OPTIMISING LINE GRAPH ASPECT RATIO

A thesis
submitted in partial fulfilment
of the requirements for the degree
of
Doctor of Philosophy in Psychology
at
The University of Waikato
by

James Stewart Anderson McEwan

The University of Waikato

1994

ABSTRACT

The aim of these studies was to find a way of selecting aspect ratios for line graphs so that they would provide the data analyst using graphical methods with the most information with the least distortion. If an ideal aspect ratio could be discovered it would help scientist make the best possible use of their research data. Cleveland and McGill (1987) theorised that the ideal aspect ratio could be selected by adjusting the aspect ratio until the line segments that make up a graph's data path were positioned as close as possible to 45°. Cleveland and McGill went on to report a study which supported their theory. Experiments 1-3 here failed to replicate the results of Cleveland and McGill. Further analysis showed that the measure of accuracy used by Cleveland and McGill (1987) and in Experiments 1- 3 was corrupted by a confounding. This confounding was between response bias and accuracy, and as a result nothing could be concluded about Cleveland and McGill's theory from the data.

Given the problems with the previous accuracy measure, a new technique was required to test Cleveland and McGill's theory of ideal aspect ratio. A "signal detection" style procedure called the "behavioural detection model" was used which produced separate measures of accuracy (log *d*) and response bias (log *c*). Using these measures it was found that, as the angle of a pair of lines increased from 0° to 45°, accuracy improved. For all angles greater than 45° accuracy was poor. These results partially support the theory of ideal aspect ratio suggested by Cleveland and McGill (1987). Like Cleveland and McGill's theory it seems the data path should be positioned as close as possible to 45°; Unlike Cleveland and McGill's theory the results here suggest that the data path should never be greater than 45°. These conclusions still need further testing to see if the results here apply to lines with negative slopes, as well as positive slopes, and how well the theory works in applied settings.

Heather McEwan

1965–1993

"You showed us the value of an act of faith:
God grant you peace"

ACKNOWLEDGMENTS

Barry Parsonson	For the use of your computer, for battling with the university research committee to get the equipment we needed, for writing encouraging thing to the higher degrees committee even when they might not have been warranted, for allowing this thesis to become something other than what we initially imagined.
Bernard Guerin	For supervising a student who refuses to do ANOVAs, for stopping what ever you were doing and listening irrespective, for teaching me about the mutants.
Mary Foster	For smiling sweetly ever time I protested that mine was the hardest thesis in the world to write up, for hours spent trying to understand what I was saying badly, for never actually turning the monitor on its side.
Bill Temple	For being the person who said I could do it, for providing a crash course in signal detection.
John Perrone	For nodding politely when I attempted to explain what I was doing, for not sending me away when I talked about pigeons in the middle of a discussion on perception.

To all those listed above, Tina McAdie, and Shirly McEwan, for putting so much of their valuable time into reading many drafts of this thesis in the hope of shaping my limited writing skills - they are still very limited but that's my fault. Finally, to all those unnamed students who made this possible.

ETHICS

All experiments presented here were approved by an independent research ethics review panel. In all experiments the subjects were informed that this research had been approved by an ethics panel, that they were free to leave the experiment at any time without prejudice and that the experiment involved no deception or adversity. Subjects were advised that the 2% towards their course would be available irrespective of their performance on the experimental task and was available should they withdraw from the experiment. In accordance with recommended practice, the subject was given a full explanation of the purpose of the experiment and its broader relevance to psychology. The subject was also offered access to their experimental results.

CONTENTS

Chapter 1 - Introduction

Measurement and quantification have provided the basis for empirical sciences from the time of Galileo to the present day (Kline, 1953). As Johnston and Pennypacker (1980) stated: "Measurement is the cornerstone of all scientific activity. The history of science is coextensive with the history of measurement of natural phenomena, for, without measurement, science is indistinguishable from naturalistic philosophy." (p. 56). Science therefore systematically measures a phenomenon, repeatedly and in controlled circumstances. Those systematic measurements produce data which must be analysed if scientists are to make useful statements about the measured phenomena of interest. Such scientific data analysis can be defined as the manipulation and presentation of a data set in one or more ways so that patterns in the data become observable to the scientist, enabling the discovery, or the confirmation, of those patterns. Tukey and Wilk (1988) described the purpose of data analysis as "to seek through a body of data for interesting relationships and information and to exhibit the results in such a way as to make them recognisable to the data analyser and recordable for prosperity" (p. 2)

The procedures used to perform these manipulations and presentations of data are numerous and varied and can be categorised in a number of different ways. One obvious categorisation is the distinction between graphical and numeric data analysis. Numeric methods of data analysis are characterised by the use of numbers to encode, manipulate and present information about a data set. This information typically indicates some aspect of the data set, for example, the amount of spread in the scores (Standard deviation), the degree to which two set of numbers covary (correlation). Numeric methods are typically data reducing, one or a small number of values stand for all of the numbers in the data set, for example, the arithmetic mean (Cleveland, 1985).

Graphical methods of data analysis, which is the topic of this thesis, are characterised by the use of graphical elements to encode, manipulate and present information about a data set. These graphical elements include: line length, area, slope, relative position, colour and texture (Cleveland & McGill, 1984). Unlike numeric methods, graphical methods are not normally data reducing, for example, instead of presenting the mean and standard deviation to describe a set of numbers, the graphical method would present a histogram showing all of the data, which is an exhaustive description of the data set. Sometimes the distinction made above, between graphical and numeric data analysis, is not always clear, for example, stem and leaf diagrams and other forms of 'semi graphic' techniques do not fall clearly into either group (Tukey, 1977). However, because the 'semi graphic' approach is not data reducing, it also can be considered a graphical technique

Development of graphical methods

Funkhouse (1937) provides an extensive and very scholarly review of the history of graphs, providing a useful historical context for modern graphical procedures. Fienberg (1979) and Beniger and Robyn (1978) also provide historical reviews, though briefer than that of Funkhouse (1937). The use of graphical elements to convey information goes back at least to the time of the Mesopotamians who produced clay maps about 4000 B.C. From Mesopotamia to the Renaissance in the seventeenth century there are many examples of graphical elements used to convey information. None of these examples was designed to analyse data, instead they were used to either show the relative positions of places in very simplistic maps (Tufte, 1983), or theoretical functions in mathematics (Funkhouse, 1937). The start of measurement and data collection in science came with the Renaissance and the development of measurement devices which enabled many reliable measurements to be made, for example, the development of the barometer.

Modern graphical data analysis methods are considered to have started with the "Commercial and political atlas" by William Playfair in 1786 (Tufte, 1983). From that publication through until the end of the 19th Century was a period of rapid development in which most of the graphical analysis techniques used today first appeared. These techniques include graph paper, bar charts, pie charts, log paper, cartograms, cumulative frequency graphs (Fienberg, 1979). Graphical data analysis in the early part of the 20th century did not produce many new developments, instead the period was characterised by further refinement and standardisation of graphical methods. Also at this time the use of graphical techniques became widespread and generally acceptable in scientific publications (Beniger & Robyn, 1978).

Graphical methods in research

Today there is widespread use of graphical methods in science for analysing data, presenting empirical findings, and illustrating theoretical models. Cleveland (1984) analysed 57 scientific journals which fell into three broad categories of science: natural, mathematical, and social. For each journal, 50 articles published between 1980 and 1981 were analysed. The amount of graphical content was measured by the area dedicated to graphs as a fraction of the total area used by an article. Although this measure has its limitations, for example, it does not deal well with a small number of large graphs, it seems in practice an acceptable approach. Cleveland reported that the fractions ranged from 0.310 for the *Journal of Geophysical Research* to 0.0 for *The Journal of Social Psychology*, with a Median fraction of 0.066.

Cleveland noted that the use of graphical methods was much greater in the natural science journals than in the mathematical or social science journals. Cleveland (1984) argued that the mathematical journals might not use graphs extensively because mathematicians have very little empirical data to present, unlike natural scientists. However, the social sciences do have empirical data to present but presumably choose not to use graphical methods. Why this is so was not clear to Cleveland. The general conclusion from Cleveland's review was that graphical methods play a major role in contemporary scientific data analysis, especially in the natural sciences.

A closer inspection of Cleveland's (1984) data suggests a reason why the social sciences choose not to use graphical methods. The data for psychology journals shows that *The Journal of Experimental Psychology* and *Perception and Psychophysics* have a graph-to-area ratio which is about the same as the median for the natural science journals. In contrast, *The Journal of Social Psychology* had a graph ratio of 0.0, that is, no graphs were used in 50 separate articles. This result is probably due to the nature of their research. Research in this area typically involves the examination of many variables or groups at the same time. The data sets resulting from these sorts of studies are generally not amenable to traditional graphical methods of data analysis. As a result, only those branches of the social sciences, or psychology, that examine limited numbers of variables are amenable to using traditional graphical methods, for example, psychophysics or behaviour analysis.

Graphical methods in behaviour analysis

Behaviour analysis is possibly the branch of psychology with the greatest use of graphical methods for data analysis. To show this, the number of graphs, rather than the area of graphs as used by Cleveland, was measured in the two key journals of behaviour analysis. In 1992, *The Journal of Experimental Analysis of Behavior* had 1.5 graphs to every page of research reports. *The Journal of Applied Behavior Analysis* had 0.75 graphs for every page of research reports. These results show quite clearly that behaviour analysts make very extensive use of graphical methods, while at the same time they make very little use of statistical tests. Looking at the number of research reports that use statistical tests we see that *The Journal of Experimental Analysis of Behavior* in 1992 had one in five reports that used statistical tests. *The Journal of Applied Behavior Analysis* had one in 15 reports using statistics. Yet every report that presented statistical analyses also presented graphs. Therefore, it seems safe to say that behaviour analysts depend extensively on graphical methods of data analysis.

This extensive dependency on graphical methods by behaviour analysts has not gone unchallenged. Many writers have suggested that this dependency on graphical methods, to the exclusion of statistical methods, has made behaviour analysts susceptible to drawing the wrong conclusions from their data. Parsonson and Baer (1992) and Yeaton (1982) present critical reviews of the case against relying upon graphical methods for analysing behaviour analytic data. Parsonson and Baer (1992) note that the objections fall into two main categories: first, that the conclusions drawn from graphical data analysis do not necessarily agree with those obtained using statistical tests, and therefore, that graphical analysis must be wrong; and second, that observers do not necessarily draw the same conclusions from the same graphical data, and therefore, graphical data analysis is unreliable and should not be used.

In response, Parsonson and Baer (1992) argued that many behaviour analysts do not accept the assumption that the statistical test should be held up as a reference against which graphical techniques should be compared. They note instead that statistical tests are not definitive arbiters of truth and that there are more variables than just the data itself which affect the conclusions drawn about a particular data set. Parsonson and Baer (1992) also argue that agreement among observers is not necessarily a criterion for truth, and that a diversity of conclusions drawn from a data set might well be beneficial. In short, the objections offered against using graphical data analysis in behaviour analysis have not led to changes in the data analysis practices of behaviour analysts. As a result, the graphical data analysis approach remains dominant.

In summary, graphical methods are used extensively in the natural sciences and within some fields of psychology. Within psychology, behaviour analysis is probably the greatest user of graphical methods, being almost exclusively dependent on them for data analysis. Because of this dependency, it seems important that behaviour analysts, and any other scientists who use graphical methods, should have an understanding of them.

Research on graphical methods

Research on graphical methods is rather limited in both amount and depth, with only a small amount of work on the topic having been done in the last 100 years. What work has been done is spread across a number of different academic disciplines: Applied maths, psychology, Education, Political Science, Economics, Cartography and Management Science. The fact that this work is spread across a number of disciplines has led to unnecessary duplication of research. Further, the quality of the research is sometimes compromised because the researchers are not trained in the sort of research they are attempting, which is primarily the study individual performance on perceptual tasks.

A number of major review papers provide a good starting point for the present review of graphical methods. MacDonald-Ross (1977) presents both a catalogue of the types of graphs that could be used and an excellent overview of the experimental research into graph usage up to that time. Wainer and Thissen (1981) also present an extensive catalogue of graphical data analysis techniques using examples from psychology. There are also a number of interesting books which provide recommendations on graphical selection and usage, for example, Bertin (1981), Cleveland (1985) and Tufte (1983, 1991). However, as will be detailed later, most of the recommendations that come from these writers are based on personal opinion and not on experimental research.

The research on graphical methods falls into three broad areas. First, when should graphs be used, that is, when are graphs the most appropriate method to present data for analysis? Most of the work in this area has looked at whether graphs or tables are a better approach to data presentation. Second, assuming that a graph is the best approach, what type of graph should be used for the best presentation of a particular set of data? For example, are pie charts superior to bar graphs? Third, having selected a type of graph, does the way the graph is drawn affect the conclusions reached from that graph? The little research done in this last area has looked at the effects of coloured verses black and white graphs on their interpretation.

Graph verses tables

The first of these three broad areas that has attracted experimental investigation raises the question: "Are graphical displays of data more effective than tabular displays ?" Early work by Wasburn (1927), Carter (1947) and Vernon (1950) suggested that tables were more effective than graphs at presenting data. Feliciano, Powers and Kearl (1963) suggested that this finding only occurs because of the nature of the task used in these studies, which required subjects to report, or recall, specific values from the table, not to describe trends or patterns in the data. MacDonald-Ross (1977) argued that the graphs used in these experiments were not of an adequate quality to allow for valid comparisons of graph types.

More recent work comparing tables and graphs by Lucas (1981), Watson and Driver (1983) and Dickson, DeSanctis and McBride (1986), concluded that, when subjects were required to report on specific values, tables were better than graphs, providing faster and more accurate reporting of values.

However, when subjects were required to report on patterns or trends in the data, graphical methods were better than tabular methods (Weiner & Reiser, 1976). It seems reasonable to argue that the experimental task of reporting on specific values does not constitute data analysis, given that the definition of data analysis presented at the start of this introduction was "the manipulation and presentation of a data set in one or more ways so that patterns in the data become observable". The conclusion here is that tables are superior to graphs only when used to read off values but not when used for data analysis.

Comparison of graphical methods

MacDonald-Ross (1977) reviewed most of the experimental research on the use of graphical methods up to that date. He attempted to determine, from the more than one hundred published papers available at the time, which graphical method would be the most effective at presenting quantitative data. He concluded that "No one graphic format is universally superior to all others, though some are so unsatisfactory that they can be virtually discarded..." (p. 400). He went on to note that which graphical method is best seems to depend on the type of data being presented, what is intended by the graphical presentation, and what the observer is supposed to do with the information presented. These considerations are complicated by other variables, such as the amount of experience the observer has with graphs and data analysis. Further complications come when subjects have no experience using a particular type of graph, for example, a box and whisker plot. Results are also confounded if the subjects are not familiar with either the data being plotted or what that data might represent. MacDonald-Ross (1977) went on to note that even when all of these confounding variables are controlled for, the presentation quality of the graphs used in research have varied so much that any conclusions about the strengths or weakness of any particular chart should still be treated with caution.

Sparrow (1989), apparently ignorant of MacDonald-Ross (1977), performed an extensive study comparing six different graphical methods: table, pie chart, bar chart, stacked bar chart, line graph, and multi-line graph. For each type of presentation he looked at six different kinds of interpretative task, for example trend tasks, in which subjects had to report if there was change over time in some measured variable. Instead of establishing any single sort of presentation as being superior, as he predicted, he discovered that the effectiveness of a graphical method varied with the nature of the experimental task and the data set being used. Thus, a decade after MacDonald-Ross (1977), researchers were still attempting to discover which graphical method was superior and were still confronting the same limitations previously identified by MacDonald-Ross.

In response to the difficulties noted above, some recent researchers have taken a new approach to the question which graphical method is best. Cleveland and McGill (1984) argued that the essence of the graphical method is the use of graphical elements to encode and present information about a data set. Therefore, the best graphical method is the one for which the graphical elements provide the most effective encoding of numbers.

Cleveland and McGill (1984) called the use of these graphical elements the "elementary perceptual tasks" (p. 532). For example, when comparing two values using bar-graphs the "elementary task" is the comparison of the lengths of the different lines, or if using pie charts, the "elementary task" is the comparison of the different angles that make up the slices of the pie chart. Studying these elementary tasks in isolation removes the problem noted earlier about confounding variables, for example, the experience of the subject.

Cleveland and McGill (1984) were the first to use this new approach. In their experiment, subjects were asked to compare two examples of a graphical element and report what percentage the smaller was of the larger. For example, looking at the elemental code of line length, subjects would be asked what percentage was the shorter line of the longer line. The measure of error, or accuracy, was the difference between the judged percentage difference and the actual percentage difference. For example, if the shorter line was 60% of the longer line and the subject reported that it was 70%, then the error was 10%. They used this technique to examine position on a common scale, position on a common but non-aligned scale, length, angle, slope, area volume, colour saturation and the colour hue. They found that accuracy across graphical elements decreased in the order just given, with position on a common scale leading to the most accurate judgements and colour hue the least accurate judgements.

Simkin and Hastie (1987) conducted a similar experiment but used only three of the graphical elements: position on a common scale, position on a common but non-aligned scale, and angle. They asked subjects to do the same task Cleveland and McGill (1984) had used, and they also asked them to guess which of the two graphical elements had the greater magnitude. Unlike Cleveland and McGill, they also varied the stimulus presentation duration, using either one second or as much time as the subjects required. Simkin and Hastie (1987) concluded that the findings of Cleveland and McGill (1984) still generally held. The major difference they found came when subjects had as much time as they wanted. Under this condition, subjects were more accurate than when they had only one second, however, this did not change the ranking of the graphical elements.

This research seems to have been useful in producing a clear ranking of the "elemental tasks of graphical perception". What remains unclear is whether the ranking of graphical elements translates directly into rankings among the different graphical methods in practice. Presumably, the best way to confirm the validity of these findings is to see whether the ranking holds in real applications. However, as previously noted, this is very difficult due to the many confounding variables that can affect performance of subjects. Therefore, the notion that elemental task performance reflects performance on real graphs is untested, and perhaps is un-testable.

Design issues for graphical methods

The third major area of investigation into graphical methods concerns the decisions that are made about how exactly graphs will be drawn: should the data points be included in the data path of a line graph; should a segmented bar chart use grey scaling or different colours to identify segments of the bar; should one graph be used to present a number of data paths or should a number of graphs be used each with only one data path; and what aspect ratio should be used? Extensive advice on these sorts of graphical design considerations is available, for example Bertin (1981), Cleveland (1984), Dickson (1963), Truran (1975), Tufte (1983, 1991), and many others. Although these authors provided much advice, none of it appears to be based on empirical research. This is perhaps not surprising given that a search for the empirical research in this area shows that almost none has been done. The research into graphical design that has been done is either about the use of colour or the choice of aspect ratio.

Colour

The little research that has been done on colour has been concerned with either maps or bar graphs. For example, Phillips and Noyes (1980) compared the effect of colour with black and white "texture" patterns to encode values onto maps. They concluded that colour is more effective than patterns if the colours used provide adequate contrast between the coloured symbol and the background. Dooley and Harkins (1970) found that subjects seem to be able to recall information from a bar chart more accurately if it was presented on a colour display, compared to a black and white display. Cleveland, Harris and McGill (1983) discovered that using colour on statistical maps (where regions are coloured to indicate some value) can lead to optical illusions in which some regions appear larger than others, especially when using high saturation colour. It is not clear whether colour would affect the visual interpretation of line graphs. Although these studies provide an interesting start to the study of colour in graphical methods, most graphs are still presented in black and white, especially in the scientific journals. Therefore, the application of these studies is limited at this time.

Aspect ratio

The other graphical design issue for which there is some research is the choice of aspect ratio. It is this issue which is the main focus of this thesis. Loosely speaking, the aspect ratio of a graph is a description of the shape of the graph: whether the graph is square, shallow and wide, or deep and narrow. Cleveland and McGill (1987) provide the following definition of aspect ratio. An aspect ratio is composed of two parts, which are mathematically combined, the shape parameter and the scaling parameter. The shape parameter is the physical length of the Y axis, in some measure, divided by the physical length of the X axis in the same measure. This is shown in the following equation:

$$C = \frac{l_y}{l_x} \tag{1}$$

where C is the shape parameter, l_y is the physical length of the Y axis and l_x is the physical length of the X axis. Figure 1-1 shows three example graphs of the same set of data. In Plot 2, the length of the Y axis is 4.8 cm and the length of the X axis is 7.8 cm, making the shape parameter C = 0.61. For Plot 1, C = 1.94, and for Plot 3, C = 0.11.

The second part of the aspect ratio is the scale parameter. It is the range of scale units on the Y axis, the maximum Y value minus the minimum Y value divided by the range of scale units on the X axis (Cleveland & McGill, 1987). This is shown in the following equation:

$$D = \frac{n_y}{n_x} \tag{2}$$

where D is the scale parameter, n_y is the range of the Y axis and n_x is the range of the X axis. In Figure 1-1, for all the graphs the Y axis has a scale range of 15 and the X axis has a scale range of 30, making the scale parameter 0.5 in all three cases.

The aspect ratio combines the shape and scale parameters for any line graph to produce a value that indicates the amount of physical length per measurement unit of the X axis as a ratio of the amount of physical length per measurement unit of the Y axis. This is given by the equation:

$$E = \frac{n_y \div l_y}{n_x \div l_x} \tag{3}$$

where E is the aspect ratio, $n_y \div l_y$ is the scale range of the Y axis divided by the length of the Y axis and $n_x \div l_x$ is the scale range of the X axis divided by the length of the X axis.

Using the preceding definition of aspect ratio Plot 2 in Figure 1-1 has an aspect ratio of 0.8 calculated using Equation 3, as follows:

$$E = \frac{15 \div 4.8}{30 \div 7.8} = 0.8$$

Effects of changing aspect ratio

The choice of aspect ratio can affect the impression a person gets from a line graph. Cooper, Heron and Heward (1987) note, for example, "... that variations in aspect ratio will highlight or minimise the apparent variability in the data set " (p. 124). Re-plotting the data of Plot 2, with an aspect ratio of 0.24 produces a data path that gives the impression that the data are highly variable, as shown in Plot 1. At the other extreme, re-plotting the data of Plot 2, with an aspect ratio of 5.0, produces a data path that might give the impression that the data are stable, as shown in Plot 3.

Extreme aspect ratios not only affect the overall impression of variability, they also make it difficult to see variations within a graph. Cleveland and McGill (1987) argued that this is because extreme aspect ratios make it difficult for people to see a difference between the slopes of the line segments that make up the data path. As a consequence, variation in the data may not be readily apparent, and as a result people might fail to notice important patterns in the data being inspected. Figure 1-2, taken from Cleveland and McGill (1987), shows the same two line pairs plotted with three different aspect ratios. The graphs at the right and the bottom have extreme aspect ratios while the large central plot has a less extreme aspect ratio of 0.50. Cleveland and McGill argue that this figure demonstrates that detecting that the two lines are of different slopes is easier with the aspect ratio used in the larger frame than with either of the two extreme aspect ratios.

Cleveland and McGill (1987) argue that the reduction in line slope discriminability becomes greater with more extreme aspect ratios. If this is true then it can be argued that the opposite must also hold. If extreme aspect ratios reduce sensitivity to line slopes then there is presumably an aspect ratio at which line slope sensitivity is at its greatest and, thus the observer is most accurate at discriminating variations in the data. Given the problems of extreme aspect ratios, it seems important to select an aspect ratio that avoids these biases and, if Cleveland and McGill (1987) are correct, improves the ability of people to detect variations in the data. Therefore, the selection of a ideal aspect ratio would aid in slope comparison, which in turn should make the graphical analysis of data less error prone and more able to extract the maximum amount of information.

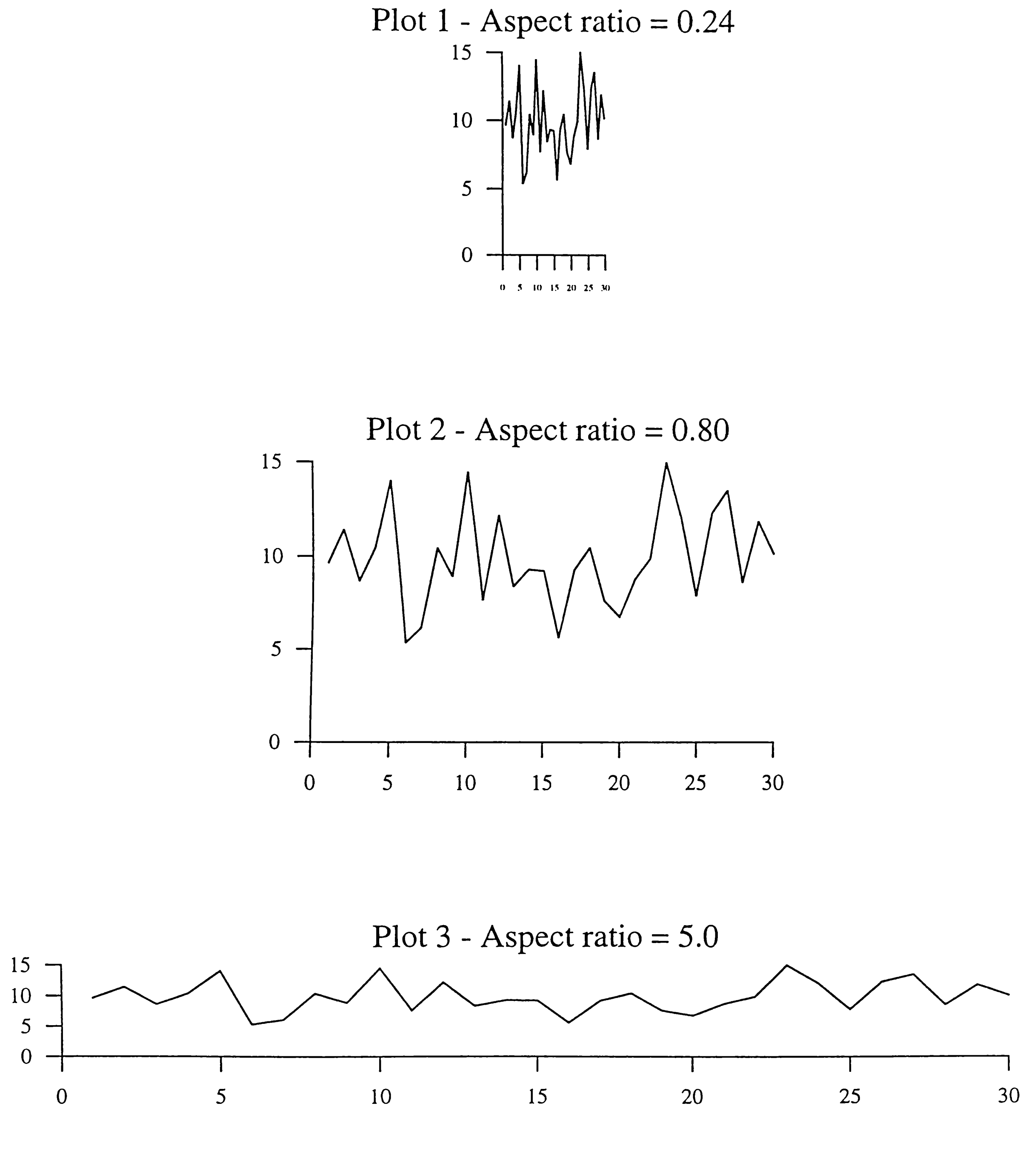


Figure 1-1 Shows a hypothetical data set displayed on three graphs with three different aspect ratios.

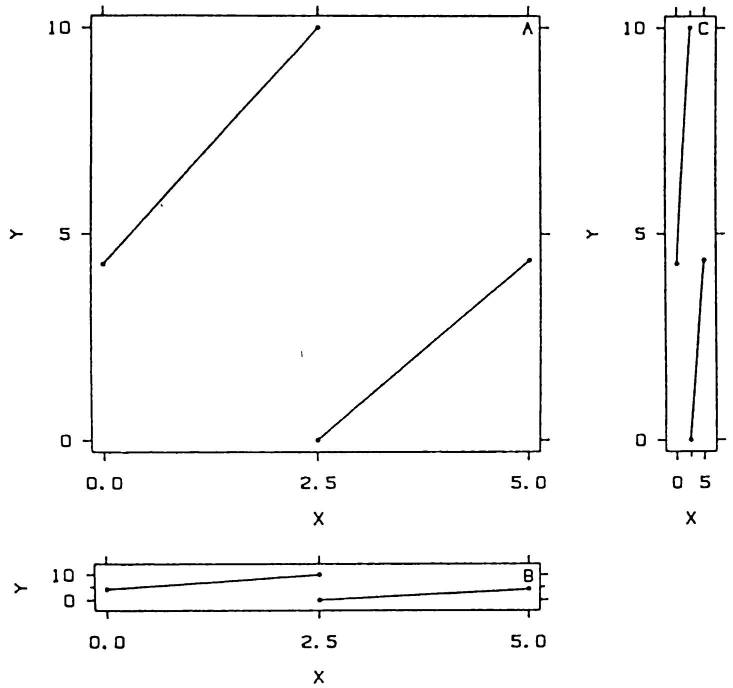


Fig. 5. Orientations of line segments. The same two pairs of points are graphed in each panel. The orientation resolution (difference in angles) of the two segments is greater in Panel *A* than in Panels *B* and *C*. Consequently we can see a slope difference in Panel *A* but not in *B* and *C*. A theorem shows that the orientation resolution of two line segments with positive slopes is maximized when the shape parameter is chosen to make the mid-angle of the two segments equal to 45°.

Figure 1-2. The effect of different aspect ratios on the discriminability of line segment slopes reproduced from Cleveland and McGill (1987) Figure 5.

Selection of aspect ratio

There are three approaches to the problem of selecting an ideal aspect ratio. All of these address only the shape parameter, and ignore the scale parameter. This is because there is very little opportunity to vary the scale parameter in graphs. Scale parameters are largely dictated by the range of the data values being plotted. If the scale parameter range was made smaller than the data range then data points would be left off the plot, although this is sometimes done when a researcher chooses not to include a small number of extreme data points. Alternatively, setting the scale parameter larger than the range of the data also does not appear to serve any purpose. As a result, the scale parameter is almost always treated as unalterable, making it appropriate to consider only the shape parameter when considering the aspect ratio. Hereinafter, the term "aspect ratio" will refer only to the shape parameter as defined above and not to the full definition of aspect ratio provided by Cleveland and McGill (1987).

Fixed-value aspect ratio selection

The first approach is to provide a fixed rule to be used under all circumstances, or a rule that allows for a small amount of variation. Again, such rules relate only to the selection of a shape parameter for a line graph. Parsonson and Baer (1986) argue that a "...ratio of 2:3 represents a proportion that limits distortion and has generalised applicability to the spaces in which graphs of behavioural data are most frequently displayed" (p. 155). In agreement with Parsonson and Baer (1986), Cooper, Heron, and Heward (1987) state that "...on most graphs the vertical axis can be drawn two thirds of the horizontal axis" (p. 124). Johnston and Pennypacker (1980) do not prescribe a shape parameter but indirectly suggest one, stating that "...by far the most common practice is an aspect ratio of 5:8" (p. 345). Tufte (1983) is even more vague, arguing that graphs should be wider than higher because "Our eye is naturally practised in detecting deviations from the horizon" (p.185).

From this collection of rules about shape parameters, plus others cited in Cleveland and McGill (1987), the general conclusion is that a graph should be wider than it is high. The reasons for these rules are not clear and the justifications, when these writers have provided them, have been inconsistent with each other and unconvincing to this writer. For example, Johnston and Pennypacker's (1980) suggestion of a particular shape parameter because it is "common practice" does not seem a well-reasoned justification nor do they provide any other evidence in support of their suggestion. Most importantly, with the exception of Parsonson and Baer (1986), none of the explanations for using a particular shape parameter show how this parameter would reduce biases at extreme aspect ratios; nor do they show how this shape parameter might improve sensitivity to line slopes. These concerns are not to suggest that there is necessarily something wrong with the recommended fixed-shape parameters but that the explanations and evidence for those shape parameters are either missing or unconvincing.

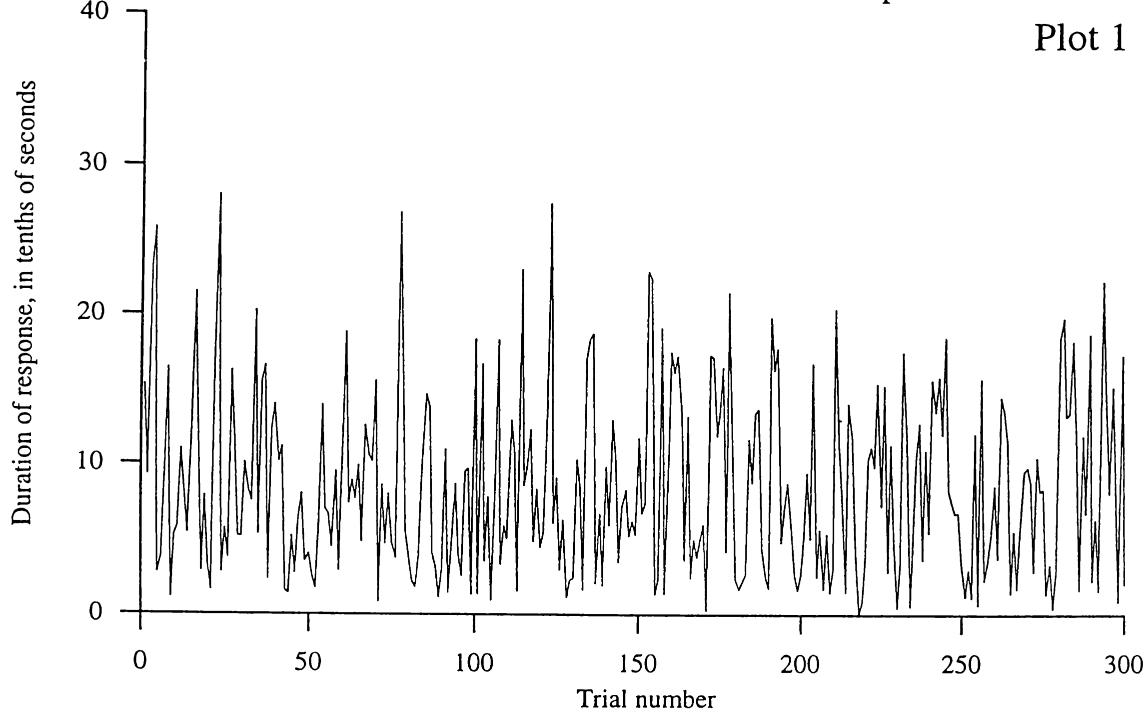
A more significant objection to the use of a fixed-shape parameters is that with some combinations of the fixed-shape parameter and data, the graph may still present the data in an extreme manner. For example, Figure 1-3, Plot 1 has the shape parameter recommended by Johnston and Pennypacker (1980) of 5:8, yet data in this set appears variable, even though the data were generated using a small standard deviation. This apparent variation comes about because the scale parameter, Equation 2, may still have an extreme range even when the shape parameter, Equation 1, has been fixed, so that the overall "aspect ratio", as defined in Equation 3, becomes extreme. Therefore, the fixed-shape parameter approach is limited because it does not take into account the scale parameter. In Plot 2 of Figure 1-3 the first 100 data points from Plot 1 are replotted with a different aspect ratio, possibly suggesting less variability than in Plot 1.

Ad hoc aspect ratio selection

The second approach to the problem of selecting an aspect ratio is to claim that no useful rule about the shape parameter can be provided for a line graph; that the range of data sets and their associated scales is so great that no rules can address all possible situations. (1987) cite a number of writers who take this approach, for example, Brinton (1914) and Meyers (1970; cited in Cleveland & McGill, 1987). This approach does not address the basic aim of finding an ideal aspect ratio but merely avoids the question. However, proponents of this approach are correct to argue that prescribing a single fixed-shape parameter for all data sets is not desirable.

Data derived aspect ratio selection

The third approach to selecting a shape parameter is defined not by a fixed aspect ratio but by the data. With this approach, for every set of data the shape parameter should be varied until the line segments that make up the data path are positioned, on average, as close as possible to a particular slope. In any data graph some line segments will have positive slopes, heading upward, and others will have negative slopes, heading downward. Using this data derived approach we are only concerned with the absolute value of the slopes in selecting a shape parameter, therefore, there is no need to consider whether the lines have negative or positive slopes. Because there is variation in most



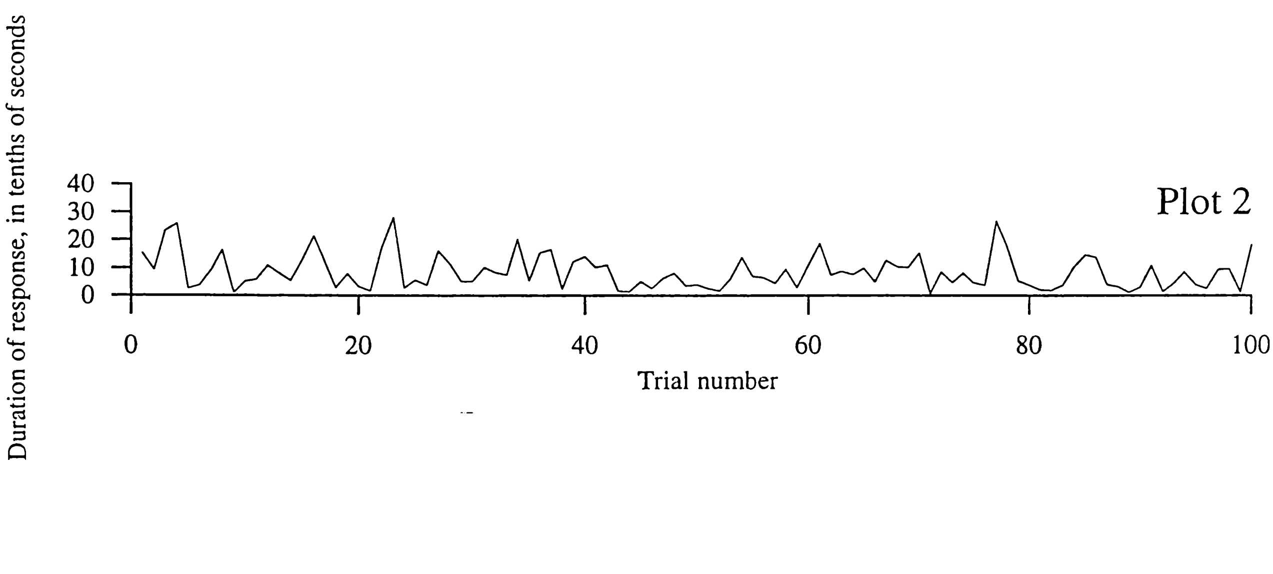


Figure 1-3 Plot 1 shows a hypothetical graph with an aspect ratio (shape parameter) of 5:8 as recomended by Johnston and Pennypacker (1980). Plot 2 shows the first 100 data points form Plot 1 with an aspect ratio of 1:10.

data sets that are plotted, every line segment in the data path will be positioned at a slightly different slope. Therefore, the shape parameter should be varied until the line segments of the data path are, on average, positioned at a particular slope, that is, so that the slopes of the largest number of line segments are as close as possible to the ideal slope. This approach takes into account the scale parameter, unlike the fixed-shape parameter approach.

Cleveland and McGill (1987) report that a number of writers since the turn of the century have made suggestions for the ideal slopes of the line segments, all of which are expressed here as angles (in degrees). Bertin (1973 and the 1983 English translation of this) suggested an angle of exactly 70°; before this Hall, (1958) recommended a range of 30° to 60°; Weld (1947) suggested angles ranging from 35° to 45°; and Von Huhn (1931) suggested an angle somewhere between 30° and 45°, all of these are cited by Cleveland and McGill (1987). There is no single consistent recommendation from these writers, and most prefer to suggest a range of values. The use of a range of values obviously prevents extreme aspect ratios from distorting the amount of apparent variability in the data path, for example, Graphs 1 and 3 in Figure 1-1. Unfortunately, using a range of values does not address the problem of finding the single optimal aspect ratio at which line slope sensitivity is at its greatest and at which the observer is most sensitive to variations in the plotted data.

Bertin (1977 and the 1981 English translation of this), suggested that "the elementary angles must be of average openness" (p. 116). Even though what is intended by this quotation, which was translated from the French original, is not clear, the diagrams accompanying the quotation provides clarification. They suggest that Bertin intended the shape parameter to be set so that the typical, or average, angle in degrees between any two consecutive line segments of the data path was neither too large nor too small. Even though Bertin, in his 1981 English text, makes no attempt to recommend an ideal slope for the average line segment in a data path, it can be concluded from his remark that he would recommend an angle of 45°, or a line segment with a slope of 1.0. The reason for this is that the angle in degrees between two consecutive line segments, measuring the smaller of the two angles or the "inner" angle, can range from 180° to 0°. The angle that is half way between these two extremes is 90°: this is the angle of "average openness" referred to by Bertin (1981). The only way to keep the angle between the line segments at 90° is to set a shape parameter that causes each line segment to be positioned, on average, at 45°.

This conclusion appears to contradict the report by Cleveland and McGill (1987) that Bertin in both 1967 and 1983 recommended a line segment angle of 70°. Further investigation shows that the 1983 reference is to an English translation (published in 1983) of the original French text (published in 1967), not more recent support by Bertin for line segments at 70°. Bertin's (1981) text, cited directly above, was published in English in 1981 but was originally published in French in 1977, four years after the 1983 translation (1973 in French) text cited by Cleveland and McGill (1987) was originally published. It is therefore possible that in the period between the 1973 and 1977 (original French texts), Bertin's position on this matter changed and the conclusion that he would support a data path line segment angle of 45°, or a line slope of 1.0, cannot be rejected outright because of an earlier reference to 70°.

A second, indirect recommendation, in support of data-driven selection of an aspect ratio, comes from Tukey (1977) who suggested that "...smoothly changing curves can stand being taller than wide, wiggly curves need to be wider than tall" (p.129). A reasonable interpretation of this remark might be as follows: "smoothly changing curves" can be taken to mean data paths made up of line segments which have shallow slopes, and that "taller than wide" means a shape parameter greater than 1.0. It is then possible to argue that Tukey (1977) suggested that when the data path is made up of shallow line segments, the shape parameter should be adjusted so that the line segments are positioned with a higher slope. If "wiggly curves" means line segments that are steep and "wider than tall" means a shape parameter of less than 1.0, Tukey (1977) is suggesting a shape parameter that makes steep lines shallower. If the above interpretations are accepted then we can again argue, as for Bertin (1977/1981), that Tukey (1977) might support setting the shape parameter of a graph so that the average slope of the line segments is 1.0, or at 45°. The reason for this is that only when line segments have a slope of 1.0 is it no longer possible to make any further adjustments to the shape parameter based on Tukey's original recommendation.

This data derived approach to aspect ratio, or shape parameter selection, seems to be the most desirable as it avoids the problems of the first approach associated with a fixed-shape parameter and still provides a rule, which the second approach did not. Using this approach to select a shape parameter requires that the ideal line segment slope needs to be very clearly established. As noted above, some writers have suggested a range of values with no particular theoretical rationale. This is not helpful in establishing an ideal line segment slope. Tukey and Bertin have both made suggestions that can be used as a basis of arguments for recommending a line segment slope of 1.0, but their suggestions are based on experience and not any theoretical justification or experimental research.

Ideal aspect ratio

Definition of ideal aspect ratio

Cleveland and McGill (1987) argue that the optimal line slope for a line segment is a slope of 1.0 and support this argument by providing both a theoretical justification and experimental data. They argue that the fundamental task people perform when they look at a line graph is to compare the slopes of the line segments in the data path. This task requires people to make a comparative judgment about how much steeper one line is than another line. For example, judging that one line is twice as steep as another or that one line is half as steep as another. Cleveland and McGill (1987) also argue that this comparative, or relative, judgement of two lines is done more accurately when the absolute difference between the slopes of the two lines is at its greatest. For example, if two pairs of lines have the same relative slopes, perhaps one line is half as steep as the other in each pair, but the absolute difference, in degrees, between one pair of lines is 10° and between the other pair is 20° then people will be more accurate at discriminating the relative slopes of the latter pair of lines.

Theoretical justification for ideal aspect ratio

If Cleveland and McGill's (1987) argument is true then a method is required to establish when the absolute difference between a pair of lines is at its greatest. This is what Cleveland and McGill do mathematically after initially defining two terms, illustrated in Figure 1-4. First, they define the "mid-angle", as the angle that lies half way between the angle of the steeper line in the pair and the shallower line, or mid-way between the pair of lines. The mid-angle can also be defined as the average angle of the two angles of the two line segments, where the angle is between the line segment and the horizontal. For example, if a pair of lines had the angles 10° and 20° the mid-angle would be 15°. Second, they define the "orientation resolution" as the absolute difference between the line angles. Mathematically, orientation resolution is the size of the angle between the pair of lines. For example, if a pair of lines had the angles 37° and 70°, the orientation resolution would be 33°, Figure 1-4.

Cleveland and McGill (1987) found that for any pair of lines when the mid-angle is at 45°, the orientation resolution, or the absolute difference between the two lines measured in degrees, is at its greatest. This can be seen clearly in Figure 1-5 which shows the mid-angle plotted against orientation resolution. From this we can see how, for any pair of lines with any fixed true percent, the orientation resolution changes with mid-angle. They argue that having the greatest possible orientation resolution provides the best possible

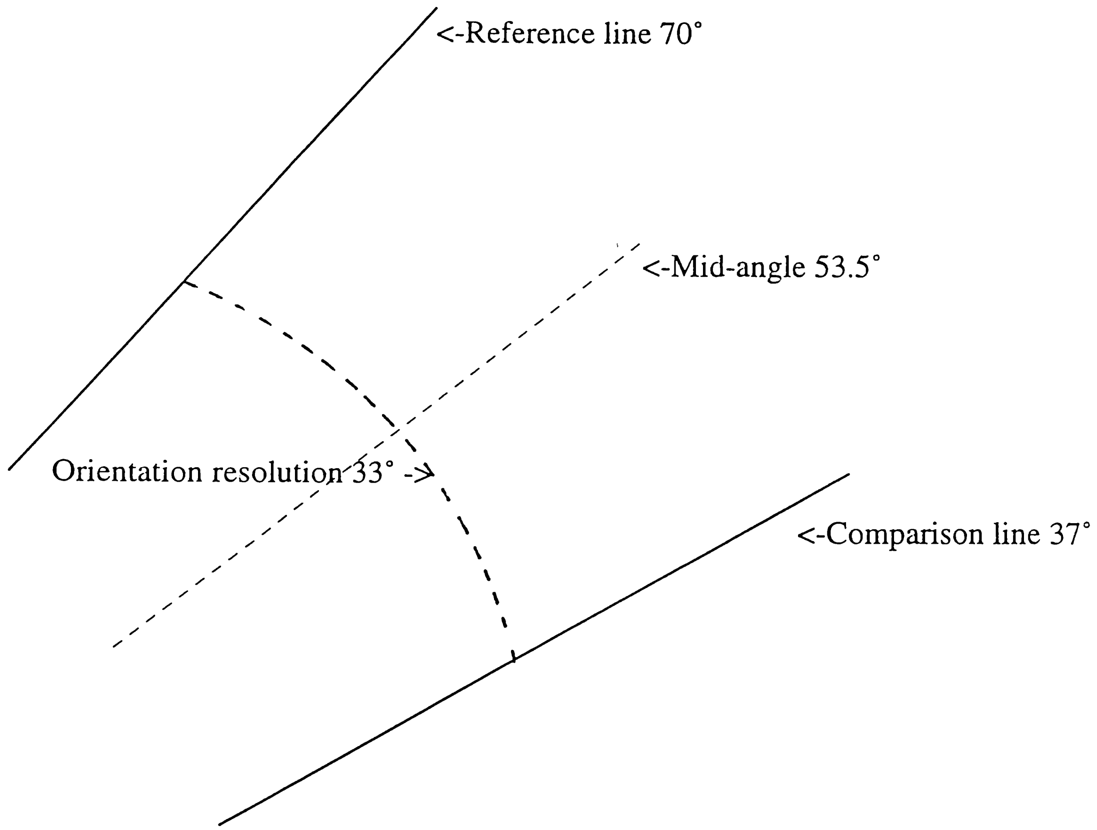


Figure 1-4 Illustrates for any line pair the reference line, comparison line, mid-angle and orientation resolution.

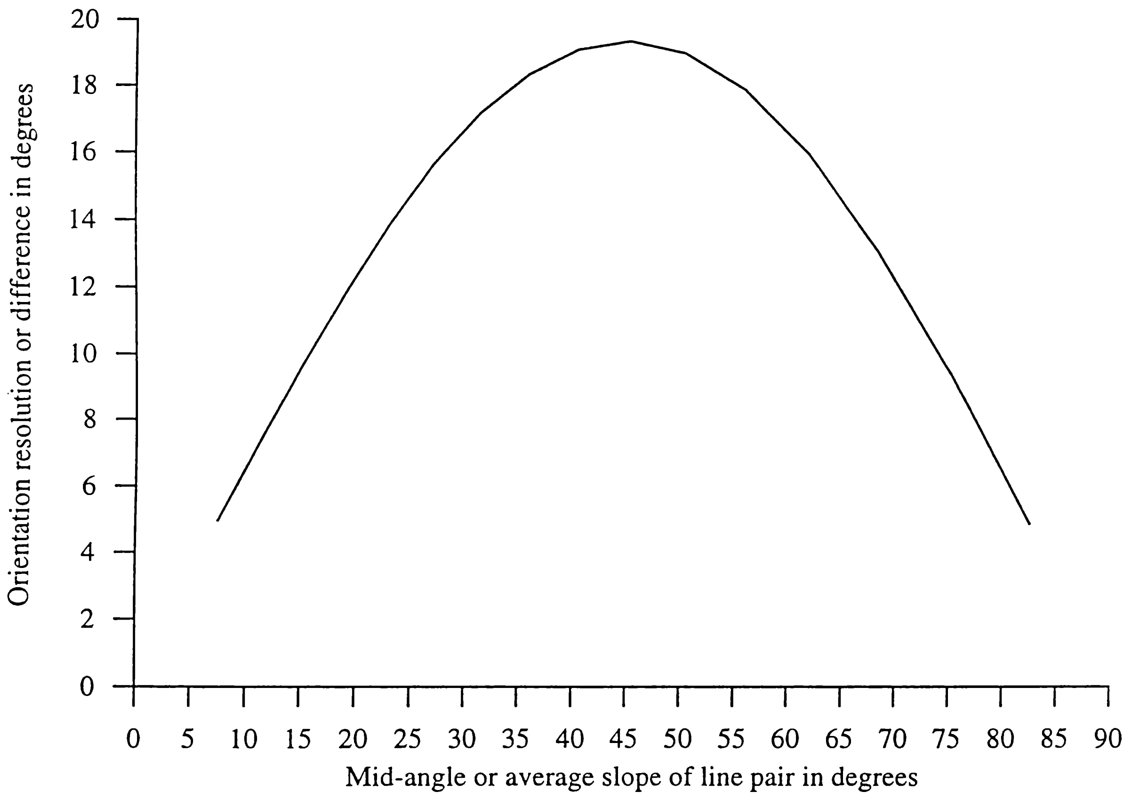


Figure 1-5 Shows that for any line pair, where the slopes of the lines are a constant percentage of each other, that the orientation resolution peaks at 45°.

line-slope discrimination, and that the greatest orientation resolution comes when the mid-angle is 45° for any single pair of lines, that is, for any true percent. From this it follows that the shape parameter of line graphs should be set so that the average line segment in the data path is at 45°, or has a slope of 1.0. They provide an example of this maximal orientation-resolution theory applied to sun-spot data, which is reproduced here in Figure 1-6. They note that the increase in sun-spot activity occurs at a greater rate than the decrease in sun-spot activity. This pattern in the data is clearly observable in the top graph, plotted using their procedure for selecting a shape parameter, but is not observable in the bottom graph, plotted using a more conventional shape parameter.

Experimental justification for an ideal aspect ratio

Cleveland and McGill (1987) followed up this theory with a supporting experiment, which forms the basis of the first three experiments reported in this thesis. In this experiment, 16 subjects worked on a line slope comparison task. They were presented with a pair of sloping line segments and were required to report what percentage the slope of one line was of the other line, for example 60% or 90%. Cleveland and McGill (1987) varied the true percentage difference between the two lines from 50 true percent to 100 true percent and simultaneously varied the mid-angle of the line pair from 8° to 55°. Cleveland and McGill's theory makes predictions about the discriminability of line pairs with mid-angles between 0° and 90°. It predicts that accuracy will increase as the mid-angle increases towards 45° and will then decrease as the mid-angle moves towards 90°. Because of this symmetry around 45°, as mid-angle and therefore orientation resolution changes, Cleveland and McGill argued that there was no need to test experimentally mid-angles greater than 55°, as performance with these mid-angles would be the same as that seen below 55°.

The results from this experiment are reproduced in Figure 1-7, where the mean absolute errors (i.e., the averages of the absolute differences between the true and the judged percents) are plotted against the mid-angles, with separate graphs for the results at each of the different true percents. The mean absolute error decreased as the mid-angle approached 45° in the manner predicted by Cleveland and McGill (1987), though they recognise that at high true percent the lowest mean absolute error occurs a little below 45°. This pattern in the results seems to hold for all the true percent values studied. However, as the true percents became larger the overall amount of error decreased. Cleveland and McGill (1987) concluded that their results showed very clearly how, irrespective of the relative difference of the line pairs (the true percents), performance improved as the mid-angle got closer to 45°. They also argued

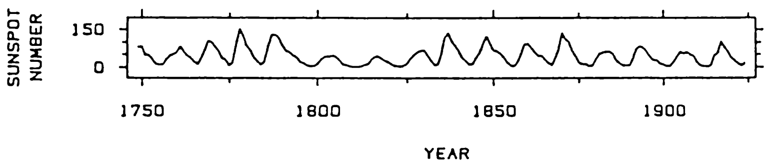


Fig. 2. Sunspot numbers. The shape parameter is 0.065. We can see from this graph that the sunspot numbers rise more rapidly than they fall.

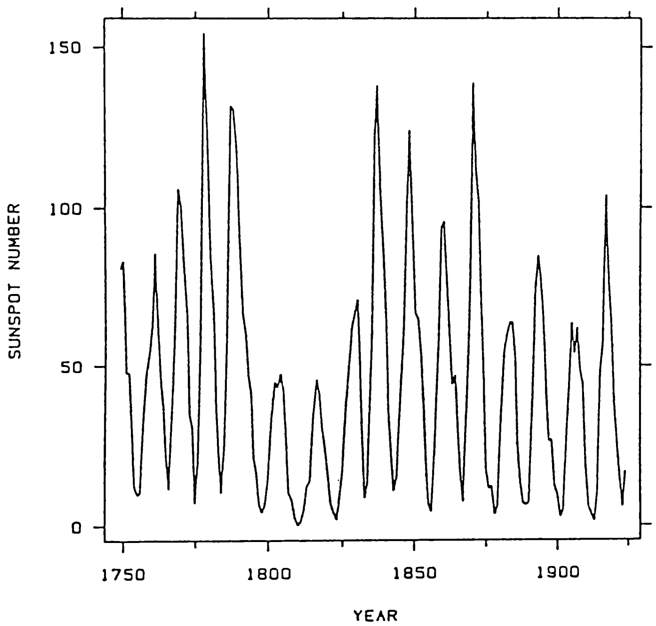


Fig. 3. Sunspot numbers. The shape parameter is 1. We can no longer see the more rapid rise than fall because the change in the shape parameter has made our visual decoding of slopes less accurate than in Fig. 2.

Figure 1-6. is a copy of Cleveland and McGill (1987) application of their theory to data on the number of sun spots. Fig 2 is based on Cleveland and McGill's equation. Fig3 is based on a square aspect ratio, or a shape ratio of 1.

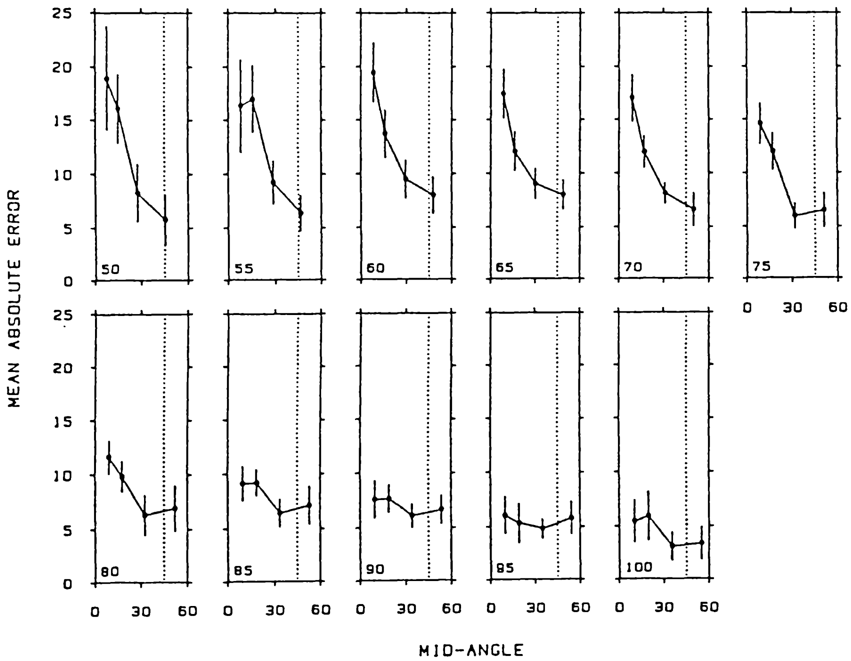


Fig. 6. Results of an experiment. Subjects judged what percent one slope was of another. Average error is graphed against mid-angle for 11 true percents, shown in the lower left corners of the panels. Errors are at a minimum when the mid-angle is in the vicinity of 45°, which is shown by the dotted lines.

Figure 1-7 is a copy of some results from Cleveland and McGill (1987), illustrating mean absolute error plotted against mid-angle across true percent categories.

that the results showed there was a dependency on true percent, with the range of the mean absolute errors decreasing as the true percent increased.

Cleveland, McGill, and McGill (1988) reported the same experiment, this time using a different analysis of the same data. In this second presentation of these results they reported the mean absolute error by true percent, as they did in 1987, but plotted against orientation resolution, not mid-angle. This result is reproduced in Figure 1-8. Again Cleveland et al. (1988) conclude that the results varied with variation in the true percent, with less overall error at higher true percents than lower true percents. They also concluded that the amount of error decreased as the orientation resolution increased.

Cleveland and McGill's (1987) theory for choosing an ideal aspect ratio predicts that it is the orientation resolution that controls line slope discrimination performance, not simply the mid-angle. Both the orientation resolution and the mid-angle vary with true percent, so that at 50 true percent the orientation resolution tends to be large, 5° to 20°, while at 100 true percent the orientation resolution is 0°. Closer inspection of Figure 1-5 shows that Cleveland and McGill (1987) are correct, that at any true percent, as the mid-angle increases, and thus the orientation resolution increases, the error decreases. Examination of the data across true percents shows that as the true percent increases, and hence orientation resolution decreases, the overall error decreases. This can be seen at 100 true percent, when the lines are at the same angle, and as a result the orientation resolution was zero, the error is at the lowest. If 100 true percent is discounted as an extreme value that might have atypical properties and 95 true percent is examined, then similar results are seen. For example, comparing the first data points, where the reference line is 10°, of the 50 and 95 true percent graphs the error at 50 true percent is 19% and at 95 true percent is 6%. The orientation resolution at 50 true percent is 5.0° at 95 true percent is 0.49°. Therefore when the orientation resolution is ten times larger (at 50 true percent) but the amount of error is three times greater than at 95 true percent. This is exactly opposite to the prediction of Cleveland and McGill's (1987) theory. These two findings contradict each other; the first finding suggesting increasing orientation resolution is correlated with decreasing error and the second that decreasing orientation resolution is correlated with decreasing error. A close examination of the data in Figure 1-8 shows that the results both support and contradict Cleveland and McGill's (1987) theory.

Closer inspection of the data in Figure 1-8 shows a similar sort of contradiction. Again Cleveland and McGill (1987) are correct, in that, within true percent, as the orientation resolution (and hence mid-angle) increases, the amount of error decreases. However, in the 95 true-percent plot, (where the

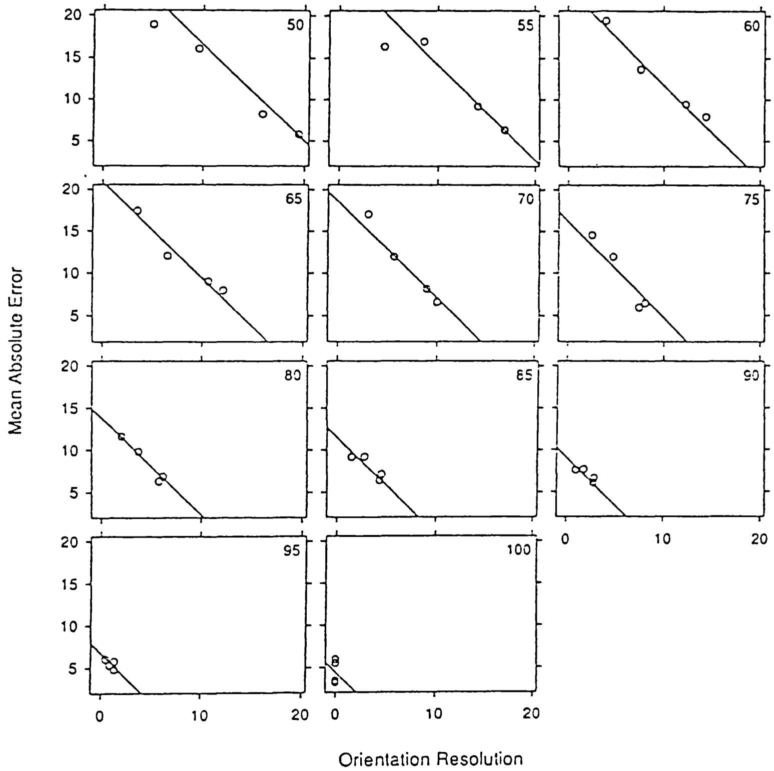


Figure 7. Experiment. In the experiment, subjects judged what percent the slope of one line segment was of the slope of another segment. (All slopes were positive.) Means (across replications and subjects) of the absolute errors for judgments of the 44 pairs of line segments are graphed against the orientation resolutions of the pairs. For each panel, the four values that are graphed are for pairs where the true percent is equal to the value shown in the upper right of the panel. The experiment supported the hypothesis. Judgment error was modeled as a linear function of orientation resolution and true percent. The line segments show the fitted model.

Figure 1-8 is a copy of some results from Cleveland and McGill's (1988), illustrating mean absolute error plotted against orientation resolution across true percent categories.

orientation resolution is almost zero as can be seen by the clustering to the left-hand side of each of the graphs), the overall error is very low at about 5%. In the 60 true-percent plot at the point where the orientation resolution is 15° there is an error of 8%. Overall, the functions relating mean absolute error to orientation resolution are lower the higher the true percent. So again, when we compare results across true percent, error decreases as orientation resolution decreases and the results do not support the theory of Cleveland and McGill (1987). When examined within each true percent the functions trend downwards as the orientation resolution increases and do support their theory.

It might be argued by Cleveland and McGill that the comparison across true percent is inappropriate because the data are already analysed by true percent because of their dependency on true percent. There are two difficulties with this defence. First, there is nothing in the theory about the importance of orientation resolution to line slope discrimination that would suggest a dependency on true percent, as the orientation resolution is the variable that controls error. Second, the only reason Cleveland and McGill give for analysing the data by true percent is that Cleveland found a dependency on true percent in an earlier study (Cleveland, 1984), but they provide no explanation for why such a dependency exists. Logically, it appears that the only effect true percent should have is that, when combined with mid-angle, it defines the orientation resolution for any pair of lines. So it remains unclear whether it is reasonable to draw conclusions based on these comparisons across true percents or not.

Replication of Cleveland and McGill (1987)

So far we have seen that the selection of an ideal aspect ratio is important for the reliable analysis of graphical data. It appears that the best approach to the selection of an ideal aspect ratio is to adjust the aspect ratio so that the line segments that make up the data path lie as close as possible to a particular angle. Cleveland and McGill (1987) have provided a strong, well-reasoned theory for selecting an aspect ratio that places line segments as close as possible to 45°, backing up that theory with experimental results that in part support their theory. If they are correct in their theory, that people do comparison tasks when looking at line graphs, and if their experimental results are correct, then this work has extensive ramifications for the estimated trillion plus line graphs produced annually (Tufte, 1983). As a finding, it is also very valuable to those who rely on graphical data analysis when analysing research data, for example, behaviour analysts, who make extensive use of graphical data analysis and rely predominantly on line graphs.

Given the importance of Cleveland and McGill's work, it was the purpose of the following experiment to replicate as closely as possible their experiment, and to show again the improved line slope sensitivity when lines have a slope approaching 1.0. It was hoped that these results could be used as a basis for further experimentation to clarify some important features of aspect ratio selection and hence aid in the appropriate selection of aspect ratios for line graphs.

Chapter 2 - Experiment 1

Method

Subjects

There were six subjects in this experiment, three males and three females, identified here as ElSl to ElS6. All were volunteers from a first year psychology course at the University of Waikato. Each subject received a 2% credit towards their first year psychology course, which was not contingent upon performance, for participating in the experiment.

Apparatus

The stimuli were presented on a high resolution 340 mm black and white monochrome monitor (Wyse WY-700). This monitor was used because it reduced the jaggedness of the stimulus lines, which otherwise might have provided subjects with an unintended cue to performing the experimental task. It is not possible to tell if the resolution of this monitor is comparable to that used by Cleveland and McGill (1987). The instructions were presented on a 310 mm amber monochrome monitor (Forefront VM-1400). Both monitors were driven by a computer (Commodore PC 10-III). Subjects responded by using the numeric keypad of the computer keyboard. The experimental control software was written in "Turbo C++" and the stimuli were prepared using the graphics program "OPlot". No apparatus was used to fix the position of the subject's head.

Stimuli

The stimuli were similar to those used by Cleveland and McGill (1987). Each stimulus comprised a pair of positively sloping lines with the reference line positioned in the top left-hand quarter of the screen and the comparison line positioned in the bottom right hand quarter of the screen. As noted before, Cleveland and McGill's theory made predictions about mid-angles from 0° to 90°, but they only examined experimentally mid-angles from 0° to 55°. Following Cleveland and McGill (1987), four reference lines were used with angles of 55°, 35.5°, 19.7° and 10.1°. These lines are the same as those used by Cleveland and McGill (1987) and from here on will be refereed to as 55°, 35°, 19° and 10°.

Each of the reference lines was paired with each of 11 different comparison lines. The slope of each comparison line was a graded percentage of the slope of the reference line. The percentages ranged from 50% to 100% in steps of 5%, giving the 11 comparison lines. Therefore, the stimulus set comprised 44 stimuli made up of 11 comparison lines for each of the four reference lines.

TABLE 2-1

The geometry of the lines making up each of the 44 stimuli in the stimuli set. Given is the angle of the reference line, the comparison line and the mid-angle in degrees. Also given is the slope of the reference line and the comparison line, the orientation resolution (the difference between the angle of the reference line and the angle of the comparison line in degrees), and the slope of the comparison line as a percentage of the slope of the reference line.

Reference line angle in degrees	Comparison line angle in degrees	Mid-angle in degrees	Reference line slope	Comparison line slope	Orientation resolution in degrees	Comparison, as a percent of reference line slope
55.0	55.00	55.00	1.42	1.42	0.00	100
	53.60	54.30	1.42	1.35	1.39	95
	52.11	53.55	1.42	1.28	2.88	90
	50.51	52.75	1.42	1.21	4.48	85
	48.80	51.90	1.42	1.14	6.19	80
	46.96	50.98	1.42	1.07	8.03	75
	44.99	49.99	1.42	0.99	10.00	70
	42.87	48.93	1.42	0.92	12.12	65
	40.59	47.79	1.42	0.85	14.40	60
	38.14	46.57	1.42	0.78	16.85	55
	35.52	45.26	1.42	0.71	19.47	50
35.5	35.50	35.50	0.71	0.71	0.00	100
	34.12	34.81	0.71	0.67	1.37	95
	32.69	34.09	0.71	0.64	2.80	90
	31.22	33.36	0.71	0.60	4.27	85
	29.71	32.60	0.71	0.57	5.78	80
	28.14	31.82	0.71	0.53	7.35	75
	26.53	31.01	0.71	0.49	8.96	70
	24.87	30.18	0.71	0.46	10.62	65
	22.16	28.83	0.71	0.42	12.33	60
	21.42	28.46	0.71	0.39	14.07	55
	19.62	27.56	0.71	0.35	15.87	50
19.7	19.70	19.70	0.35	0.35	0.00	100
	18.78	19.24	0.35	0.34	0.91	95
	17.86	18.78	0.35	0.32	1.83	90
	16.92	18.31	0.35	0.30	2.77	85
	15.98	17.84	0.35	0.28	3.71	80
	15.03	17.36	0.35	0.26	4.66	75
	14.07	16.88	0.35	0.25	5.62	70
	13.10	16.40	0.35	0.23	6.59	65
	12.12	15.91	0.35	0.21	7.75	60
	11.14	15.42	0.35	0.19	8.55	55
	10.14	14.92	0.35	0.17	9.55	50
10.1	10.10	10.10	0.17	0.17	0.00	100
	9.60	9.85	0.17	0.16	0.49	95
	9.10	9.60	0.17	0.16	0.99	90
	8.60	9.35	0.17	0.15	1.49	85
	8.11	9.10	0.17	0.14	1.98	80
	7.60	8.85	0.17	0.13	2.49	75
	7.10	8.60	0.17	0.12	2.99	70
	6.60	8.35	0.17	0.11	3.49	65
	6.10	8.10	0.17	0.10	3.99	60
	5.59	7.84	0.17	0.09	4.50	55
	5.08	7.59	0.17	0.08	5.01	50

Table 2-1 includes the slopes and angles for the reference lines and the comparison lines and Figure 2-1 presents annotated examples of the stimulus line pairs.

The reference lines and the comparison lines were drawn starting in the bottom left-hand corner of their respective quarters of the screen. The horizontal extent of both the reference line and the comparison line were kept equal for each line pair. However, because the slope of one of the four reference lines was greater than 1.0, the vertical extent of that reference line and its associated comparison lines was less than that of the other stimuli, (Figure 2-1, A).

Procedure

The experiment was conducted in a small, windowless room that was illuminated by ceiling lights. Once the subject was seated at a table in front of the two computer screens, the experimenter asked the subject to listen while the following instructions, replicating where possible Cleveland and McGill (1987), were read aloud:

> In this experiment the left screen will show a pair of sloping lines for 2.25 seconds. At the end of the 2.25 seconds the pair of sloping lines will disappear and you will be asked to estimate what percentage the slope of the bottom right-hand line is of the slope of the top left-hand line. You are then asked to type in your percentage estimate using the number pad on the far right-hand side of the keyboard, for example 100, 75, 50. After you have typed in your estimate push the <Enter> key, which is also on the number pad, this will start the display of the next pair of lines. Though some lines may appear longer than others, do not worry, it is the judgement of slope that we are concerned with here. Some of the lines appearing on the screen may look a little jagged, please try your best to ignore this. The top line will always be the steeper of the two lines presented. Thus your estimate of percentage should never exceed 100% and it should never be less than 50%. The percentage you type in will be your estimate of what fraction the bottom right slope is of the top left slope. For example if you consider that the bottom line was as steep as the top line your estimate would be 100%, if you considered that the bottom line had about half as much slope as the top line your estimate would be 50%. It is important to remember that it is very difficult to judge slopes perfectly, this is why we are trying to see how close to correct your judgements are. We would like you to give the best estimate you can. Please see the following four pages for examples of line pairs and the correct percentage of slope the bottom line is of the top line. If you have any further questions please ask the experimenter.

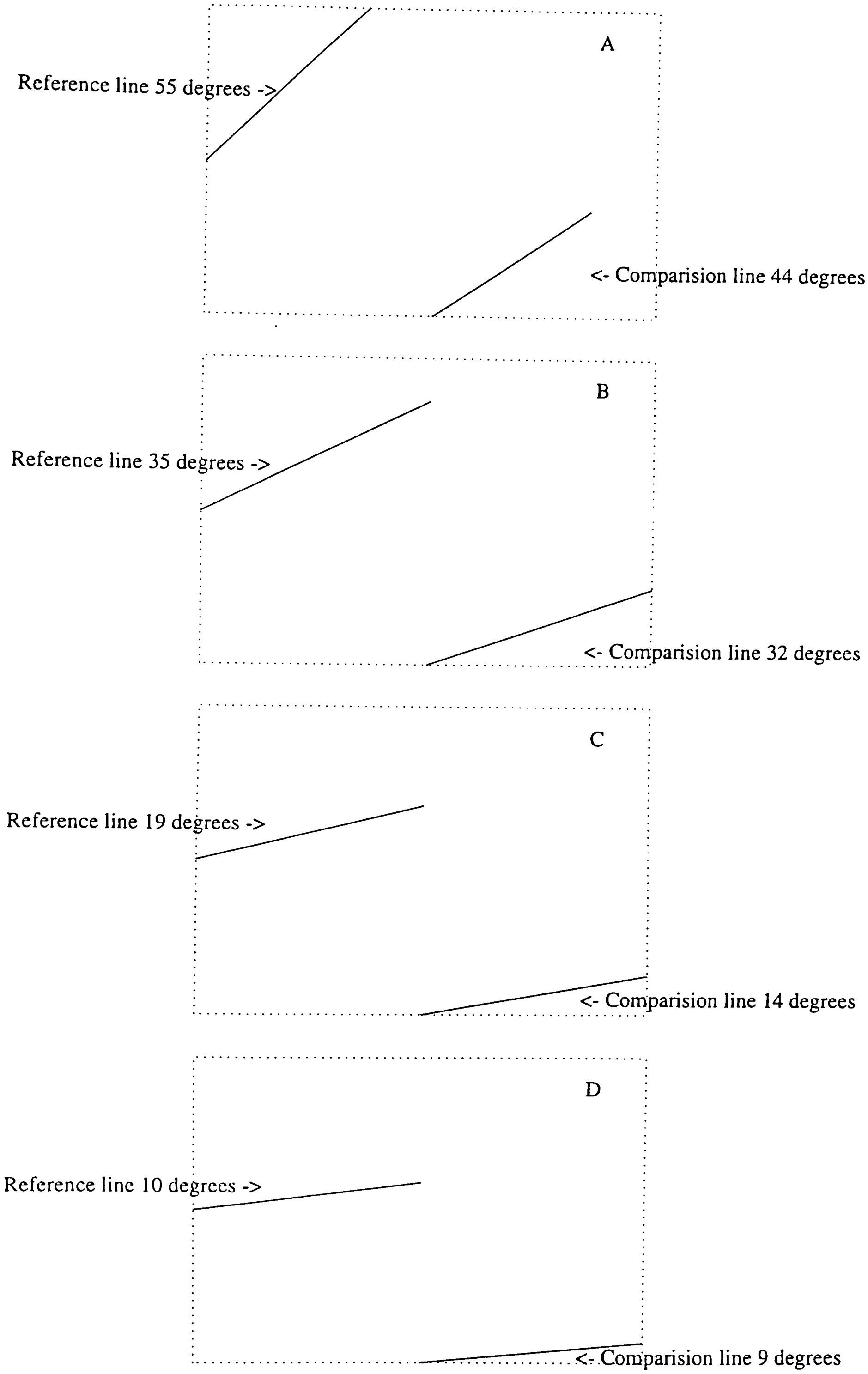


Figure 2-1. Four annotated examples of stimuli. The dotted box respresents the edge of the computer screen. In examples A and C the comparision line has 70% of the slope of the reference line, in the other examples it is 90%.

The subject was then shown four example stimuli on four A4 sheets of paper, these were held vertically in front of the subjects. Figure 2-2 presents these example stimuli on a single page. For each example, the subject was shown what percent the comparison line's slope was of the reference line's slope. Any questions from the subject were then answered by the experimenter, by repeating the relevant part of the instructions. The experiment was then started, and after the completion of four trials by the subject the experimenter left the room.

Each trial started with the presentation of a stimulus on the right hand, high resolution, computer screen for 2.25 seconds after which that screen went blank until the next trial. The words "Please enter your estimate" appeared on the left-hand, amber-monochrome, computer screen as soon as the right-hand screen went blank. The subject then entered, using the keyboard-number pad, an estimate of what percentage slope the bottom line was of the top line. The subject was required to use number keys only, and they could edit these numbers before pressing the <Enter> key to record their estimate and end the trial. When the subject had recorded their estimate, by pressing the <Enter> key, the left hand screen went blank and a new stimulus appeared on the right-hand screen for the next trial. There was no arranged feedback on the accuracy of the subject's response. At the completion of each trial the computer recorded the dependent measures for that trial, these are the subject's estimate of the percentage one slope is of another and the delay between the end of the stimulus presentation and the entering of the judged percent. The computer also recorded the independent variables of actual true percent and mid-angle.

The stimulus used in each trial was selected randomly from the set of 44 stimuli (Table 2-1) without replacement. The complete stimulus set was presented 12 times, each time using a different random sequence. Overall, a session comprised 528 trials (44 stimuli $\times$ 12 presentations). A typical session lasted about one hour during which the experimenter checked progress every quarter of an hour.

At the end of the session subjects were asked informally for their reactions to the experiment and then the purpose and relevance of the experiment was explained and any further questions were answered.

Results

For clarity of presentation and ease of comparison, the figures presented for this thesis all use the same aspect ratios and scales. As a result, some data points are occasionally not within the scale and are not plotted. Instead, the numerical value of the missing data point is shown on the figure. To further improve the clarity of presentation, the line graphs in this thesis do not have markers for each data point. All the data presented in this thesis are available by FTP from the writer at the internet address 'jmcewan@waikato.ac.nz'

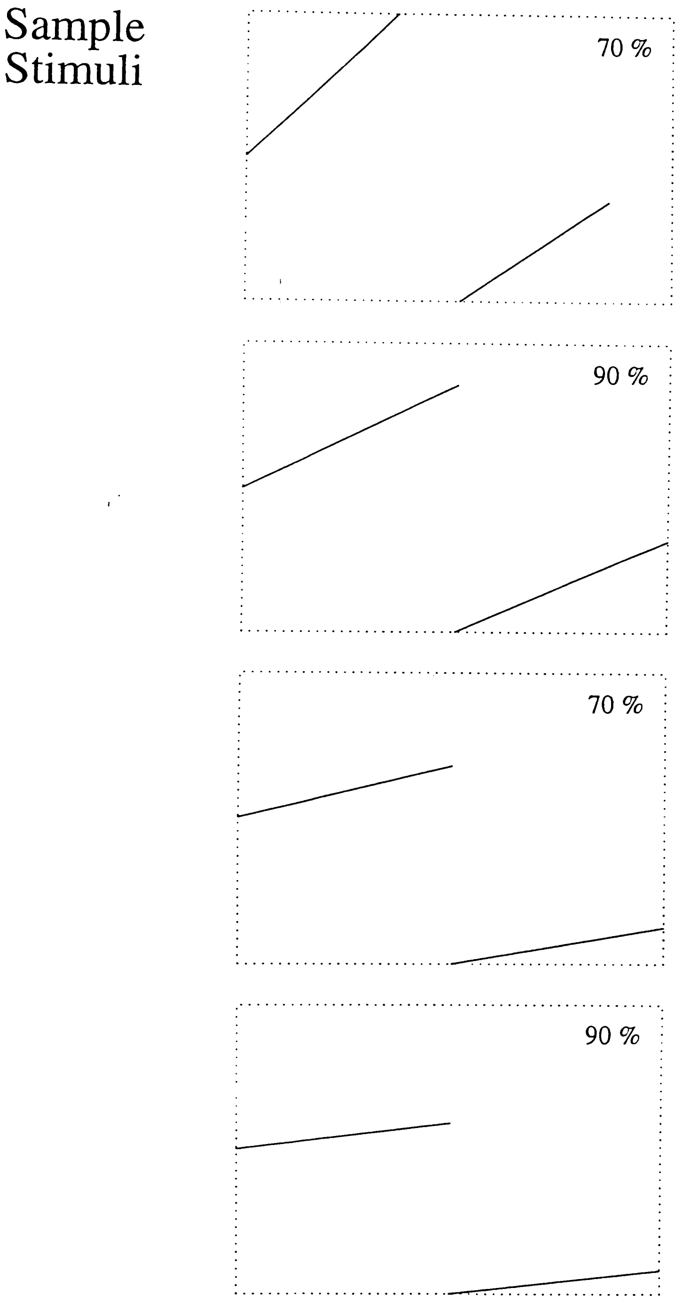


Figure 2-2. The sample stimuli presented to subjects after they had listened to the instructions. The dotted box represents the edge of the the sheet of paper the stimuli were presented upon. Only sample stimuli include true percent.

The measure of accuracy used here and by Cleveland and McGill (1987) is the mean-absolute error. This measure is calculated by taking the percentage the subject estimates away from the true percentage. The absolute value of this error, or difference, is then calculated giving a measure of the magnitude of the error irrespective of the direction of that error. The arithmetic mean is then calculated across the 12 absolute error values from the 12 replications per session. This resulting mean-absolute error is expressed as a percentage, because it is based on one percentage subtracted from another.

Figure 2-3 presents the group means of the absolute error, that is, the arithmetic mean of the absolute difference between a subject's estimate of, and the true, percentage of slope, across the 12 replications and the six subjects. This mean-absolute error is plotted in Figure 2-3 against mid-angle for each true percent of slope as a solid line. The results from Cleveland and McGill (1987) are replotted on Figure 2-3 as a light dotted line.

From Figure 2-3, it can be seen that the mean-absolute error generally decreased as the mid-angle increased towards 45° for the true percents from 50% to 65%. The maximum error, which is at the smallest mid-angle, decreased over these true percents, while the minimum error, which is at the largest mid-angle, stayed roughly constant. From 70% to 90% the mean-absolute error stayed at about 10% irrespective of the changes in both the mid-angle and the true percent of slope. The overall level of mean-absolute error increased at 95% and further increased at 100%. For these two true percents, the mean-absolute error again decreased as the mid-angle increases towards 45°. The maximum error at these true percents, which is at the smallest mid-angle, was larger at 100% than at 95%, while the error at the largest mid-angle remained constant. The pattern, then, is that subjects were making more errors when the mid-angle was smaller, but not when the two lines were around 90 true percent.

While group means were similar to those of Cleveland and McGill (1987) in some ways, the individual data showed very different patterns. Figure 2-4 is similar to Figure 2-3 except that the mean-absolute errors for each subject are plotted separately. The results from Cleveland and McGill (1987) are replotted on Figure 2-4 as a light dotted line. For Subjects E1S1, E1S5 and E1S6, the mean-absolute error up to 90% changed in a similar way to that seen for the group results. For subjects E1S2, E1S3 and E1S4, the mean-absolute error up to 90% stayed around 10% and there were no clear patterns in the mean-absolute error as the mid-angle approaches 45° or as the true percent changes from 50 to 85.

For 90 to 100 true percent, the mean-absolute error for Subject E1S2 generally stayed at about 5% irrespective of true percent or mid-angle. For 90 to 100 true percent the mean-absolute error for Subjects E1S1, E1S3 and E1S5 generally

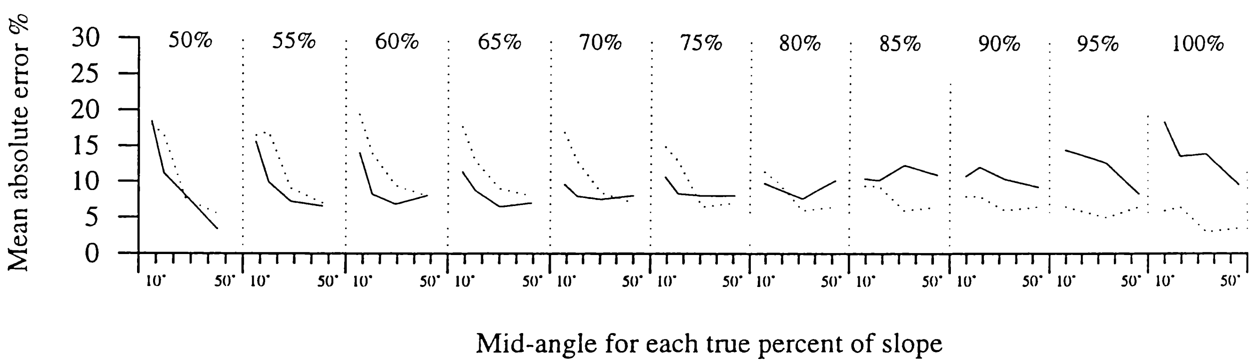


Figure 2-3. The arithmetic mean of the absolute error in estimates, averaged across the 12 replications of the stimulus set and the six subjects, plotted against the mid-angle for each true percent. The light dotted lines are the results from Cleveland and McGill (1987).

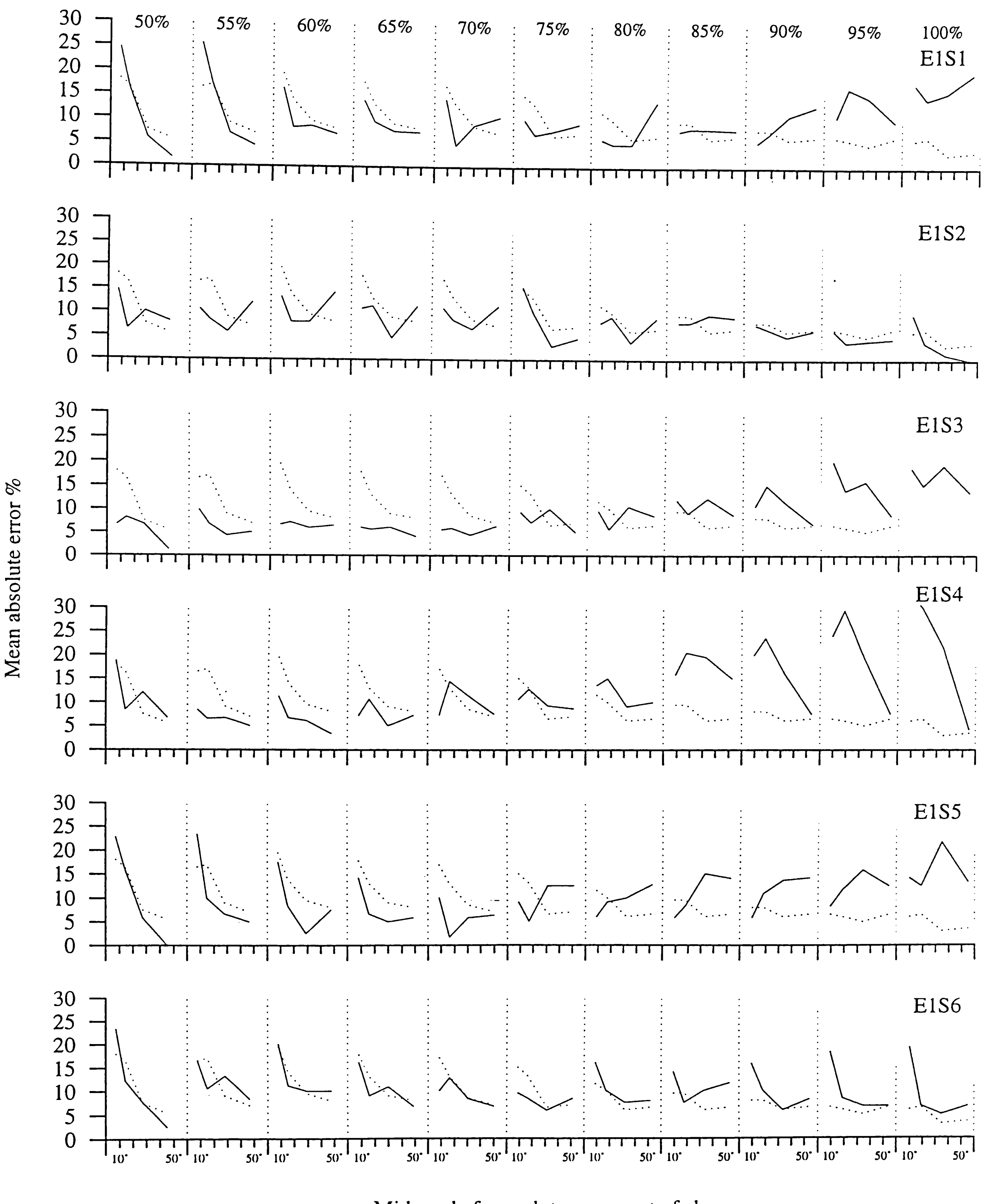


Figure 2-4. The arithmetic mean of the absolute error in estimating slope, across the 12 replications, plotted against the mid-angle for each true percent of slope, plotted for each of the six subjects. The light dotted lines are the results from Cleveland and McGill (1987).

increased as the true percent increased, however there was no systematic variation in mean-absolute error with changes in mid-angle. For Subject E1S6, the mean-absolute error at the lowest mid-angle increased as the true percent increased from 90 to 100, the mean-absolute error at the other mid-angles remained constant at about 10% irrespective of mid-angle or true percent. For subject E1S4, as the true percent increased from 90 to 100 the mean-absolute error at the lower two mid-angles increased and mean-absolute error at the upper two mid-angles decreased.

Another mean-absolute error can be obtained by averaging the absolute errors from the trials with the four mid-angles that all had the same true percent across all six subjects. Figure 2-5 shows the mean-absolute error data for each of the 12 replications of the stimulus set, plotted against true percent. This figure shows that for the group as a whole, over all the replications, the mean-absolute error tended to stay around 10%. Replications 1 to 3, inclusive, showed that the mean absolute error tended to decline for 55 to 85 true percent. No systematic variation as a function of the percentage of slope was apparent for replications 4 and 5. Replications 6 to 12, inclusive, also showed no systematic variation as a function of the replication number. However, for these replications the mean-absolute error increased from around 10% at 55 true percent to 17% at 95 true percent of slope.

Figure 2-6 is similar to Figure 2-5 except that the data for each subject is plotted separately. For all subjects, the data showed no systematic variation as a function of the replication number. For Subjects E1S1, E1S2 and E1S6 there is also no effect as a function of true percent. For Subjects E1S3, E1S4 and E1S5 the mean absolute error increased as a function of increasing true percent of slope, for replications 4 to 12, 4 to 12 and 9 to 12 respectively.

In Figure 2-7, the overall mean-absolute error is plotted against orientation resolution for each true percent for the group. The orientation resolution is the difference between the two stimulus lines measured in degrees of angle from the their respect horizontals, for example when one line is at 55° and the other is at 50° the orientation resolution is 5°. The data from 100 true percent are not shown because the orientation resolution at 100% is zero. Inspection of Figure 2-7 shows that the mean-absolute error decreased as the orientation resolution increased for each true percent from 50 to 75. For true percents greater than 75 there appears to be no systematic variation due to orientation resolution. The difference between the largest and smallest mean-absolute error decreased to almost zero as the true percent increases from 50 to 75, this difference remained between five and zero for true percents from 75 to 95.

Figure 2-8 shows for each of the six subjects the mean-absolute error plotted against orientation resolution for each of the true percents of slope. Comparisons with the group data in Figure 2-7 show that the mean-absolute error from Subjects E1S1, E1S5 and E1S6 at true percents from 50 to 75 look similar to the group data.

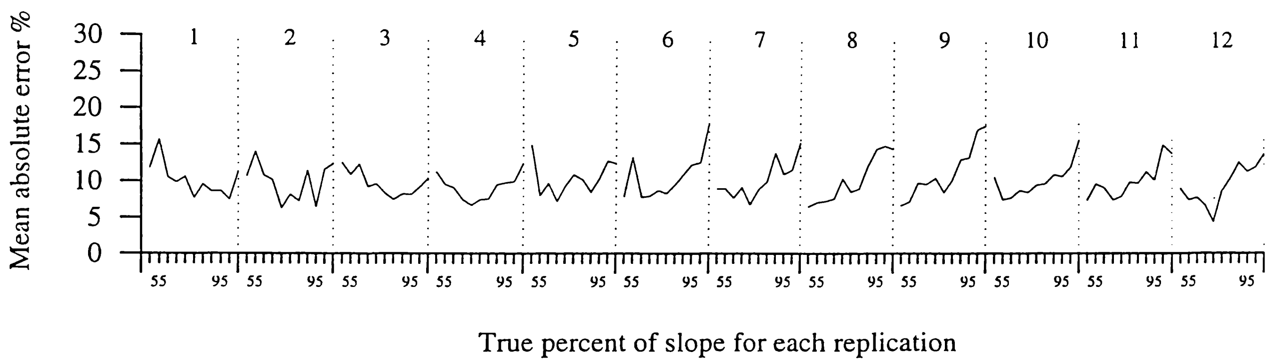


Figure 2-5. The arithmetic mean of the absolute error in estimating slope, across the four mid-angles, at any given true percent, across the six subjects, plotted against true percent of slope for each of the 12 replications.

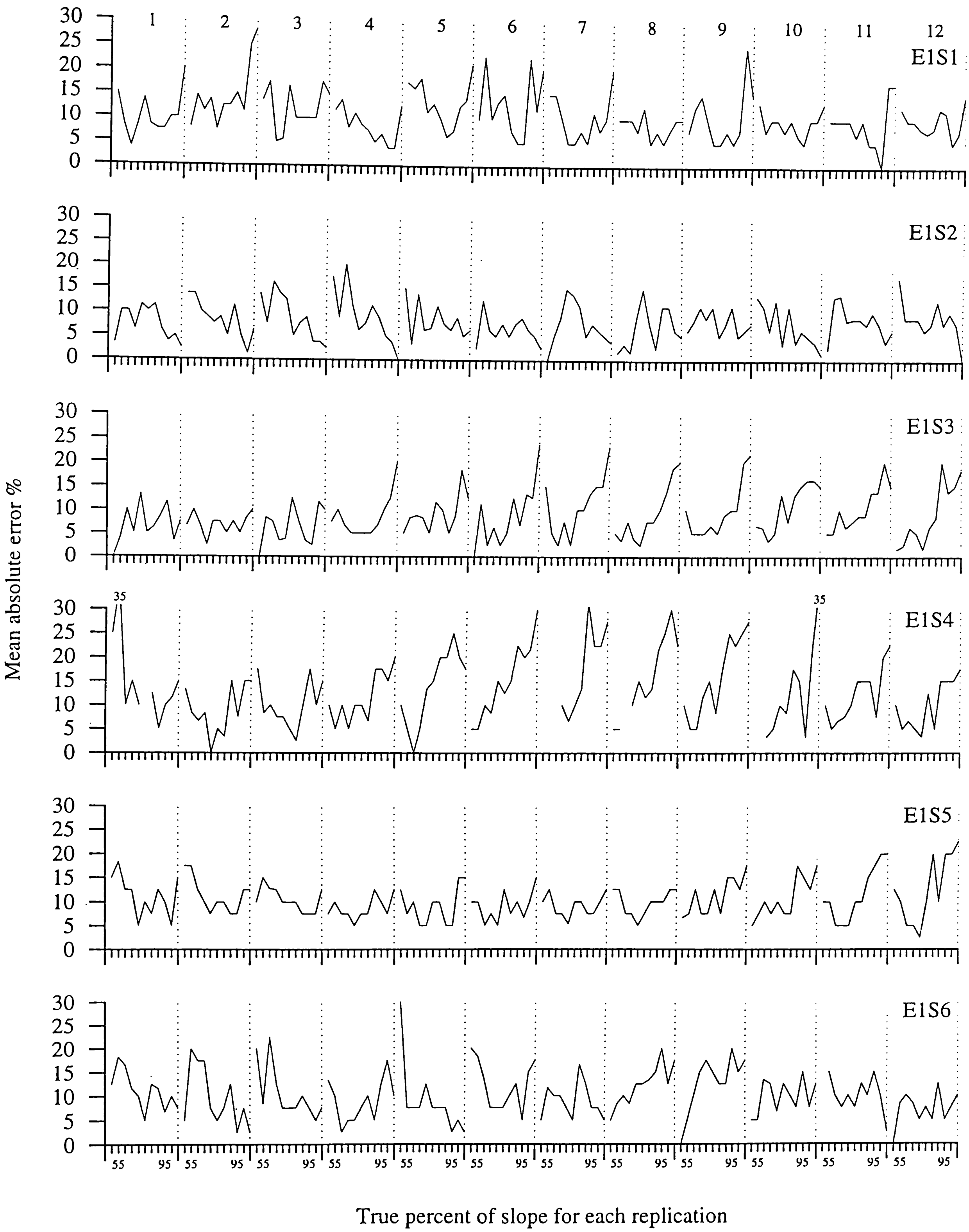


Figure 2-6. The arithmetic mean of the absolute error in estimating slope, across the four mid-angles at any given true percent plotted against true percent of slope for each of the 12 replications, plotted for each of the six subjects.

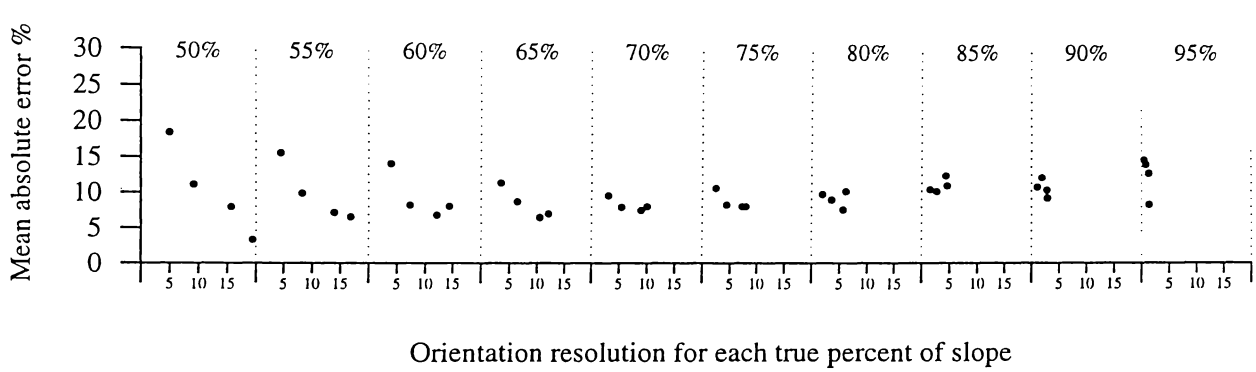


Figure 2-7. The arithmetic mean of the absolute error in estimating slope, across the 12 replications and the six subjects plotted against the orientation resolution for each true percent of slope.

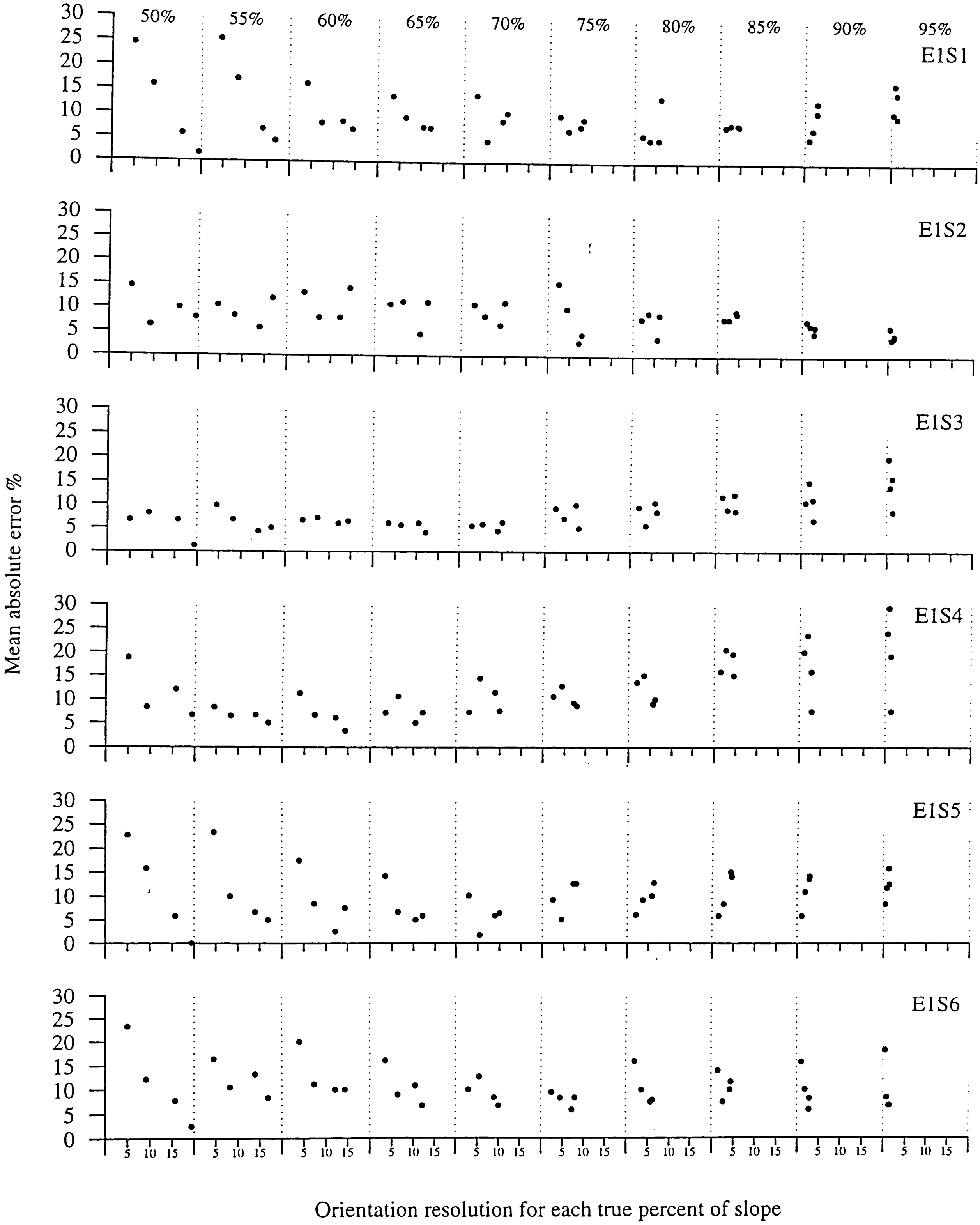


Figure 2-8. The arithmetic mean of the absolute error in estimating slope, across the 12 replications plotted against the orientation resolution for each true percent of slope, plotted for each of the six subjects. Some data points are hidden by other data points e.g E1S1 at 85%.

For these three subjects when the true percent is greater than 75 there was no systematic variation in the mean-absolute error with orientation resolution. The data from Subjects E1S2, E1S3 and E1S4 showed no systematic trends across orientation resolution or true percent.

The mean delay to responding was plotted against mid-angle for each true percent, for each of the 12 replications, and for both the group and individual subjects. These graphs showed that almost all subjects responded at the first opportunity, that is, after the stimuli were removed at the end of 2.25 seconds. As a result the mean delay to responding was between 2.5 s and 3.0 s, with almost no exceptions, and no trends were visible with changes in mid-angle, true percent or replication. Figure 2-9 is presented here as an example; it shows the mean delay to responding broken into four groups based on the mid-angle group and within this the true percent.

Discussion

Cleveland and McGill (1987) proposed that as the mid-angle between a pair of positively sloping lines got closer to 45° subjects would become more accurate at judging the slope of one line as a percentage of the slope of another line. Cleveland and McGill's results supported their theory and showed that overall accuracy for a true percent increased as the mid-angle increased. The results from this experiment shown in Figure 2-3 clearly replicate those of Cleveland and McGill (1987) for true percents from 50 to 90. Unlike Cleveland's results, however, overall accuracy decreased for 95 and 100 true percent, though accuracy still increased with increasing mid-angle.

There was a large amount of intra-subject variation in the present data, and how this compares to Cleveland and McGill's (1987) data is not clear because they did not report individual subject's results. Accuracy at smaller true percents increased with increasing mid-angle as predicted by Cleveland and McGill for three subjects, while for the other three there was no change in accuracy as a function of true percent at smaller mid-angles. The results for Subject E1S2 showed the greatest accuracy at the largest true percents, as did Cleveland's results, while for all other subjects in this experiment overall accuracy decreased at the largest true percents, contrary to Cleveland's results.

Cleveland et al. (1988) re-reported the data from their 1987 paper and proposed an alternative analysis for those data. They showed that as the orientation resolution, the size of the difference between two sloping lines measured in degrees, increased, subjects became more accurate at judging the slope of one line as a percentage of the slope of another line. Again their results supported this theory and showed that overall accuracy for a true percent increased as the mid-angle increased. Results from the present experiment, when analysed as a function of orientation resolution,

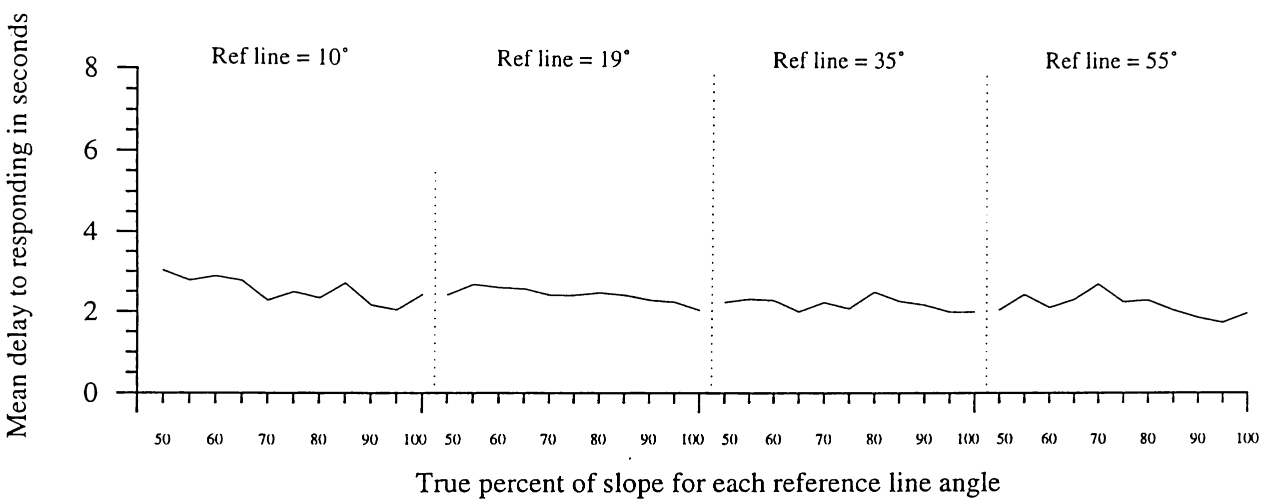


Figure 2-9. The arithmetic mean of the delay to responding, across all 12 subjects, plotted against the true percent, by the reference line angle.

replicated those seen for a mid-angle analysis, with accuracy increasing with increasing orientation resolution for each true percent. Once more, unlike Cleveland's results, these results showed an overall decrease at larger true percents for the group. There was intra-subject variation in the results for the orientation resolution similar to that seen for the individual subject's mid-angle results.

One possible cause for the group results in this experiment not being like Cleveland and McGill's might be that the subjects were first-year psychology students, whereas there subjects were "statisticians, engineers and technical typists" (p. 294). The experimental task of "estimating the percentage one line's slope is of another line's slope" (p. 294) is one that technical people might understand better, and be able to do more readily, than first year psychology students. However, it is not likely that this difference caused the results of this experiment to differ from those of Cleveland and McGill (1987). If subjects were unable to do the task then they would not have been able to get consistently high levels of accuracy at some combinations of true percent and mid-angle. Although this explanation might account for those group results from the present experiment where accuracy was low, the explanation can not also account for the results where the accuracy was high. Further, there is nothing in the theory of Cleveland and McGill (1987) that suggest that the effect might not be true of first year students.

Another possible explanation for the difference from Cleveland's group results is that subjects became fatigued during the long experimental session in which the same task was repeated over 500 times. Unlike their procedure, where subjects were told their overall mean absolute error at the end of each of the 12 replications, no feedback was provided here to the subjects as to how well they were doing. If fatigue was a significant factor, it is likely that performance on later trials would be very inaccurate and therefore could be distorting the overall results. This explanation is unlikely for a number of reasons. First, the accuracy plotted against true percent across replications showed no variation due to the replication number, showing that accuracy effectively remained the same across the duration of the session. Second, no variations in latency to responding were found across the session, suggesting that if there was fatigue it did not exhibit itself in an increasing delay to responding. Third, to account for the differences from the results of Cleveland and McGill (1987) any inaccuracies brought on by fatigue would have to have occurred only for larger true percents. The suggestion that fatigue could have selectively affected performance only at larger true percents this is possible but does not seem highly probable.

Last, another possible explanation for why performance was markedly different at larger true percents than expected is because the stimuli used at those true percents were unlike those used by Cleveland and McGill. This explanation is possible because we have very little information about the exact configuration of their stimuli.

However, it does not seem a likely explanation for a number of reasons. First, the stimuli at larger true percents differ from those at smaller true percents only in the slopes of the lines. Yet we get results comparable to theirs at the smaller true percents, suggesting that the smaller true percent stimuli are the same as Cleveland and McGill's. Therefore, it seems unlikely that the stimuli at larger true percents are different from theirs, though this might still be possible. Second, if the results were different from Cleveland and McGill's at larger true percents because the stimuli used in this experiment are not like theirs, then why does accuracy of some subjects continue to improve in the way predicted by their group results?

A problem of finding an adequate explanation of why the results of this experiment differ from those of Cleveland and McGill (1987) can be seen in the results for each subject analysed separately. These results showed a number of unexpected outcomes. First, none of the individual subject results were properly reflected in the group results. Second, the results from one subject, E1S4, showed none of the patterns in responding predicted by Cleveland and McGill's theory or seen in the group results of either this experiment or those of Cleveland and McGill (1987), when the mean absolute error is plotted against the mid-angle. Third, the trend towards having lower accuracy at low mid-angles at small true percents, as is seen in the group results of this and Cleveland and McGill's experiment, is completely absent from the results of three of the subjects. Fourth, none of the individual subjects in this experiment have results that are the same as the group results from Cleveland and McGill (1987). How this compares with the results of the individual subjects in Cleveland and McGill's experiment cannot be judged as Cleveland and McGill present no individual subject data.

There does not appear to be any single explanation that is adequate to explain all of the intra-subject variation seen in the results of this experiment. For an explanation to account adequately for all the results it would have to be able to predict the absence of any systematic effects across subjects or as a function of mid-angle and true percent, and yet still be able to account for the obtained results. That is, the explanation would have to be able to explain different results, from different subjects, under the same experimental conditions.

The large amount of variability in the data suggests failure to achieve good experimental control over the subjects' behaviour. If, as suggested, we do not have adequate experimental control, then we cannot expect to see the control by the stimulus line angle in the subjects' results. Therefore, the next step would be to find a way to increase experimental control. The two areas of the Cleveland and McGill (1987) experiment that are not well described and which may have been replicated incorrectly in Experiment 1 were the instructions given to the subjects and the format of the stimuli. In the next experiment, the instructions to the subjects were enhanced.

Chapter 3 - Experiment 2

Results from Experiment 1 showed a large amount of variability in the data, both within and across subjects, and no systematic control over performance as a function of the stimulus dimensions: mid-angle, true percent or orientation resolution. Johnston and Pennypacker (1980), like many others, point out that such variability suggests that there is no reliable experimental control over the behaviour of subjects. Only when the performance on the task is stable and reliable can variation in the data due to the systematic variation of the stimuli be observed independently of other sources of control. Greater experimental control is required over the behaviour of subjects in these line slope judgement tasks before control by the stimulus mid-angle can be examined.

Having failed to replicate the results of Cleveland and McGill (1987), the writer attempted to make contact with Dr. Cleveland to get clarification on a number of aspects of the experiment. He replied that if the results had not replicated his "you must have done something seriously wrong". When asked for clarification of a number of procedural details he indicated that the research was done some time ago and that he could not recall those details. As a result we had to rely on the limited information reported in Cleveland and McGill (1987).

Many aspects of procedure used in Experiment 1, such as the experimental setting and stimulus presentation, were constant across subjects and should not have led to variability in the data between and within subjects. One aspect of the experiment that might have been a source of variability was the experimental task itself. The task, following Cleveland and McGill (1987), was to estimate the percentage that one line's slope was of another line's slope. This task might have been difficult for subjects to do if they had no history of working with lines, or had worked with lines described only by angles in degrees and did not have much experience with slope as a description of a line.

As previously discussed, Cleveland and McGill's (1987) experiment used subjects with technical backgrounds, which the subjects in Experiment 1 lacked. If it is reasonable to suggest that subjects with technical backgrounds would be familiar with the ideas of slope and percent, then Cleveland and McGill's subjects would have found the experimental task understandable. This was supported by Cleveland and McGill who state, "...subjects capable of understanding and carrying out the instructions of the experiment" (p. 298). Without a technical background and familiarity with slope and true percent it is possible that the student subjects would find the experimental task difficult to understand. This suggestion that a subject's experience, or history, may affect performance on this type of task is also offered by Rochlin (1955).

To overcome this difficulty, subjects could be provided with more detailed instructions to make the task more understandable and to enable them to perform the task more accurately. There is some suggestion in the literature that instructions can affect how a task is done. For example, Schiano and Tversky (1992) had subjects redraw from memory a sloping line segment and found different results depending on whether the line segment was described as part of a graph or as part of a map. Case and Fantino (1989) showed that they could change the observing responses of subjects to fit with the experimenter's instructions to the subjects.

Another reason for trying to improve the instructions to subjects is that it is also possible that the instructions given were inadequate when compared with those used by Cleveland and McGill (1987). However, it is unclear just what instructions Cleveland and McGill (1987) used or how detailed they were, therefore it is not possible to tell if the subjects in Experiment 1 were given instructions that were sufficient to replicate Cleveland and McGill's (1987) study.

The second experiment was performed to examine whether the instructions could influence performance on a line slope judgement task. One angle judgement experiment which successfully used instructions was by Wenderoth (1984). This experiment used instructions to assist subjects to improve their accuracy on a similar experimental task to the present one. In Wenderoth's experiment, subjects were asked to attend only to the part of the stimuli that was believed to be important in doing the task. By following these instructions subjects did better on the task than other less well informed subjects had on a similar task.

In the second experiment here subjects were given a "Rule" or set of instructions intended to help them perform the task more accurately. They were told that to get a good estimate of the percentage of slope that one line is of another line, they should compare the vertical distance between the top, or end of the line, to an imaginary horizontal line that intersected with the start of each sloping line. This "Rule" required subjects to compare the lengths of two "imaginary" vertical lines. This task should be easier than comparing line slopes, given Cleveland and McGill's (1984) finding that subjects were better at comparing the lengths of lines than comparing the slopes of lines. To make the effect of the "Rule" clear each session of Experiment 2 had a "Standard" phase or condition the same as Experiment 1 followed by the "Rule" phase or condition.

Method

Subjects

There were six subjects in Experiment 2, two males and four females, identified here as E2Sl to E2S6. As in Experiment 1, all subjects were volunteers from a first year psychology course at the University of Waikato and each subject received a 2% credit towards that course, which was not contingent upon their performance, for participating in the experiment.

Stimuli

The stimulus set used in Experiment 2 was a subset of that used in Experiment 1. Only stimuli where the comparison line was 50, 60, 70, 80, 90 or 100 percent of the reference line were used. This subset of stimuli was used as it provided a range of values from 50% to 100%, yet had half as many stimuli and as a result could be presented twice within the same experimental session. The stimulus set of 24 stimuli was approximately half the size of that used in Experiment 1. Table 3-1 gives the details of the particular line pairs used.

Procedure

The computer, monitors and software were the same as those used in the previous experiment. The procedure for this experiment was similar to that used in Experiment 1, differing only in that the session was divided into two equal length conditions. Condition 1, termed the "Standard" condition here, used the same procedure as Experiment 1 except that there were 244 trials rather than 528. When Condition 1 was completed there was no discussion and the experimenter immediately started Condition 2. Condition 2 was labelled the "Rule" condition and it differed from the "Standard" condition only in that at the beginning of this condition subjects were asked to listen while the following instructions were read aloud:

> I would like you to repeat what you have just done so far in the experiment, this time I am going to give you some instructions to help you improve your accuracy at judging line slopes. In this experiment another way to compare slopes is to compare the vertical distance between the top, or end of the line, and an imaginary horizontal base from which each line starts. For example, if the vertical distance of the comparison and reference lines were the same then the percentage of slope would be 100%, or if vertical distance of the comparison line is half that of the reference line then the percentage of slope would be 50%.

When the instructions were completed the subject was again shown the four example stimuli seen at the start of Condition 1, Figure 2-2. For each example the experimenter demonstrated the position of the imaginary horizontal baseline at the bottom of each line and the vertical distances between it and the top or end of the line. Apart from these differences, the procedure was identical to that used in Experiment 1.

TABLE 3-1

The geometry of the lines making up each of the 44 stimuli in the stimulus set. Given are the angle of the reference line, the comparison line and the mid-angle in degrees, the slope of the reference line and the comparison line, the orientation resolution (the difference between the angle of the reference line and the angle of the comparison line) in degrees, and the slope of the comparison line as a percentage of the slope of the reference line.

Reference line angle in degrees	Comparison line angle in degrees	Mid-angle in degrees	Reference line slope	Comparison line slope	Orientation resolution in degrees	Comparison, as a percent of reference line slope
55.0	55.00	55.00	1.42	1.42	0.00	100
	52.11	53.55	1.42	1.28	2.88	90
	48.80	51.90	1.42	1.14	6.19	80
	44.99	49.99	1.42	0.99	10.00	70
	40.59	47.79	1.42	0.85	14.40	60
	35.52	45.26	1.42	0.71	19.47	50
35.5	35.50	35.50	0.71	0.71	0.00	100
	32.69	34.09	0.71	0.64	2.80	90
	29.71	32.60	0.71	0.57	5.78	80
	26.53	31.01	0.71	0.49	8.96	70
	22.16	28.83	0.71	0.42	12.33	60
	19.62	27.56	0.71	0.35	15.87	50
19.7	19.70	19.70	0.35	0.35	0.00	100
	17.86	18.78	0.35	0.32	1.83	90
	15.98	17.84	0.35	0.28	3.71	80
	14.07	16.88	0.35	0.25	5.62	70
	12.12	15.91	0.35	0.21	7.75	60
	10.14	14.92	0.35	0.17	9.55	50
10.1	10.10	10.10	0.17	0.17	0.00	100
	9.10	9.60	0.17	0.16	0.99	90
	8.11	9.10	0.17	0.14	1.98	80
	7.10	8.60	0.17	0.12	2.99	70
	6.10	8.10	0.17	0.10	3.99	60
	5.08	7.59	0.17	0.08	5.01	50

Results

For ease of comparison the aspect ratios and scales of the figures here are the same as those used in Experiment 1. The measure of accuracy used here is the mean-absolute error, which is the same measure used in the previous experiment. This measure is calculated by subtracting the subjects' estimate from the true percentage the comparison line is of the reference line, taking the absolute value of the difference and taking the average across all 12 presentations.

Figure 3-1 presents the group mean-absolute error, calculated by averaging the mean-absolute errors from the six subjects, plotted against mid-angle for each true percent of slope. The mean-absolute error for the "Standard" condition is plotted using the solid line and the "Rule" condition is plotted using the dashed line. Generally, the mean-absolute error decreases as the mid-angle increases towards 45° for the true percents 50 and 60. At a true percent of 70 the mean-absolute error decreases from 15 to 5 true percent between 10° and 30° on both conditions, on the "rule" condition the error stays at about 5% between 30° and 50° for the "Standard" condition the error increases slightly. The maximum error, which is at the smallest mid-angle, decreases over these true percents while the minimum error, which is at the largest mid-angle, stays roughly constant. At a true percent of 80 the mean-absolute error stays at about 5% irrespective of changes in the mid-angle. The overall level of the mean-absolute error increases for the stimuli at 90% and further increases for those at 100%. For these two true percents the mean-absolute error increases and then decreases as the mid-angle increases towards 45°. Unexpectedly, these results are the same for both the "Standard" and the "Rule" conditions, with a very close similarity between the two conditions and no systematic difference between conditions. Again, these results are similar to those of Cleveland and McGill (1987) except for the larger than expected overall error at 90 and 100 true percents.

Figure 3-2 is similar to Figure 3-1 except that the mean-absolute errors for each subject are plotted separately. It shows that for all subjects, other than E2S6, the mean-absolute error changes as a function of true percent and mid-angle in a similar way to that seen for the group results. Results for Subject E2S6 differ from the rest, with the mean-absolute error staying around 10% with no clear patterns as a function of changes in the mid-angle or true percent. For all subjects, these results are the same for both experimental conditions, there being no consistent differences in the data from the "Standard" and "Rule" conditions.

The mean-absolute error used above was calculated by taking the mean of the 12 absolute error values from the 12 replications for each of the four mid-angles across the 11 different true percents. In the next two figures, the mean-absolute error was calculated by taking the mean of the 11 different true percents for each of the

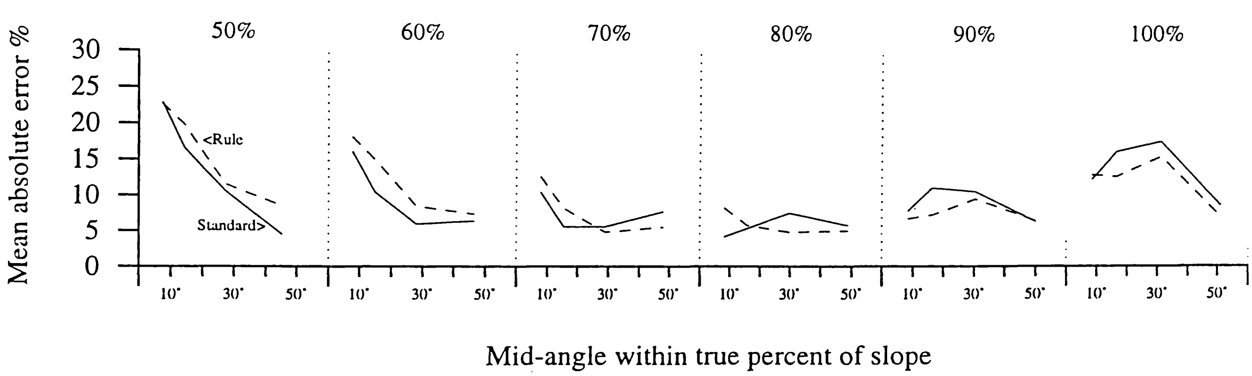


Figure 3-1. The arithmetic means of the absolute errors in line slope estimates, averaged across the 12 replications of the stimulus set and the six subjects, and plotted against the mid-angle for each true percent of slope. The solid line is for the standard condition, the broken line is for the Rule condition.

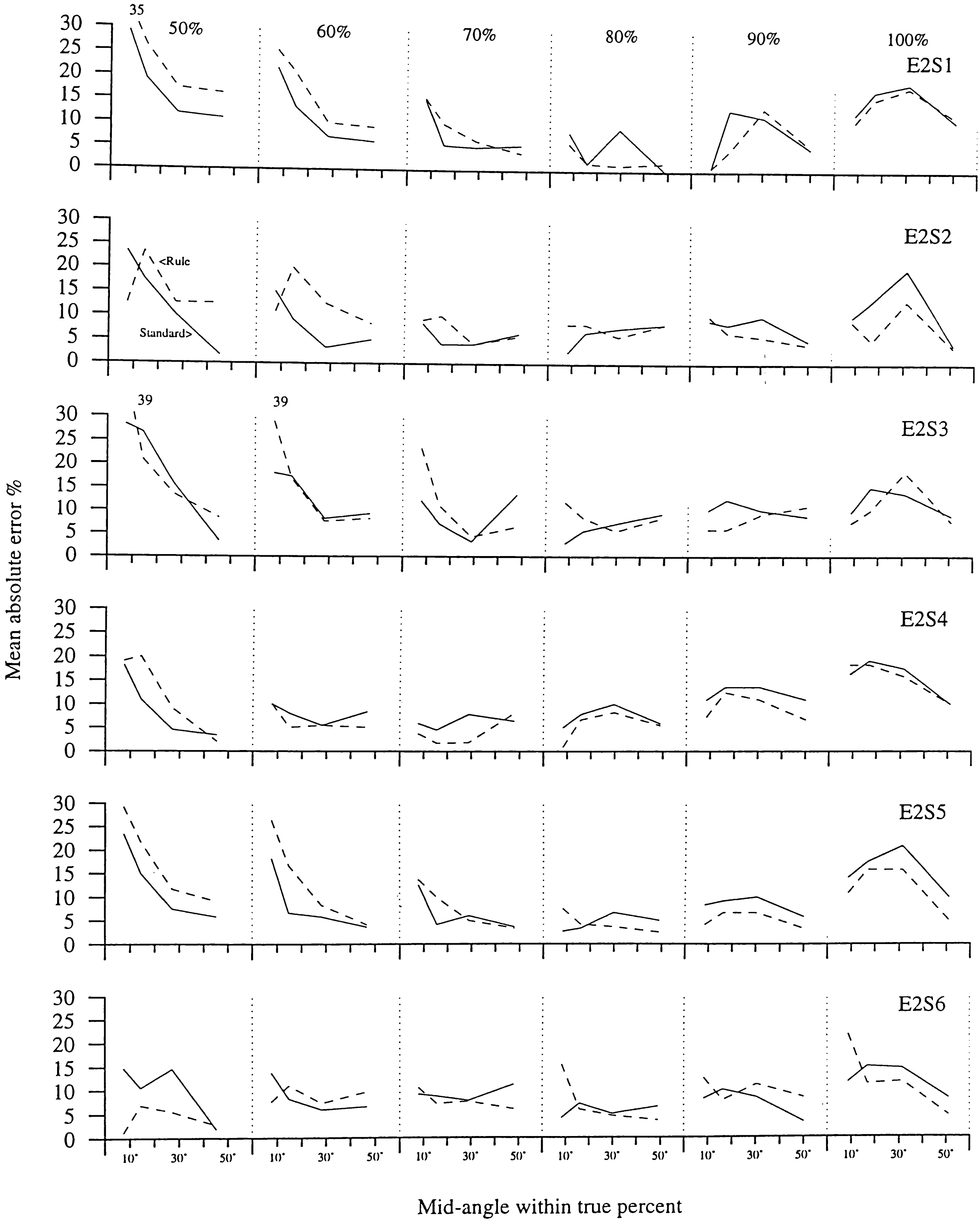


Figure 3-2. The arithmetic means of the absolute error in line slope estimates, averaged across the 12 replications of the stimulus set plotted against the mid-angle for each true percent of slope, plotted for each of the six subjects. The solid line is for the Standard condition and the broken line is for the Rule condition.

four mid-angles across the 12 replications. This was done to identify any changes in performance over the session. Figure 3-3 presents the group mean-absolute error, calculated by averaging the mean-absolute error from the six subjects, plotted against mid-angle for each of the 12 replications. It shows that for the group as a whole, over all the replications, the mean-absolute error tends to vary as a function of the percentage of slope, starting at about 15% at 50 true percent and falling to about 5% at 80 true percent and then returning to approximately 15% at 100 true percent. There appears to be no systematic variation due to replication number. These results are the similar for both the "Standard" and the "Rule" condition, with the exception that the error is lower in the "Standard" condition at 50 true percent for replications 1-6.

Figure 3-4 is similar to Figure 3-3 except that the mean data for each subject are plotted separately. For all subjects, the data show no systematic variation as a function of the replication number. The figure shows that with each replication the data path is a "U" shape which reflects the data seen in Figure 3-2 where error is at its lowest at 80%, and also forms a "U" shaped data path. For all subjects these mean-absolute errors change as a function of true percent in a similar way to that seen for the group results, though this is perhaps less clear for Subject E2S6. These results are the same for both experimental conditions, there being no consistent differences in the data from either condition.

The orientation resolution, or the difference in degrees between the reference and comparison lines, within true percents was plotted against mean-absolute error, and was calculated in the same way as used for Figures 3-1 and 3-2. These graphs were done for both the group and the individual subjects and therefore were similar to Figures 2-7 and 2-8. As noted in Experiment 1, the results for the orientation resolution analysis are necessarily the same as the mid-angle analysis, Figure 3-1 and Figure 3-2, except that there is increasing compression of the X axes as the true percent increases. Because the X axis compression is the only difference, as was pointed out in Experiment 1, this analysis added nothing extra to the understanding of the data so is not included here.

For both the group and the six individual subjects graphs were prepared of the delay to responding in seconds as a function of mid-angle and true percent. These graphs were essentially the same as those in Figure 2-9 in Experiment 1. The delay graphs for Experiment 2 are not presented here since examination of the graphs showed that the mean delay to responding was between 2.5 s and 3.0 s, with almost no exceptions, and no trends were visible with changes in mid-angle, true percent or replication.

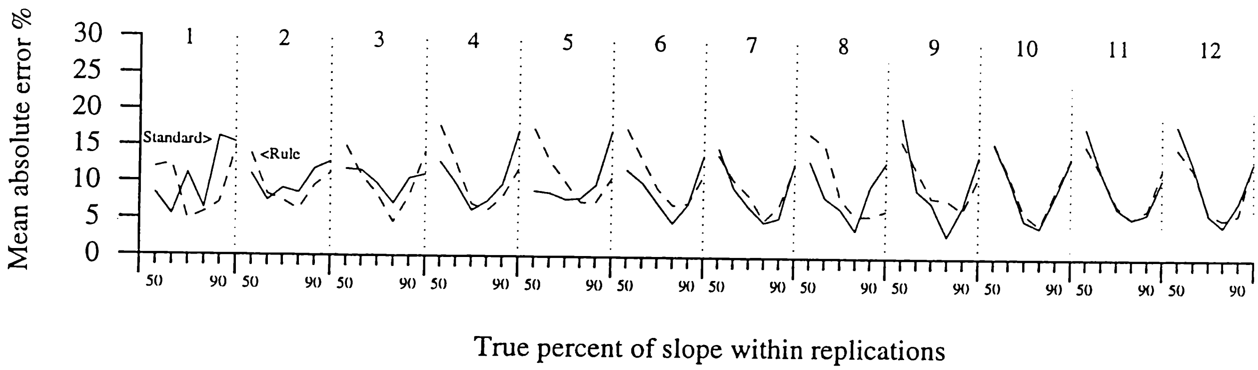


Figure 3-3. The arithmetic mean of the absolute error in estimating slope, across the four mid-angles, at any given true percent, across the six subjects plotted against true percent of slope for each of the 12 replications. The solid line is for the Standard condition and the broken line is for the Rule condition.

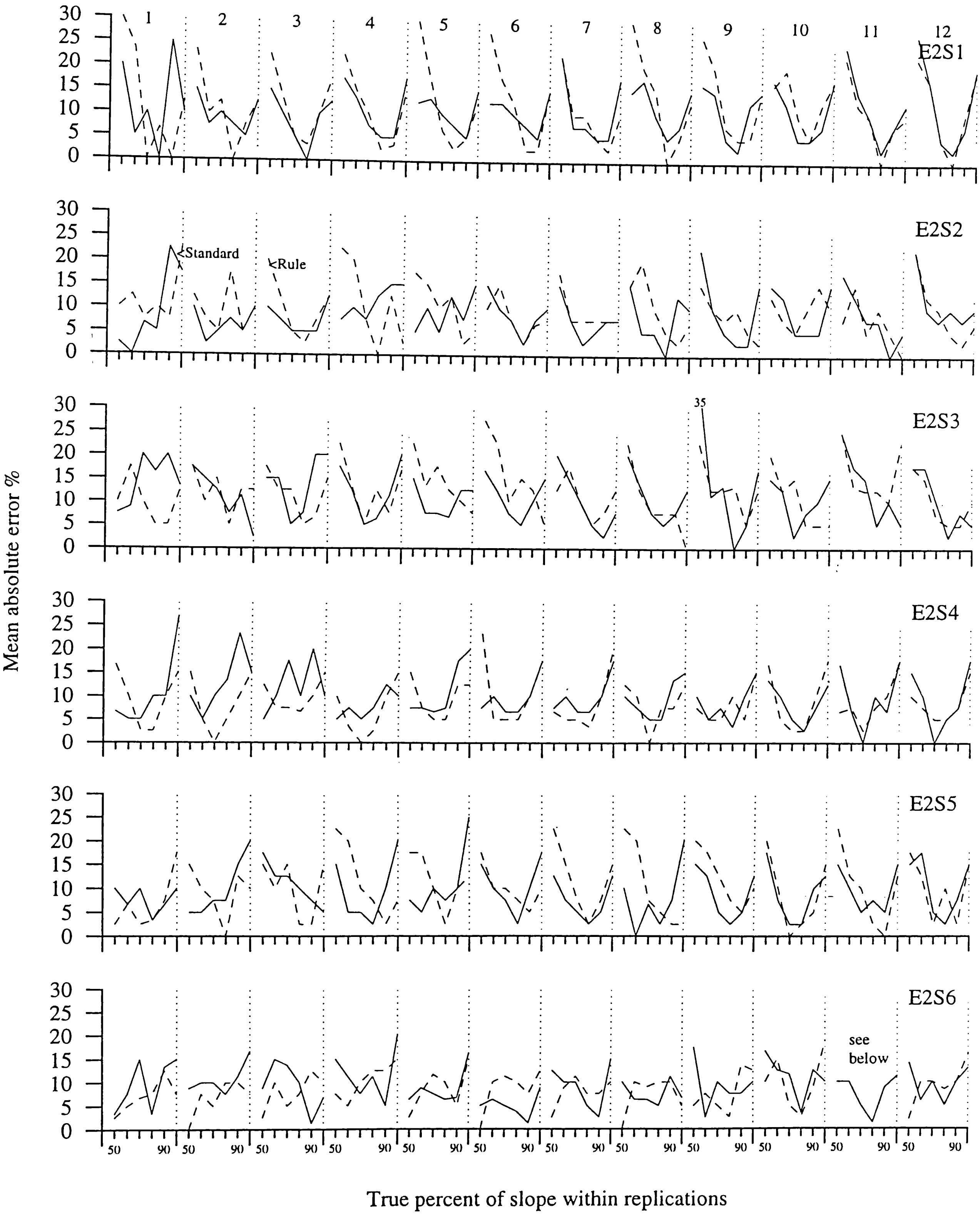


Figure 3-4. The arithmetic means of the absolute errors in line slope estimates, averaged across all the true percents, for each mid-angle, plotted against true percent of slope for each of the 12 replications. The solid line is for the Standard condition and the broken line is for the Rule condition. Note - the data for E2S6 in replication 11 are the same in both conditions.

Discussion

The purpose of this experiment was to provide a rule, or instruction, that would help subjects complete the experimental task more accurately. This was done in an attempt get increased experimental control over the subjects' judgement behaviour. From inspection of the group results it appears that data from both the "Rule" and the "Standard" conditions were very similar. This suggests that for the group as a whole the "Rule" condition had no evident effect on the subjects' performance on the task. This is unexpected, as it was anticipated that the "Rule" would make these results more like those of Cleveland and McGill (1987). The individual subject results generally reflect the group results, therefore the conclusions based on group results are supported by the results from individual subjects.

Why this "Rule" has failed to change the performance of subjects is not clear, although there is a number of possible explanations. First, subjects may not have understood the instructions and as a result could not use the "Rule" to improve performance, although there was no indication of this from the subjects during their debriefing. Second, subjects may have understood the instructions but still found the task very difficult, perhaps because of the short time for which the stimuli were presented or the way the line pairs were presented. Third, it is possible that subjects might stop using the "Rule", even if the "Rule" was an effective way to improve their performance. This could happen because there was no feedback to subjects at the completion of each trial indicating whether they were correct, so they would not know if following the "Rule" was effective. It is even more likely that subjects would stop using the "Rule" if doing the task in the manner described in the "Rule" required more "effort" than not following the "Rule".

Why the "Rule" condition failed to provide the anticipated experimental control is not clear. Perhaps another approach is required to make the task easier for subjects to do. Two other possible sources of task difficulty were noted at the end of Experiment 1, the duration of stimulus presentation and the relative position of the stimulus lines. It is possible to give subjects more time to look at the stimuli in the hope of improving performance on the task. However, this is not likely to be effective, given that Cleveland and McGill (1987) noted that giving subjects less time to look at the stimuli reduced the variability in the resulting data. The argument that the duration of stimulus presentation is important is also not supported by an unreported replication of Experiment 1 in which subjects had as much time to observe stimulus as they wanted, yet the results still did not replicate Cleveland and McGill (1987). The other possible approach, mentioned at the end of Experiment 1, is to change the relative position of the stimulus lines, in the hope that this might make the stimuli more like those used by Cleveland and McGill (1987).

Chapter 4 - Experiment 3

As noted at the end of Experiment 1 there are two aspects of Cleveland and McGill's experiment which could not easily be replicated due to the lack of a detailed description in their published studies, and these two aspects might be responsible for the current failure to replicate Cleveland and McGill's results. The first aspect concerned the type and extent of the instructions provided to subjects. In Experiment 2, Experiment 1 was repeated with extended instructions, but this did not result in a replication of Cleveland and McGill's results. The second aspect of concern was how Cleveland and McGill had configured their stimulus lines.

The two sloping lines in the stimuli used in Experiments 1 and 2 were positioned in diagonal quarters of the screen. This positioning was selected based on Figure 1-2 taken form Cleveland and McGill's (1987) paper that illustrated the effects of differing aspect ratios on the slopes of lines. This figure was used because no other information could be found suggesting where to place the lines in order to replicate the Cleveland and McGill's (1987) experiment. Cleveland and McGill did not indicate in either their 1987 or their 1988 paper what arrangement they had used, and when asked he was unable to recall the positioning (W.S. Cleveland, personal communication, May, 1992). It was possible that the stimulus configuration used in Experiments 1 and 2 was different from Cleveland and McGill's, resulting in a task being more difficult for the subjects in these present experiments. This observation was also made by a discussant of the Cleveland and McGill paper (1987, p. 213), Dr D.J. Hand, who argued that the way the stimulus lines are presented in Figure 1-2 are less than ideal. If detecting a small slope difference between two lines is of paramount importance then presenting stimuli in the way they appear in Figure 1-2 is inappropriate.

In the next experiment, Experiment 3, the stimuli were changed in an attempt to make the experimental task easier for subjects, in the hope that this would give better experimental control over their behaviour. Another way of selecting a new stimulus configuration was required. To this end, the last ten years of a major perception journal, *Perception and Psychophysics,* was reviewed looking particularly at the 55 papers that used lines as part of the stimuli. This provided no clear recommendations for a stimulus configuration. Many articles that used pairs of lines presented them using a common origin. This configuration was rejected because when lines were at 100 true percent, that is the same angle, they would lie over each other, which would make it impossible to present a line pair at 100 true percent.

It was decided that the stimuli in the experiment would be arranged beside, or next to, each other so that the origins of both lines were at the same vertical position but at different horizontal positions. Experimental trials using this new stimulus configuration are said to be in the "Beside" condition. To see clearly the effects of the new stimulus configuration, each session of Experiment 3 had a "Standard" condition, the same as the "Standard" condition used in Experiment 2, followed by the "Beside" condition.

Method

Subjects

There were six subjects in Experiment 3, one male and five females, identified here as E3Sl to E3S6. As in previous experiments, all were volunteers from a first year psychology course and received a 2% credit, that was not contingent upon performance, towards that course for participating in the experiment.

Stimuli

There were two stimulus sets used in Experiment 3. One set was the same as that used in the "Standard" condition of the two previous experiments. The other stimulus set was similar to those used in the previous experiments, except that the reference line was positioned in the bottom left-hand quadrant of the screen instead of the top left-hand quadrant. Thus the reference line and the comparison line were positioned next to each other. Figure 4-1 presents four annotated sample stimuli, these stimuli are similar to those seen in Figure 2-1, with the exception that the reference lines are shifted in position.

Procedure

The computer, monitors and software were the same as those used in previous experiments. The procedure for this experiment was the same as that used in Experiment 2. The first part of the session involved a "Standard" condition, which was the same as Condition 1 in Experiment 2. The second part of the session, Condition 2, was termed the "Beside" condition. The data used in this condition were the same as those in Condition 2 in Experiment 2 except that the new stimulus format was used, and at the start of each condition the subject was asked to listen while the following instructions were read aloud:

> I would like you to repeat what you have just done so far in the experiment, this time the relative position of the lines will be changed, thus helping you improve your accuracy at judging line slopes.

The subject was then shown four sample stimuli on four A4 sheets of paper, Figure 4-2 presents these sample stimuli on a single page. Apart from these differences the procedure was identical to that used in Experiment 1.

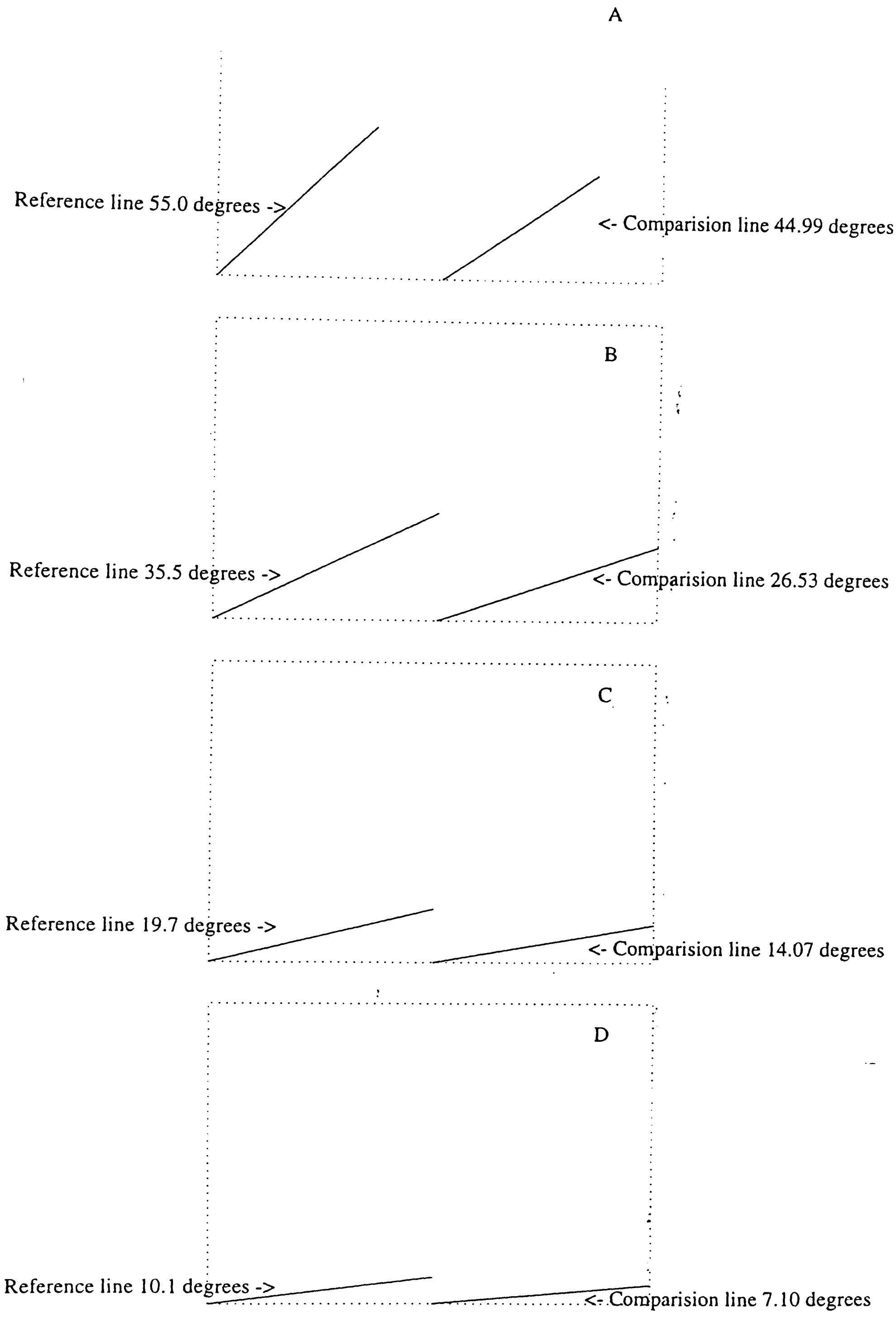


Figure 4-1. Four annotated examples of stimuli. The dotted box respresents the edge of the computer screen. In examples A and C the comparision line is 70% of the slope of the reference line, in the other examples it is 90%.

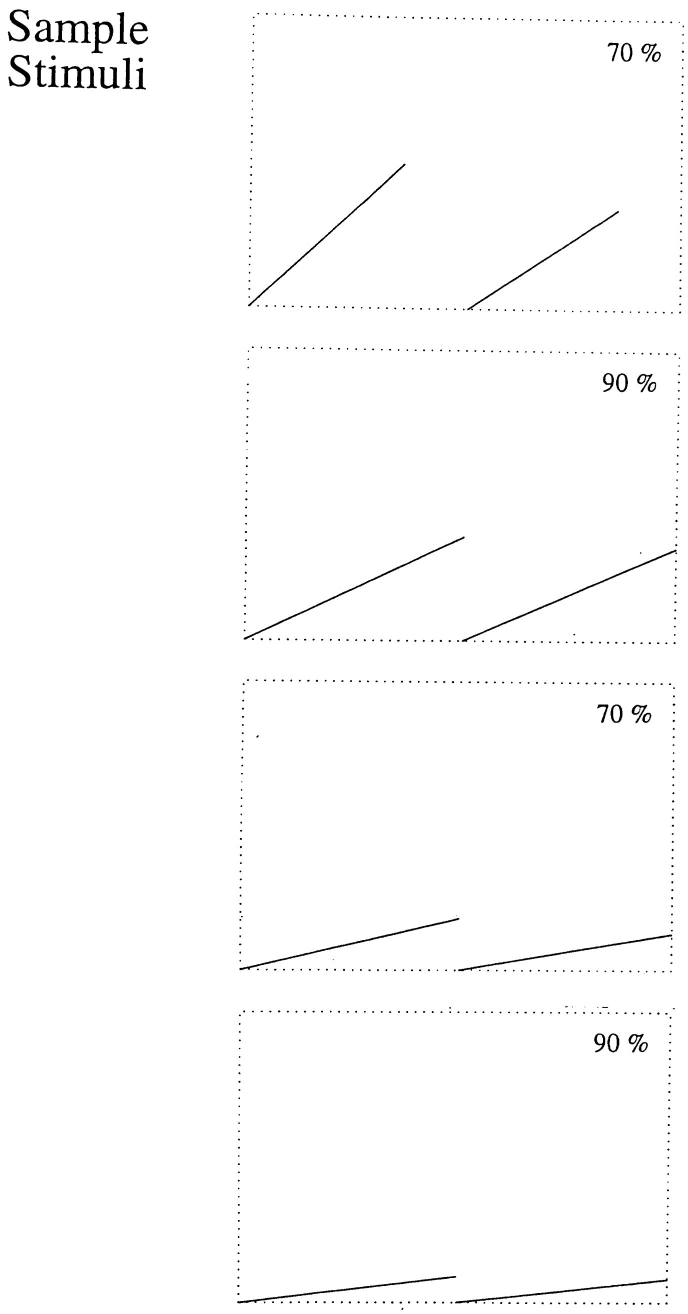


Figure 4-2. The sample stimuli presented to subjects after they had listened to the instructions. The dotted box represents the edge of the the sheet of paper the stimuli were presented upon. Only these sample stimuli include true percent on them.

Results

Figure 4-3 presents the group mean-absolute error for the "Standard" condition plotted using the solid line and the "Beside" condition is plotted using the dashed line. Generally, the errors decrease as the mid-angle increases towards 45° for all true percents and the overall size of the mean-absolute error decreases gradually as the true percent increases from 50 to 100. These results are similar for both the "Standard" and the "Beside" condition, except that the mean-absolute error during the "Beside" condition is lower at higher true percents.

Figure 4-4 is similar to Figure 4-3 except that the mean-absolute errors for each subject are plotted separately. It shows that for subjects E3S1 and E3S2 the mean-absolute error changes as a function of true percent and mid-angle in a similar way to that seen for the group results, and the group results from Cleveland and McGill (1987). This is true across both experimental conditions. For Subjects E3S4 and E3S6, the mean-absolute error generally stays between 5% and 15% and there is no clear pattern as a function of changes in the mid-angle or true percent, nor are there any notable differences between the "Standard" and the "Beside" condition data. The data from Subject E3S5 is similar to that seen for Subjects E3S4 and E3S6 except that there is a greater range in the mean-absolute error. The mean-absolute error for Subject E3S3 during the "Beside" condition shows no clear pattern as a function of changes in the mid-angle or true percent, typically falling within a narrow band between 5% and 10%. For this subject, the mean-absolute error during the "Standard" condition decreases as the mid-angle increases and generally increases as the true percent increases from 50 to 100 true percent.

Overall, there was a lot of across subject variation in results, and the data from the two conditions were very similar, with the exception of those from E3S3. The mean-absolute error shown in the next two figures was calculated by taking the mean of the 11 different true percents for each of the four mid-angles across the 12 replications. This was done to look for any changes in performance over the session. Figure 4-5 shows this mean-absolute error for the "Standard" condition using the solid line and the "Beside" condition using the dashed line. These mean-absolute errors for each of the 12 replications of the stimulus set are plotted against true percent. This figure shows that for the group as a whole, over all the replications, the mean-absolute error in the "Beside" condition tends to vary as a function of the percentage of slope, with the mean-absolute error starting at about 15% at 50 true percent falling to about 5% at 100 true percent. There appears to be no systematic variation due to replication number. These results are similar for both the "Standard" and the "Beside" condition, except that the mean-absolute error during the "Standard" condition does not fall as low as for the "Beside" condition as it approaches 100 true percent.

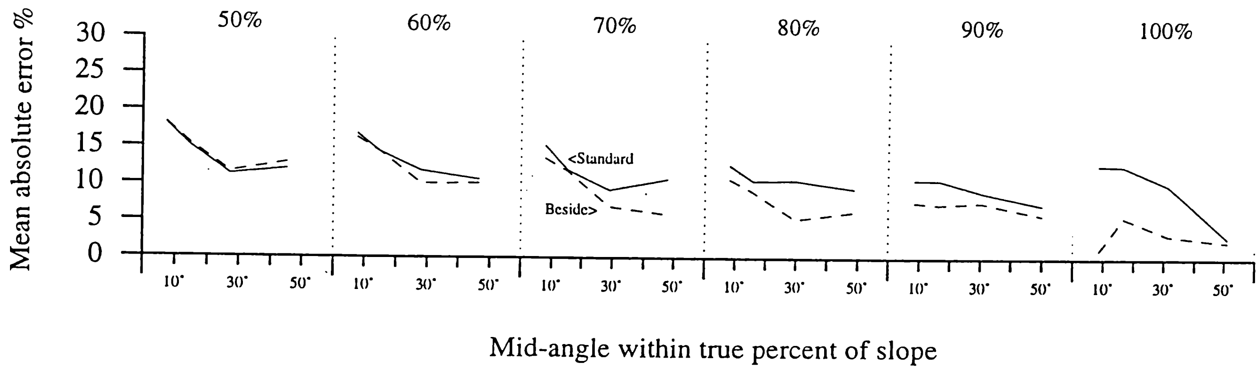


Figure 4-3. The arithmetic mean of the absolute error in estimates, averaged across the 12 replications of the stimulus set and the six subjects, plotted against the mid-angle for each true percent. The solid line is for the Standard condition and the broken line is for the Beside condition.

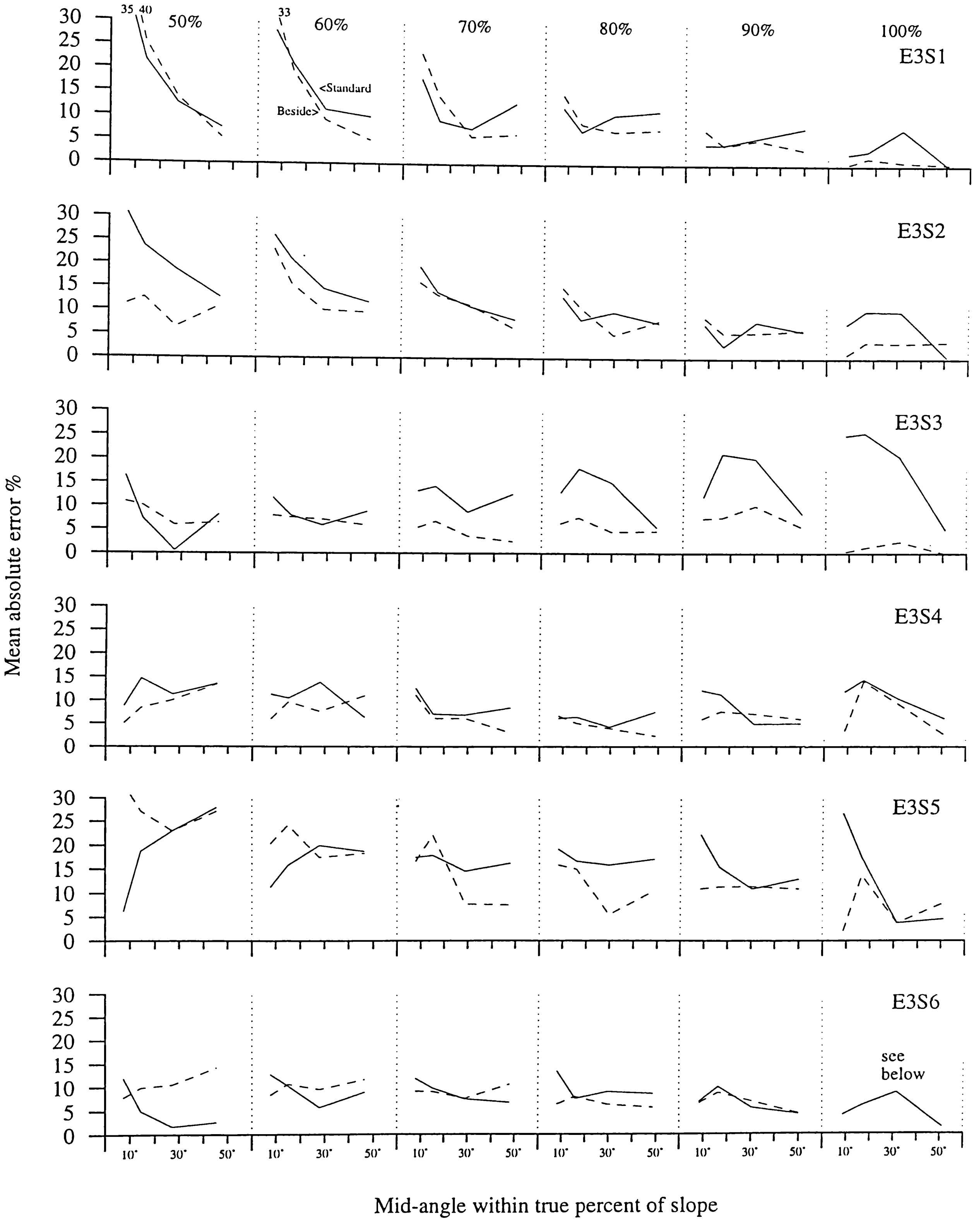


Figure 4-4. The arithmetic mean of the absolute error in estimating slope, across the 12 replications plotted against the mid-angle for each true percent of slope, plotted for each of the six subjects. The solid line is for the Standard condition and the broken line is for the Beside condition. Note - all the data for E3S6 at 100% during the Beside condition has Zero error.

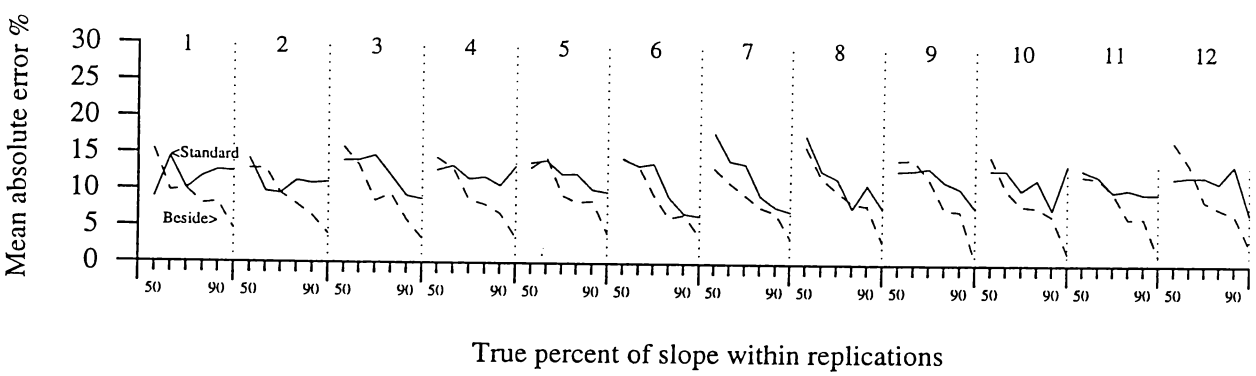


Figure 4-5. The arithmetic mean of the absolute error in estimating slope, across the four mid-angles, at any given true percent, across the six subjects plotted against true percent of slope for each of the 12 replications. The solid line is for the the Standard condition the broken line is for the Beside condition.

Figure 4-6 is similar to Figure 4-5 except that the mean data for each subject are plotted separately. For all subjects the data show no systematic variation as a function of the replication number. It shows that for Subjects E3S1, E3S2 and E3S5 the mean-absolute error changes as a function of true percent in a similar way to that seen for the group results. For Subjects E3S4 and E3S6, this mean-absolute error generally stays around 10 and there is no clear pattern as a function of changes in the true percent or replication. This mean-absolute error for Subject E3S3 during the "Beside" condition shows no clear pattern as a function of changes in the true percent or the replication number, being constant below 5%. For this subject the mean-absolute error during the "Standard" condition increases as the true percent increases but does not vary systematically with replication number.

Discussion

The purpose of this experiment was to help subjects perform the experimental task more accurately by locating the stimulus lines so that the lines would be easier to compare. The group results suggest that the "Beside" condition led to an overall, though small, improvement in accuracy on the task and only at the higher true percents. However, this conclusion was not supported by the individual subject's results that showed that subjects, did not perform consistently better on the "Beside" condition and thus performances were generally about the same in both conditions. For Subject E3S3 accuracy was much greater on the "Beside" condition. This greatly improved accuracy for one subject on the "Beside " condition probably accounts for the improvement seen in the group results. These results suggest that the re-arranged stimulus lines in the "Beside" condition did not lead to the subjects behaving differently from the way they did for the "Standard" condition, except for Subject E3S3. Contrary to what was anticipated, the "Beside" condition had no effect on five out of six subjects at all, neither making the results less variable, as was anticipated, nor making the result vary systematically in some other unanticipated way.

The group results from this experiment tend to be shallow, varying little with mid-angle or true percent. These group results are unlike those seen in the previous two experiments or in Cleveland and McGill (1987). The shallow pattern seems to come about because: Subjects E3S4 and E3S6 generate shallow data paths with little variability; Subject E3S5 produces shallow data, though highly variable; and because data from Subjects E3S2 and E3S3, when averaged, would also produce a shallow data path. Like Experiment 2, the group results reflect the averaging across data sets that are not similar but instead have opposite trends.

There was a large amount of between-subject variation in the present data, as there was in Experiment 1, so again only some of the individual results reflect the group results. In this experiment, only Subjects E3S4 and E3S6 produce shallow

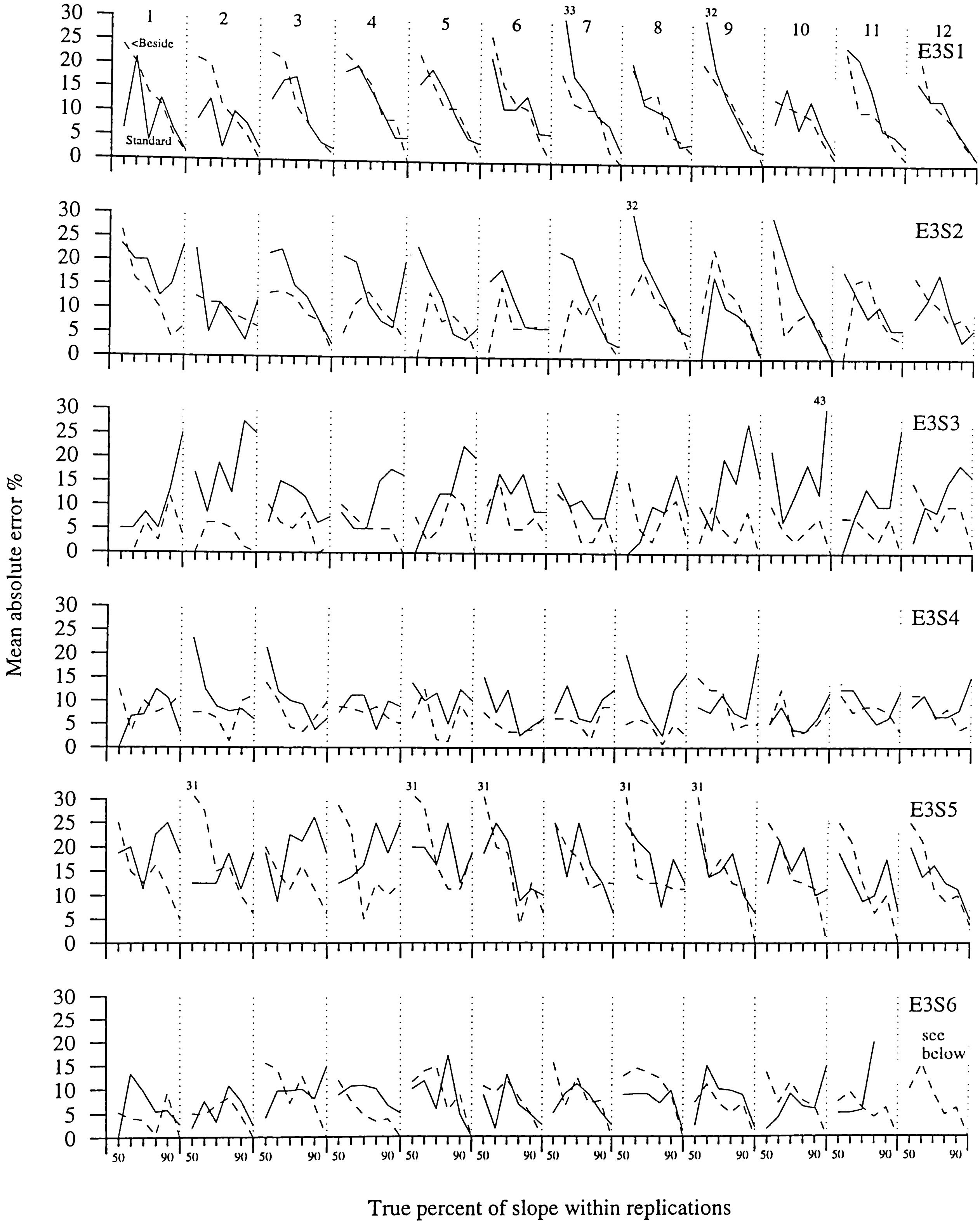


Figure 4-6. The arithmetic means of the absolute errors in line slope estimates, averaged across all the true percents, for each mid-angle, plotted against true percent of slope for each of the 12 replications. The solid line is for the Standard condition and the broken line is for the Rule condition. Note - there is no data for E3S6 in replication 12 during the Standard condition.

data paths that are similar to the averaged group results. No subject in this experiment had data similar to the group results of Experiments 1 and 2, where the error was higher at the largest and smallest true percents.

These experiments were conducted in an attempt to replicating Cleveland and McGill's (1987) findings in which overall error, within true percent, continued to decrease as the mid-angle increased. Two subjects in this present experiment are the first to have data that replicates those of Cleveland and McGill's group data. Unfortunately, success in replicating Cleveland and McGills findings for these two subjects provides no indication of the conditions that led to the reliable replication. This is because for these subjects the same data paths occur for both the "Beside" condition and the "Standard" condition, yet "Standard" condition in Experiment 3 was the same as both Experiment 1 and the "Standard" condition in Experiment 2. Hence, it was not the changed stimuli that accounted for their results. This suggests that the "Standard" condition provides adequate instructions and an adequate stimulus configuration for results from subjects to replicate Cleveland and McGill (1987). It remains unclear then why the other 16 subjects who also went through the "Standard" trials did not produce results like Cleveland and McGill's. This result also suggests that the two experimental variations employed in Experiments 1 and 2 do nothing to improve the chances of replicating Cleveland and McGill (1987).

Chapter 5 - Discussion of Experiments 1-3

Combining the results from Experiments 1 to 3

The first three experiments were conducted in an attempt to replicate Cleveland and McGill's (1987) finding that line slope discrimination improved as the mid-angle of a line pair approached 45°. This mid-angle is also the point at which the orientation resolution of two lines is at its maximum. In Experiment 1, the group results were similar to those of Cleveland and McGill (1987), except that the amount of error at 100 true percent was somewhat greater than expected. In Experiments 2 and 3 attempts were made to try to increase accuracy on the experimental task, but these attempts were ineffective. In Experiment 2, subjects initially undertook the same task as in Experiment 1 and then repeated it with the aid of more detailed instructions. These instructions were intended to help subjects attend to certain features of the stimuli on the assumption that this would lead to improved performance. Group results in both conditions of Experiment 2 were similar to each other and similar to those seen in Experiment 1. In Experiment 3, subjects again did a very similar task to that in Experiment 1 and then repeated it with the stimulus lines configured next to each other instead of being in opposing corners of the screen. Group results for both conditions were again similar to each other but were more like those of Cleveland and McGill than those for the previous two experiments. However, the amount of error at low true percents was smaller than that seen in the results of Cleveland and McGill. Thus, the group results from these three experiments did not fully replicate the results of Cleveland and McGill (1987).

Individual subject results

Examination of results from all the individual subjects, across all three experiments, showed that the group results, generally, did not reflect the results seen for the individual subjects. Further, the individual subjects' results showed that subjects were generally just as accurate in both the "Standard" condition and the conditions designed to improve performance. Within each of the three experiments the results from individual subjects were often markedly different from each other. Out of the eighteen subjects in the three experiments only two of them, E3S1 and E3S2, had results that looked like the group results of Cleveland and McGill. The results of these two subjects were the same, during both the "Standard" condition and the "Beside" condition, and were similar to the group results of Cleveland and McGill (1987). That there were two successful replications of the group results of Cleveland and McGill (1987) suggests that the procedure used in Experiment 1 and in the "Standard" Condition of Experiments 2 and 3 was sufficient to allow replication of their results. The question that remains is why only two of the eighteen subjects had results that replicated the group results of Cleveland and McGill (1987), given that all subjects experienced the same experimental procedure, the "Standard" condition.

Results in common over three experiments

In an attempt to explain why only two subjects responded as predicted, the data from all eighteen subjects from the "Standard", "Beside" and "Rule" conditions were treated as a single group. The individual results were then inspected for any common patterns across all the conditions. Inspection of Figures 2-4, 3-2 and 4-4 suggests that there are four general patterns in the data. These patterns are seen in the way the amount of error varies across true percents. The following are descriptions of the response patterns, although the results from the subjects were more variable than the exact values used to describe the patterns suggest here.

1. In the first pattern the largest absolute error was at 50 true percent and the lowest at 100 true percent. This is seen for Subjects E3S1, E3S2 and the group results of Cleveland and McGill (1987).
2. In the second pattern the largest absolute error was at 100 true percent and the lowest at 50 true percent. This is seen for Subjects E1S3 and for E1S4 and E3S3 in the "Standard" condition.
3. In the third pattern, the largest absolute error occurs equally at 50 true percent and 100 true percent with the lowest error at 75 true percent. This is seen for Subjects E1S1, E1S5, E1S6 and E2S1 to E2S5.
4. In the fourth pattern the same absolute error occurs over all true percents. This is seen in the data of Subjects E1S2, E2S6 and E3S4 to E3S6.

Because these four patterns occur across experiments and conditions they cannot be a function of the different experimental procedures. They are generally consistent for any individual in both experimental conditions of Experiments 2 and 3. What ever the response pattern, it occurred consistently throughout the session. For example, for Subject E3S1 (Figure 4-4) the error decreased as the true percent increased and this pattern remained consistent across all 12 replications. No explanation of these four patterns of results is readily apparent from the data analyses presented so far, nor is any theoretical explanation likely to account for the occurrence of four very different response patterns under the same experimental conditions. Given the lack of any obvious explanation it was decided that further data analysis was necessary.

Response bias

Response bias analysis

One feature of the data that was investigated was the presence of response bias. Response bias would be shown by the subjects using one or more response options more often than would be expected. For these experiments, a possible response pattern would be for subjects to use all the responses from 50 to 100 equally. A bias would be shown if they had a tendency to use a limited range of these responses. Analysis of the data examining response bias suggested an explanation of the four different patterns of results outlined previously. For this analysis the distribution of the different responses the subjects made were examined, for example, how often they reported 50% or 100% irrespective of the experimental condition, or accuracy. To do this analysis all the subjects' judged true percents were rounded to the nearest multiple of ten. This rounding made little difference to the data as most subjects had responded using multiples of ten. Counts of the number of times a subject used each of the response "bins", 50 to 100, were made separately for each of the four reference line angles. These were then converted to percentages of the responses at a given reference line angle that fell in a particular response bin. The analysis was done separately for each subject and where there were two experimental conditions the analysis was done separately for each condition.

Figure 5-1 presents examples of these results, in this case for each of the six subjects in Experiment 1. It should be noted that these percentages do not give any indication of whether the responses were correct or not. While it is possible to present these data for both correct and incorrect responses, this adds nothing to the understanding of the data since the number of correct trials is directly related to the number of response in that response bin. This is because in unused response bins there are no correct responses and in heavily used response bins, those towards which the behaviour is biased, the probability of responding correctly to stimuli that have the same true percent increases as the response bias becomes high. The horizontal dotted line in Figure 5-1 indicates the percentage of responses we would expect in each response bin if a subject had used each bin equally often and showed no consistent bias but responded randomly within and across bins.

Inspection of Figure 5-1 shows that data from Subject E1S2 approximated the no-bias line at all four reference line angles with a small deviation from it when the reference line was at 55°. All other subjects allocated the majority of their responses to a small number of response bins when the reference line was 10°, 19° and 35°. Where responding was biased towards a particular bin the percentage of responses allocated to that response bin tended to peak at that value and reduce symmetrically around that value, as seen in the data from Subject E1S5 at reference line 35°.

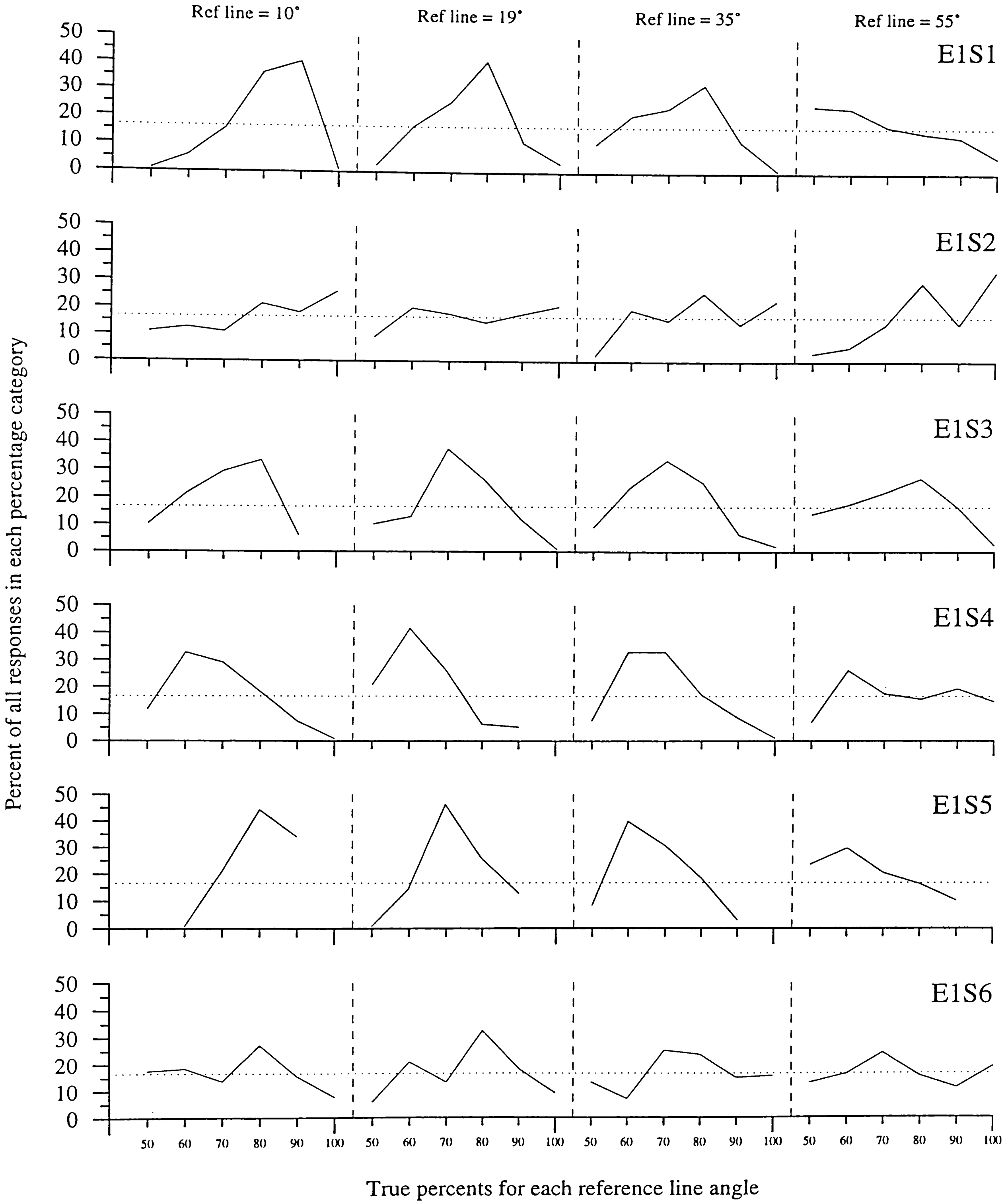


Figure 5-1. The percentage of all responses that are in each response bin of judged true percent split by the four reference line angles.

Different subjects showed bias towards different response bins, for example, E1S4 tended to use 60% and E1S1 tended to use 80%. When the reference line was 55° all subjects, except perhaps E1S2, used a greater number of bins and as a result their data paths were closer to the no-bias line.

Figure 5-2 is the same as Figure 5-1, except the data path with the solid line shows the data from the "Standard" condition and the broken line those from the "Rule" condition in Experiment 2. The results here are very similar to those seen for five of the subjects in Experiment 1, showing response bias when the reference line was between 10° and 35°. As in the previous results, different subjects showed bias towards different response bins. When the reference line was 55° subjects again showed less bias. As seen for the previous data analyses, the data here are similar for both experimental conditions.

Figure 5-3 is similar to Figure 5-2, except the solid line shows the data from the "Standard" condition and the broken line those from the "Beside" condition. Subjects E3S1 and E3S2 showed stronger biases towards the higher percentages than seen for any other subject, these biases are extreme at 10° and then reduce as the reference-line angle increases, until at 55° the data path is close to the no-bias line. Their data are similar across the two experimental conditions. Subject E3S6 showed little response bias, except for a slight tendency to respond more in the 100% bin, irrespective of reference-line angle or experimental condition. The remaining three subjects' data were variable and did not vary consistently with reference line angle, and they showed no strong overall bias towards any particular response bin. For two of these subjects, E3S3 and E3S4, the data from the "Standard" and "Beside" conditions were similar. The results from Subject E3S5 were very disorderly.

It appears that at lower mid-angles the responding of many of the subjects was biased, in that it fell into a small number of response bins. At higher mid-angles these biases were reduced. Why this should be so is not clear, presumably it is not a function of mid-angle because the responding of some subjects was not biased even at the lowest mid-angles. Both the bin a subject showed the most bias towards and the strength of that bias varied across subjects. Across subjects there are four distinct response bias patterns: no bias, bias to 50%, bias to 100% and bias to 75-80%. There were four patterns of mid-angle performance and four patterns of response bias. It is possible that these two measures are related. To examine the relation between the mid-angle measure, mean absolute error, and the response bias, a simulation of extreme response bias was done so that the effect of bias on the mean absolute error could be observed.

Response bias simulation

The simulation was done by preparing three data files in which the subjects' estimates, or judged percent, were replaced with values of either 50, 75 or 100.

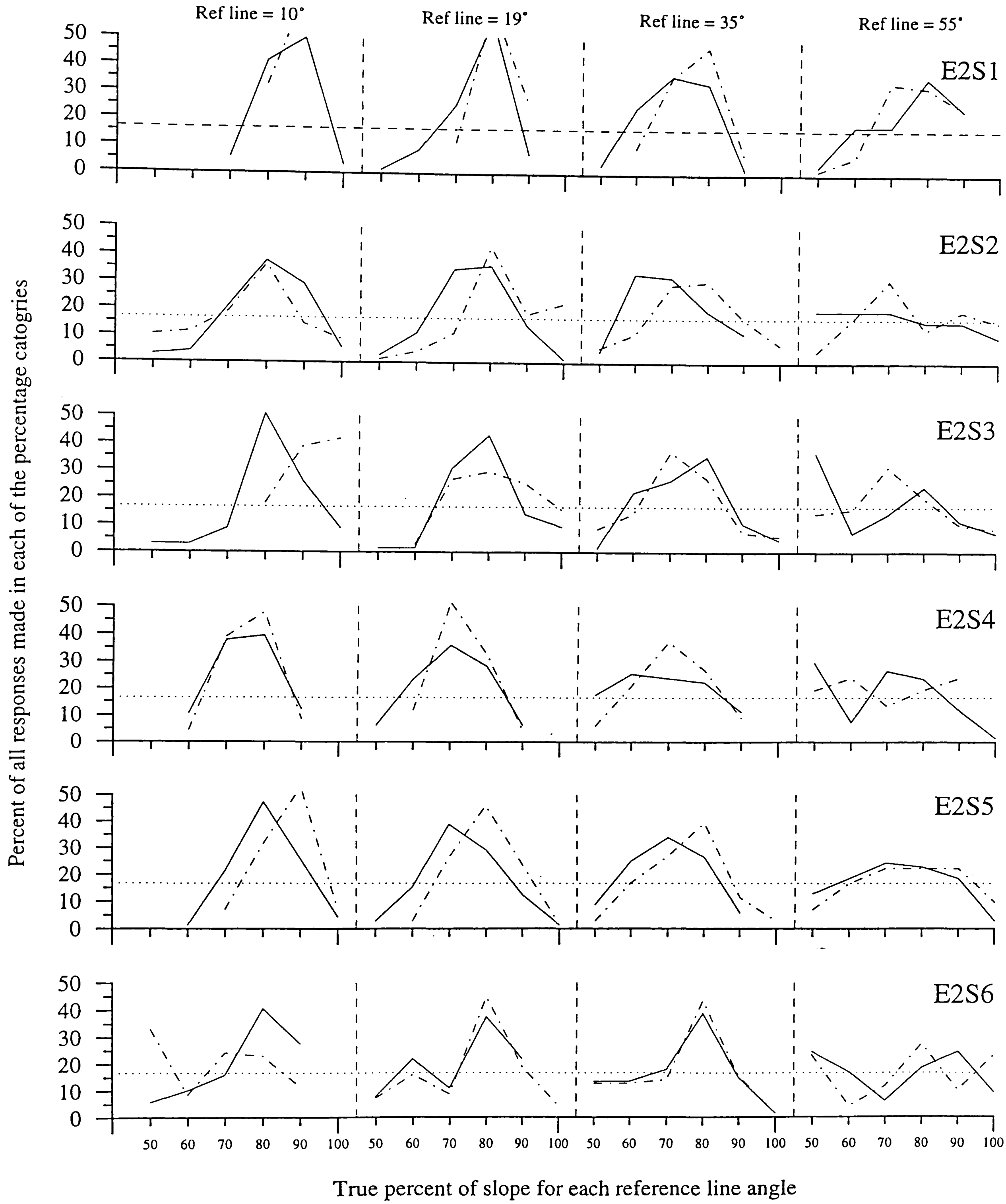


Figure 5-2. The percentage of all responses that are in each response bin of judged true percent, split by the four reference line angles. The solid line is the Standard condition and the broken line is the Rule condition.

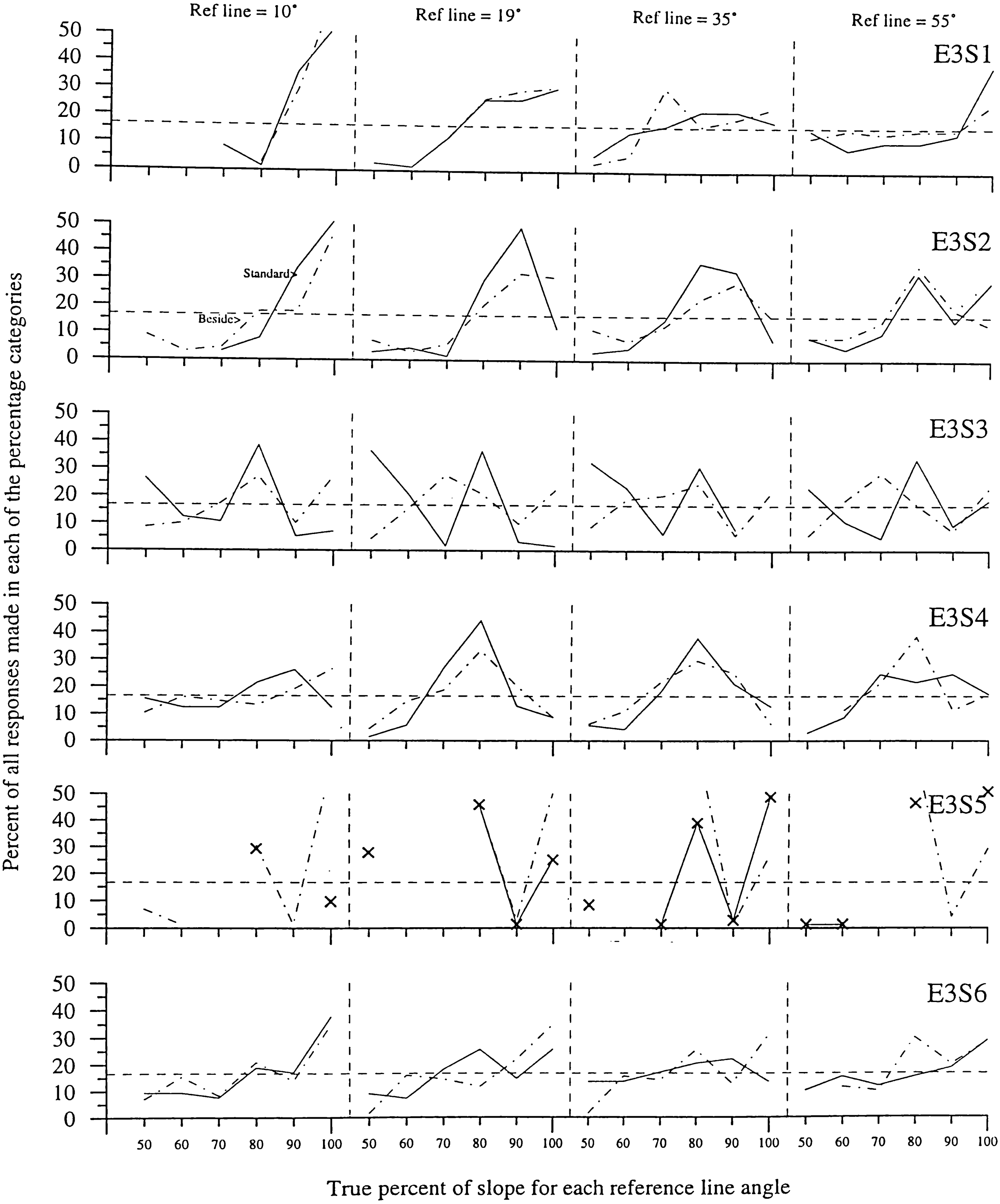


Figure 5-3. The percentage of all responses that are in each response bin of judged true percent, split by the four reference line angles. The solid line shows the Standard condition and the broken line shows the Beside condition. In the Standard condition for E3S5 too many values are out of range thus markers show where in rage data points are avaliable.

These data files were then analysed and plotted in the same way as all previous mid-angle analyses in this thesis.

The results of this simulation appear in Figure 5-4. When all the simulated responses were 100%, performance was perfect at 100 true percent and the mean absolute error increased smoothly as the true percent decreased, until at 50 true percent the mean absolute error was 50%. When all simulated responses were 50%, the mean absolute error increased smoothly from 0% at 50 true percent to 50% at 100 true percent. When the simulated responses were all 75%, the mean absolute error was 0% at 75 true percent, the error then increased smoothly and at equal rates as the true percent both increased towards 100 and decreased towards 50. At both 50 and 100 true percent, the mean error was 25%.

Response bias prediction from simulation

The results of this simulation may seem counter-intuitive. If all the responses were 100% then, obviously, when the stimulus was at 100 true percent the simulated response would always be correct. However, if all the responses were 100 true percent then all the stimuli presented that were less than 100 true percent would never be correctly identified, and thus all of these responses would be equally wrong. However, as this simulation shows, the amount of error for stimuli that are not 100 true percent is variable and progressively increases the greater the difference between the simulated response and the true percent of the stimulus. This variation is introduced as a function of the measure even though there was no variation in the simulated response.

These simulated results are an inevitable product of how the mean absolute error (MAE) is calculated. It is calculated from the following:

$$MAE = \frac{\sum |true\ percent - judged\ percent|}{12}$$

The mean absolute error is the arithmetic average of the absolute difference between the true percent of the presented stimuli and the judged percent from the subject. In this simulation the mean absolute error at any true percent is simply the difference between the actual true percent and the estimated true percent. Given this, we can calculate all the simulated mean absolute errors. For example, in the simulation when all responses are 100% and when the stimulus is 50 true percent the difference will be 50%. With a simulated response of 100% when the true percent of the stimulus is 75 then the difference, which is 25%, is the mean absolute error. In both of these cases the simulated responses are wrong but the degree of error varies. It seems that the measure of error used by Cleveland and McGill (1987), and used here, confounds the ability of a subject to identify a stimulus correctly with the magnitude of their bias towards a particular response.

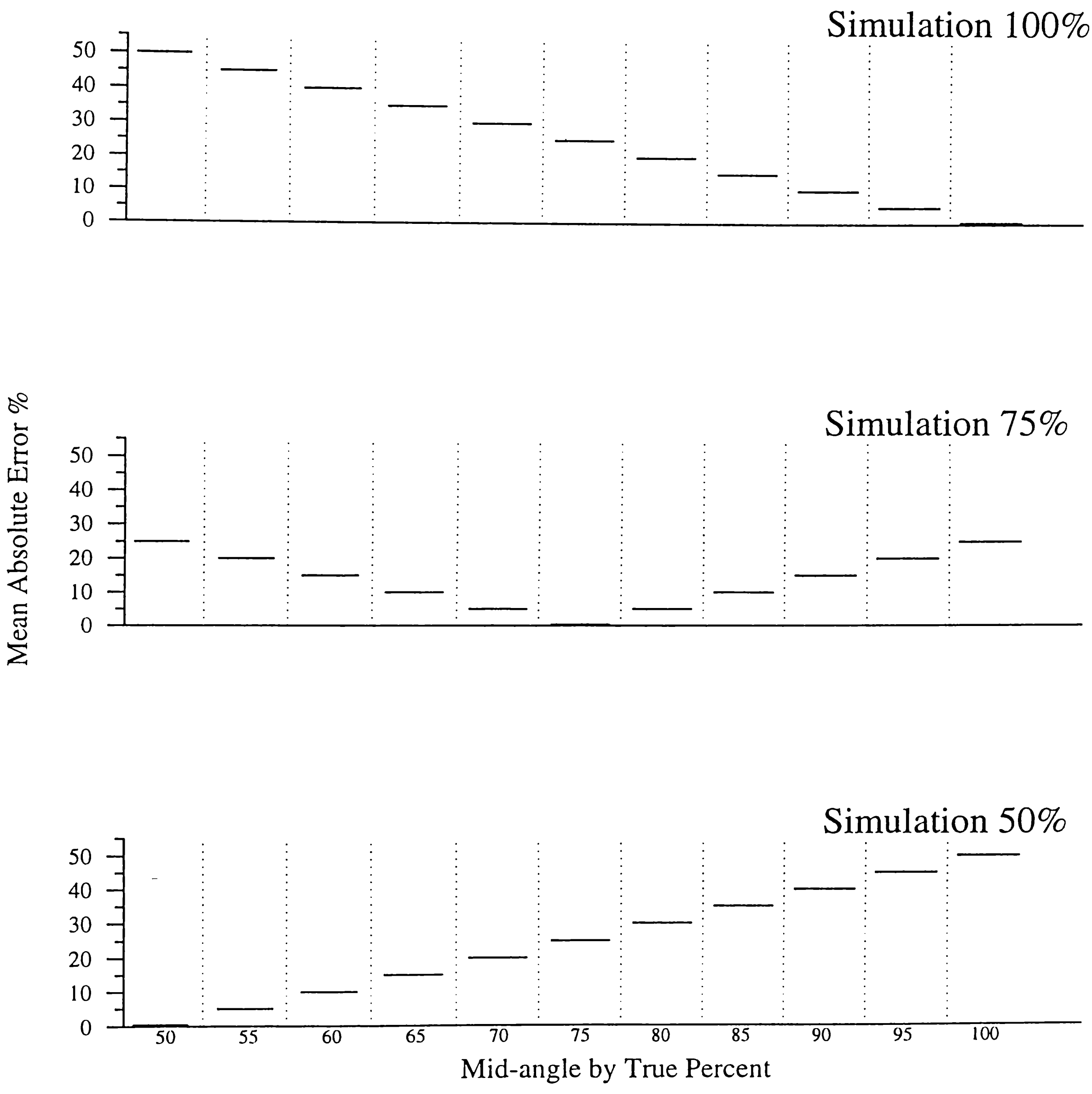


Figure 5-4. The arithmetic mean of the absolute error in estimates, plotted against the mid-angle, for each true percent. Results from simulated responding at 100%, 75% and 50% exclusively.

Response bias and mid-angle error

Response bias predictions of mid-angle error

It was noted above that the changes in mean absolute error fall loosely into four categories. The results of the simulation match the first three of those four categories: the results from the 100% simulation are similar to those in Category 1, those from the 75% simulation are similar to those in Category 3, and those from the 50% simulation are similar to those in Category 2. The data classified as Category 4, in which there was no variation in mean absolute error with true percent, reflect the absence of a strong response bias. It is possible, therefore, to predict in which result category a subject's mid-angle analysis falls into by finding out which response bin, if any, towards which they were biased. The opposite would be also true, that response bias could be predicted from the mid-angle analysis category. These predictions are possible because the response to which a subject is biased will have the least amount of error in the mid-angle analysis. For example, Subjects E3S1 and E3S2 both show strong biases towards responding 90% and 100%, as seen in Figure 5-3. From this it is possible to predict that the mid-angle analysis will show low error at 100 true percent and high error at 50 true percent for both of these subjects, and this is confirmed in Figure 4-4. Predicting from mid-angle to bias can be illustrated with the data from Subject E1S4. For this subject the mid-angle analysis shows lowest error at 55 to 60 true percent and the highest at 100 true percent. The prediction from this is that the response bias will be towards 55% and 60%, and this is confirmed in Figure 2-4. It seems that for many subjects the data are not a function of differing discriminative abilities but a function of different response biases.

For those subjects who showed no particular bias in their responding, that is, they responded evenly across all bins, the mid-angle analysis should show that no true percent had an especially high absolute error. To put this another way, when there is no response bias there is no variation in the absolute error across true percents. This can be seen in the data of Subject E1S2, who showed little bias, in Figure 5-1 by the absence of high levels of error at any particular true percent.

Response bias variation, predictions of mid-angle error

Although it appears that variations in response bias can contribute to the within-subject variation in absolute error across true percents, such variations do not immediately appear to account for the reason that error often still decreases within a true percent as a function of increasing mid-angle. Closer examination, however, suggests that response bias can also account for the within-true percent variation as a function of mid-angle. For all subjects in Experiments 2 and 3, the first data points of each of the separate data paths going from 50 to 100 true percent in the mid-angle analysis corresponds to the data points in the data path from 50 to 100 true percent in the response bias graphs at 10°. The second data points of each of the mid-angle data paths correspond to the 19° bias plot, the third points in the mid-angle data paths correspond to the 35° bias graphs and the last data points in each mid-angle data path to the 55° bias plot. For Experiment 1 there were eleven not seven true percents used, as a result the first data points from the first two data paths from the mid-angle analysis correspond to the first data point in the 10° data path in the bias analysis, etc. There is a direct correspondence between the data point in the bias analysis and the mid-angle analysis.

As noted earlier, when there is a clear response bias towards one response bin then there will be large variations in the degree of error across true percent, i.e., the data in Figures 5-1 to 5-3 would appear clearly 'peaked'. For example, the data from Subject E2S5 for bias at 10° in the "Standard" condition, shows bias peaking at 80 true percent and decreasing rapidly towards 60 and 100 true percent. It follows that the lowest error will be at 80 true percent and the highest at both 50 and 100 true percent. From this it is possible to predict that the first data point in each of the data paths of the mid-angle analysis (Figure 3-2) will follow the reverse pattern to the bias, the error should decrease, as true percent increases, to a minimum at 80 true percent, then increasing again as the true percent increases to 100. This is confirmed in Figure 3-2.

The response bias analysis in Figures 5-1 to 5-3 showed that the extent of any biases reduced with increasing mid-angle. It has now been shown that when the response bias was small there was little variation in the absolute error. Again, the data from Subject E2S5 for bias at 55°, in the "Standard" condition, show a comparatively flat data path with a peak at 70 true percent and decreasing gradually towards both 50 and 90 true percent. From these data it is possible to predict that the last data point in each of the data paths, for each of the true percent graphs, in the mid-angle analysis, should follow a particular pattern. This pattern was generally low but decreased slightly as the true percent increased to 80 and then increased again as true percent increases to 100, and again this is confirmed in Figure 3-2. The same process of prediction can be undertaken from the bias at 19° and 35° for the centre two data points of each data path in the mid-angle analysis. Thus, it can be seen that the changes in bias as a function of reference-line angle are directly related to the shapes of the functions relating mean absolute error to mid-angle by true percent.

It is possible to argue that reduced bias with increased mid-angle indicates improved discrimination performance. In the absence of extraneous sources of control, a bias in favour of a particular response, for example 50%, suggests that the subject is unable to differentiate the stimulus that would be correctly associated with that response, the 50 true percent stimuli, from any other stimulus. If this were the case, then decreasing bias might indicate that the subject is better able to discriminate and thus is less inclined to use the one response for most stimuli. Although it can be argued that the reducing response bias with increasing mid-angle shows some control by mid-angle on discrimination performance, it is not the case that bias itself can be used as measure of accuracy on this experimental task. The only conclusion that we can draw with any confidence about the effect of mid-angles on the performance of subjects is that bias reduces as the mid-angle increases. This is because bias confounds with other possible indicators of performance accuracy, but is not confounded by itself.

In summary, three experiments produced group results that partially replicate the group results of Cleveland and McGill (1987), but the results from only two individual subjects replicated Cleveland and McGill group results. Instead, four general patterns of error were noted. An analysis of how subjects distributed their responding across the possible responses showed that these four general patterns were a function of the different ways that responding was biased and were not controlled by the mid-angle of the stimulus. Further, it was argued that the patterns within true percent for the mid-angle analysis for individual subjects can be predicted by the amount and type of response bias.

Response bias and Cleveland and McGills' mid-angle error

The bias analyses for the subjects who had mid-angle results like those of Cleveland and McGill (1987), Subjects E3S1 and E3S2 in Figure 5-3, show that both had an extreme response bias towards 100% at 10°. These biases then reduced smoothly as the mid-angle increased, until the data paths showed almost no bias at 55°. As argued above, the high levels of error at 50 true percent and the low levels of error at 100 true percent in the mid-angle analysis are a function of a response bias towards 100%. These data suggest that Cleveland and McGill's (1987) results are also a product of a strong response bias towards 100% across their 16 subjects, not a product of control by mid-angle as they originally suggested. Against this it might be argued that their results are from 16 subjects and it is unlikely that all 16 would have had the same response bias. In fact, that all had the same bias seems unlikely given the individual results obtained in the present experiments.

Cleveland and McGill (1987), as reported previously, present group averages. It may be that the group data reflect the fact that some of the 16 subjects may have shown strong biases towards 100%, and this is reflected in the group data. Not every subject need have shown a response bias that favours 100%, so long as across the 16 subjects 100% was more favoured on average. As an example of such an averaging effect, when the data from the 18 subjects from the present experiments are averaged for a mid-angle analysis in the way that Cleveland and McGill (1987) did, the result is shown in Figure 5-5. This result is comparable with that of Cleveland and McGill

(1987) shown as the dotted line in Figure 5-5 and is very similar to their data. The main exception is an increase in the error at 100 true percent, most of which can be accounted for by extreme data of Subject E1S4 which has a distorting effect on the mean. The current result is a product of a range of different response biases across the 18 subjects, as was seen earlier. Therefore, it might well be the case that the group results that Cleveland and McGill (1987) present also are the product of a range of biases with an average bias towards 100% and not the result of all 16 subjects showing the same bias.

The present data suggest that a strong argument can be made that Cleveland and McGill's (1987) results may be confounded by response bias. However, as a result of the lack of information about their experiment it is impossible to prove that this is the case. What is indisputable is that Cleveland and McGill's (1987) data analysis relies on a measure of error that subtracts the actual true percent of the stimuli from the judged true percent and confounds the measure of the accuracy of a subject's performance on this discrimination with the magnitude of the stimulus. This is potentially a problem that makes their results uninformative, even if there was only the smallest amount of response bias. As noted in the Introduction, Cleveland and McGill's (1987) own explanation of their results is inconsistent with their data, and if the argument that bias is responsible for the group results is rejected, then there would be no obvious explanation for their results.

In the Introduction it was noted that Cleveland and McGill (1987) had done separate data analyses for each true percent of the stimuli, yet their theory did not require this analysis. Perhaps response bias can be used to explain why they needed to do the analysis by true percent. They state that the analysis was done by true percent because in previous research Cleveland (1984) had found a dependency on true percent in the results. The argument can now be made that this dependency on true percent in their data comes about both because of the presence of a response bias and because of the use of a measure that confounds accuracy with the stimulus magnitude. For example, the tendency to much lower error at 80 true percent in the mid-angle results is probably a function of response bias towards 80%, not to any special property of the 80 true percent stimuli that lead to higher accuracy. Cleveland and McGill (1987) claim that this dependency on true percent was discovered by Cleveland (1984), but that paper shows that Cleveland used the same measure of error in 1984 as was used in 1987. Given this, there is every possibility that the results in the 1984 study were also confounded by the problems noted for this measure of error and by the presence of response bias. The data reported here suggest that the dependency they reported is not a function of the true percent of the stimulus but an artifact of response bias.

Conclusion

As a result of confounding the way error is measured with a tendency for subjects' responding to be biased at low mid-angles, it is not possible to conclude, with confidence, anything about the theory of orientation resolution proposed by Cleveland and McGill (1987). As a result, their theory has still not been successfully examined experimentally, even though a number of attempts have been made here and by Cleveland and McGill (1987).

Nothing in the results examined here suggests that there is anything wrong with the theory that to maximise orientation resolution, the aspect ratio of a graph should be varied so that, on average, the line segments are positioned at 45°. With maximum orientation resolution, people will be most sensitive to variations in the data (Cleveland & McGill, 1987). Experiments here showed that bias reduces with increasing mid-angle and that this might suggest improved discriminative performance as mid-angle increases. This result may provide some indirect evidence to support their theory. Nothing found here refutes the arguments put forward in the Introduction that there is a need for a reliable way of selecting an aspect ratio. Therefore, it still seems important to attempt to investigate the orientation resolution theory of Cleveland and McGill (1987) empirically. To do this requires an approach that deals with the problem of bias and avoids any confounding accuracy with the magnitude of the stimulus. It also requires a measure which still addresses the question of line slope sensitivity.

Chapter 6 - Introduction to Experiments 4-7

It has been argued here that neither Cleveland and McGill's experiment nor those reported above tested the theory that line slope discrimination is controlled by orientation resolution. The main reason for this was because the measure of error used, the "mean absolute error", is confounded with response bias, making the results of these experiments uninterpretable. In order to test Cleveland and McGill's theory empirically, it is necessary to find a measure of performance which enables accuracy on that task to be assessed independently of any responses bias, or the tendency to favour one response over another. Nevin (1991) noted that there are two recognised methods that enable accuracy and bias to be differentiated, signal-detection theory (Green & Swets, 1966) and the more recent behavioural-detection theory (McCarthy & Davison, 1984; McCarthy, 1991).

Signal-detection theory

Green and Swets (1966) presented the theory of signal-detection. This theory was based on earlier work in signal processing theory. A signal-detection task usually involves two types of trials, one in which a signal (or one of the two stimuli, Stimulus 1) is presented and the other in which no signal (or the other of the two stimuli, Stimulus 2) is presented. The subject generally is required to report either if the signal was present or if it was absent (or which of the two stimuli was presented). On each trial there are two possible outcomes depending on the type of trial. These are shown in Figure 6-1. On Signal (Stimulus 1) trials the subject may report a signal and be correct (W, termed a correct detection) or they may report no signal and thus be incorrect (X, termed a miss). On No-signal (Stimulus 2) trials the subject can report no signal and so be correct (Z, termed a correct rejection) or report a signal and be incorrect (Y, termed a false alarm).

Signal-detection analysis provides an index of the sensitivity of responding to the signal (or the difference between the stimuli) termed d' (d prime). The signal-detection analysis was developed to remove the confounding effects of motivational and attentional deficits that were assumed to lead to "false alarms", that is, subjects reporting the stimulus present when it was actually absent. Blough and Blough (1977) noted that approaches to measuring discrimination, prior to signal-detection, assumed that the stimuli themselves and the observer's sensory system were the only sources of control over behaviour. Signal-detection theory provided a critical development by also recognising the control over the subjects' responses by variables such as payoffs, instructions and stimulus presentation probability. These variables are referred to as biasers because they are said to explain why one response is favoured over another, irrespective of the stimulus arrangement

		Responses	
		Left	Right
Stimuli	S_1	W Hit	X Miss
	S_2	Y False Alarm	Z Correct Rejection

Figure 6-1 The four possible response conditions of a Yes - No Signal-detection task. Left (reporting present or same) and Right (reporting absent or different) are the two possible response alternatives. S_1 and S_2 denote the two discriminative stimuli. W, X, Y and Z denote the number of events (responses and reinforcers) in each cell. (Davison & McCarthy, 1988)

(Nevin, 1991). The signal-detection technique has been used extensively with animal subjects (Blough & Blough, 1977) and in over 1000 experiments using human subjects (Swets, 1992).

Although extensively used, signal-detection analysis does have limitations. Blough and Blough (1977) argued that results from both animal and human studies often do not conform with the strict predictions of the signal-detection model and, as a result of this, analysis methods have been developed to accommodate these departures from the model. However, very few attempts have been made to explain these departures. Blough and Blough (1977) suggested that this lack of explanation is because the departures from the model challenge one of the fundamental assumptions of the signal-detection model: that the sensitivity to the presence of the signal (or to the stimulus difference) is independent of the effects of the non-sensory or biasing variables. They pointed out that the truth of this assumption had not been established. More recently, McCarthy (1991) also suggested that the correctness of this assumption can be readily challenged given the data, especially from experiments in which a particular response of the organism is explicitly reinforced or biased (e.g., McCarthy & Davison, 1984).

Even if it is accepted that the results from Signal-detection are still useful, irrespective of the concern outlined above, there are still problems with using signal-detection as a technique. This is because to use signal-detection properly a lot of data must be collected, typically this will require many thousands of experimental trials (Blough & Blough, 1977). To get this number of trials with human subjects takes a great deal of time and may lead to compliance problems. Both Blough and Blough (1977) and Green and Swets (1966) have suggested that, in practise, if there is no notable bias in the responses of subjects then percentage correct maybe just as effective as d' as a measure of sensitivity. Irwin and Terman (1970) went further and argued that proportion correct is reasonably insensitive to the effects of bias. Blough and Blough (1977) argue that, although proportion correct is contaminated by the bias from non-sensory variables, it remains a viable measure for many studies of sensory performance.

Behavioural-detection theory

A second approach to measuring performance independent of bias was developed separately from the theory of signal-detection. It has its origins in the study of behaviour and its consequences, as suggested by Skinner (1953), and in experimental work with animals, especially in the study of choice behaviour. This behavioural detection model is an extension of the matching law, first described by Herrnstein (1961). The typical matching law experiment involves animals responding under two concurrently available variable-interval (VI) schedules. These schedules maintain behaviour by aperiodically providing a consequence (termed the reinforcer), typically food. The animals can allocate their behaviour to either schedule and thus, in effect, choose between them. Herrnstein (1961) reported that animals allocated responses to the two schedules proportionally to the relative rate of reinforcer delivery obtained across those two schedules. Herrnstein expressed this matching relation in the following equation:

$$\frac{P_1}{P_1 + P_2} = \frac{R_1}{R_1 + R_2} \tag{3}$$

where P_1 and P_2 refer to responses on the two alternative schedules, and R_1 and R_2 refer to the reinforcements obtained on each schedule. This states that the relative response rate on one alternative of the concurrent schedules equals the relative reinforcement rate provided by that schedule, this has been termed the strict matching law.

It has been found that not all concurrent schedule data are well described by the strict matching law. As a result Baum (1974, 1979) suggested an extension of this termed the generalised matching law, which described the data better. There are two major deviations visible when data are plotted using log ratios rather than proportions correct. One is response bias, or a tendency to perform one response more often than the other, irrespective of the arranged reinforcement contingencies. The other is a result of the fact that organisms do not always allocate responding between the alternatives in exactly the same proportion as the obtained reinforcers. Baum suggested the following equation, which takes into account these two deviations:

$$\log\left(\frac{P_1}{P_2}\right) = a \log\left(\frac{R_1}{R_2}\right) + \log k \tag{4}$$

where P_1, P_2, R_1, and R_2 are as in Equation 3, and a and log k are parameters which measure sensitivity to reinforcement rate and the bias, respectively. When data that conform to strict matching (Equation 3) are plotted as ratios and on log-log co-ordinates they will be well described by a straight line that has a slope of 1.0 and that passes through the origin. However, generally data from concurrent schedules have been found to have a slope, a in Equation 4, less than 1.0 (termed undermatching) suggesting less than perfect sensitivity in behaviour to reinforcement rate changes (Davison & McCarthy, 1988). The y intercept gives the degree of response bias, log k.

Davison and Tustin (1978) proposed the basic formulation of the behavioural-detection model based on the generalised-matching law. This model treats the signal-detection task, shown in Figure 6-1, as two concurrent reinforcement-extinction schedules, each with a distinctive stimulus. In the presence of one stimulus (S_1) responses of one type (P_w, e.g., "yes" or "same") produce reinforcement, while the second type of responses (P_x, e.g., "no" or "different") are not reinforced, they are said to be in extinction. In the presence of the other stimulus (S_2) the second type of responses (P_z,"no" or "different") are reinforced while the first type (P_y,"yes" or "same") are not, they are said to be in extinction (McCarthy, 1991).

If the two stimuli (S_1 and S_2) are indistinguishable to the subject then, according to the generalised-matching law, the distribution of responses across the two alternatives will be determined by the reinforcement available from each alternative. The control of behaviour is then effectively from the two concurrent variable-interval schedules. However, if the stimuli become more discriminable, responding will shift towards one response (P_w) during S_1 and the other response (P_z) during S_2. Davison and Tustin give the following equations which describe behaviour during S_1 and S_2:

Presence of (S_1): $$\log\left(\frac{P_w}{P_x}\right) = a_{r1}\log\left(\frac{R_w}{R_z}\right) + \log c + \log d \qquad (5)$$

Presence of (S_2): $$\log\left(\frac{P_y}{P_z}\right) = a_{r2}\log\left(\frac{R_w}{R_z}\right) + \log c - \log d \qquad (6)$$

where P and R refer to the number of responses made and the number of reinforcers obtained respectively, and w, x, y and z refer to the four cells of the matrix in Figure 6-1, and a_{r1} and a_{r2} are measures of reinforcer sensitivity in the presence of the two stimuli.

Log c is a measure of inherent bias, that is, of any consistent tendency to perform one response rather than the other irrespective of stimulus or reinforcement changes. Log d is a measure of the discriminative performance of the animal. Log d is added in one equation and subtracted in the other as it measures the shift from one response to the other as the stimuli become more discriminable. Provided a_{r1} and a_{r2} are equal, subtracting Equation 6 from Equation 5 results in the following equation for log d, the measure of performance accuracy:

$$\log d = .5\left(\log\left(\frac{P_w}{P_x}\right) - \log\left(\frac{P_y}{P_z}\right)\right) \qquad (7)$$

this shows that log d does not involve log c, inherent bias. Therefore, log d is a bias-free measure of stimulus discriminability as reflected in the animal's performance. Davison and McCarthy (1980) showed that the assumption that a_{r1} equals a_{r2} holds. Adding the two equations (5 and 6) gives the following equation:

$$\log\left(\frac{P_w}{P_x}\right) + \log\left(\frac{P_y}{P_z}\right) = 2a_r \log\left(\frac{R_w}{R_z}\right) + 2\log c \qquad (8)$$

Rearranging this gives:

$$.5 \log\left(\frac{P_w . P_y}{P_x . P_z}\right) = a_r \log\left(\frac{R_w}{R_z}\right) + \log c \qquad (9)$$

The measure on the left-hand side of this equation, termed by McCarthy (1991) log *b*, is the measure of total *response bias*, and is generally considered to be independent of stimulus discriminability (McCarthy, 1991). The expression on the right-hand side of this equation is made up of log *c*, or inherent bias, and the reinforcement bias, or the bias in responding due to changes in the obtained reinforcement ratio. In situations where the reinforcement rates under the two stimuli are equal ($R_w = R_z$) the ratio of these rates will be equal to 1.00 and the logarithmic ratio of these is zero. Substituting zero for the reinforcement term on the right-hand side of Equation 9 makes the response bias, log *b*, equal to only log *c*, inherent bias.

McCarthy (1991) showed that the generalised matching law, and the resulting log *c* and log *d*, could be applied to data from human subjects on detection tasks. She reanalysed data from several studies and showed that log *c* and log *d* provided independent measures of discrimination and bias. In these procedures there are normally no formally arranged reinforcers. She assumed that for these human experiments that the signal presentation ratio (SPR) could be used as the best estimate of the reinforcer ratio in Equation 9. She based this assumption on the fact that McCarthy and Davison (1984) had found that the obtained reinforcer ratio covaried with SPR. So if the signal was presented on half the trials the SPR would be 1.0 and it would be assumed that the reinforcement ratio was also 1.0. Consequently it would be assumed that there was no bias resulting from reinforcers and the inherent bias, log *c*, would provide the measure of response bias. This method of analysis, therefore, seems to provide measures which overcome some of the problems with the measure used by Cleveland and McGill (1987). Log *d* provides a measure of discrimination performance which is independent of response bias and log *c* provides a measure of inherent bias which is independent of discrimination performance.

In conclusion, it appears that there is a problem with the underlying assumption of signal-detection theory and in addition, there are some practical problems with obtaining enough data to make such an analysis possible. In light of this it has been suggested by several writers that proportions correct may well be more viable as a measure of discriminative performance than those provided by a signal-detection analysis, even though it is not a bias-free measure. Further, the behavioural-detection model seems to be able to provide useful independent measures of both discriminative performance and response bias.

Chapter 7 - Experiment 4

It was noted previously that the signal detection approach has a number of problems with it, especially when used with human subjects. These problems have led some writers to suggest that in many circumstances "proportions correct" may be just as informative as a measure of discrimination performance as signal detection, even though it is not a bias-free measure (Blough and Blough, 1977). The behavioural detection approach also has problems when used with human subjects, however these problems can be overcome if reinforcement is treated as equal across two response options when there are no experimentally arranged consequences for responding.

As a result of these arguments presented previously, further experiments here will use both proportions correct, which is not a bias-free measure but requires no assumptions about reinforcement, and log d, which is bias-free but requires assumptions about reinforcement. Because of this choice of discrimination measurements, the experimental task used in further experiments is limited to only two possible responses. Given that Cleveland and McGill's theory argues that orientation resolution, and thus mid-angle, affects line slope discrimination, it seems that the best possible experimental task is a "same - different" task in which subjects report whether the two stimulus lines have different slopes or the same slope. The assumption is that at some orientation resolutions, pairs of stimulus lines that have different slopes will be described by the subject as having the same slope. Such erroneous responses would suggest that for a line pair at that particular orientation resolution, subjects are insensitive to, or unable to discriminate, variations in the slopes of lines.

Having established suitable measures of discrimination and a suitable experimental task to investigate Cleveland and McGill's theory, it seemed important to investigate, parimetrically, all of the predictions of their theory. Cleveland and McGill (1987) assumed that performance was symmetrical around 45°, and that the results would be the same with mid-angles from 45° to 90° as they were from 0° to 45°. This is a reasonable argument if it is accepted that orientation resolution is the only thing that controls line slope discrimination performance. It is not reasonable to assume accuracy would be the greatest at 45° if anything other than orientation resolution affects line slope discrimination, for example the angle of the line segments with respect to the edge of the computer screen may also exert some influence. Instead of relying on this assumption, it was decided to experimentally test line pairs with mid-angles greater than 45°.

In Cleveland and McGill's' experiment, they varied the true percent as well as the mid-angle, yet provided no convincing explanation for doing this. In the discussion of Experiments 1 to 3, the explanation Cleveland and McGill gave was presented by this writer as being an unfortunate result of Cleveland and McGill's misinterpretation of the data from an earlier experiment by Cleveland (1984). If this argument is correct then it is not necessary to vary true percent when trying to vary orientation resolution. This is because true percent can be kept constant and the mid-angle alone can be varied. While for any given line pair the orientation resolution is at its greatest when the mid-angle is 45°, the size of that maximum orientation resolution is dependent on the true percent. When the true percent is very large, almost 100%, the maximum orientation resolution will be very small, less than 1°, when the true percent is small, for example 50%, the maximum orientation resolution will be large, about 20°. Therefore when doing a same-different task with line pairs that have a low true percent and, therefore, a high orientation resolution, there is a very good chance that subjects will get all trials correct. Put differently, the difference between the line slopes is so great that it is obvious at any mid-angle that they are different. This would lead to a "ceiling effect" in which it is not possible to observe increasing accuracy with increasing orientation resolution because there is no further room for improvement. At the other extreme, if the true percent is very high, close to 100%, and thus the maximum orientation resolution is small, then it is possible that subjects would be unable to do the task accurately irrespective of the mid-angle. That is, the difference in slopes of the lines is so small that they always appear the same. This would lead to a "floor effect" in which it is not possible to observe increasing accuracy with increasing orientation resolution because even at the maximum orientation resolution the task can not be done accurately.

Because some true percents make investigating the effects of orientation resolution difficult, there is a need to discover, experimentally, which true percents would allow the investigation of the effect of orientation resolution. Therefore, there is a need for a true percent which is not substantialy confounded by either "floor" or "ceiling" effects, and at which accuracy will vary with mid-angle. In the following experiment, therefore, subjects were presented with the stimulus set used by Cleveland and McGill (1987) and in Experiment 1, and asked to judge whether the slopes of the lines were the same or different. It was predicted that as the true percent increases from 50% to 95% accuracy would decline, and that as the angle of the reference line increases from 10° to 55° accuracy at any given true percent would increase. So the goal here was to find the true percent at which accuracy was neither extremely high or low.

Method

Subjects

There were six subjects in this experiment, four females and two males, identified here as E4C1Sl to E4C1S6. All were volunteers from a first year psychology course at the University of Waikato. Each subject received 2% credit towards their first year psychology course for participating in the experiment, and this was not contingent upon their performance.

Apparatus

The experimental hardware and software were the same as those specified in Experiment 1. With the stimuli presented on a high resolution 340 mm black and white monochrome monitor (Wyse WY-700) and the instructions presented on a 310 mm amber monochrome monitor (Forefront VM-1400). Both monitors were driven by a computer (Commodore PC 10-III). Subjects responded by using the numeric keypad of the computer keyboard where the (1) key was relabelled "Same" and the (3) was relabelled "different"

Stimuli

The stimuli were the same as those used in Experiment 1. Each stimulus comprised a pair of positively sloping lines with the reference line positioned in the top left-hand quarter of the screen and the comparison line positioned in the bottom right hand quarter of the screen. Following Cleveland and McGill (1987), four reference lines were used with angles of 55°, 35.5°, 19.7° and 10.1°. Each of the reference lines was paired with each of 11 different comparison lines. The slope of each comparison line was a graded percentage of the slope of the reference line. The percentages ranged from 50% to 100% in steps of 5% giving the 11 comparison lines. Therefore, the stimulus set comprised 44 stimuli made up of 11 comparison lines for each of the four reference lines. All stimuli are described fully in Table 2-1.

Procedure

The experiment was conducted in a small windowless room that was illuminated by ceiling lights. Once the subject was seated at a table in front of the two computer screens, the experimenter asked the subject to listen while the following instructions were read aloud:

> In this experiment the left screen will show a pair of sloping lines for 2.25 seconds. At the end of the 2.25 seconds the pair of sloping lines will disappear and you will be asked to report if you think the two lines had the same slope or different slopes. Put another way, are the lines as steep as each other? You are then asked to type in your response using the number pad on the far right-hand side of the keyboard, using the (1) key that is marked "same" if you thought they were the same or the (3) key marked "different" if you thought they were different. After you have entered your response push the <Enter> key, which is also on the number pad, this will start the display of the next pair of lines. Though some lines may appear longer than others, do not

worry, it is the judgement of slope that we are concerned with here. Some of the lines appearing on the screen may look a little jagged, please try your best to ignore this. It is important to remember that it is very difficult to judge slopes perfectly, this is why we are trying to see how well you can make the judgements. We would like you to give the best estimate you can. Please see the following four pages for examples of line pairs and the correct response. If you have any further questions please ask the experimenter.

The subjects were then shown four example stimuli on four A4 sheets of paper. Figure 2-1 presents these example stimuli on a single page. For each example, the subjects were told whether they were the same or different. Any questions from the subject were then answered by the experimenter, by repeating the relevant part of the instructions. The experiment was then started, and after the successful completion of four trials by the subject the experimenter left the room.

Each trial started with the presentation of a stimulus on the right hand, high resolution, computer screen for 2.25 seconds after which that screen went blank until the next trial. The words "Please enter your response" appeared on the left-hand, amber-monochrome, computer screen as soon as the right-hand screen went blank. The subject then entered the response, using the keyboard-number pad. When the subject had recorded the response, by pressing the <Enter> key, the left hand screen went blank and a new stimulus appeared on the right-hand screen for the next trial. There was no arranged feedback on the accuracy of the subject's response. At the completion of each trial the computer recorded the two dependent measures for that trial: the subject's response, same or different, and the delay between the end of the stimulus presentation and the entering of the response. The computer also recorded the independent variables of actual true percent and mid-angle.

The stimulus used in each trial was selected randomly from the set of 44 stimuli (Table 2-1) without replacement. The complete stimulus set was presented 12 times, each time using a different random sequence. Overall a session comprised 528 trials (44 stimuli $\times$ 12 presentations), and a typical session lasted about one hour during which the experimenter checked progress every quarter of an hour.

At the end of the session subjects were asked informally for their reactions to the experiment and then the purpose and relevance of the experiment was explained and any further questions were answered.

Results

Figure 7-1 presents proportions correct at each of the 11 true percents for each of the four reference line angles, for each of the six subjects. The measure proportion

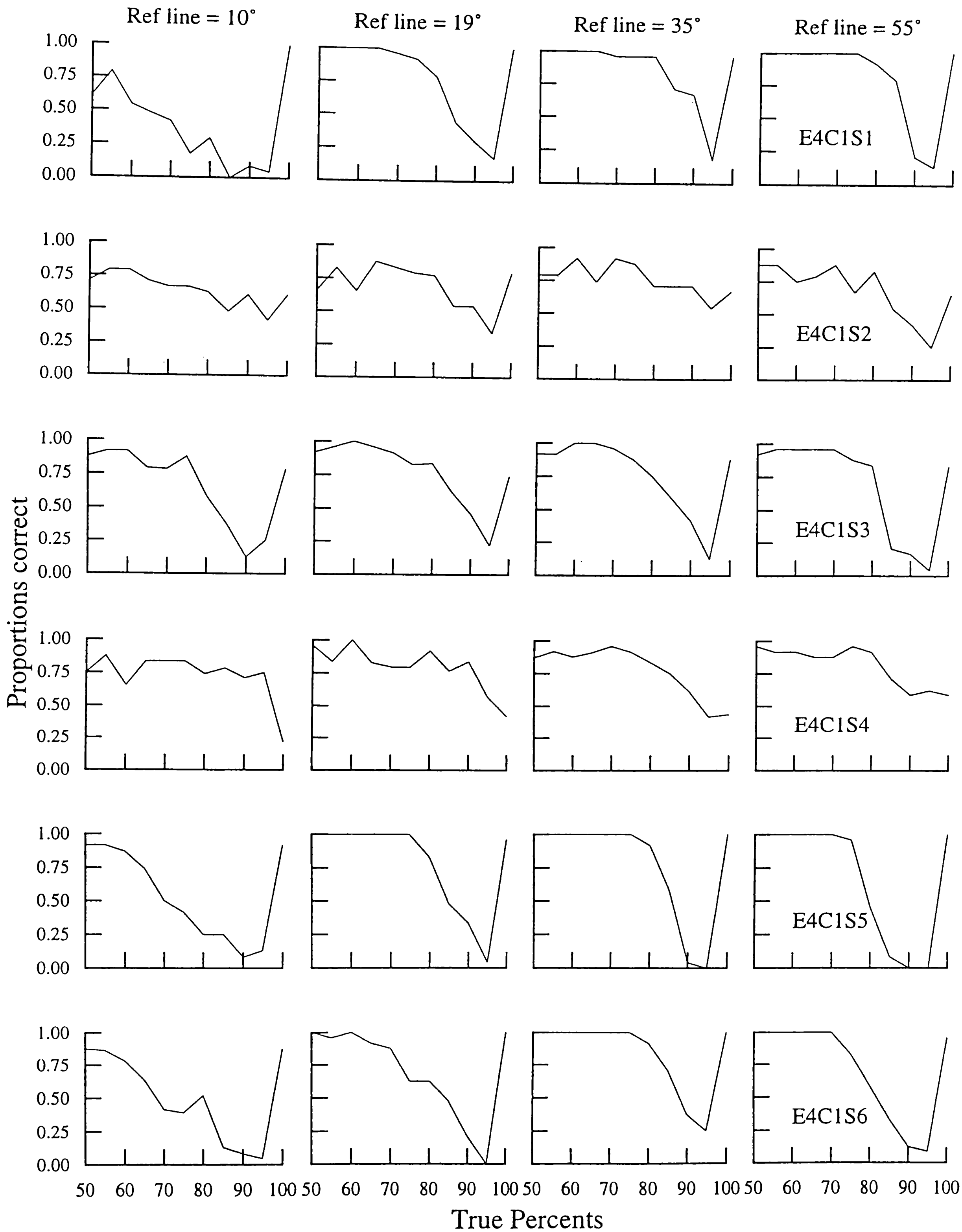


Figure 7-1. The proportion correct at each true percent split by reference line angle, for all six subjects.

correct, is:

$$\frac{\text{number of correct trials}}{\text{number of trials}} \tag{9}$$

For true percents from 50% to 95% the stimulus lines were different, thus a correct response was "Different"; for 100 true percent the stimulus lines had the same slope, thus a correct response was "Same".

Overall, irrespective of the reference line angle, as the slopes of the two lines became similar, that is, as the true percent increased from 50% to 95%, accuracy decreased. When the slopes of the two lines were the same, that is, the true percent was 100%, then accuracy was generally high. Subjects E4C1S2 and E4C1S4 were less accurate than the other subjects overall. As a result, the patterns in responding seen for other subjects are not clear for these two subjects.

Overall, for all subjects, accuracy on "different" trials, those from 50 to 95 true percent, generally improved as the reference line angle increased. On "same" trials, at 100 true percent, accuracy was generally very good irrespective of the reference line angle, with the exception of subjects E4C1S2 and E1C1S4, and the accuracy did not vary with reference line angle. On "Different" trials with low true percents, for example 50% to 60%, accuracy was very good irrespective of reference line angle. When the true percent was between 65% and 95% accuracy generally increased with increasing reference line angle.

Discussion

It was expected that as the true percent increased the accuracy would decrease, and that as the reference line angle increased overall accuracy across all true percents would also increase. Generally these two expectations were supported by the results. No attempt was made to use a log *d* or log *c* measures with these data, as it was not clear what effect having an unequal stimulus presentation probability of ten to one would have on the accuracy of the these two measures.

Contrary to expectations, the increase in overall accuracy was not a linear function of an increasing mid-angle. Instead, accuracy improved markedly between reference line angle 10° and 19°, and then there was a comparatively small improvement when the reference line was 35°. Yet there is the same increase in orientation resolution between 10° and 19°, as there is between 19° and 35°, of 2° at 80 true percent. This suggests that there is a non-linear relation between changes in orientation resolution and changes in accuracy, as measured by proportions correct. This result was not expected, as there was nothing in Cleveland and McGill's theory that would have predicted it.

Subjects differed markedly in how well they were able to perform the experimental task. For example, Subject E4C1S5 was 100% correct for true percents from 50 to 70 and at 100% with a reference line of 55°, but with the same reference line Subject E4C1S2 never achieved 100% accuracy, irrespective of the true percent. This result was surprising as there was nothing to suggest that subjects should perform in a markedly different fashion on this sort of experimental task. If this result is true, it suggest two things. First, that there will be no single true percent that will necessarily avoid "ceiling" and "floor" effects for all subjects. Second, it may not be possible, in the long term, to find one single ideal aspect ratio for drawing graphs, simply because of how much each individual varies in their sensitivity to line slope.

It is possible that at some true percents most subjects will be able to do the experimental task well enough to allow the effect of the orientation resolution to be observable, yet some subjects may not be able to do the task under the same conditions. Given this variation in the ability of people to do the experimental task it would be better to select a small number of true percents across a range, instead of a single true percent, as was planned. It seems the range of true percents from 65% to 95% provides the best option for examining control by orientation resolution as the accuracy at these true percents is generally never 0.00 or 1.00 irrespective of reference line angle.

Having discovered a suitable set of true percents, which hopefully avoid "ceiling" and "floor" effects, it is now possible to go ahead and do the experiment envisaged earlier, and study the effect of orientation resolution, or mid-angle, on discrimination accuracy for angles from 5° to 85°. This experiment may also be able to provide more information about the suggestions coming from Experiment 4, that discrimination accuracy changes in a non-linear way with changes in orientation resolution, and that subjects differ markedly in their ability to do this kind of discrimination task.

Chapter 8 - Experiment 5

The purpose of Experiment 4 was to establish a single true percent that would allow the study of orientation resolution and its effect on line slope discrimination at mid-angles from 5° to 85°, without the confounding of "ceiling" effects (resulting from too easy a task) and low accuracy (resulting from too hard a task). However, the results suggested that no single true percent would prove ideal for all subjects, as subjects differed in how accurately they could perform the experimental task for stimuli of any given true percent. It was decided that a small number of true percents should be chosen from a range of true percent values within the range that would avoid strong "ceiling" effects and avoid producing performance so poor that the effects of increasing orientation resolution could not be seen. The data suggested this was from 65 to 95 true percent. Thus the values 65, 75, 85 and 95 were selected for use in Experiment 5.

As was suggested in Experiment 4, in this next experiment a "same-different" task would be used, making it possible to calculate log d, discrimination performance, and log c, response bias, as well as the proportions correct, as measures of discrimination performance. Also noted previously, was the need to examine experimentally all the mid-angles from 5° to 85° instead of simply assuming that changes in performance would be symmetrical around 45° as Cleveland and McGill (1987) had done. To keep the stimulus presentation probability equal, a "same" and a "different" stimulus would be presented for each of the seventeen mid-angles for each of the four true percents. A "same" stimulus was one with reference and comparison lines with the same slope, and a "different" stimulus was one with a comparison line slope that was one of the chosen percentages of the reference line slope.

With seventeen mid-angles, two stimuli at each mid-angle, four true percents, and the need for a reasonable number of replications to get a reliable estimate of performance, in this case 40, the total number of trials would have been 5440. It would have taken a subject over eight hours to complete an experimental session with this number of trials, and it was felt that this could lead to problems with experimental compliance. Hence, the number of trials was reduced by presenting individual subjects with a stimulus set with only one true percent, as a result there would be 1360 trials and these could be completed in about two hours.

The purpose of this experiment was to test Cleveland and McGill's theory that line slope discrimination is a linear function of orientation resolution which results from changes in the mid-angle for any given true percent less than 100. Accordingly it was expected that accuracy would increase as the mid-angle moved from 5° to 45°, would peak at 45° and then decrease as the mid-angle increased to 85°. Across the four selected true percents it was expected that subjects would generally be less accurate when the true percent was greatest, that is, when the difference between the slopes of the two lines was smallest, subjects would be less able to differentiate "different" from "same" stimuli. However, there was some suggestion from the data in Experiment 4 that the results might not be as expected. First, in Experiment 4 it appeared that accuracy on the task increased with orientation resolution in a non-linear manner, improving more rapidly at lower mid-angles, for example, around 10°, than at the higher mid-angles, for example, around 30°. Second, subjects appeared to vary in their ability to do the task, thus at some true percents some subjects might be unable to do the task while others can.

Method

Subjects

There were 24 subjects in this experiment, 15 females and 9 males, identified here by the notation E5CxSy, where x is the condition number and y is the subject number. Subjects 1, 5, 9, 13, 17 and 21 were in Condition 1, Subjects 2, 6, 10, 14, 18 and 22 in Condition 2, Subjects 3, 7, 11, 15, 19 and 23 in Condition 3, and Subjects 4, 8, 12, 16, 20 and 24 in Condition 4. All were volunteers from a first year psychology course at the University of Waikato. Each subject received 2% credit towards their first year psychology course for participating in the experiment, this was not contingent upon performance. Subjects were assigned to experimental conditions arbitrarily.

Apparatus

The experimental hardware and software were the same as those specified in Experiment 1.

Stimuli

Each stimulus comprised a pair of positively sloping lines with the reference line positioned in the top left-hand quarter of the screen and the comparison line positioned in the bottom right-hand quarter of the screen, as for the stimuli in Experiment 1. The reference lines and the comparison lines were drawn starting in the bottom left-hand corner of their respective quarters of the screen. The horizontal extent of both the reference line and the comparison line were kept equal for each line pair when the mid-angle was equal to or less than 45°. When the mid-angle of the stimuli was greater than 45°, the horizontal extent of the lines covaried with the mid-angle, with the horizontal extent becoming smaller as the mid-angle became larger, however the horizontal extent remained the same for any line pair.

There were four stimulus sets, each used in one of the four experimental conditions: the first set used in Condition 1 had comparison lines which were 65% of the slope of the reference lines, the stimulus set for Condition 2 had comparison lines which were 75% of the slope of the reference lines, for Conditions 3 and 4 the percents of slope were 85% and 95% respectively. In each of the four stimulus sets there were 34 stimuli, half the stimuli were the "same", that is, the comparison line's slope and the reference line's slope were the same, the remaining stimuli were "different", that is, the comparison line's slope differed from the reference line's slope by a fixed percentage, depending on the experimental condition. Each of the stimulus sets had 17 "different" stimuli with mid-angles ranging from 5° to 85° in steps of 5° and, for each of these "different" stimuli, there was a corresponding "same" stimulus. Mid-angles below 5° and above 85° were not used, as they could not be plotted with these stimulus line positions. Table 8-1 describes all the stimuli.

Procedure

The experiment was conducted in a similar manner to Experiment 4 including the same instructions to subjects. The only exception was that the complete stimulus set for that experimental condition was presented 40 times, each time using a different random sequence. Overall, a session comprised 1,360 trials (34 stimuli and 40 presentations). A typical session lasted about 2 hours, during which the experimenter checked progress every quarter of an hour.

Results

Figure 8-1 presents proportions correct, calculated using Equation 9, for each of the 17 reference line angles. The solid line indicates the data path for trials where the slopes of the stimulus lines were "different" and the broken line for trials where the slopes were the "same". One plot is presented for each of the 24 subjects. Plots are positioned on the page in columns based on the experimental condition.

The proportions correct for subjects in Condition 1, 65 true percent, on "same" trials were generally above 0.75. These proportions correct did not vary systematically with the variation in mid-angle. For three of the subjects, E5C1S1, E5C1S13 and E5C1S17, proportions correct were lower when the mid-angle was below 45° than when it was above. On "different" trials, for all subjects, proportions correct increased rapidly as the mid-angle increased from 5° to 15°, from 15° to 45° proportion correct was above 0.85. Between 45° and 55° proportion correct fell very quickly to 0.25 or less, and generally remained at this level for mid-angles from 45° to 85°. Across subjects performance was similar on both "Same" and "Different" trials.

The proportions correct for subjects in Condition 2, 75 true percent, on "same" and "different" trials were generally the same as the results seen in Condition 1, again between 45° and 55° proportion correct fell quickly to 0.25 or less. The only exception was subject E5C2S14, where the proportions correct on both "same" and "different" trials was around 0.5 and showed no systematic variation with changes in

TABLE 8-1

The geometry of the lines making up each of the 44 stimuli in the stimuli set. Given is the angle of the reference line, the comparison line and the in degrees, the slope of the reference line and the comparison line, the orientation resolution (the difference between the angle of the reference line and the angle of the comparison line in degrees), and the slope of the comparison line as a percentage of the slope of the reference line.

Comparison as a percent of reference line slope	Reference line angle in degrees	Comparison line angle in degrees	Reference line slope	Comparison line slope	Orientation resolution in degrees	Mid-angle in degrees
65	5	5.000	0.087	0.087	0.000	5.000
	5	3.254	0.087	0.056	1.745	4.127
	10	10.000	0.176	0.176	0.000	10.000
	10	6.538	0.176	0.114	3.461	8.269
	15	15.000	0.267	0.267	0.000	15.000
	15	9.880	0.267	0.174	5.120	12.440
	20	20.000	0.364	0.364	0.000	20.000
	20	13.310	0.364	0.236	6.689	16.655
	25	25.000	0.466	0.466	0.000	25.000
	25	16.862	0.466	0.303	8.137	20.931
	30	30.000	0.577	0.577	0.000	30.000
	30	20.570	0.577	0.375	9.430	25.285
	35	35.000	0.700	0.700	0.000	35.000
	35	24.472	0.700	0.455	10.528	29.736
	40	40.000	0.839	0.839	0.000	40.000
	40	28.608	0.839	0.545	11.391	34.304
	45	45.000	1.000	1.000	0.000	45.000
	45	33.023	1.000	0.650	11.976	39.011
	50	50.000	1.191	1.191	0.000	50.000
	50	39.762	1.191	0.774	12.237	43.881
	55	55.000	1.428	1.428	0.000	55.000
	55	42.870	1.428	0.928	12.129	48.935
	60	60.000	1.732	1.732	0.000	60.000
	60	48.387	1.732	1.125	11.612	54.193
	65	65.000	2.144	2.144	0.000	65.000
	65	54.344	2.144	1.393	10.655	59.672
	70	70.000	2.747	2.747	0.000	70.000
	70	60.752	2.747	1.785	9.247	65..376
	75	75.000	3.731	3.731	0.000	75.000
	75	67.596	3.731	2.425	7.403	71.298
	80	80.000	5.670	5.670	0.000	80.000
	80	74.822	5.670	3.685	5.177	77.411
	85	85.000	11.426	11.426	0.000	85.000
	85	82.333	11.428	7.427	2.666	83.666

Table 8-1 (Continued)

Comparison as a percent of reference line slope	Reference line angle in degrees	Comparison line angle in degrees	Reference line slope	Comparison line slope	Orientation resolution in degrees	Mid-angle in degrees
75	5	5.000	0.087	0.087	0.000	5.000
	5	3.754	0.087	0.065	1.245	4.377
	10	10.000	0.176	0.176	0.000	10.000
	10	7.533	0.176	0.132	2.466	8.766
	15	15.000	0.267	0.267	0.000	15.000
	15	11.363	0.267	0.201	3.637	13.181
	20	20.000	0.364	0.364	0.000	20.000
	20	15.628	0.364	0.273	4.731	17.634
	25	25.000	0.466	0.466	0.000	25.000
	25	19.276	0.466	0.349	5.723	22.138
	30	30.000	0.577	0.577	0.000	30.000
	30	23.413	0.577	0.433	6.586	26.706
	35	35.000	0.700	0.700	0.000	35.000
	35	27.706	0.700	0.525	7.293	31.353
	40	40.000	0.839	0.839	0.000	40.000
	40	32.183	0.839	0.629	7.816	36.091
	45	45.000	1.000	1.000	0.000	45.000
	45	36.870	1.000	0.750	8.130	40.935
	50	50.000	1.191	1.911	0.000	50.000
	50	41.790	1.191	0.893	8.209	45.895
	55	55.000	1.428	1.428	0.000	55.000
	55	46.966	1.428	1.071	8.033	50.983
	60	60.000	1.732	1.732	0.000	60.000
	60	52.410	1.732	1.299	7.589	56.205
	65	65.000	2.144	2.144	0.000	65.000
	65	58.128	2.144	1.608	6.871	61.564
	70	70.000	2.747	2.747	0.000	70.000
	70	64.112	2.747	2.060	5.887	67.056
	75	75.000	3.731	3.731	0.000	75.000
	75	70.339	3.731	2.798	4.660	72.669
	80	80.000	5.670	5.670	0.000	80.000
	80	76.769	5.670	4.252	3.230	78.384
	85	85.000	11.426	11.426	0.000	85.000
	85	83.346	11.426	8.570	1.653	84.173

Table 8-1 (Continued)

Comparison as a percent of reference line slope	Reference line angle in degrees	Comparison line angle in degrees	Reference line slope	Comparison line slope	Orientation resolution in degrees	Mid-angle in degrees
85	5	5.000	0.087	0.087	0.000	5.000
	5	4.253	0.087	0.074	0.746	4.626
	10	10.000	0.176	0.176	0.000	10.000
	10	8.524	0.176	0.149	1.476	9.262
	15	15.000	0.267	0.267	0.000	15.000
	15	12.830	0.267	0.227	2.169	13.915
	20	20.000	0.364	0.364	0.000	20.000
	20	17.190	0.364	0.309	2.809	18.595
	25	25.000	0.466	0.466	0.000	25.000
	25	21.621	0.466	0.396	3.378	23.310
	30	30.000	0.577	0.577	0.000	30.000
	30	26.139	0.577	0.490	3.860	28.069
	35	35.000	0.700	0.700	0.000	35.000
	35	30.760	0.700	0.595	4.239	32.880
	40	40.000	0.839	0.839	0.000	40.000
	40	35.497	0.839	0.713	4.502	37.748
	45	45.000	1.000	1.000	0.000	45.000
	45	40.364	1.000	0.850	4.635	42.682
	50	50.000	1.191	1.191	0.000	50.000
	50	45.369	1.191	1.013	4.630	47.684
	55	55.000	1.428	1.428	0.000	55.000
	55	50.519	1.428	1.213	4.480	52.759
	60	60.000	1.732	1.732	0.000	60.000
	60	55.814	1.732	1.472	4.185	57.907
	65	65.000	2.144	2.144	0.000	65.000
	65	61.251	2.144	1.822	3.748	63.125
	70	70.000	2.747	2.747	0.000	70.000
	70	66.819	2.747	2.335	3.180	68.409
	75	75.000	3.731	3.731	0.000	75.000
	75	72.503	3.731	3.172	2.496	73.751
	80	80.000	5.670	5.670	0.000	80.000
	80	78.280	5.670	4.819	1.719	79.140
	85	85.000	11.426	11.426	0.000	85.000
	85	84.123	11.426	9.712	0.876	84.561

Table 8-1 (Continued)

Comparison as a percent of reference line slope	Reference line angle in degrees	Comparison line angle in degrees	Reference line slope	Comparison line slope	Orientation resolution in degrees	Mid-angle in degrees
95	5	5.000	0.087	0.087	0.000	5.000
	5	4.751	0.087	0.083	0.248	4.875
	10	10.000	0.176	0.176	0.000	10.000
	10	9.509	0.176	0.167	0.490	9.754
	15	15.000	0.267	0.267	0.000	15.000
	15	14.281	0.267	0.254	0.718	14.640
	20	20.000	0.364	0.364	0.000	20.000
	20	19.074	0.364	0.345	0.925	19.537
	25	25.000	0.466	0.466	0.000	25.000
	25	23.893	0.466	0.433	1.106	24.446
	30	30.000	0.577	0.577	0.000	30.000
	30	28.744	0.577	0.548	1.255	29.372
	35	35.000	0.700	0.700	0.000	35.000
	35	33.631	0.700	0.665	1.368	34.315
	40	40.000	0.839	0.839	0.000	40.000
	40	38.560	0.839	0.797	1.439	39.280
	45	45.000	1.000	1.000	0.000	45.000
	45	43.531	1.000	0.950	1.568	44.265
	50	50.000	1.191	1.191	0.000	50.000
	50	48.547	1.191	1.132	1.452	49.273
	55	55.000	1.428	1.428	0.000	55.000
	55	53.607	1.428	1.356	1.392	54.303
	60	60.000	1.732	1.732	0.000	60.000
	60	58.711	1.732	1.645	1.288	59.355
	65	65.000	2.144	2.144	0.000	65.000
	65	63.856	2.144	2.037	1.143	64.428
	70	70.000	2.747	2.747	0.000	70.000
	70	69.037	2.747	2.609	0.962	69.518
	75	75.000	3.731	3.731	0.000	75.000
	75	74.249	3.731	3.545	0.750	74.624
	80	80.000	5.670	5.670	0.000	80.000
	80	79.485	5.670	5.387	0.514	79.742
	85	85.000	11.426	11.426	0.000	85.000
	85	84.738	11.426	10.855	0.261	84.869

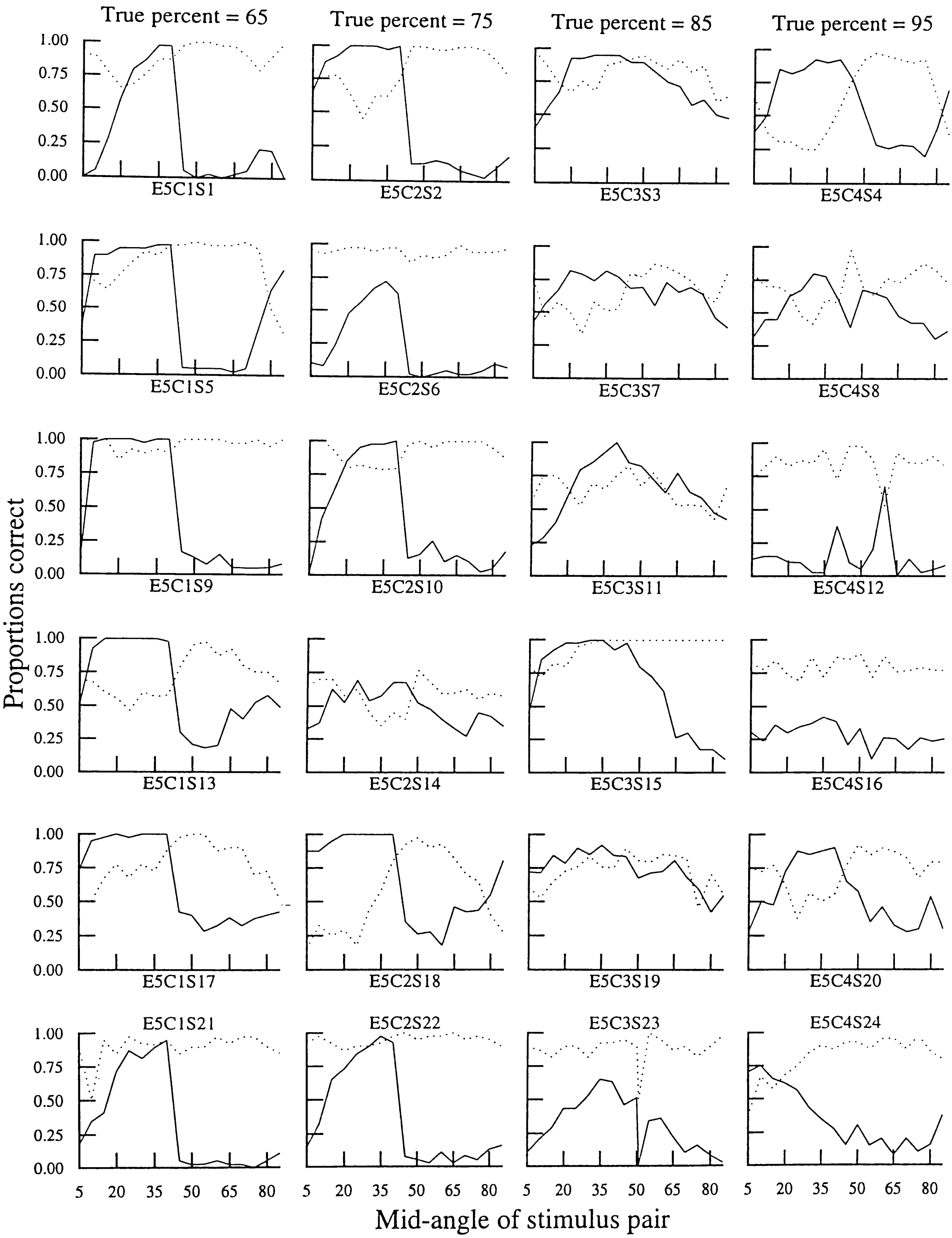


Figure 8-1. The proportion correct at each mid-angle split by true percent, for all 24 subjects. The solid line is for Different trials and the broken line is for Same trials.

mid-angle. Across subjects performance again was similar on both kinds of trials, except for E5C2S14 who, unlike the others, had a proportions correct of .5 irrespective of the trial type or the mid-angle.

In Condition 3, 85 true percent, the data differed from those in Condition 1 and 2. The proportions correct for all subjects on "same" and "different" trials were generally similar, with values around 0.75. Proportion correct increased between 5° and 15°, remained high to around 70°, and decreased again between 70° and 85°. The only exception was for Subject E5C3S15 for whom the proportion correct on "same" trials was 1.0 when the mid-angle was above 45° and the proportion correct on different trials decreased from 1.0 at 45 to 0.01 at 85°. Across subjects performance was varied, with Subjects E5C3S15 and E5C3S23 being both unlike each other or the four remaining subjects.

For subjects in Condition 4, 95 true percent, the proportions correct on both kinds of trial showed no consistent effects across subjects. For Subject E5C4S4, performance was similar to that seen for subjects in Condition 1, though more exaggerated. The results for Subjects E5C4S8 and E5C4S20 were similar to those seen for Subject E5C2S14 in Condition 2, where proportions correct on both types of trial was around 0.5. For the remaining three subjects proportions correct did not vary systematically with mid-angle, on "same" trials proportions were generally about 0.25 and "different" trials were about 0.75. Across subjects here there is almost no consistency in the paterns of performance on either trial type.

Figure 8-2 presents log *d*, a measure of discrimination performance, calculated using Equation 7 for each of the 17 reference line angles, one plot is presented for each of the 24 subjects. Higher positive values indicate increasing accuracy on the discrimination task. Negative values indicate that subject inaccuracy is lower than chance, that is, the subject was making the incorrect response to a stimulus on over half the trials. Again plots are grouped into columns based on the experimental condition. For Subjects in Condition 1 log *d* generally increased from 0.0 to 1.5 or 2.0 as the mid-angle increased from 5° to 45° and then decreased back towards 0.0 as the mid-angle increased towards 85°. For Subjects E5C1S13 and E5C1S21 the drop in log *d* when the mid-angle was greater than 45° was rapid with log *d* returning to 0.0 when the mid-angle was 55°. In Condition 2, for subjects E5C2S2, E5C2S6 and E5C2S10, the results are very similar to those seen in Condition 1. For Subject E5C2S14 log *d* was close 0.0 irrespective of the mid-angle. For subjects E5C2S18 and E5C2S22 the results are similar to those for E5C1S13, with a rapid drop in log *d* after 45°. For Condition 3 log *d* was generally around 0.0 and did not vary with mid-angle, the exceptions being a small increase in log *d* for Subject E5C3S11 at around 45°, and for Subject E5C3S15 log *d* was generally above 1.0 and peaked at 35°. In Condition 4 for all subjects log *d* remained around 0.0 irrespective of the mid-angle.

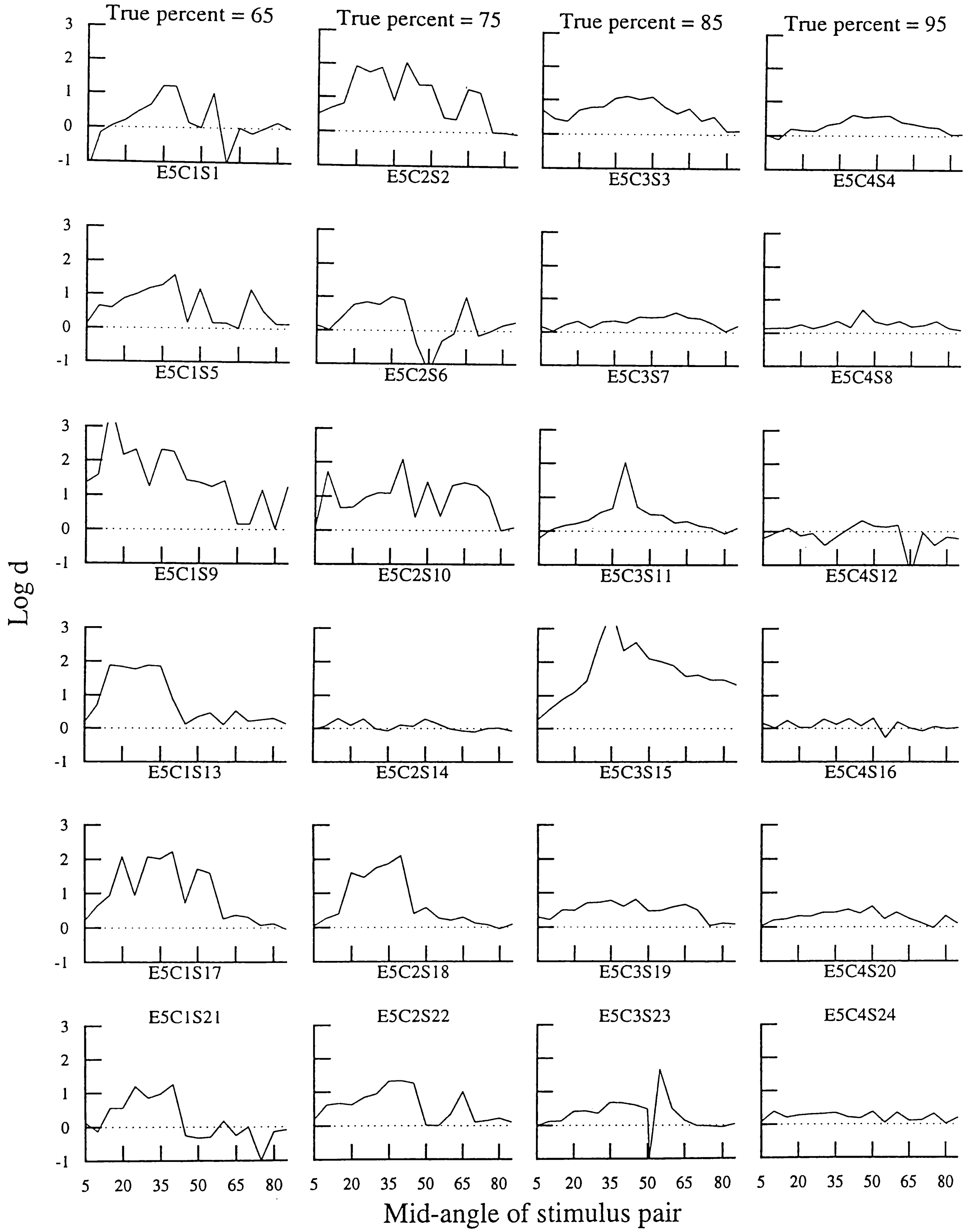


Figure 8-2. Log d at each mid-angle split by true percent, for all 24 subjects. Increasing postive values indicate increasing accuracy on the discrimination task.

Figure 8-3 presents log *c*, an estimate of the bias towards a particular response, calculated using Equation 8 for each of the 17 reference line angles, one plot is presented for each of the 24 subjects. Positive values of log *c* indicate a bias in responding towards the "same" response, negative values indicate a bias towards the "different" response. For all subjects, in all conditions, the pattern is similar, at mid-angles between 5° and 45° there was either no response bias or a response bias towards the "different" response, when the mid-angle was greater than 45° there was a bias towards the "same" response. This pattern was present both when responding had large biases, for example E5C1S5, and when responding had very little bias, for example E5C2S14. Only three subjects, E5C4S8, E5C4S12, and E5C3S11, had results that do not follow this pattern. For these three subjects log *c* was variable but was centred around 0.0 and showed no systematic variation with mid-angle. The general pattern across the four conditions is for the amount of bias to decline as the true percent of slope increased from 65 to 95.

Discussion

Discrimination

Experiment 5 is the first to use the alternative measure of line slope discrimination, log *d,* and of bias, log *c* as described in the introduction to Experiments 4-7. These measures are considered to be more satisfactory than that used by Cleveland and McGill's (1987). It is also the first time the full range of mid-angles from 5° to 95° have been tested. This is unlike the work of Cleveland and McGill (1987) and the first three experiments, which only tested mid-angles from 5° to 45°.

True percent and discrimination

It was anticipated that as the true percent got larger, subjects would generally become less able to do the experimental task correctly. This would come about because at very high true percents the difference in slope between the two lines of the "different" stimuli would be so small that subjects would erroneously identify "different" stimuli as being the same. Values of log *d* around 0.0 suggest that a subject is not discriminating between the "same" and the "different" stimuli, higher values of log *d*, for example 1.0, suggests that subjects are getting 10:1 correct and log *d* 2.0 suggests that subjects are getting 100:1 correct, therefore clearly discriminating between the two types of stimuli (Davidson & McCarthy, 1989).

In Figure 8-2 it can be seen that, as expected, as the true percent became greater the number of subjects with very low log *d* values, indicating poor performances, increased. This result is also reflected in Figure 8-1, which shows that at the larger

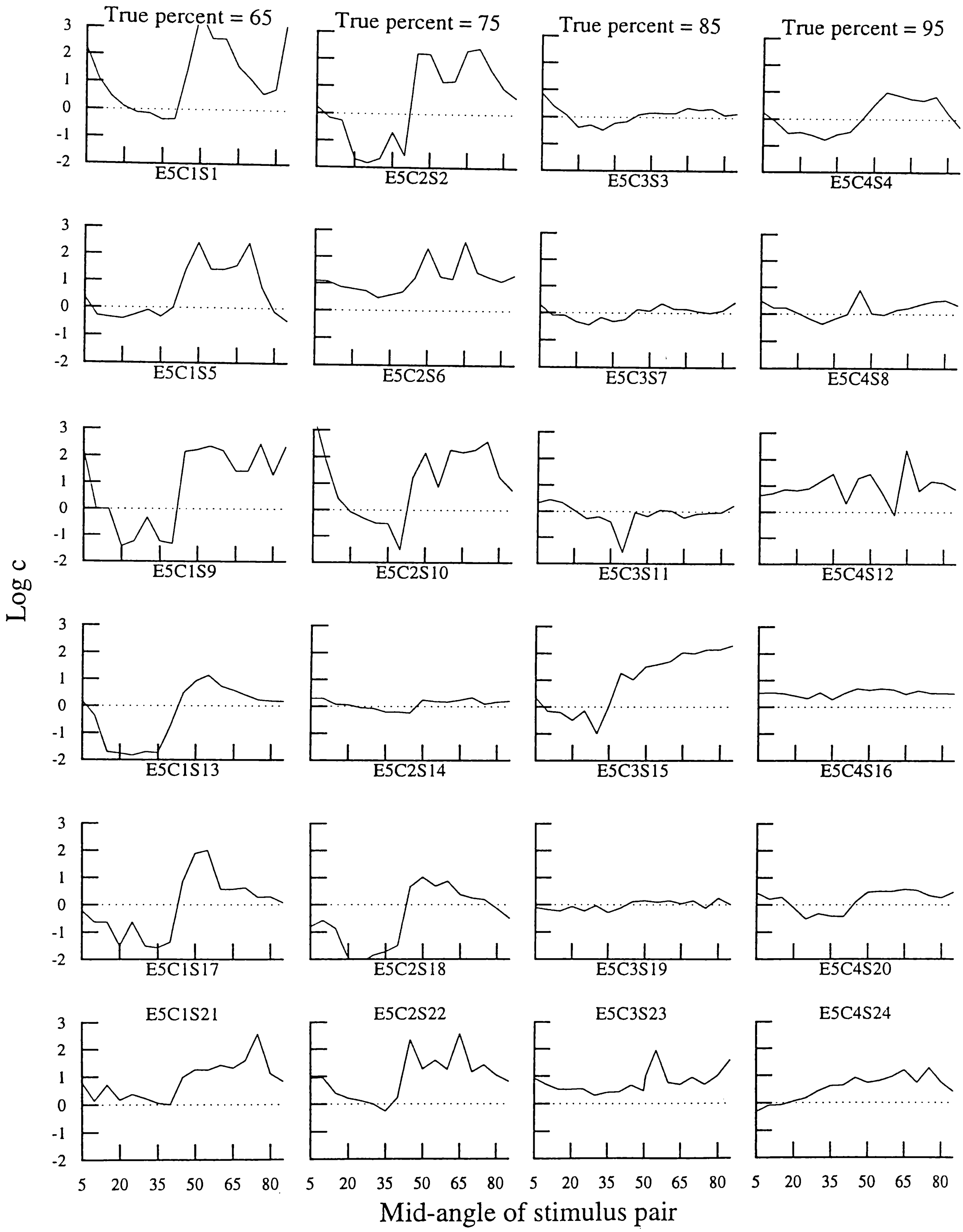


Figure 8-3. Log c at each mid-angle split by true percent, for all 24 subjects. Postive values indicate a bias towards the Same response, negative values indicate bias towards the Different response.

true percents the proportions correct tend to be 0.75 or lower, also indicative of poor performances. Therefore, it seems reasonable to suggest that performance on this kind of line-slope-discrimination task is constrained by the choice of true percent. As anticipated in Experiment 4, at high true percents in Conditions 3 and 4 subjects were generally unable to perform the discrimination task well enough at the maximum orientation resolution to enable improved performance with increasing orientation resolution to be seen. This suggests the task was too hard at these values. This effect is seen for all subjects at 95 true percent and for most at 85 true percent.

It was also suggested, on the basis of the results from Experiment 4, that the true percent at which subjects could no longer discriminate would be different for different subjects. This suggestion is supported by the data presented here, where for Subject E5C2S14 in Figure 8-1 log d is close to zero, yet the remaining five subjects in Condition 2 were all able to do the discrimination task very well. Also supporting the suggestion that different subjects would be unable to do the task at different true percents, is the performance of E5C3S15 (Figure 8-1), who was still able to discriminate well at 85 true percent while the remaining five subjects in Condition 3 did not.

Since all the other subjects in Condition 2 were able to do the task it might be suggested that Subject E5C2S14 should have been able to do it. If this was so then this subject could have been "malingering" in some way and not attending to the stimuli, thus these data do not reflect an inability to do the task. However, this does not seem a likely explanation because the pattern of changes in proportions correct on "different" trials and "same" trials (Figure 8-1) is similar to that seen for the other subjects in Condition 2. Though the patterns in the data of Subject E5C2S14 are not as marked as those of the other subjects, they are still visible and suggest that the subject was probably not malingering or responding totally randomly. Whether or not a subject with poor performance would have performed better at a lower true percent cannot be determined here since there were data from each subject at only one true percent. Perhaps the best way to confirm the suggestion, that true percent exerts control over discrimination performance, would be to conduct an experiment in which each subject is exposed to all four experimental conditions. However, as noted in the introduction to this experiment, the number of trials and thus the session length, makes this sort of experiment difficult to do.

Mid-angle and discrimination

Cleveland and McGill (1987) argued that discriminability was a function of orientation resolution, or mid-angle at any one true percent. This suggestion is supported by the data in Figure 8-2 where log d increases with increasing mid-angles from 5° to 45°, and thus with increasing orientation resolution, for the low true percents. This pattern is generally not seen for the higher true percents in

Conditions 3 and 4. It was suggested in Experiment 4 that the rate of improvement in discrimination would be greater at smaller mid-angles, however, there is nothing in these present results to support this suggestion. The rate of improvement in discrimination performance between 5° and 45° was slower in Condition 2 than in Condition 1. This is consistent with Cleveland and McGill's argument that discriminative performance is a function of orientation resolution, since in Condition 1 at 15° the orientation resolution is 5.1° whereas in Condition 2 at 15° the orientation resolution is 3.6°. In fact the orientation resolution is smaller in Condition 2 than Condition 1 for all mid-angles. Thus the orientation resolution theory would predict lower discriminative performance in Condition 2 than in Condition 1, for any given mid-angle.

Cleveland and McGill argued that because discriminative performance is a function of orientation resolution, and because orientation resolution decreases as the mid-angle increases from 45° to 90°, then the discrimination performance would decrease as mid-angle increases. They go on to suggest that, because orientation resolution increases at the same rate between 5° and 45° as it decreases between 90° and 45° (that is, orientation resolution peaks at 45° and decreases symmetrically around 45° towards 0° and 90°), discriminative performance should also increase from a minimum at 0° to a maximum at 45° then decrease, symmetrically, to a minimum at 90°. Because this argument seemed reasonable to Cleveland and McGill (1987) they did not study mid-angle values above 55°, and as a result did not provide any empirical support for their argument.

The results presented in Figure 8-2 do not support Cleveland and McGill's suggestion. These results show that log *d* was generally lower overall for mid-angles above 45° than for those below 45°, and that, having reached peak discrimination performance at 45°, the performance then drops rapidly between 45° and 55°. This result can also be seen in Figure 8-1 where proportions correct on both kinds of trial were generally very good up to a mid-angle of 45°, after 45° proportions correct for "same" trials remained good, but proportions correct for "different" trials fell from approximately 1.0 to below 0.10 when the mid-angle was around 45°. In short, subjects were reporting that all stimuli with mid-angles above 45° had line pairs with the same slope. That is, subjects were not able to discriminate differences in line slopes when the line pairs had a mid-angle greater than 45°, even though their performance was close to perfect for line pairs with mid-angles just below 45°.

This rapid fall in accuracy above 45° appears to be a very strong effect which does not correspond with Cleveland and McGill's orientation resolution theory, nor is it a result that might have been expected. This result cannot be readily seen in log *d* for Conditions 3 and 4 because of the overall poor performance at these high true percents. There is, however, some suggestion that this pattern still may be present at these high true percents, with the appearance of small peaks followed by rapid falls in log *d* around 45°. This suggestion is supported by the results in Figure 8-1 in which there is a tendency for proportions correct on "different" trials to be higher below 45° than above, for all conditions, even in Condition 4.

Log d as a measure of discrimination

As noted previously, log *d* is not without some problems. One such problem experienced here, was what to do when performance on one type of trial is perfect. This is a problem because when performance is perfect, that is when one of the values in Equation 7 is zero, it is not possible to calculate a value for log *d* (McCarthy 1991). There are three ways to address this problem. First, collect more data in the hope that as the amount of data increases an error will become more likely. This approach is costly as it requires further data collection and it may take a long time before an error occurs. The second approach is to treat that data point as a missing value when plotting it. This approach is easier than the first, however it fails to show to the reader that the organism performed very accurately under those conditions and by being missing may inadvertently give the opposite impression. The third approach, used here, is to replace the zero value with a very small value (here 0.01), thus enabling the calculation to proceed. This approach avoids the problems of the previous two approaches, but has the disadvantage that when subjects had perfect scores the log *d* values became very high and difficult to plot on a sensible scale. For example, Subject E5C3S15 obtained 100% correct on both types of trial at 35° and as a result log *d* for this mid-angle was very large.

Log *d* values of zero indicate that the subject is correct on only half of the trials and is not discriminating. In the data presented in Figure 8-2 log *d* occasionally falls below zero. Under these conditions, the subject is saying "different" more than 50% of the time in the presence of "same" stimuli and is saying "same" more than 50% of the time in the presence of "different" stimuli. These negative log *d* values do not represent discriminative performance below chance. If a subject is unable to discriminate then in theory they should be correct exactly 50% of the time on both kinds of trial, however, in practice there will be some variation around this 50% figure, both below and above. It is the variation below that leads to log *d* values below zero.

Response bias

Cleveland and McGill did not discuss response bias. It has been assumed here that, at some orientation resolutions, pairs of stimulus lines that have different slopes will be described by the subject as having the same slope. These erroneous responses would suggest that, for a line pair at that particular mid-angle and orientation resolution, subjects are insensitive to, or unable to discriminate variations in the slopes of lines. As a consequence, it was suggested that at these values, when the stimuli were indiscriminable, responding would be biased towards the "same" response. This suggestion is consistent with classical psychophysical theory.

Bias, or a tendency to favour one response over another, is measured here by log *c* calculated using Equation 8, and presented in Figure 8-3. Negative values of log *c* indicate a tendency towards using the "different" response, positive values indicate a tendency towards the "same" response. At higher true percents (Conditions 3 and 4) there was very little indication of bias. At lower true percents there were strong biases in favour of the "same" response above 45°, and either no bias below 45° or some bias in favour of the "different" response. For Subjects E5C4S4 and E5C4S20 there was a suggestion that the same bias pattern seen at lower true percent might be present at the higher true percents, but this is a less clear result. There was very little bias both above 45° and at the higher true percents.

Earlier, it was suggested that when a subject was unable to discriminate the difference between the slopes of two lines, they would report that the lines were the "same". This would result in a large bias towards the "same" response, which was seen at low true percents for mid-angles greater than 45°. This bias to "same" indicates a failure to discriminate, and is seen in the log *d* results from trials with high mid-angles at low true percents. At high true percents discriminative performance was very poor but, contrary to expectations, this was not accompanied by a strong bias towards the "same" response. Instead, the subjects' responses were allocated evenly across the two response types, as can be seen in the log *c* results which show a tendency to stay around zero.

These results show that poor discriminative performance can be associated with two very different response patterns. In one (at high mid-angles at low true percents) performance became extremely biased towards the "same" response, suggesting the different stimuli were being treated as same stimuli. This pattern had been expected. In the other (at high true percents) the performance became random and there was no bias. This pattern was not expected.

Bias and Discrimination

It is not clear why there were two different "patterns" of poor performance. The classical theory of psychophysics was based on the assumption that subjects' responses were under the control of only the stimulus (Woodworth & Schlosberg, 1972). As pointed out previously, based on this classical approach, it was expected that when the stimuli could not be discriminated subjects' responding would become biased towards the "same" response. For example, if a very quiet tone was played for a subject to detect, it might be expected the subject would report that the tone was absent and, of course, when the tone was absent the subject would also report it absent. Under conditions of low discriminability a subject would be expected to be biased towards the response "absent". Accordingly, the assumption made in the introduction to this experiment, that poor discrimination would lead to extreme bias in favour of the "same" response, was appropriate to the classical model. This classical model can readily account for the present results when the mid-angle was above 45° at low true percents, but it cannot explain the results seen at high true percents. This is because the model does not account for the situation in which a subject says a stimulus is present when it is absent. The classical model would not predict that a subject would report that the stimulus line slopes are different when they are, in truth, the same.

With the development of the theory of signal detection, mentioned earlier, it was recognised that the subjects' responses are controlled by both the stimulus and variables which can bias behaviour (Green & Swets, 1966). Under the signal detection model it is accepted that a subject can report that a stimulus is present when it is absent, because of a bias or tendency to make that response, irrespective of whether the stimulus is absent or not. The bias is assumed to be the result of sources of control such as the pay-off for that response or the probability of occurrence of the stimulus. If there are no biasing variables present then the control is assumed to be by the stimulus only. Stimuli which are not easily discriminated would be expected to result in equal probabilities of the response "same" on same trials and of the response "same" on different trials. When there are no extraneous biasers present then, for these data, it would be expected that these probabilities would be 0.5. The signal detection model can account for the pattern of responding above 45° at high true percents.

The theory of signal detection still does not explain why there was a strong bias towards the "same" response when there was low accuracy in responding at mid-angles greater than 45°. A similar result was seen in Experiment 4. In that experiment, in which in each session a range of different true percents was presented, the stimuli with 95 true percent resulted in low discriminative performance and a strong bias towards the "same" response. One similarity between the sessions in these two experiments, where this result was found, was that they contained clearly discriminable stimuli. In Experiment 5, where accuracy was poor and no bias was seen, the sessions contained only very similar stimuli.

Signal detection has normally been applied to situations in which only two stimuli (e.g., signal present and signal absent) are presented in any session and thus the physical discriminability of the stimuli remains constant across a session. The conditions with high true percents in Experiment 5 are closest to this procedure, having very similar stimuli throughout the session, and the results from this session appear to be consistent with a signal detection analysis. At low true percents the sessions contained a wide range of physically different stimuli (as in Experiment 4) and thus were unlike the typical signal-detection procedure. It is not clear from signal detection theory whether or not this might be expected to influence the results.

Here it seems to have done so by biasing the behaviour. As a consequence of this bias these data are like those predicted by "classical" perception theory.

The suggestion that the presence of discriminable stimuli in the experimental session effects bias on trials with indiscriminable stimuli would be further supported if data from a single subject across all four experimental conditions were similar to those seen in this experiment across subjects. Data generated under these conditions would also help to support the earlier suggestion that subjects differ in how high the true percent has to be before their discrimination performance becomes poor.

Chapter 9 - Experiment 6

In the data from Experiment 4 there were some suggestions that subjects differed in how well they performed at any single combination of true percent and mid-angle. For example, Subject E4S5 performed notably better than Subject E4S3 at 85 true percent when the reference-line angle was 35°. The data from Experiment 5 suggested further that there might be variations across subjects in how well they performed for any given set of conditions. For example, in Condition 2 Subject E5C2S14 had very low accuracy while the remaining subjects in that condition had high accuracy, and in Condition 3 Subject E5C3S15 had high accuracy while the remaining subjects in that condition had low accuracy. Although these results clearly indicate that subjects do differ in their performance on the same task and under the same conditions there is no clear indication of why this occurs. It was suggested, in Experiment 5 that these differences were related to the subjects' abilities to discriminate between the line pairs. It could well be that the true percent at which a subject can no longer discriminate between the line pairs is different for each subject, that is, one subject may become unable to discriminate at 65 true percent while for another this may only occur at 95 true percent. The data from Experiment 5 can not directly support this explanation because each individual served in only one condition. As a result it is not possible to rule out the explanation that some subjects were not able to do the task at all, as distinct from the suggestion that they were not able to do it at certain true percents. It was suggested, therefore, that an experiment in which each subject did the task at a number of different true percents would address this question.

It was found in Experiment 5 that low discriminative performance occurred at mid-angles greater than 45°, irrespective of the true percent, and also at high true percents. When the true percent was high, both types of response were used equally often. When the mid-angle was above 45°, at lower true percents, responding was biased towards the "same" response. It was argued that it might be the presence of clearly discriminable stimuli, at low true percents and low mid-angles, that gave rise to this bias towards the "Same" response. To investigate this, data could be collected from the same subject both when clearly discriminable stimuli are present and when they are absent from the experimental session. Stimuli at only 65 true percent would appear to have been clearly discriminable, based on the accuracy of the subjects in the previous experiment and, in general, those at 95 true percent appear to have been indiscriminable. If the data from a single subject across a session with 65 true percent and with 95 true percent were similar to those seen for the individuals in Experiment 5, this would suggest that the change in bias was a function of the true percent and not some inter-subject variation.

In Experiment 5 discriminative performance was generally good when the mid-angle was below 45° and decreased when the mid-angle moved above 45°. This decrease in performance between 45° and 55° was very rapid, with proportions correct on "different" trials falling from about 1.0 to 0.0 when the mid-angle increased from 45° to 55°. This result was surprising, it was strong, and there was nothing in Cleveland's theory (which suggest performance should be symmetrical about 45°) or in any theory the writer has found that would predict it. It is important because it has significant implications for the question of selecting an ideal aspect ratio, since it suggests very poor discriminative performance at these higher mid-angles. Therefore it would be desirable to attempt further replication to see if the effect was robust.

In designing Experiment 5 it had been decided that subjects need not experience all four true percents. There was no clear reason for having them do so, and it was envisaged that all four would be run in a single session, making the session about eight hours long. However, in light of the discussion, both here and previously, it appeared that having individual subjects perform across all four experimental conditions would provide some very useful data. It was decided that, in the next experiment, each subject would have four separate sessions on four separate days, each with one of the four true percents. This approach has the advantage of reducing both the risk of fatigue and the subject compliance problems. In addition the results can be compared with those from Experiment 5. It was anticipated that Experiment 6 would produce results, within subjects, that generally were the same as those seen between conditions in Experiment 5, including: decreasing performance with increasing true percent and at mid-angles above 45°, bias to the "same" response at low true percents above 45°, and unbiased-poor performance at high true percents.

In summary, we can extend the useful data already collected in Experiment 5, by having subjects perform under all four different true percent conditions. This should enable us to test if subjects stop being able to do the experimental task accurately at different true percents, e.g. do some subjects stop being accurate at 75% while others are still accurate at 95%. It also makes it possible to check if within subjects the bias towards "Same" responses above 45° at low true pecents is absent at high true percents. Lastly, further replication to confirm this unexpected drop in accuracy above 45°, irrespective of the true percent, would be highly desirable. Therefore, within subject results here will serve to confirm and extend a number of important observations made in Experiment 5.

Method

Subjects

There were six subjects in this experiment, one female and five males, identified here as E6Sl to E6S6. All were self selected from a first-year psychology course at the University of Waikato. Each subject received a cash payment for participating in the experiment, which was not contingent upon performance. Subjects were also guaranteed the payment should they withdraw before the completion of the experiment, although none did.

Apparatus

The experimental hardware and software were the same as those used in Experiment 4.

Stimuli

The stimuli were the same as those used in Experiment 5.

Procedure

The experimental session was similar to Experiment 5, except that subjects were not debriefed until the end of their fourth experimental session. Each subject participated in four experimental sessions and served under one experimental condition per experimental session. All subjects went through the experimental conditions in the same order, that being 85 true percent (referred to as Condition 3), 95 true percent (Condition 4), 65 true percent (Condition 1) and finally 75 true percent (Condition 2). Subjects completed one experimental session per day, and all completed the experiment in four days. All subjects took part in the experiment over the same five day period.

Results

Figure 9-1 presents the proportions correct, calculated using Equation 9, for each of the 17 reference-line angles. The solid lines show the data from trials in which the slopes of the stimulus lines were "different" and the broken lines show those from trials in which the slopes were the "same". Four plots are presented for each of the six subjects. Plots are positioned on the page in columns, based on the true percent, and in rows, based on the subject number.

The proportions correct from the "same" trials with 65 true percent (Condition 1) were generally above 0.75 when the mid-angle was above 45° and below 0.75 when the mid-angle was below 45°. For Subject E6S3 it remained above 0.75 irrespective of mid-angle. On "different" trials proportions correct generally

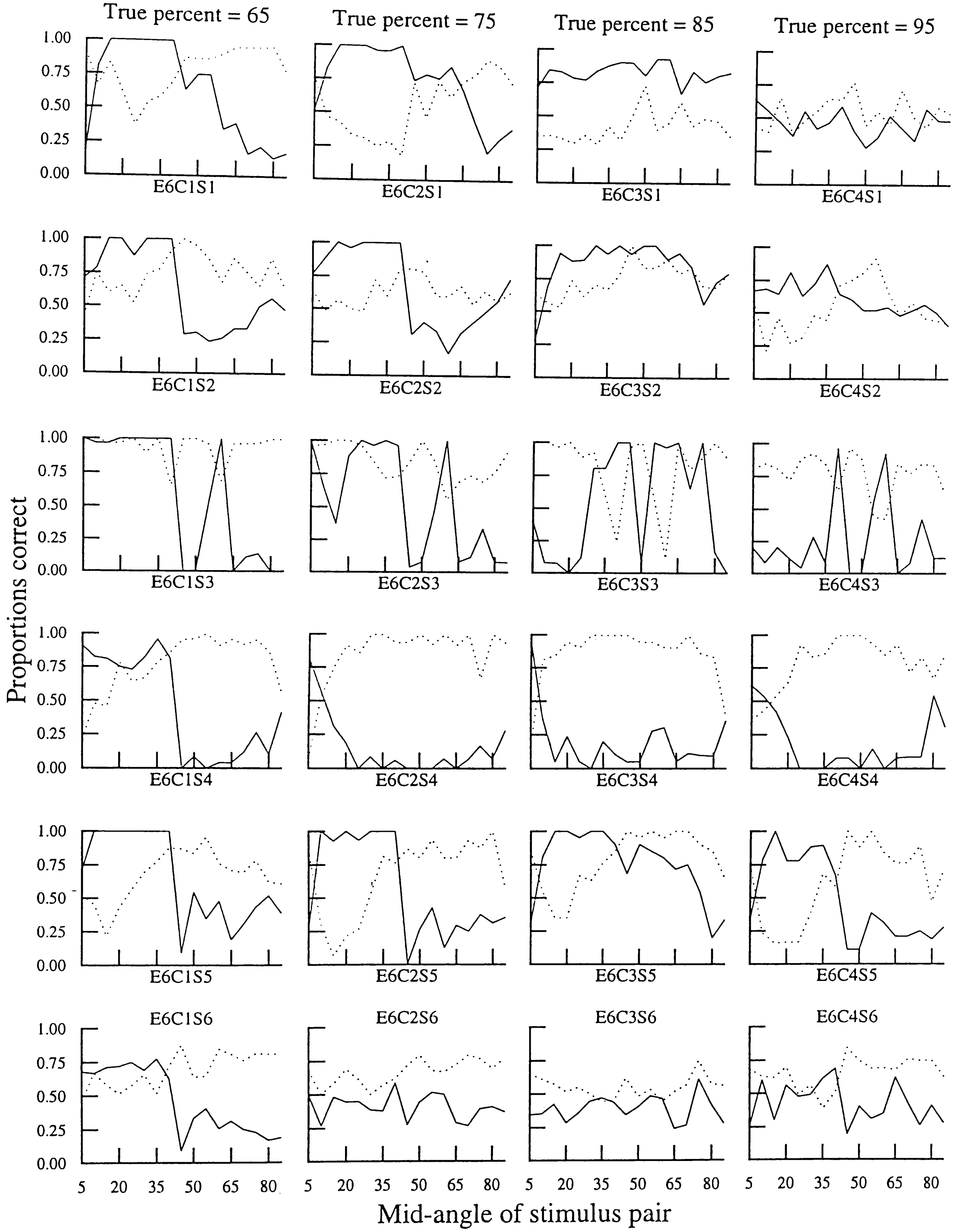


Figure 9-1. The proportion correct at each mid-angle split by true percent, for all six subjects. The solid line is for Different trials and the broken line is for Same trials.

increased rapidly as the mid-angle increased from 5° to 15°. From 15° to 45° proportions correct were above 0.85, except for Subject E6S6, for whom it was about 0.75. Between 45° and 50° proportions correct fell very quickly to 0.25 or less, except for Subject E6S1 for whom it decreased gradually between 45° and 85°. Proportions correct on "different" trials generally remained below 0.25 for mid-angles from 45° to 85°. The exception to this was for Subject E6S4, at 55° was 1.00 for this subject and it returned to 0.00 at the next mid-angle.

The proportions correct on "same" and "different" trials in Condition 2, with 75 true percent, were generally similar to those seen in Condition 1 on both types of trial, except for two subjects. For Subject E6S4 the proportion correct on "different" trials was about 0.01 both above and below 45°. For Subject E6S6 on "different" trials it was about 0.25 both above and below 45°.

The proportions correct in Condition 3, 85 true percent, on "same" and "different" trials were generally similar to those in Condition 2 for Subjects E6S6 and E6S4. Proportions correct for Subject E6S1 was 0.25 on "same" trials and 0.75 on "different" trials, and values did not systematically vary with mid-angle. Subjects E6S2 and E6S5 had proportions correct which were generally around 0.75 irrespective of the mid-angle or the trial type. The results from both kinds of trial were extremely variable for Subject E6S3.

In Condition 4, 95 true percent, the proportions correct on both kinds of trial showed no consistent effects across subjects. For Subjects E6S1, E6S2, and E6S6 the proportions correct on both kinds of trial were around 0.50. For Subject E6S4 results were very similar to those seen in Condition 2 and 3, except that performance was further reduced on both kinds of trial. For Subject E6S3, on "same" trials, the proportion correct was about 0.75 and on "different" trials it was about 0.10 (except that at 40° and 60°, for "different" trials, it was about 1.00).

Figure 9-2 presents log d, or the measure of discriminative performance, calculated using Equation 7 for each of the 17 reference-line angles. Four plots are presented for each of the six subjects. Increasing positive values indicate increasing accuracy on the discrimination task, negative values indicate that subject inaccuracy is lower than chance, that is, subjects are making the incorrect response to a stimulus on over half the trials. Again plots are grouped into columns based on the experimental condition and rows based on the subject.

For Subjects E6S1, E6S2 and E6S5 in Condition 1, log d generally increased from 0.0 to 1.5 or 2.0 as the mid-angle increased from 5° to 45° and then it dropped back towards 0.0 as the mid-angle increased towards 85°. This drop in log d at 45° was very rapid with log d returning to 0.0 when the mid-angle was 55°. For Subjects E6S4 and E6S6 log d was close to 0.0 at most mid-angles, being slightly higher when the mid-angle was below 45° than when it was above it. For Subject E6S4 log d was small for mid-angles below 45°, at 45° it fell to 0.0, after 45° it fluctuated around 0.0.

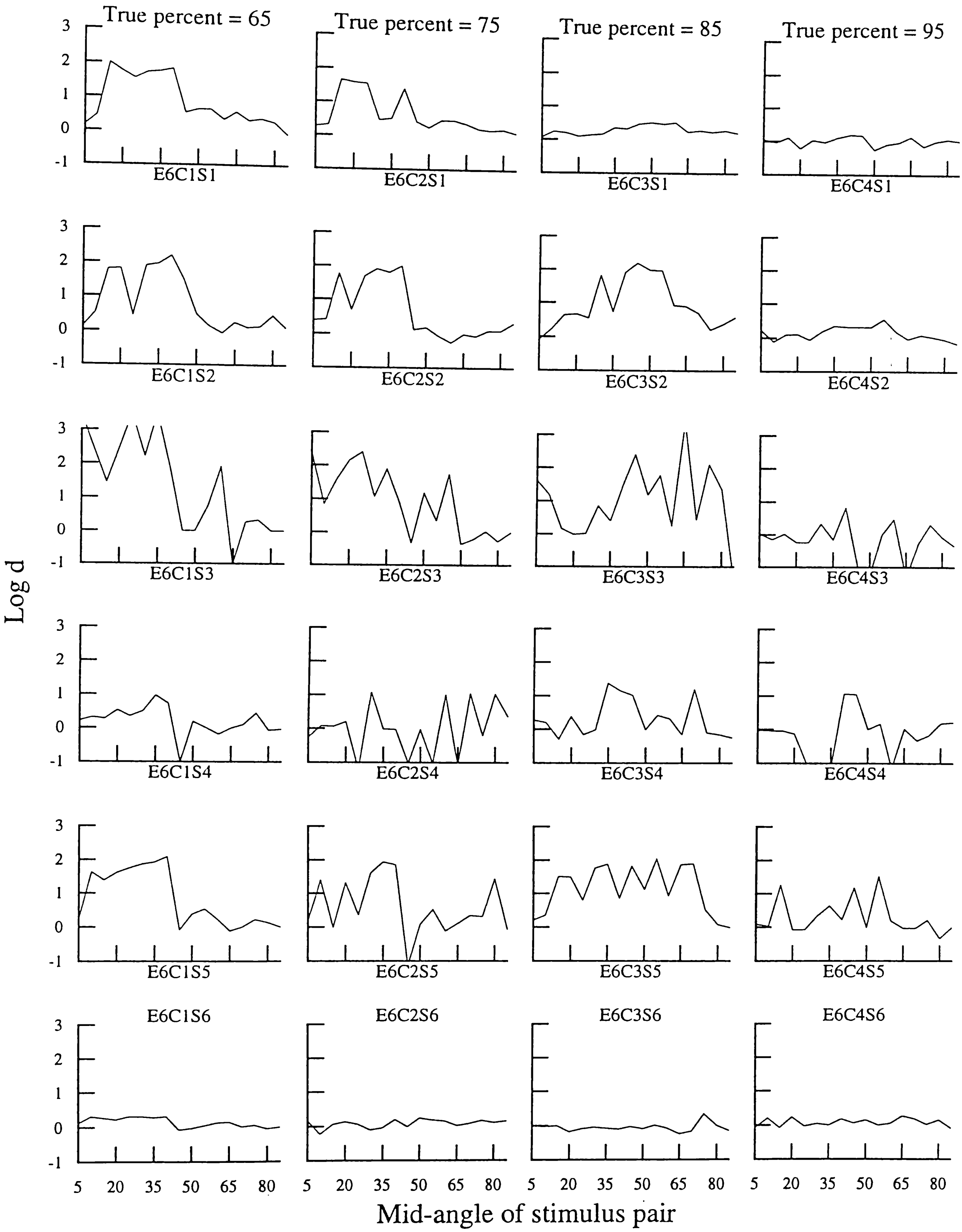


Figure 9-2. Log d at each mid-angle split by true percent, for all six subjects.

In Condition 2 the results were very similar to those seen in Condition 1, except that, for all subjects, log *d* was lower. For Subjects E6S4 and E6S5 log *d* was more variable than it had been in Condition 1. In Condition 3, for subjects E6S1 and E6S6, log *d* was generally around 0.0 and did not vary with mid-angle. For Subject E6S4 log *d* was around 0.0 but highly variable. Subjects E6S2, E6S3 and E6S5 had log *d* values generally around one, being smaller at very low and very high mid-angles. In Condition 4, for all subjects, log *d* was around zero and did not vary systematically with the mid-angle.

Figure 9-3 presents log *c*, or bias towards a particular response, calculated using Equation 8 at each of the 17 reference-line angles, four plots are presented for each of the six subjects. Positive values of log *c* indicate a bias in responding towards the "same" response, negative values indicate a bias towards the "different" response. Generally across all subjects and conditions the results were similar, at mid-angles between 5° and 45° there was either no response bias, or a response bias towards the "different" response. When the mid-angle was greater than 45° there was a bias towards the "same" response. This pattern was present both when there was a large bias, for example, E6S1 in Condition 1, and when there was a small bias, for example E6S6 in Condition 1.

Discussion

It was predicted here that all subjects would be able to do the experimental task accurately at some true percent, but the true percent at which they could no longer do the task would differ across subjects. This prediction was supported by these results. Most subjects were able to do the task at 65 true percent. One subject, E6S3, was still able to do the task at 85 true percent, while others were not able to, e.g., E6S1. These results confirm that it would be very difficult to find a single true percent that would be ideal for studying orientation resolution, as had been attempted in Experiment 4, therefore the choice of four different true percents in Experiment 5 and Experiment 6 seems justified.

It was also predicted that the data here, as in Experiment 5, would show a response bias to the "same" response at low true percents above 45° and reduced bias at high true percents. This is what was found, and although this result is not as clear as that seen in Experiment 5, there is no doubt that it is still present to some degree. The low accuracy on "different" trials above 45° at low true percents indicates a strong bias towards using the "same" response. The tendency for proportions correct, on both kinds of trial, to be about 0.5 at high true percent shows that responding was not biased at these values. The biases seen for a subject at low true percents above 45° disappeared at high true percents. These results are important because they

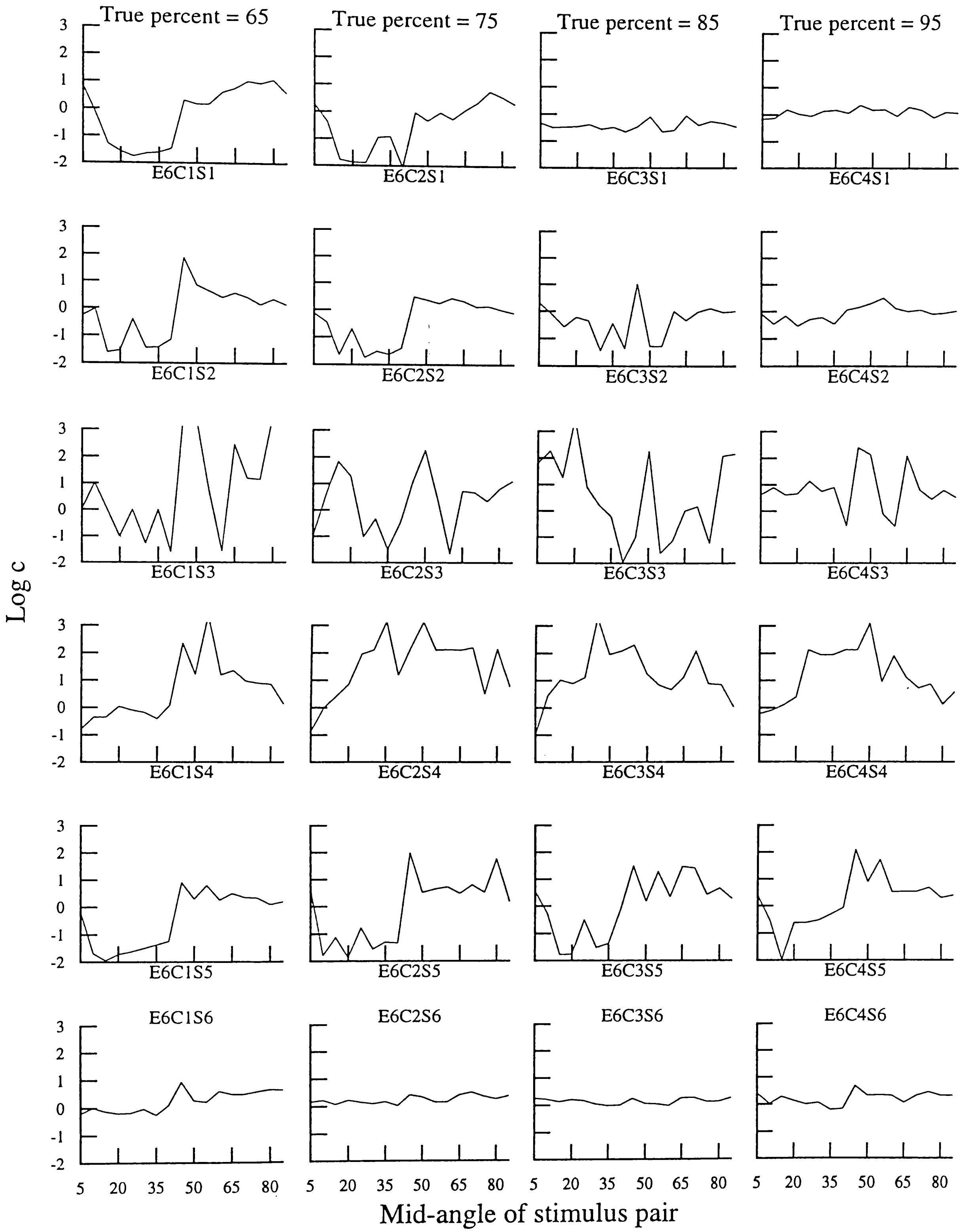


Figure 9-3. Log c at each mid-angle split by true percent, for all six subjects.

indicate that the large biases were a product of a variable that changed with the experimental conditions and not of a general disposition in a subject's responding.

The variable that was changed across these experimental conditions was the true percent, a measure of the difference between the slopes of the lines. At 65 true percent the slopes of the line pairs were very different and at 95 true percent they were very similar. The suggestion given in Experiment 5, that it was the presence of clearly different stimuli in the experimental session (as at 65 true percent) that produced the bias above 45°, is not challenged by the results here and needs further investigation.

It was found that for lines below 45° accuracy was very high and that accuracy above 45° was very low. In Experiment 5 it was noted that the very poor line slope discrimination found with lines above 45° was not expected. Such low accuracy suggests that Cleveland and McGill's theory that accuracy in line slope discrimination is controlled by orientation resolution applies only to line segments with slopes less than 45°. The fall in accuracy around 45° was fast and clear. It was noted previously that such a strong effect needed further replication to be sure that it was a reliable result. The results from Experiment 6 confirm that this effect is replicable and that it occurs when the true percent is varied for an individual, as well occurring across different subjects.

The very poor line slope discrimination above 45° is contrary to the expectation that performance would be more accurate at 45° and that accuracy would decrease symmetrically as the mid-angle increases and decreases from this value. This result is also inconsistent with the findings from research on the "Oblique Effect". This effect suggests that performance on line-slope discrimination tasks, similar to those reported here, does vary as the mid-angle varies. The Oblique Effect shows that performance decreases as the mid-angle increases from 0° to 45°, that is as the mid-angle moves away from the horizontal. It also shows that performance increases again as the mid-angle decreases from 45° to 90°, that is as the mid-angle moves towards the vertical. The Oblique Effect suggests that performance improves as the mid-angle gets closer to vertical, or horizontal, axis and is worst at 45°. Appelle (1972) reviewed research on the Oblique Effect and reported that this effect has been observed repeatedly in experiments using: human subjects, non-human subjects, a wide range of experimental procedures, and a range of stimulus configurations. The Oblique Effect seems to be reliable and well established.

The literature on the Oblique Effect suggests that it is the result of mid-angle changes, Cleveland and McGill's (1987) model would not predict this effect but would predict the completely opposite result: Cleveland and McGill (1987) predicted that people will have the greatest sensitivity to line slope variation at 45° while the Oblique Effect shows that people seem to have the least sensitivity to line slope variation at 45°. The Oblique Effect findings suggest that as the mid-angle increases from 0° to 45° accuracy will decrease. The results from the last two experiments are the opposite of this, with accuracy increasing, not decreasing, as the mid-angle increased. The Oblique Effect also suggests that as the mid-angle increases from 45° to 90° accuracy will increase. However, the results here show low accuracy above 45° irrespective of mid-angle, which is inconsistent with the Oblique Effect findings.

Results from these two experiments are clearly not predicted by either Cleveland and McGill's model or the Oblique Effect, but they appear to be reliable, having been observed with 30 different subjects on many hundreds of trials. It is not clear what gave rise to these results. It seems highly unlikely that it was something about the subjects, given that similar results were observed for 30 different subjects. Therefore, it seems more likely that these results were a function of the experimental procedure or of the way the sloping lines were configured in the stimuli. It seems highly unlikely that the experimental procedure used here was the cause of these results as it does not differ markedly from the procedures used by others investigating the Oblique Effect or others doing visual perception research, for example, Wenderoth and van der Zwan (1989), or Day (1973). It is possible that the unexpected results here were a product of some feature of the stimuli, therefore it seems important to extend the previous experiments by using a number of different stimulus configurations in order to discover if the results seen in Experiments 5 and 6 are product of the stimulus configuration.

Chapter 10 - Experiment 7

At the end of the previous experiment it was argued that neither the experimental procedure nor the choice of experimental subjects was likely to have been responsible for the experimental results found here. It was suggested that some specific feature of the stimulus configuration is likely to have been responsible and, if so, alternative configurations of the reference and comparison lines might produce different results. Therefore, the purpose of this experiment is to test this out. There are many possible stimulus configurations and so it was necessary to select some that were considered most likely to influence the subject's accuracy. Seven stimulus configurations were chosen for study here, the reasons for their selection are outlined below.

Examination of the stimulus configuration used in previous experiments, showed that there was at least one aspect of the stimulus lines that changed around a mid-angle of 45°. This was the way the stimulus lines intersected with the edges of the computer screen, as can be seen in the examples which were given in Figure 2-1. For all mid-angles, the left-hand end of the stimulus line in the top left-hand quadrant reaches the side of the screen and the left-hand end of the stimulus line in the bottom right-hand quadrant reaches the bottom of the computer screen. When both the lines have slopes of less than 45° (and thus the mid-angle is less than 45°), as in (B-D) in Figure 2-1, the right-hand end of the top left-hand line does not reach the top of the screen but the right-hand end of the bottom right-hand line reaches the side of the screen. When the mid-angle is above 45°, the right-hand end of the top left-hand line reaches the top of the screen, but the right-hand end of the bottom right-hand line does not reach the side of the screen (as in Figure 2-1, A). When the mid-angle is 45°, as depicted in the standard configuration with parallel lines in Figure 10-1, the right hand-ends of both lines reach the edge of the computer screen. When the mid-angle is close to 45° then, for a few of the line pairs used, both lines reach the edges of the screen. Table 8-1 shows that, in the previous experiment, this occurred for three of the seventeen line pairs at 65 true percent, two at 75 true percent, one at 85 true percent and one at 95 true percent.

This change in the way the lines intersected with the screen edges is correlated with the very clear change in the performance of many subjects around 45°. It is, therefore, possible that the change in performance might be a result of this change in the way the lines are 'framed' by the screen. One condition of this present experiment therefore involved drawing vertical and horizontal lines, and so dividing the screen into four clear quadrants, to act as a frame for all stimulus line pairs (as shown in Figure 10-1, Condition 6). Performance with stimuli presented in this way was compared with performance with the standard configuration.

Another aspect of stimulus configuration that might have been important is the length of the stimulus lines. There are some data to support this suggestion. Rochlin (1955) required subjects to adjust a comparison line until it was parallel with a reference line. Subjects participated in two conditions in which the length of the stimulus lines were ¼ inch in one and 1 inch in the other. Rochlin found that subjects were less accurate when the lines were ¼ inch than when the lines were 1 inch. Thus Rochlin's findings suggest that the length of the stimulus line might also effect accuracy on a line-slope-discrimination task. The task used by Rochlin (1955), adjusting one line until it was parallel with another, is somewhat similar to that used here in previous Experiments.

In the stimulus set used previously, the horizontal extent of both lines in any pair was kept equal, as suggested by Cleveland and McGill (1987). It was this that gave rise to the way the lines intersected with the edges of the screen as mid-angle changed as discussed above. This also had an effect on the lengths of the lines, with both the comparison and the reference lines becoming longer as the mid-angle increases from 5° to 45°. As the mid-angle increases above 45° the lines decrease in length. As a result lines are generally longest around 45° and get shorter as the mid-angle increases and decreases from this. Although the stimuli used in previous experiments here contained a range of line length, this was coincidental, and it was not the purpose of these studies to look at the effect of line length. Any contributions of line length to the data are confounded and cannot be assessed separately from the effects of mid-angle.

In order to see whether line length might be contributing to the results seen for previous experiments, the next experiment included a condition in which performance of subjects with line pairs shorter than those used previously (as in Figure 10-1, Condition 1) was compared with their performance on the standard stimuli used previously. If line length effected performance, as it did for Rochlin's (1955) subjects, then the performance with the shorter lines should be worse than that with the standard lines. However, shortening the lines in the way shown in Figure 10-1, Condition 1, also has the effect of completely removing any interaction between the ends of the lines and the screen edges. Any differences in performance with the short and standard stimulus sets could also be attributed to this change. Accordingly, a further condition was included in which the stimuli were short and, in addition, squares were draw around them (as in Figure 10-1, Condition 4). This stimulus configuration provided both the short stimuli and a frame. Performance with these stimuli was compared with performance on the standard stimuli.

It was suggested in the Introduction to Experiment 3 that the relative position of the stimulus lines might alter the difficulty of the task. As was argued, such a line-slope-comparison task may well be made easier if both line segments started from the same baseline, i.e., were vertically aligned, as shown in Figure 10-1,

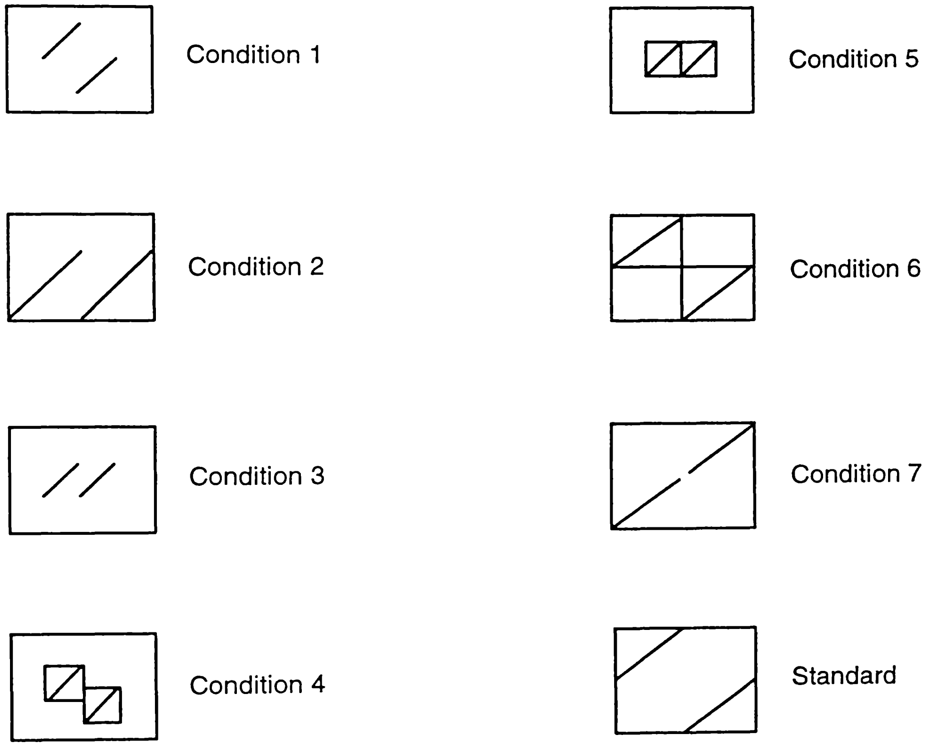


Figure 10-1 Sketches of the standard stimulus configuration, and the seven alternative stimuli (1-7). Sketches show an example of a "same" stimuli at 45• mid-angle in all conditions.

Condition 2. Although this alternative stimulus configuration was used in Experiment 3, the performance measure used there was flawed, as outlined in the discussion following Experiment 3, and so it is not clear what effect line position had on performance. A further change in position that might make the task easier would be to move the lines to the top right and bottom left quadrants (as in Figure 10-1, Condition 7). Both of these alignments were included in the present experiment, one condition with each in which performance with the new alignment was compared with performance on the standard stimulus set.

One effect of these different alignments is that they alter the way the lines interact with the screen edges. In order to see whether this had an effect, two further conditions were included: one in which the lines were short and vertically aligned but never touched the screen edges (as in Figure 10-1, Condition 3); and another in which the lines were short, vertically aligned and contained within a square frame (as in Figure 10-1, Condition 5). Performance on these two configurations was also compared with performance with the standard configuration.

Reconfiguring the stimuli used in the previous experiments in a number of different ways should show whether or not the results from those experiments were a product of the specific stimulus configuration. The present experiment involved seven conditions, each with a different group of subjects, in which the way the lines interacted with the screen edges, the provision of an explicit frame, the length of the stimulus lines, and the relative positions of the lines were varied. Some conditions involved varying more than one aspect of the standard stimulus, e.g., Condition 5 (Figure 10-1) in which the alternate stimulus set had shorter , vertically aligned, lines with a square frame. In Experiment 6 the true percent was varied and it was found that overall accuracy was greatest at 65 true percent and the unusual patterns in the data were very clear. It was decided that stimuli in which the comparison line was 65 true percent of the reference line would be used (together with the equivalent parallel lines). In addition, since each subject's performance was to be measured with two different stimulus sets, a reduced set of line pairs was required. For each true percent in Experiment 6 there were 34 line pairs in each stimulus set. Here only those where the angle of the reference line was a multiple of ten were used in order to reduce the size of the stimulus set and to enable subjects to work through two sets of stimuli.

Method

Subjects

There were 42 subjects, six subjects in each of seven conditions. Those in Condition 1 are identified here as E7C1Sl to E7C1S6, those in Condition 2 as E7C2S1 to E5C2S6, and this naming pattern is repeated for the subjects in the remaining five conditions. All subjects were volunteers from a first-year psychology course at the University of Waikato. Each subject received 2% credit towards their first-year psychology course for participating in the experiment, this was not contingent upon performance. Subjects were assigned to the experimental condition being run at the time they joined the experiment.

Apparatus

The experimental hardware and software were the same as those specified in Experiment 1.

Stimuli

In this experiment there were seven conditions, in each condition there were 16 stimuli with the standard-stimulus and 16 with the alternative-stimulus configuration. There was one standard-stimulus set and there were seven different alternative-stimulus sets, one for each experimental condition. Of the sixteen stimulus line pairs in each of the eight sets of stimuli, eight were pairs of parallel lines and eight were pairs of lines with different slopes. The geometry of the lines was the same for each stimulus set and the details of this are presented in Table 10-1. These stimuli were a subset of those used in Experiment 6 (using 100 and 65 true percent where the reference lines angles were multiples of 10). There was a different configuration for the lines in each of the alternative stimulus sets. These are shown in Figure 10-1, together with the standard configuration, for pairs of parallel lines with mid-angles of 45°. In summary the complete stimulus set used in a condition comprised 16 standard stimuli and 16 stimuli from one of the alternative stimulus configurations. Half of these 32 stimuli had parallel lines and the other half had line pairs with slopes that differed by 65 true percent.

Procedure

The experiment was conducted in a similar manner to Experiment 4. The exception was that, as in Experiments 5 and 6, the complete stimulus set for an experimental condition (a mixture of stimuli with the standard configuration and one of the alternative configurations) was presented 40 times, each time using a different random sequence. Overall, a session comprised 1,280 trials (32 stimuli and 40 presentations). A typical session lasted about 2 hours, during which the experimenter checked progress every quarter of an hour. There were six subjects in each of the seven conditions each working for one session with their allocated stimulus set.

Results

The left-hand side of Figures 10-2 to 10-8 present, for each subject, log d, the measure of discrimination calculated using Equation 7 for each of the reference-line angles, for the trials with the standard stimuli (the solid line) and for those with the alternative stimuli (the dotted line).

TABLE 10-1

The geometry of the lines making up each of the 16 stimuli in the stimulus set. Given is the slope of the comparison line as a percentage of the slope of the reference line, the angle of the reference line and the comparison line in degrees, the slope of the reference line and the comparison line, the orientation resolution (the difference between the angle of the reference line and the angle of the comparison line in degrees), and the mid-angle in degrees.

Comparison as a percent of reference line slope	Reference line anglein degrees	Comparison line angle in degrees	Reference line slope	Comparison line slope	Orientation resolution in degrees	Mid-angle in degrees
65	10	10.000	0.176	0.176	0.000	10.000
	10	6.538	0.176	0.114	3.461	8.269
	20	20.000	0.364	0.364	0.000	20.000
	20	13.310	0.364	0.236	6.689	16.655
	30	30.000	0.577	0.577	0.000	30.000
	30	20.570	0.577	0.375	9.430	25.285
	40	40.000	0.839	0.839	0.000	40.000
	40	28.608	0.839	0.545	11.391	34.304
	50	50.000	1.191	1.191	0.000	50.000
	50	39.762	1.191	0.774	12.237	43.881
	60	60.000	1.732	1.732	0.000	60.000
	60	48.387	1.732	1.125	11.612	54.193
	70	70.000	2.747	2.747	0.000	70.000
	70	60.752	2.747	1.785	9.247	65..376
	80	80.000	5.670	5.670	0.000	80.000
	80	74.822	5.670	3.685	5.177	77.411

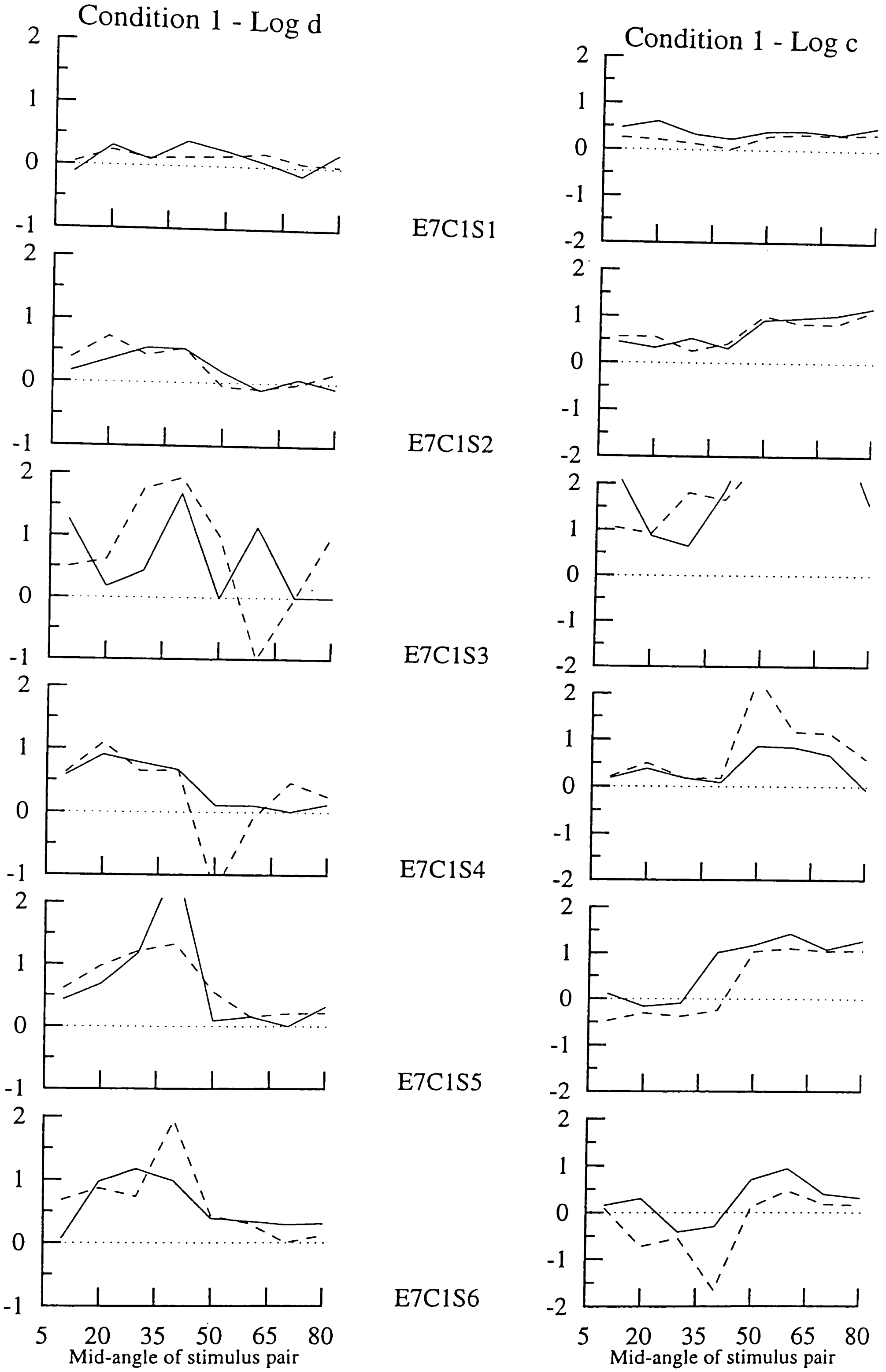


Figure 10-2 For the six subjects Log d and Log c are plotted againest mid-angle. The solid line shows data from the Standard condition, the broken line shows data from Condition 1.

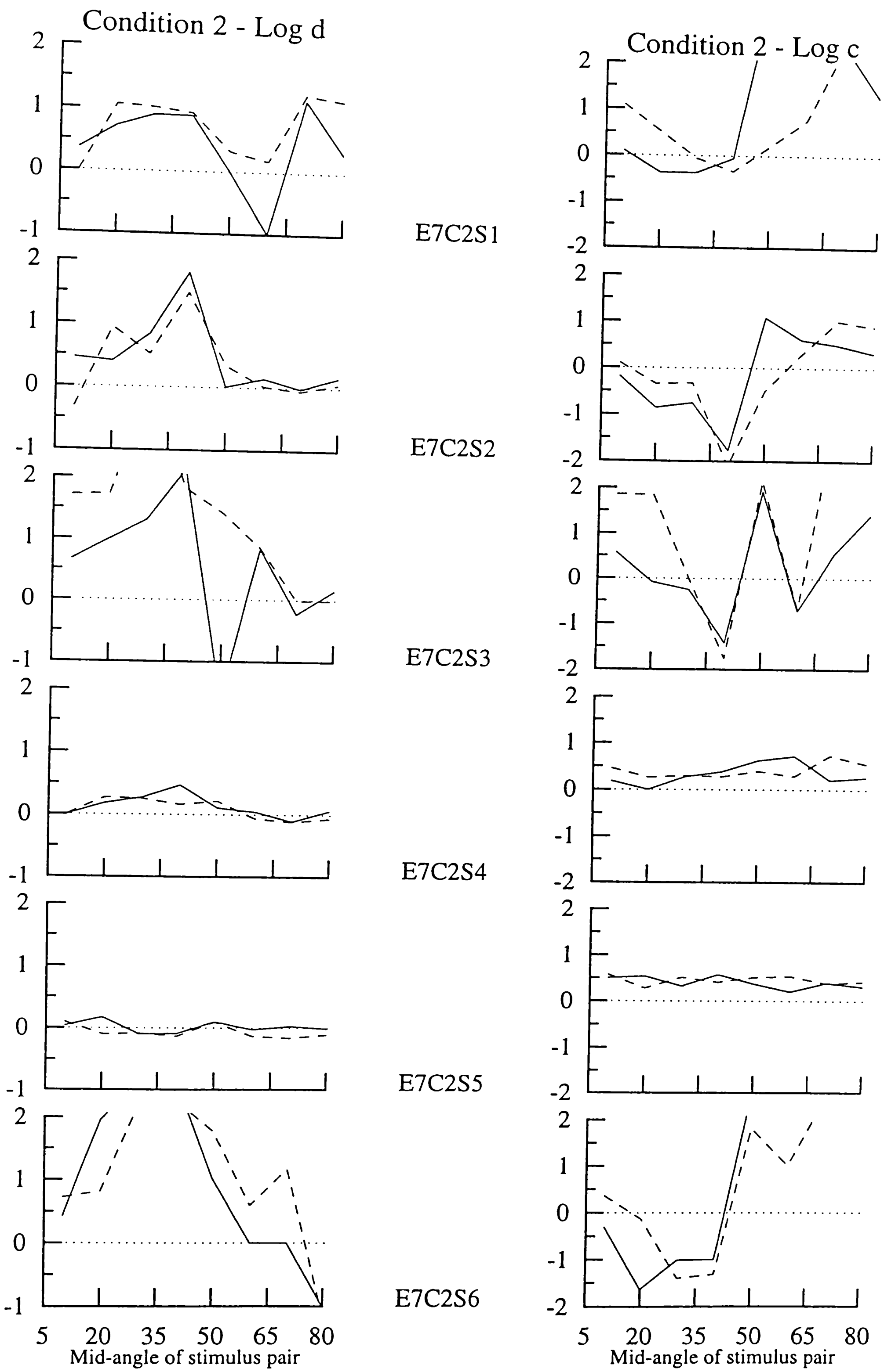


Figure 10-3 For the six subjects Log d and Log c are plotted againest mid-angle. The solid line shows data from the Standard condition, the broken line is for data from Condition 2.

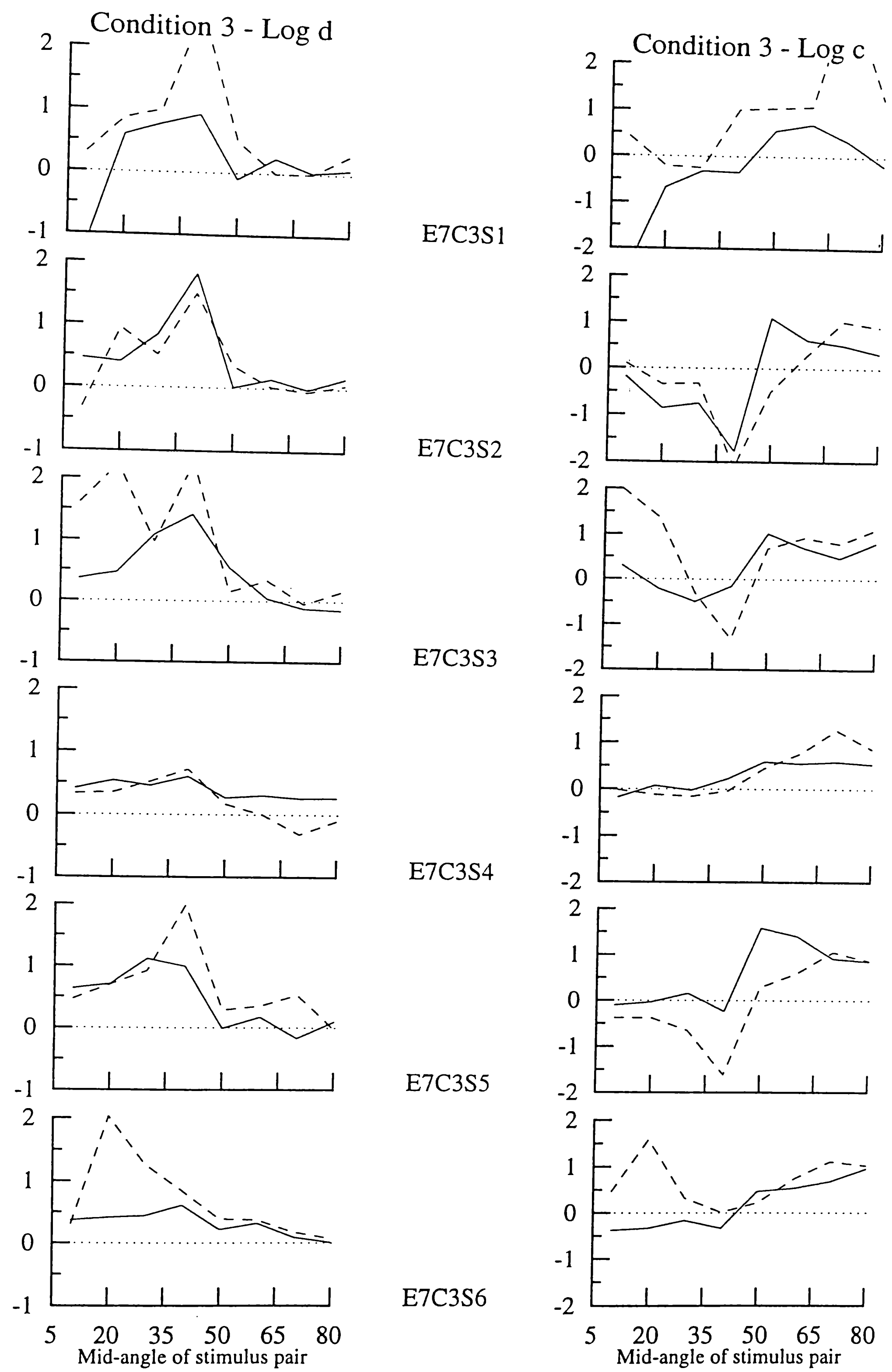


Figure 10-4 For the six subjects Log d and Log c are plotted againest mid-angle. The solid line shows data from the Standard condition, the broken line shows data from Condition 3.

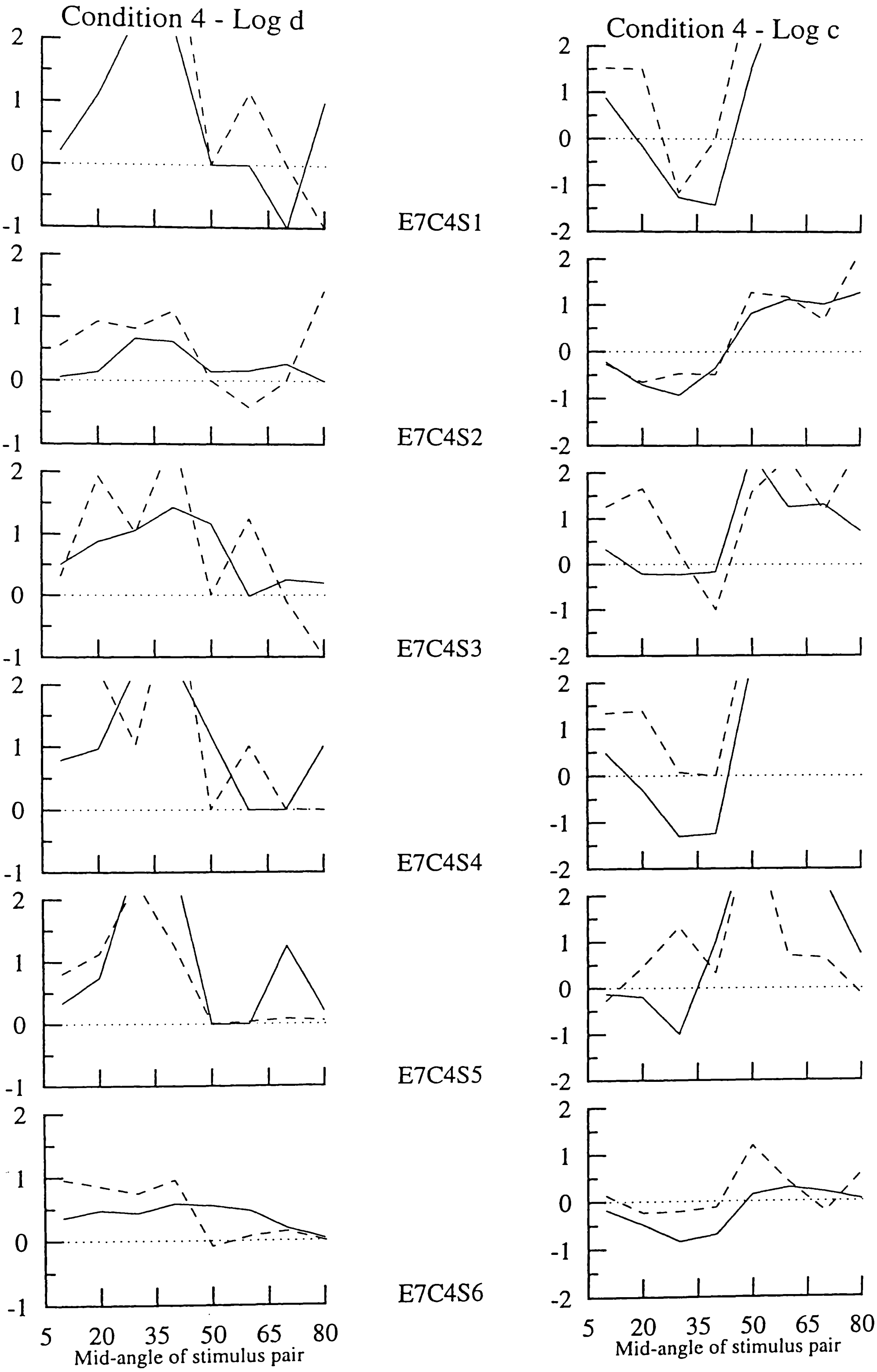


Figure 10-5 For the six subjects Log d and Log c are plotted againest mid-angle. The solid line shows data from the Standard condition, the broken line shows data from Condition 4.

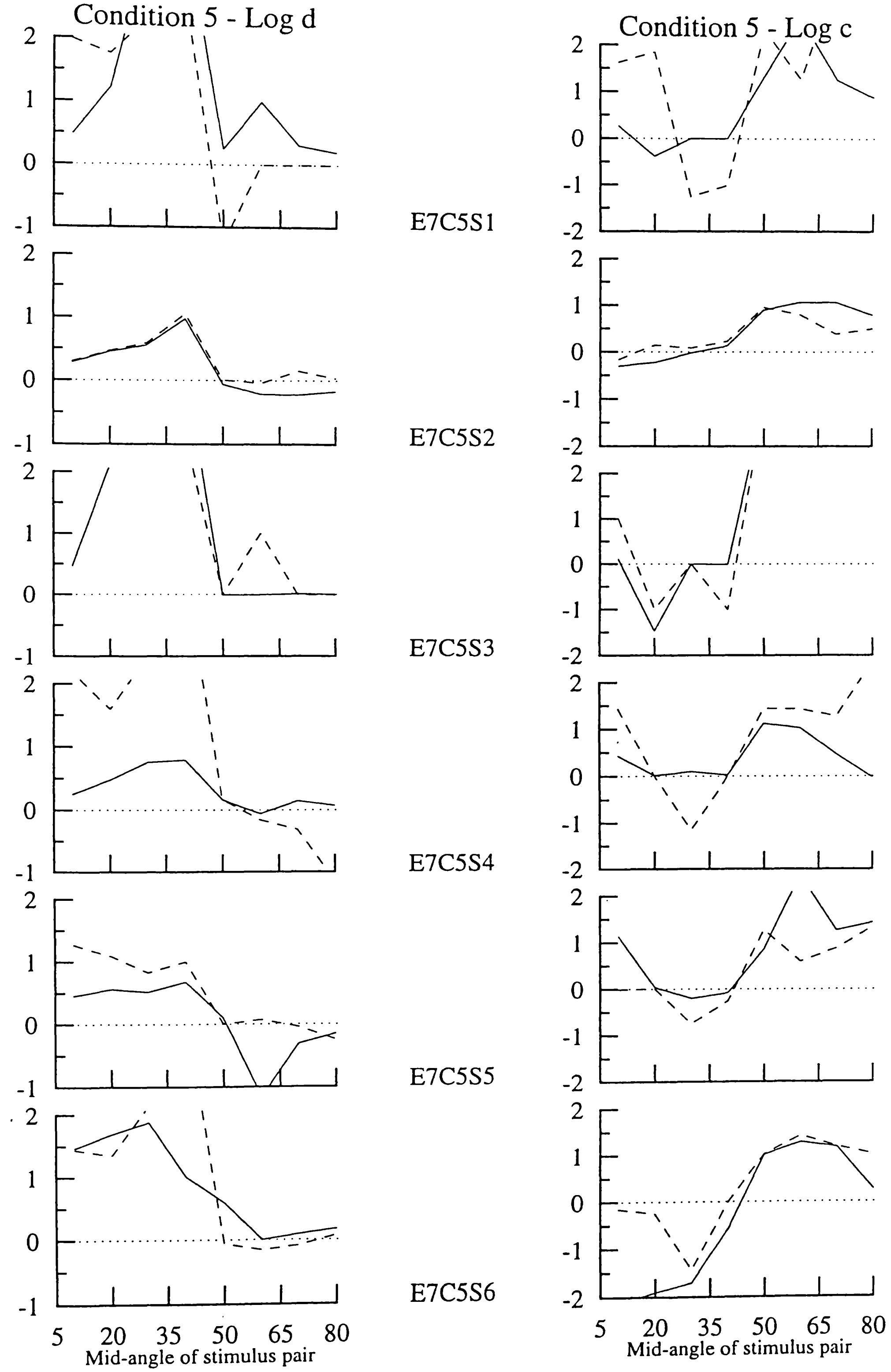


Figure 10-6 For the six subjects Log d and Log c are plotted againest mid-angle. The solid line shows data from the Standard condition, the broken line shows from Condition 5.

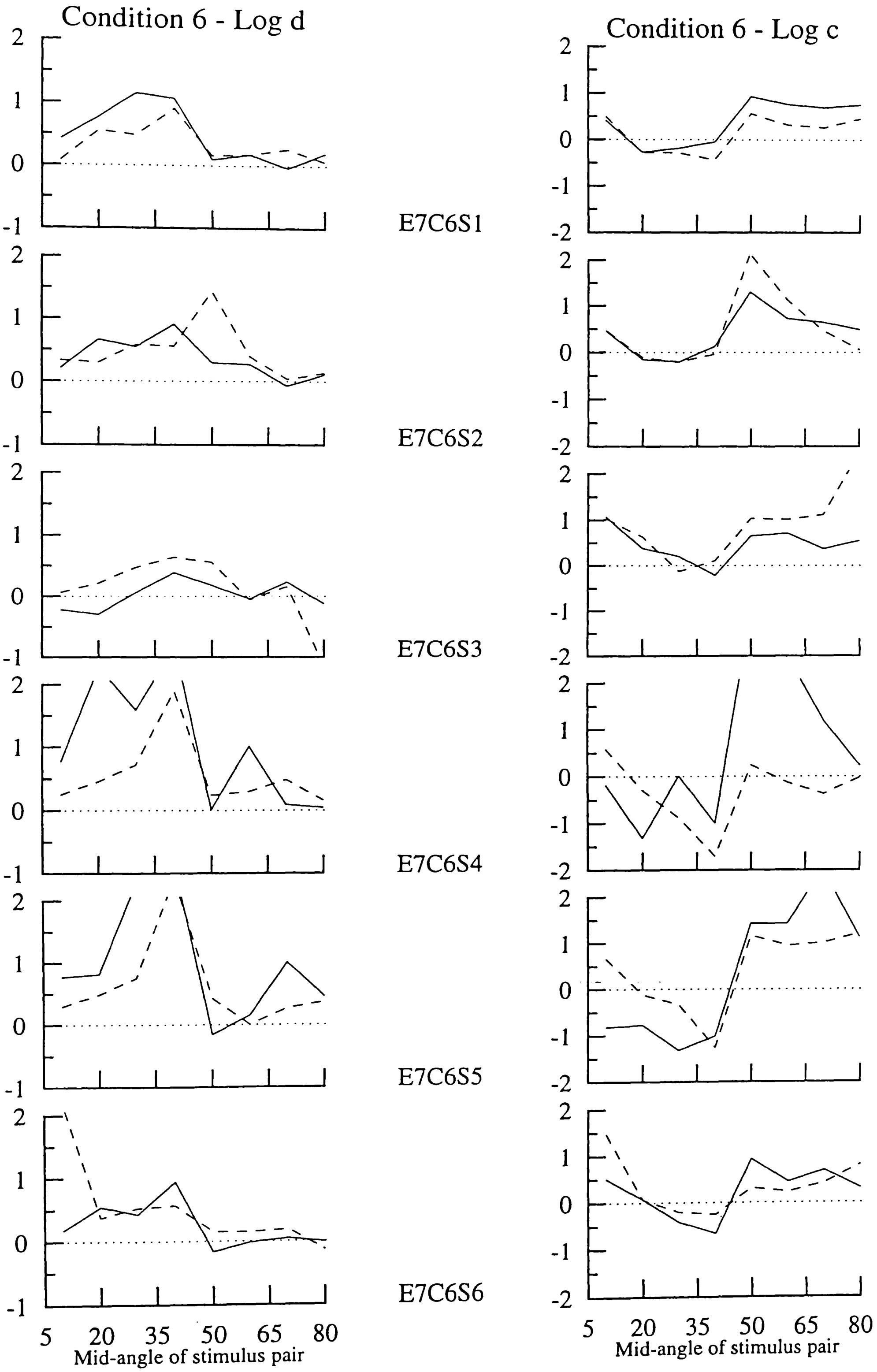


Figure 10-7 For the six subjects Log d and Log c are plotted againest mid-angle. The solid line shows data from the Standard condition, the broken line shows from Condition 6.

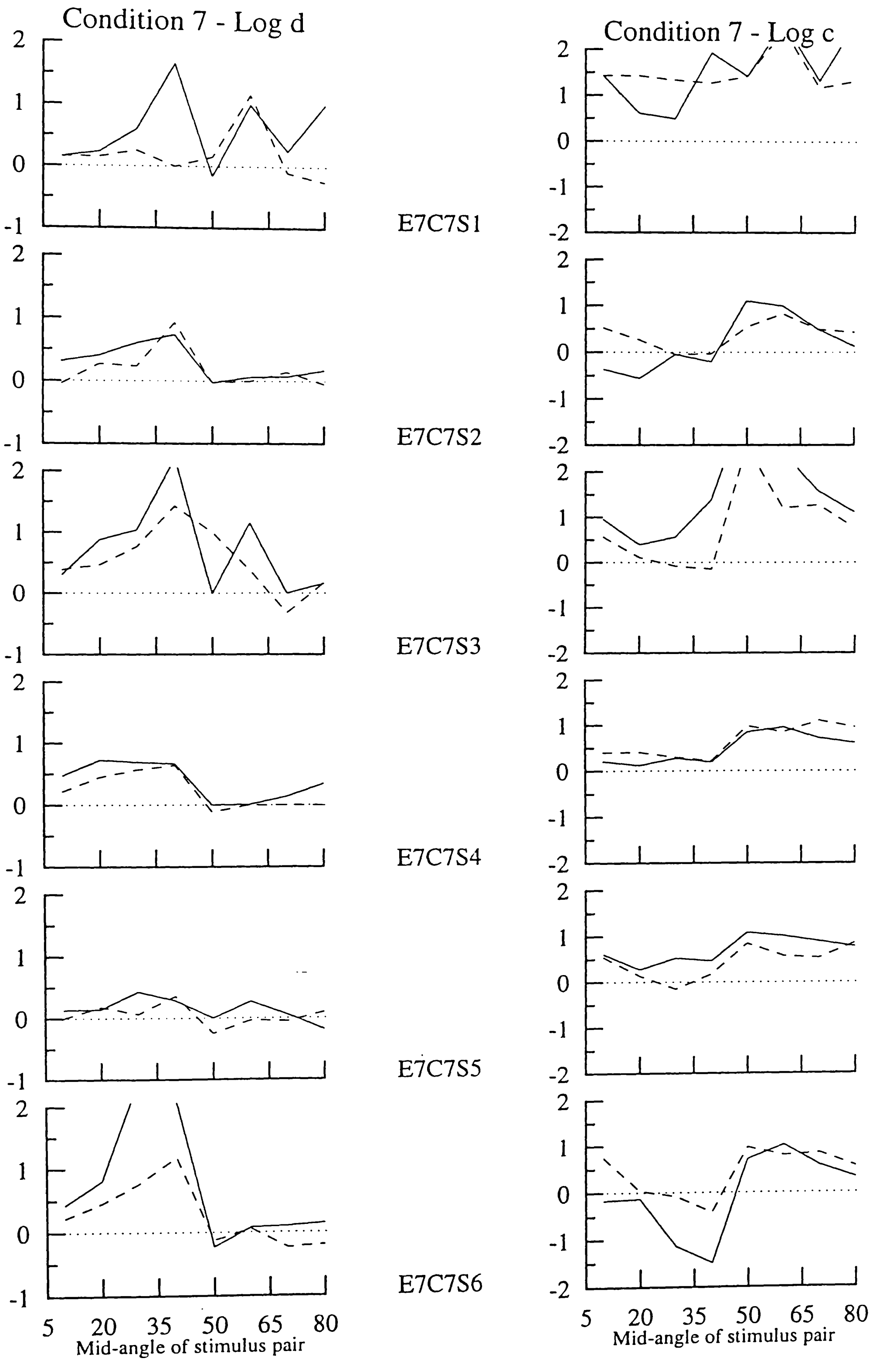


Figure 10-8 For the six subjects Log d and Log c are plotted againest mid-angle. The solid line is for data from the Standard condition, the broken line is for data from Condition 7.

As previously, increasing positive values indicate increasing accuracy on the discrimination task, negative values indicate that the performance is lower than chance, that is, subjects are making the incorrect response to a stimulus on over half the trials. Each figure shows the data from a different condition.

Inspection of these figures suggests that there was a limited number of patterns in the way log *d* changed with mid-angle. In one pattern log *d* increased as the mid-angle increased from 5° to 45° from around 0.0 to a peak above 1.0. As the mid-angle increased further log *d* decreased suddenly to around 0.0, or lower, and generally stayed there with further increases in mid-angle. Therefore, accuracy, as measured by log *d*, increased a lot with mid-angle to 45° and then decreased suddenly to chance levels with mid-angles of over 45°. This pattern, called here Pattern A, can be seen clearly in the data of Subject E7C2S2 in Figure 10-3, for both stimulus sets. Table 10-2 present a summary of the patterns seen in the data for each subject. Some are marked with a '?' to indicate that the data are closest to the designated pattern but some part of the data path deviates from it. The table shows that 36 of the 84 data paths follow Pattern A and another 4 follow most of this pattern, making 40 in total.

Another pattern in the data, called here Pattern B, is very similar to Pattern A with the exception that the increase in log *d* as the mid-angle increased from 5° to 45° is not as great, peaking at less than 1.00. The sudden drop still occurs around 45° and log *d* stays around 0.0 with further increases in mid-angle. This pattern is illustrated by the data from Subject E7C5S2. Table 10-2 shows that it can be seen for 30 of the data paths, with a further 3 following most of this pattern, making 33 altogether.

In the third and last pattern, called here Pattern C, log *d* stays around zero regardless of the mid-angle. Accuracy, as measured by log *d*, was at chance levels and the subjects appear not to have been able to do the task. The data of Subject E7C2S5 illustrate Pattern C. This pattern is seen for eight data paths with a further three following it to some extent, making eleven in total (Table 10-2).

For 33 of the 42 subjects the data paths from both stimulus sets follow the same patterns (Table 10-2). Subjects with similar data paths for both stimulus sets occur across all conditions and there appears to be no tendency for the data paths from the alternative and standard stimulus sets to differ systematically in any condition. This can also be seen in the figures where the dashed and solid lines generally map onto each other. Occasionally, one line is slightly higher or lower for one subject in a condition but there is always another subject for whom the opposite is true. These findings suggest that none of the alternative-stimulus configurations tried made the task either easier or more difficult than the standard configuration. They also show that most subjects could do the task with increasing accuracy as mid-angle increased to 45° but when it increased beyond 45° they could not do the task.

TABLE 10-2

The changes in log *d* with changes in mid-angle classified into three types, as described in the text, for all 6 subjects in the 7 different conditions, on both the standard and the alternative conditions.

CONDITION	SUBJECT	PATTERN standard condition	PATTERN alternative condition
Condition 1	E7C1S1	C	C
	E7C1S2	B	B
	E7C1S3	A?	A?
	E7C1S4	B	B
	E7C1S5	A	A
	E7C1S6	B	A
Condition 2	E7C2S1	B?	B?
	E7C2S2	A	A
	E7C2S3	A?	A
	E7C2S4	C	C
	E7C2S5	C	C
	E7C2S6	A	A
Condition 3	E7C3S1	B	A
	E7C3S2	A	A
	E7C3S3	A	A
	E7C3S4	B	B
	E7C3S5	B	A
	E7C3S6	B	A?
Condition 4	E7C4S1	A	A
	E7C4S2	B	B
	E7C4S3	A	A
	E7C4S4	A	A
	E7C4S5	A	A
	E7C4S6	B	B
Condition 5	E7C5S1	A	A
	E7C5S2	B	B
	E7C5S3	A	A
	E7C5S4	B	A
	E7C5S5	B	B
	E7C5S6	A	A
Condition 6	E7C6S1	B	B
	E7C6S2	B	B
	E7C6S3	C?	B
	E7C6S4	A	A
	E7C6S5	A	A
	E7C6S6	B	C?
Condition 7	E7C7S1	B?	C?
	E7C7S2	B	B
	E7C7S3	A	A
	E7C7S4	B	B
	E7C7S5	C	C
	E7C7S6	A	B

The right-hand side of Figures 10-2 to 10-8 present, for each subject, log *c*, the measure of the bias towards a particular response calculated using Equation 8 for each of the reference-line angles, for the trails with the standard stimuli (the solid line) and for those with the alternative stimuli (the broken line). As previously, values above 0.0 indicate the subjects' responding was biased towards reporting that the lines were parallel and values below 0.0 that the responding was biased towards reporting the lines were different. For 36 of the 42 subjects log *c* changed similarly on both standard-stimulus trails (the solid lines) and alternative trials (the dotted line). Log *c* was 0.0 or less at mid-angles between 5° and 45°. As the mid-angle increased beyond 45° log *c* increased suddenly to some positive value. The size of this increase varied across subjects but not systematically so in any condition. For the alternative-stimulus log *c* followed much the same pattern for the same 36 subjects. It was equal to or less than 0.0 between 5° and 45°, above 45° it increased suddenly (how large it became varied across subjects) and it then remained that size as the mid-angle increased further. For two subjects (E7C1S3 and E7C7S1) the pattern was similar to that outlined above except that the log *c* values were always positive. For these 38 subjects responding was hardly biased at all when the mid-angle was between 5° and 45° and became very biased towards reporting the lines were parallel as mid-angle increased above 45°. For the four remaining subjects (E7C1S1, E7C2S4, E7C2S5, and E7C7S5) log *c* stayed at around 0.0 for all mid-angles. Thus log *c* followed similar patterns for most subjects and for both stimulus sets, regardless of the condition, and there is, therefore, no evidence of any effects of the stimulus configurations on response bias. Response bias did, however, vary with mid-angle.

Proportions correct on the four different trial types are not presented here. They were, however, plotted and examination showed that the patterns in the way log *d* and log *c* changed corresponded to the changes in proportions correct for both stimulus sets and trial types (as in the previous experimental results here). These changes were for proportion correct to be high on "Same" trials across all mid-angles. On "Different" trials proportion correct was high below mid-angles of 45° and then dropped suddenly when the mid-angle was around 45° to remain low for further increases in mid-angle. These changes were the same for both the standard and alternative stimulus trials and for all conditions, as shown by log *c*, overall there was little response bias below 45° and a strong bias towards responding "Same" above 45°, this reflected a tendency for subjects to use the response "Same" on "Different" trials at these larger mid-angles.

Discussion

The majority of experimental results above conform with Pattern A or B. As noted previously, the major difference between the two patterns is whether the peak in accuracy at around 45° is near one or closer to two. It can be argued that even though these two patterns of results are differentiated, this may perhaps be unnecessary. This is because, when using the log *d* analysis with 40 trials, a subject only needs to get one or two more trials correct to move from a log *d* of one to a log *d* of two. As a result, the difference between Pattern A and Pattern B results maybe due to only two or three extra trials being correct at 45°. This differentiation of the two patterns, based on the outcome of only one or two trials of the 1,280 trials in an experimental session may not be warranted. It could be argued that Pattern A and B should be seen as one, so that 74 of the 84 results in this experiment should be considered here as being basically the same sort of result.

This experiment was done, in part, to see what effect the interaction between the edge of the screen and the stimulus lines might have on the accuracy of subjects' responses. To do this the stimuli were configured so that the stimulus lines were either never in contact with the edges of the computer screen, Condition 1, or continually in contact, Condition 6. Results from Conditions 1 and 6 show that both increasing and decreasing the amount of framing had no effect on the results. As a consequence there appears to be no support for the suggestion that subjects' accuracy was effected by the interaction between the stimulus lines and the edge of the computer screen.

This experiment also examined what effect the length of stimulus lines might have on the accuracy of subjects' responses. To do this the stimuli were configured so that the stimulus lines were either the same length, as in previous experiments, or shorter than previously. Conditions 1 and 3 had shorter lines and the "Standard" Condition had a line length the same as those used previously in the "Standard" Condition. Results from Conditions 1 and 3 show that using lines which were shorter had no overall effect on the results. Thus, again there appears to be no support in these results for the suggestion that subjects' accuracy was effected by the length of the line.

The final aim of this experiment was to see what effect the relative positions of stimulus lines might have on the accuracy of subjects' responses, especially as it had already been argued in Experiment 3 that when lines had a common baseline they should be easier to compare. To do this the stimuli were configured so that they had a common baseline, as well as the configuration used in previous experiments. Conditions 2, 3 and 5 had a common baseline for the stimulus lines, the "Standard" Condition had the relative line position used previously. Again, these results show that the relative position of the lines had no effect on the results and there appears to be no support in these results for the suggestion that the subjects' accuracy was affected by the length of the lines.

Overall, the changes in log d with mid-angle were similar for all the different stimulus configurations used here. The manipulations of line-segment length, relative position and the presence of a frame did not systematically effect discriminative performance. For all subjects there were no clear differences between the ways log d changed with mid-angle for the standard stimuli and for the alternative stimuli. It seems that line-slope discrimination was not affected by any of the stimulus manipulations.

The changes in discriminative performance across mid-angles seen here are similar to those seen in Experiments 5 and 6, suggesting that it was not the specific stimulus configuration used in those experiments which led to that particular variation in discrimination with mid-angle. The finding that discrimination increases with mid-angle to 45° and then drops suddenly to 0.0 with increased mid-angle has now been obtained from a large number of the 72 subjects studied here. This suggests that this finding is robust against changes in the stimulus configuration and is highly replicable.

For both the alternative-stimulus and standard-stimulus trials, there was either no bias or one towards responding "different" below a mid-angle of 45°. Both of these changed to large biases towards responding "same" above 45°. The changes in log c with mid-angle were similar across experimental conditions and were the same as in the previous two experiments. Therefore, the presence of the alternative stimuli in the experimental session did not affect the response bias.

In conclusion, it seems that if a subject can do the task at all they become more accurate at reporting whether lines are parallel or not as the mid-angle increases from 5° to 45°. When the mid-angle is greater than 45° subjects are unable to report correctly whether a pair of lines are parallel or not, instead they report that all line pairs are parallel.

Chapter 11 - General Discussion

In the Introduction it was argued that quantification is a critical part of empirical science, and that the quality of any empirical science is governed, in part, by the quality of the data analysis used by scientists to draw conclusions from their data. While numeric methods of data analysis are widely used, graphical methods still provide a legitimate method of data analysis. Even though graphical methods have been used for almost as long as scientists have been collecting data, and are still used extensively today, little research has been done on how best to make use of them. Many of the studies reviewed in the Introduction were poorly done and made little reference to other work in the area. The work reviewed in the Introduction falls into three categories: comparison of graphical methods with tables, comparing different graphical methods, and design variables of a graph. Of these three areas, the least work has been done on the contribution of design variables to the effectiveness of graphs.

The design variable that has had some work done on it is the selection of aspect ratio. Aspect ratio, loosely speaking, is how "low and wide" or "tall and narrow" a graph is drawn. It has been recognised that the selection of aspect ratio affects the appearance of graphs and, as a consequence, selection of aspect ratio possibly affects the conclusions people draw from the data plotted in those graphs. Because of this problem, there has been a number of attempts to find a method for selecting an aspect ratio that is either free of, or at least limits, these problems.

The approach taken in this thesis was to adjust the aspect ratio of a given data plot so that it provided the most accurate possible presentation of the data. This required that we establish the conditions under which a plot provides the most accurate presentation of the data. A number of writers have suggested that the most accurate presentation comes when the aspect ratio is adjusted so that on average the line segments that make up the data path of a line graph are as close as possible to a particular angle. The only suggestion that was supported by emprical research came from Cleveland and McGill (1987), who suggested that on average the line segments should be around 45° because at this angle the difference between any two sloping lines in the data path was greatest. Because this theory for selecting aspect ratio looked plausible, and was potentially very useful, it was decided to replicate Cleveland and McGill's (1987) research. It was envisaged that this theory could be part of a future program of applied research into graphical design variables.

The attempted replication of Cleveland and McGill's study did not produce the group, or individual, results that had been expected. In an attempt to successfully replicate Cleveland and McGill's results, two further experiments were conducted. The first examined the effects of the experimenter's instructions on the performance of subjects. The second was designed to see whether the relative position of the stimulus lines would affect performance. Like Experiment 1, the results from these two further experiments did not replicate those of Cleveland and McGill (1987). However, by doing a single subject analysis of the results it was revealed that some subjects did have results like those presented by Cleveland and McGill (1987), but that results from other subjects were nothing like those expected.

The discoveries from Experiments 1-3 stand as a strong reminder of the value of single subject analysis. It seems very clear that these discoveries were only made possible by the continued insistence that all results be analysed on an individual basis. Looking at what has happened here, it still seems that group data analysis should not necessarily be the norm, especially in a developing area like that presented here. Once more we are reminded of Sidman (1960) when he stated that "The average will be composed of individual measures which reflect non-random differential effects of all the uncontrolled factors in the situation" (p. 164). It is heartening to see that after 30 years those who ignore Sidman's basic arguments still do so at their own cost.

Further analysis revealed that the majority of subjects had a bias towards a particular response. That is, they favoured some responses over others. This led to the discovery that the measure of "mean absolute error" was corrupted by the presence of bias in responding and that with this measure it is not possible to separate bias and accuracy. "Mean absolute error" is not a bias free measure of accuracy and, when a subject has a bias towards one or more of the possible responses, the accuracy measure will start to reflect the bias, not just the accuracy. Because of this corruption, it was not possible to use the data from these first three experiments to examine the accuracy of the subjects on the task.

The "mean absolute error" measure used here is one example of a "magnitude estimate", all of which have in common the idea of measuring the difference between an actual and an estimated magnitude of some dimension of the stimuli. Yet as outlined in the discussion of Experiments 1-3, calculating the difference between actual and estimated values leads to the confounding of accuracy with bias. Therefore, all magnitude estimates including the "mean absolute error" confound accuracy of responding and bias, and should be either never used or only used if there is no response bias.

If a magnitude estimate measure can only be used in the absence of bias then it would seem incumbent upon the researcher to report that they had checked the individual data and that there was no response bias present in their data. Yet Appendix A provides a list of some of the research reports that use some form of magnitude estimate but make no mention of response bias. It is reasonable to argue that any of the data presented in the studies listed in Appendix A could have included some response bias. If this is true then we have no way of telling whether or not the results presented in these studies are corrupted by a confounding between bias and the "magnitude estimate" measure. This problem is not overcome by using group results. It only requires the results of one subject to have a response bias to corrupt the group average results for accuracy. It also follows that the same doubts should be cast over any other research reports that use a "magnitude estimate" measure and that are not guaranteed to be free of response bias.

In the Introduction it was suggested that this work by Cleveland and McGill and others had provided the first clear answer to the question of which type of graphical method is the most effective. Their approach to comparing graphical methods was to discover which "graphical elements" provide the most effective encoding of numbers. For example, the line length was said to be the "graphical element" of a bar graph, and how well people performed on a line length comparison task reflected how well line graphs encoded information. It was felt that this technique avoided the difficulties others had previously had with confounding variables when trying to find a way of comparing graphical methods.

It is very doubtful whether the conclusions from these studies can be of any use, however. The measure of error used by Cleveland and McGill, and later by Simkin and Hastie (1987), was the difference between the judged percentage difference and the actual percentage difference between two stimuli. This is a magnitude estimate measure which, as noted above, is unreliable if there is any bias in the responses of subjects. Because these authors have not checked for response biases, all previous measures of accuracy in graph perception are put in doubt. As a result it is still not possible to conclude which graphical method is superior.

In the second part of this thesis an alternative accuracy measure was used which did not confound accuracy with bias. This measure required that the experimental task be changed from an estimation task, for example, what percentage is one thing of another, and be replaced with a detection task, for example, is one thing the same as another? This enabled the use of log d as a bias free measure of accuracy and log c as a measure of bias.

The results from Experiments 5-7 had significant implications for the general question of this thesis, that the line angle at which subjects were most sensitive to variations could be established. From this, the aspect ratio of a line graph would be selected so that the line segments that go to make up the data path would be positioned, on average, at the line slope which provided the maximum discriminability. Cleveland and McGill (1987) had suggested that maximum discriminability occurred when the line angle was 45°. The results from Experiments 5-7 suggested that the aspect ratio should be set so that, generally, the line slopes that go to make up the data path are never steeper than 45°, but are as close as possible to 45°. This is because line slope sensitivity improved as the angle increased towards 45° but was very poor immediately after 45°.

Having discovered that discrimination performance above 45° is very poor, there is then some difficulty in deciding how to set a graph's aspect ratio. This comes about because of a conflict when, setting an average line angle, between keeping as many lines as close as possible to 45° while making sure as few lines as possible have angles above 45°. It is not at all clear whether this conflict can be overcome. The best suggestion is that the average slope for the data path line segment should be well below 45°, possibly even as low as 25° on average. This approach has the advantage that most of the lines are below 45° and thus very few of them are indiscriminable. It has the disadvantage that most of the lines are well below the 45° point of optimal discrimination. If this recommendation is correct, it implies that all graphs should have aspect ratios that makes them very shallow and very wide, in general more like the upper plot of Figure 1-6.

Given this result, it seems ironic that Cleveland and McGill (1987) demonstrated the problem of aspect ratio, concerning variations in sun spot activity over time (Figure 1-6), by showing an example of a "difficult" aspect ratio in which most of the line segments had slopes greater than 45°. Yet when they did their study of aspect ratio, they looked at only those lines where the angle was below 45°, because variation in accuracy was assumed to be symmetrical around 45°. However, the data from Experiments 5-7 suggest that accuracy is very poor above 45°, a result not predicted by Cleveland and McGill (1987).

Before the above recommendations on aspect ratio can be put into use two further issues must be investigated. First, all the experiments in this thesis used line slopes that had positive slopes, although many line segments in a line graph have negative slopes. Given the unpredicted difference in performance above and below 45°, it seems important to investigate emprically line segments with angles greater than 90° (negative slopes), to see if this marked change in accuracy around 45° is mirrored at 135°. If this change in accuracy is also present at 135°, and in the same direction, then the suggestion that line slopes should be kept below 45° still applies. If a similar change in accuracy does not occur at 135°, it is then very unclear what sort of recommendation should be made because the suggestion that line slopes should be under 45° would apply to only some of the line segments in a line plot. The second issue to be investigated is the ecological relevance. If it is true that the results for positive slopes also applies to negative slopes, then the recommendation for aspect ratios above needs to be tested in a more applied condition to establish how valid it is when used to do "real" graphical data analysis. This seems necessary because we do not know how strong the relationship is between the line slope discrimination tasks and performance on applied tasks, e.g., interpreting trends or seeing differences in different parts of the graph.

From Experiments 4-7 it was found that subjects differed as to how small the difference between a pair of lines was before they were no longer accurate at discriminating whether the lines had the same slope or not. That is, some subjects needed to have big differences between the slopes of lines before they were able to detect those differences; for others, small differences were accurately detected. This was seen most clearly in Experiment 6 where accuracy changed systematically with the true percent, but the true percent at which subjects were no longer accurate varied between subjects. As a consequence of this variation, it needs to be recognised that even if an ideal aspect ratio procedure was found some subjects might still be unable to discriminate patterns in the data even though the aspect ratio has been set for maximum discriminablity.

Conclusion

This thesis has shown that when using line graphs for data analysis the aspect ratio should be set with the line segments as close as possible to 45°, but not exceeding this. This ideal aspect ratio will produce the best possible data analysis performance for any individual, although the level of performance might vary markedly across individuals.

To arrive at these conclusions it was necessary to use a methodology different from that of most of the previous empirical research on graphical methods and perception. As a result, future research in these areas needs to use bias free measures of accuracy and to analyse individual subjects' data and not rely solely on averaged results.

Utilising the methodological advances demonstrated in this thesis it is now possible to examine, or re-examine, other questions of graphical data analysis. For example, the measures log d and log c could be applied to the experiment done by Cleveland and McGill (1985). In this case, an experiment would have subjects looking at pairs of "graphical elements", such as line lengths, and report whether the comparison line was, or was not, 100, 75 or 50 percent of the reference line. This would enable comparison across the different "graphical elements" at a number of slightly different magnitudes. These methodological advances can now be applied to the issues of graphical comparison and graphical design in order to improve graphical data analysis and, thus, the quality of any empirical science reliant upon graphical methods.

References

Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: The "oblique effect" in man and animals. *Psychological Bulletin, 78,* 266-278.

Baum, W. M. (1974). On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental Analysis of Behavior, 22,* 231-242.

Baum, W. M. (1979). Matching, undermatching and overmatching in studies of choice. *Journal of the Experimental Analysis of Behavior, 32,* 269-281.

Beniger, J. R., & Robyn, D. L. (1978). Quantitative graphics in statistics: A brief history. *The American Statistician, 32,* 1-11.

Bertin, J. (1983). *Semiology of graphics* (Berg, W.T. Trans.). Madison, Wisconsin: University of Wisconsin Press. (Original work published 1973)

Bertin, J. (1981). *Graphics and graphic information processing* (Berg, W.J. & Scott, P. Trans.). New York: de Gruyter. (Original work published 1977)

Blough, D., & Blough, P. (1977). Animal psychophysics. In W. K. Honig & J. E. R. Staddon (Eds.), *Handbook of operant behavior* (pp. 514-539). Englewood Cliffs, NJ: Prentice Hall.

Carter, L. F. (1947). An experiment on the design of tables and graphs used for presenting numerical data. *Journal of Applied Psychology, 31,* 640-650.

Case, D. A., & Fantino, E. (1989). Instructions and reinforcement in the observing behavior of adults and children. *Learning and motivation, 20,* 373-412.

Cleveland, W. S. (1984). Graphs in scientific publications. *The American Statistician, 38,* 270-280.

Cleveland, W. S. (1985). *The elements of graphing data.* Monterey, CA: Wadsworth & Brooks.

Cleveland, W. S., Harris, S. H., & McGill, R. (1983). Experiments in quantitative judgements of graphs and maps. *The Bell Systems Technical Journal, 62,* 1659-1674.

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.

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation and application to the development of graphical methods. *Journal of the American Statistical Association, 79,* 531-554.

Cleveland, W. S., & McGill, R. (1987). Graphical perception: The visual decoding of quantitative information on graphical displays of data. *Journal of the Royal Statistical Society (A), 150,* 192-229.

Cooper, J. O., Heron, T. E, & Heward, W. J. (1987). *Applied behavior analysis.* Columbus OH: Merrill Publishing.

Davison, M. C, & McCarthy, D. (1988). *The matching law: a research review.* Hillsdale, NJ: Erlbaum.

Davison, M., & McCarthy, D. (1989). Effects of relative reinforcer frequency on complex color detection. *Journal of the Experimental Analysis of Behavior, 51,* 291-315.

Davison, M.C., & Tustin. R. D. (1978). The relation between the generalized matching law and signal detection theory. *Journal of the Experimental Analysis of Behavior, 29,* 331-336.

Day, R. H. (1973). The oblique line illusion: The poggendorf effect without parallels. *Quarterly Journal of Experimental Psychology, 25,* 535-541.

Dickson, G. C. (1973). *Statistical mapping and the presentation of statistics.* London: E. Arnold.

Dickson, G. W., DeSanctis, G., & McBride, D. J. (1986). Understanding the effectiveness of computer graphics for decision support: A cumulative approach. *Communications of the ACM, 29,* 40-47.

Dooley, R. P., & Harkins, L. E. (1970). Functional and attention-getting effects of colour on graphic communication. *Perceptual and Motor Skills, 31,* 851-854.

Enrick, N. L. (1972). *Effective graphic communications.* Princeton, NJ: Auerbach.

Feliciano, G. D., Powers, R. D., & Kearl, B. E. (1963). The presentation of statistical information. *Audio Visual Communication Review, 11,* 32-39.

Fienberg, S. G. (1979). Graphical methods in statistics. *American Statistician, 33,* 165-178.

Funkhouse, H. G. (1937). Historical development of the graphical representation of statistical data. *Osiris, 3,* 269-405.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior, 4,* 267-272.

Irwin, R. J., & Terman, M. (1970). Detection of brief tones in noise in rats. *Journal of the Experimental Analysis of Behavior, 13,* 135-145

Johnston, J. M., & Pennypacker, H. S. (1980). *Strategies and tactics of human behavioral research.* NJ: Lawrence Erlbaum.

Kline, M. (1953). *Mathematics in western culture.* London: Penguin Books.

Lucas, H. C. (1981). An experimental investigation of the use of computer-based graphics in decision making. *Management Science, 27,* 757-768.

MacDonald-Ross, M. (1977). How numbers are shown. *AV Communication Review, 25,* 359-409.

McCarthy, D. (1991). Behavioral-detection theory: Some implications for applied human research. In M. L Commons, J. A Nevin, & M. C Davision (Eds.), *Signal detection: mechanisms, models and applications* (pp. 239-253). Hillsdale, N.J. Erlbum Associates.

McCarthy, D., & Davison, M. (1980). Independence of sensitivity to relative reinforcement rate and discriminability in signal detection. *Journal of the Experimental Analysis of Behavior, 34,* 273-284.

McCarthy, D., & Davison, M. (1984). Isobias and alloiobias functions in animal psychophysics. *Journal of Experimental Psychology: Animal Behavior Processes, 10,* 390-409

Nevin, J. A. (1991). Signal-detection analysis of illusions and heuristic's. In M. L Commons, J. A Nevin, & M. C Davision (Eds.), *Signal Detection: mechanisms, models and applications.* (pp. 257-274). Hillsdale, N.J. Erlbum Associates.

Parsonson, B. S. & Baer, D. M. (1986). The graphic analysis of data. In A. Poling & R. W. Fuqua (Eds.), *Research methods in applied behaviour analysis: Issues and advances* (pp. 157-186). New York: Plenum Press.

Parsonson, B. S., & Baer, D. M. (1992). The visual analysis of data, and current stimuli controlling it. In T. R. Kratochwill & J. R. Levin (Eds.), *Analysis of single-case research* (pp. 15-40). Hillsdale, N.J. Erlbum Associates.

Phillips, R. J., & Noyes, L. A. (1980). Comparison of colour and visual texture as codes for use as area symbols on thematic maps. *Ergonomics, 23,* 1117-1128.

Rochlin, A. M. (1955). The effect of tilt on the visual perception of parallelness. *The American Journal of Psychology, 68,* 223-236.

Schiano, D. J., & Tversky, B. (1992). Structure and strategy in encoding simplified graphs. *Memory and Cognation, 20,* 1,12-20.

Sidman, M. (1960). *Tactics of scientific research.* New York: Basic Books.

Simkin, D., & Hastie, R. (1987). An information analysis of graph perception. *Journal of the American Statistical Association, 82,* 454-465.

Skinner, B. F. (1953). *Science and human behavior.* New York: Macmillan Company.

Sparrow, J. A. (1989). Graphical displays in information systems: Some data properties influencing the effectiveness of alternative forms. *Behaviour & Information Technology, 8,* 43-56.

Swets, J. A. (1992). The science of choosing the right decision threshold in high-stakes diagnostics. *American Psychologist, 47,* 522-532.

Truran, H. C. (1975). *A practical guide to statistical maps and diagrams.* London: Heinmann Educational.

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

Tufte, E. (1991). *Envisioning data.* Cheshire, CO: Graphics Press.

Tukey, J.W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

Tukey, J. B., & Wilk, M. B. (1988). Data analysis and statistics: Techniques and approaches. In W. S. Cleveland (Ed.), *The Collected Works of John W. Tukey: Vol. 5 Graphics: 1965-1985* (pp. 1-24). Monterey, CA: Wadsworth & Brooks.

Vernon, M. D. (1950). The visual presentation of factual data. *British Journal of Educational Psychology, 20,* 174-185.

Wainer, H., & Reiser, M. (1976). Assessing the efficacy of visual displays. *Proceedings of the 136 Annual Meeting of the American Statistical Association, Social Statistics Section.* (pp. 83-88). Washington, DC: American Statistical Association.

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

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

Watson, C. J., & Driver, R. W. (1983). The influence of computer graphics on the recall of information. *MIS Quarterly, 7,* 45-53.

Wenderoth, P. (1984). The effects of angle-arm length on judgements of angle magnitude and orientation contrast. *Perception and Psychophysics, 36,* 588-544.

Wenderoth, P. & van der Zwan, R. (1984). The effects of exposure duration and surrounding frames on direct and indirect tilt after effects and illusions. *Perception and Psychophysics, 46,* 338-344.

Woodworth, R.S. (1972). *Woodworth and Schlosberg's experimental psychology* (3rd ed.). London: Methuen.

Yeaton, W. H. (1982). A critique of the effectiveness of applied behavior analysis research. *Advances in Behaviour, Research and Therapy, 4,* 75-96.

Appendix A

List of research reports that report magnitude estimates but do not make any mention of response bias.

Cleveland, W. S., Harris, S. H., & McGill, R. (1983). Experiments in quantitative judgements of graphs and maps. *The Bell Systems Technical Journal, 62,* 1659-1674.

Jordon, K., & Schiano, D. J. (1986). Serial processing and the parallel-lines illusion: Length contrast through relative spatial seperation of contours. *Perception and Psychophysics, 39,* 261-266.

Lauer, T. W., & Post G. V. (1989). Density in scatterplots in the estimation of correlation. *Behaviour and Information technology, 8,* 235-244

Schiano, D. J., & Tversky, B. (1992). Structure and strategy in encoding simplified graphs. *Memory and Cognation, 20,* 1,12-20.

Simkin, D., & Hastie, R. (1987). An information analysis of graph perception. *Journal of the American Statistical Association, 82,* 454-465.

Sparrow, J. A. (1989). Graphical displays in information systems: Some data properties influencing the effectiveness of alternative forms. *Behaviour & Information Technology, 8,* 43-56.

Tversky, B., & Schiano, D. J. (1989). Perceptual and conceptual factors in the distortions in memory for graphics and maps. *Journal of Experimental Psychology: General, 118,* 387-398.

Wenderoth, P., O'Connor, T., & Johnson, M. (1986). Expansion and contraction of outline and solid acute angles: Effects of angle magnitude, type of display, and the nature of the matching task. *Perception and Psychophysics, 39,* 261-266.