph	-.10*
	Histogram - Scatterplot	-.04
	Histogram - Suspended Bar	.03
	Line Graph - Scatterplot	.06*
	Line Graph - Suspended Bar	.13***
	Scatterplot - Suspended Bar	.07*
Increasing Linear	Histogram - Line Graph	-.10*
	Histogram - Scatterplot	-.13*
	Histogram - Suspended Bar	.003
	Line Graph - Scatterplot	.03
	Line Graph - Suspended Bar	.11*
	Scatterplot - Suspended Bar	.14**

+ $p<.1$ * $P<.05$ ** $P<.01$ *** $P<.001$ **** $p<.0001$

Experiment 2 Significant Main Effects

Table C6

Experiment 2 Significant Main Effects

Variable	Simple Comparison	Mean Difference
Variability		
	Low-High	-.29**
	Low-Intermediate	-.18***
	Intermediate-High	-.11*
Sample Size		
	Low-High	.07*
	Low-Intermediate	.03
	Intermediate-High	.04**
Graph Type		
	Histogram-Line Graph	.235+
	Histogram-Scatterplot	.07
	Histogram-Suspended Bar	.02
	Line Graph-Scatterplot	-.17*
	Line Graph-Suspended Bar	-.23*
	Scatterplot-Suspended Bar	.06
Trend Type		
	Decreasing Exponential-Decreasing Asymptotic	-.1
	Decreasing Exponential-Decreasing Linear	-.12*
	Decreasing Exponential-Increasing Asymptotic	-.10
	Decreasing Exponential-Increasing Exponential	.06+
	Decreasing Exponential-Increasing Linear	-.04
	Decreasing Asymptotic-Decreasing Linear	.02
	Decreasing Asymptotic-Increasing Asymptotic	.004
	Decreasing Asymptotic-Increasing Exponential	-.15*
	Decreasing Asymptotic-Increasing Linear	.06+
	Decreasing Linear-Increasing Asymptotic	.01
	Decreasing Linear-Increasing Exponential	.12**
	Decreasing Linear-Increasing Linear	.07*
	Increasing Asymptotic-Increasing Exponential	.16*
	Increasing Asymptotic-Increasing Linear	.06
	Increasing Exponential-Increasing Linear	-.17

+ p<.1 * P<.05 ** P<.01 *** P<.001 **** p<.0001

Experiment 2 Significant Interactions

Table C7
Experiment 2: Sample Size by Trend Type Interaction

Sample Size	Trend Comparison	Mean Difference
Low		
	Decreasing Exponential-Decreasing Asymptotic	.04
	Decreasing Exponential-Decreasing Linear	.05
	Decreasing Exponential-Increasing Asymptotic	.05
	Decreasing Exponential-Increasing Exponential	.04
	Decreasing Exponential-Increasing Linear	-.01
	Decreasing Asymptotic-Decreasing Linear	.01
	Decreasing Asymptotic-Increasing Asymptotic	-.05
	Decreasing Asymptotic-Increasing Exponential	-.08
	Decreasing Asymptotic-Increasing Linear	-.06
	Decreasing Linear-Increasing Asymptotic	.005
	Decreasing Linear-Increasing Exponential	-.08+
	Decreasing Linear-Increasing Linear	-.65+
	Increasing Asymptotic-Increasing Exponential	-.08
	Increasing Asymptotic-Increasing Linear	-.06
	Increasing Exponential-Increasing Linear	.03
Moderate	Decreasing Exponential-Decreasing Asymptotic	.12
	Decreasing Exponential-Decreasing Linear	.14*
	Decreasing Exponential-Increasing Asymptotic	.14
	Decreasing Exponential-Increasing Exponential	.05
	Decreasing Exponential-Increasing Linear	.07
	Decreasing Asymptotic-Decreasing Linear	-.02
	Decreasing Asymptotic-Increasing Asymptotic	-.02
	Decreasing Asymptotic-Increasing Exponential	-.17*
	Decreasing Asymptotic-Increasing Linear	-.05
	Decreasing Linear-Increasing Asymptotic	-.002
	Decreasing Linear-Increasing Exponential	-.19**
	Decreasing Linear-Increasing Linear	-.07+
	Increasing Asymptotic-Increasing Exponential	-.19*
	Increasing Asymptotic-Increasing Linear	-.07
	Increasing Exponential-Increasing Linear	-.12**
High	Decreasing Exponential-Decreasing Asymptotic	.13+
	Decreasing Exponential-Decreasing Linear	.15**
	Decreasing Exponential-Increasing Asymptotic	.12
	Decreasing Exponential-Increasing Exponential	.08*
	Decreasing Exponential-Increasing Linear	.07
	Decreasing Asymptotic-Decreasing Linear	.01
	Decreasing Asymptotic-Increasing Asymptotic	-.01
	Decreasing Asymptotic-Increasing Exponential	.21*
	Decreasing Asymptotic-Increasing Linear	-.07*
	Decreasing Linear-Increasing Asymptotic	-.03
	Decreasing Linear-Increasing Exponential	-.23**
	Decreasing Linear-Increasing Linear	-.08
	Increasing Asymptotic-Increasing Exponential	-.20*
	Increasing Asymptotic-Increasing Linear	-.05
	Increasing Exponential-Increasing Linear	.21*

+ p<.1 * P<.05 ** P<.01 *** P<.001 **** p<.0001

Table C8

Experiment 2: Variability by Trend Type Interaction

Variability	Trend Comparison	Mean Difference
Low	Decreasing Exponential-Decreasing Asymptotic	.07
	Decreasing Exponential-Decreasing Linear	.15**
	Decreasing Exponential-Increasing Asymptotic	.08
	Decreasing Exponential-Increasing Exponential	.03
	Decreasing Exponential-Increasing Linear	.06
	Decreasing Asymptotic-Decreasing Linear	-.08+
	Decreasing Asymptotic-Increasing Asymptotic	-.002
	Decreasing Asymptotic-Increasing Exponential	-.104+
	Decreasing Asymptotic-Increasing Linear	-.01
	Decreasing Linear-Increasing Asymptotic	-.08
	Decreasing Linear-Increasing Exponential	.183**
	Decreasing Linear-Increasing Linear	-.10*
	Increasing Asymptotic-Increasing Exponential	-.10
	Increasing Asymptotic-Increasing Linear	-.02
	Increasing Exponential-Increasing Linear	.09
Moderate	Decreasing Exponential-Decreasing Asymptotic	.06
	Decreasing Exponential-Decreasing Linear	.09*
	Decreasing Exponential-Increasing Asymptotic	.06
	Decreasing Exponential-Increasing Exponential	.09+
	Decreasing Exponential-Increasing Linear	.02
	Decreasing Asymptotic-Decreasing Linear	.03
	Decreasing Asymptotic-Increasing Asymptotic	-.002
	Decreasing Asymptotic-Increasing Exponential	.15*
	Decreasing Asymptotic-Increasing Linear	.04
	Decreasing Linear-Increasing Asymptotic	-.03
	Decreasing Linear-Increasing Exponential	-.18*
	Decreasing Linear-Increasing Linear	-.07
	Increasing Asymptotic-Increasing Exponential	-.15*
	Increasing Asymptotic-Increasing Linear	-.04
	Increasing Exponential-Increasing Linear	.11*
High	Decreasing Exponential-Decreasing Asymptotic	.16+
	Decreasing Exponential-Decreasing Linear	.09+
	Decreasing Exponential-Increasing Asymptotic	.17+
	Decreasing Exponential-Increasing Exponential	.04
	Decreasing Exponential-Increasing Linear	.04
	Decreasing Asymptotic-Decreasing Linear	.06
	Decreasing Asymptotic-Increasing Asymptotic	-.01
	Decreasing Asymptotic-Increasing Exponential	-.2*
	Decreasing Asymptotic-Increasing Linear	-.12+
	Decreasing Linear-Increasing Asymptotic	-.07
	Decreasing Linear-Increasing Exponential	-.135**
	Decreasing Linear-Increasing Linear	-.05
	Increasing Asymptotic-Increasing Exponential	-.21*
	Increasing Asymptotic-Increasing Linear	-.13*
	Increasing Exponential-Increasing Linear	.09*

+ p<.1 * P<.05 ** P<.01 *** P<.001 *** p<.0001

Post Hoc Tests for Signal Detection Analyses

Table C9

Average Sensitivity Main Effects

Variable	Simple Comparison	Mean Difference
Variability		
	Low-High	1.55***
	Low-Intermediate	0.93**
	Intermediate-High	0.62**
Sample Size		
	Low-High	-0.42**
	Low-Intermediate	-0.21
	Intermediate-High	-0.20
Graph Type		
	Histogram-Line Graph	-0.48*
	Histogram-Scatterplot	-0.33+
	Histogram-Suspended Bar	0.13
	Line Graph-Scatterplot	0.14
	Line Graph-Suspended Bar	0.60**
	Scatterplot-Suspended Bar	0.46*

+ p<.1 * P<.05 ** P<.01 *** P<.001 **** p<.0001

Significant Interactions

Table C10

Average Sensitivity: Variability by Sample Size Interaction

Level of Variability	Comparison	Mean Difference
Low	Low-High	-1 .2***
	Low-Intermediate	-0.56**
	Intermediate-High	-0.57**
Moderate	Low-High	-1 .66**
	Low-Intermediate	-1.04*
	Intermediate-High	-0.62**
High	Low-High	-1.75**
	Low-Intermediate	-1.20***
	Intermediate-High	-0.56+

+ p<.1 * P<.05 ** P<.01 *** P<.001 **** p<.0001

Table C11

Average Sensitivity: Variability x Graph Type

Level of Variability	Comparison	Mean Difference
Low	Histogram-Line Graph	-.62
	Histogram-Scatterplot	-.91**
	Histogram-Suspended Bar	-.32
	Line Graph-Scatterplot	-.29
	Line Graph-Suspended Bar	.31
	Scatterplot-Suspended Bar	.56+
Moderate	Histogram-Line Graph	-.45*
	Histogram-Scatterplot	.15
	Histogram-Suspended Bar	.59*
	Line Graph-Scatterplot	.60*
	Line Graph-Suspended Bar	1.03**
	Scatterplot-Suspended Bar	.43+
High	Histogram-Line Graph	-.35
	Histogram-Scatterplot	-.24
	Histogram-Suspended Bar	.12
	Line Graph-Scatterplot	.11
	Line Graph-Suspended Bar	.47**
	Scatterplot-Suspended Bar	.36*

+ p<.1 * P<.05 ** P<.01 *** P<.001 **** p<.0001

Table C12

Average Sensitivity: Sample Size by Graph Type Interaction

Sample Size	Comparison	Mean Difference
Low Sample Size	Histogram-Line Graph	-.53*
	Histogram-Scatterplot	.14
	Histogram-Suspended Bar	.38+
	Line Graph-Scatterplot	.67**
	Line Graph-Suspended Bar	.91**
	Scatterplot-Suspended Bar	.24
Moderate Sample Size	Histogram-Line Graph	-.50*
	Histogram-Scatterplot	-.72+
	Histogram-Suspended Bar	.27
	Line Graph-Scatterplot	-.22
	Line Graph-Suspended Bar	.23
	Scatterplot-Suspended Bar	.45
High Sample Size	Histogram-Line Graph	-.39
	Histogram-Scatterplot	-.41
	Histogram-Suspended Bar	.28
	Line Graph-Scatterplot	-.02
	Line Graph-Suspended Bar	.67**
	Scatterplot-Suspended Bar	.69**

+ p<.1 * P<.05 ** P<.01 *** P<.001 **** p<.0001

Average Relative Efficiency Main Effects

Table C13

Average Relative Efficiency Main Effects

Variable	Simple Comparison	Mean Difference
Variability		
	Low-High	0.13+
	Low-Intermediate	0.08+
	Intermediate-High	0.04
Graph Type		
	Histogram-Line Graph	-0.15**
	Histogram-Scatterplot	-0.08+
	Histogram-Suspended Bar	0.05
	Line Graph-Scatterplot	0.07
	Line Graph-Suspended Bar	0.21**
	Scatterplot-Suspended Bar	0.14*

+ p<.1 * P<.05 ** P<.01 *** P<.001 **** p<.0001

Significant Interactions

Table C14

Relative Efficiency: Variability by Sample Size

Level of Sample Size	Comparison	Mean Difference
Low	Low-High	-0.07
	Low-Intermediate	-0.11*
	Intermediate-High	0.04
Moderate	Low-High	0.18*
	Low-Intermediate	0.13+
	Intermediate-High	0.04
High	Low-High	0.29*
	Low-Intermediate	0.23**
	Intermediate-High	0.06

+ p<.1 * P<.05 ** P<.01 *** P<.001 **** p<.0001

Table C15

Average Relative Efficiency: Sample Size by Graph Type

Sample Size	Comparison	Mean Difference
Low Sample Size	Histogram-Line Graph	-.24
	Histogram-Scatterplot	.08
	Histogram-Suspended Bar	.18
	Line Graph-Scatterplot	.32*
	Line Graph-Suspended Bar	.42*
	Scatterplot-Suspended Bar	.10
Moderate Sample Size	Histogram-Line Graph	-.15+
	Histogram-Scatterplot	-.24+
	Histogram-Suspended Bar	.10
	Line Graph-Scatterplot	-.09
	Line Graph-Suspended Bar	.05
	Scatterplot-Suspended Bar	.14
High Sample Size	Histogram-Line Graph	-.07
	Histogram-Scatterplot	-.09
	Histogram-Suspended Bar	.07
	Line Graph-Scatterplot	-.03
	Line Graph-Suspended Bar	.14**
	Scatterplot-Suspended Bar	.17****

+ p<.1 * P<.05 ** P<.01 *** P<.001 **** p<.0001

BIOGRAPHY OF THE AUTHOR

Lisa A. Best was born and raised in Minto, New Brunswick, Canada on July 28, 1968. She graduated from Minto Memorial High School and attended Glendon College, York University. She received a Bachelor of Arts Degree with Specialized Honours in Psychology in 1993. In 1994 she received her Master's of Applied Psychology from the University of Arkansas at Little Rock. During 1994-95, she was employed as a Lecturer in the Psychology Department at the University of Arkansas in Little Rock. In the fall of 1995, Lisa entered the graduate program at the University of Maine in Experimental Psychology, where she worked closely with Dr. D. Alan Stubbs and Dr. Laurence D. Smith. While at the University of Maine, Lisa worked on research projects in graphical perception and the history of psychology. This research resulted in three publications and several conference presentations. During this time, she has also taught courses in General Psychology, Sensation and Perception, Methods of Research, Social Psychology, and Introduction to Computers using the Macintosh. After receiving her Ph.D. in Psychology in August, 2001, Lisa moved to Lancaster, Pennsylvania, where she is employed as an assistant professor in the Psychology Department at Millersville University. Lisa is a candidate for the Doctor of Philosophy degree in Psychology from the University of Maine in August, 2001.