DO SEPARABLE DIMENSIONS MATTER? THE EFFECT OF GRAPH COMPLEXITY ON ERRORS OF INTERPRETATION

by

Cathy Dudek
M.A. Psychology, Carleton University, 2003

A thesis submitted to
the Faculty of Graduate and Postdoctoral Affairs
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Psychology

Carleton University
Ottawa, Canada

Your file Votre référence
ISBN: 978-0-494-81559-5
Our file Notre référence
ISBN: 978-0-494-81559-5

Abstract

The purpose of this thesis was to better understand some reasons why errors occur when graph perceptual complexity varies. This question is interesting because perceived graph complexity may be more affected by factors other than the amount of information or the pattern of the data. The problem of line graph complexity was addressed here by investigating the extent to which graphical information can be efficiently processed with brief exposure times, when distracter line size, color and orientation varied. The results from four experiments showed that line graphs can be interpreted accurately in 50 ms. The results also showed that orientation, color and size are separable dimensions (Garner, 1974) in a line graph context as evidenced by a lack of interference and also by the adoption of specific processing strategies that are only available when dimensions are separable. Errors of interpretation could be understood in terms of state- and process limitations as well as in terms of parallel and serial processing mechanisms. Parallel processing was assumed to occur when target and distracter lines were consistently oriented due to 'configural superiority effects' (Pomerantz, 1981). Serial processing was expected in inconsistent distracter line configurations due to low discriminability (process limitation) between the two lines (Garner, 1974). Differences in processing strategies appear to be related to the number and nature of the stimulus set permutations as well as the nature of the distracting information. Exploring separable dimension combinations, namely orientation-color, orientation-size and color-size, helped to explain performance efficiency by identifying situations under which state- and process limitations could account for errors and by identifying processing strategies that were most likely adopted to enhance performance.

Acknowledgement

As much as completing this thesis represents a milestone in my life, I could not have done it without the support of so many people. This thesis was not an individual effort but rather an undertaking inspired and guided by some very knowledgeable, dedicated people, whom I would like to thank. First and foremost, I would like to thank my Thesis Supervisor, Prof. Gitte Lindgaard who is one of the most inspirational women I know. I am honored to know you and to have had the opportunity to spend so many years working with you. There are no words to express how much I appreciate all that you have taught me and all the support and guidance you have given me throughout the years. Thank you, Gitte, from the bottom of my heart, I thank you. I would also like to thank the members of my Doctoral Committee: Dr. Justin Hollands (external examiner) for his comments and suggestions and Prof. Ali Arya (internal examiner) for his comments. Thank you to Prof. Chris Herdman for his insights and opinions and also to Prof. Avi Parush for his guidance on this and many other projects. I very much appreciate that you always took the time to help me and to talk with me. Thank you all so much! I would also like to thank all the people who helped me by lending support, guidance and understanding throughout this program. A special thanks goes to the Late Prof. Dick Dillon, whose encouragement meant so much to me. Thank you for everything, Dick. I am also very grateful to Prof. Jo Wood for her willingness to offer her insights and her assistance on various projects and to Prof. Warren Thorngate, who advised my comprehensive project, you both taught me so much. I would also like to thank Prof. Brian Tansley for all the statistics discussions. To Dr. Kate Oakley, Bruce Tsuji and Livia Sumegi, thanks all the 'graph' discussions. I very much appreciate all the support and sometimes wine you provided. And to Sheila Narasimhan, Patricia Trbovich, Adam Bronsther, Greg Dunn, Cassandra Holmes and Rachel White, thank you for your friendship and support throughout the years. A special thanks goes to Etelle Bourassa for all the help she is always so willing to give. A big thank you goes to all the people who participated in my studies. I would also like to thank all the people who supported me, believed in me and loved me enough to put up with me through this program. My sincerest thank you goes to my husband and soul mate, Danny Chennette for his unwavering commitment to me and my pursuits, and for reminding me of how much I am loved no matter what. I love you my sweetheart and thank you for everything. Thank you to my step-children, Jessica Chennette and Matthew Chennette, Melissa Kemp and her new husband Aubrey Kemp for always believing in me and for all your support throughout the years. A very special thank you goes to my parents Joe and Denise Dudek, who have always believed in me and taught me that there is nothing that is unattainable if you are willing to work hard enough for it. Thank you both for everything you have done for me. Another special thank you goes to John and Cara Dudek and Tom and Amanda Dudek for encouraging me and for supporting my goals. Thank you to the whole Chennette family: Rose, Tom Sr., Tom Jr., Norma, Rick, Eileen, Dave, Chris, Janet, Art, Al and families, for offering so many words of encouragement. To my cousin, Susan Horvath, thank you for helping me and standing by me through it all. Thank you to Karen Korb for always being there for me even if it meant talking about graphs, and to my 'big sister' Lynn Thibaudeau for her support and for her willingness to share her wisdom with me. To Sandra McGuire, whom I am privileged to call my best friend, thank you for making me smile and for being an inspiration to us all.

Table of Contents

List of Tables

List of Figures

List of Appendices

Do Separable Dimensions Matter? Effect of Graph Complexity on Errors of Interpretation

People need to communicate information to each other in an efficient and effective manner. One way to condense vast amounts of information is to display them in a graph. According to Pinker (1990), all graphs have a common characteristic even though information may be displayed in many different graph forms. That is, "each graph tries to communicate to the reader a set of n-tuples of values on n mathematical scales, using objects whose visual dimensions (i.e., length, position, lightness, shape, etc…) correspond to the respective scales, and whose values on each dimension (i.e., an object's particular length, position and so on) correlate with the values on the corresponding scales." [p. 74].

Graphs are used to display information and the relationships between variables. Graphs however, can differ in complexity by showing multiple variables and data patterns that vary, sometimes rendering the message in a graph simple to discern, but at other times making the message complex. Complexity is defined as, "the quality of being intricate or compounded" (Webster's Online Dictionary, 2006). Although the term complexity lacks a concrete definition in the graph literature, the word complex is sometimes used to explain why errors in graph interpretation occur. When the graph becomes too complex, peoples' ability to interpret the message in a graph deteriorates, which often leads to errors (e.g., Carpenter & Shah, 1998).

Complexity has been shown to influence graph interpretation, including perceptual complexity (e.g., Lohse, 1993), the complexity of the task (e.g., Ratwani, Trafton & Boehm-Davis, 2003), and the conceptual complexity of the data (e.g., Roth, 2004). To date, much research investigating graph interpretation has focused on how graphs are constructed, but little is known about how people interpret graphical information when perceptual complexity varies. When a graph is perceptually complex, more errors of interpretation are expected to occur than when a graph is perceptually simple. Complexity, however, is an inference that cannot be directly measured.

The purpose of this thesis is to better understand how graphical information is processed by inferring perceptual complexity from performance errors occurring when stimulus exposure times are very short. This is interesting because perceived graph complexity may be more affected by factors other than the number of variables in a graph, or the pattern of the data. Perceptual complexity can be explored by manipulating visual dimensions such as line color, thickness and orientation, which could make the message in a line graph stand out or become undetectable. Graph interpretability was investigated here by focusing on visual variables that, when combined, may lead to increased perceptual complexity.

To determine the effectiveness of a graph, many different performance measures have been employed. Few studies employ subjective measures of graph interpretation such as certainty ratings (e.g., McClure, 1998; Murphy, 2000). Accuracy, response times and certainty ratings may be important indicators of perceived graph complexity; people should respond faster and be more certain when graphs are perceptually simple. These measures were thus chosen for the experiments reported in this thesis.

Combinations of visual variables can be thought of as either separable or integral (Garner, 1974). According to Garner, separable dimensions are those that are perceived independently of one another. Variation in one dimension, therefore, has no effect on the perception of the other. For example, image size and line orientation have been found to be separable in classification tasks (Garner, 1974). However, integral dimensions are not perceived independently of one another. That is, variation in one dimension does affect the perception of the other. For example chroma and value have been found to be integral dimensions in classification tasks. Exploring the separability of dimensions can help better understand when perceptual processing is most efficient and which combinations of visual variables lead to more errors. The visual dimensions of color, orientation and size are explored here because they have been found to influence sorting times in speeded classification tasks (e.g. Garner, 1974) and target identification in search tasks (Treisman & Gelade, 1980). Using a paradigm that combines target identification with classification, this thesis asks if separable dimensions matter for graph complexity and errors of interpretation.

One way to explore the effect visual dimensions have on perception is to investigate when dimensions are processed pre-attentively to signal the message in the graph. Pre-attentively means that incoming information is processed automatically. That may happen when that message is clear, leading to relatively effortless interpretation (e.g., Pinker, 1990; Winn, 1994). Under those circumstances the amount of data in the graph or the pattern in the data may have no effect on graph interpretation.

The efficiency of graph interpretation is related to the apparent complexity of the graph and to the clarity of the message in the graph. It was thus important to study how combinations of visual dimensions are processed to identify which ones may be processed pre-attentively. Some authors argue that perceptual complexity is related to how easily visual dimensions are encoded (e.g., Kosslyn, 1989). That was investigated here by varying line color, size and orientation in line graphs shown very briefly.

Outline of the Thesis

A Chapter reviewing the history of graph research describes the questions that have guided investigations to date and how these have advanced the understanding of graph interpretation. Theories of graph comprehension are discussed in the next Chapter. The major theoretical contributions are discussed with respect to how they account for pre-attentive processing of information and how certain questions drive the search for the message in the graph. These contributions point to some limitations thought to restrict processing capabilities. Next, the Chapter entitled, "Graph Complexity" discusses how different types of complexity apply to graphs and defines the factors under investigation here. The processes thought to be involved in attention are outlined in the Chapter entitled, "Processing of Visual Dimensions". It outlines the relations between dimensions, the strategies that can be adopted for efficient performance, and discusses how and how quickly visual information can be processed. A Chapter summarizing the results of two preliminary studies is then presented. These aimed to determine if graphical information could be processed quickly and accurately. Following that, a Chapter outlining the motivation for the experiments reported in this thesis is presented. It includes a model describing the relations among the concepts tested and shows how these were tested in the experiments. That is followed by Chapters detailing Experiments 1 through 4. All experiments explored which of the visual dimensions tested here, when combined, may affect graph perceptual complexity and thereby performance. All four Experiments explored the effects of perceptual complexity by limiting graph exposure time and measuring accuracy, response times and certainty when the visual dimensions of line color, line thickness and line orientation vary. The Experiments aimed at understanding how these visual dimensions combine to allow parallel processing of information and guide the graph reader to explicitly identify the message in very briefly exposed line graphs. A General Discussion, which includes a summary of main findings and contributions of the research presented in this thesis, is followed by a set of conclusions.

History

This chapter outlines the history of graph research. It documents the questions that researchers have attempted to answer with respect to the presentation and interpretation of graphical data and outlines where this thesis is positioned.

In 1786, Playfair (as cited in Hink, Euastace, & Wogalter, 1998; See Feinberg, 1979; and Wainer & Thissen, 1981, for reviews) demonstrated that line graphs could be used to show trends in statistical data. In 1915, the Joint Committee on the Standards for Graph Presentation (Brinton, 1915) recommended that common standards be applied when constructing and displaying graphs, for example, that the general arrangement of the diagram should proceed from left to right, thereby following the direction of reading in the western world. These recommendations were made primarily with respect to line graphs but with one recommendation for the use of bar graphs. The question that drove this inquiry concerned how data could be depicted.

Eleven years later, Eells (1926) asked which method best represented different components (e.g., categories) of the data. Eells, and other researchers (Croxton & Stryker, 1927; von Huhn, 1927) began to study best methods of graph depiction. The question driving their research was primarily concerning which graphical form is best to convey information accurately and efficiently.

Eells (1926) outlined various objections launched against the pie chart method of display, such as 1) the inability to read it as rapidly and easily as a bar diagram, and 2) the inability of people to interpret a circle graph accurately. The reason given for the inaccuracies was the assumption that people cannot judge areas and chords accurately and that the human eye cannot accurately compare angles and arcs. Eells (1926) tested the opinions of his predecessors by conducting the first experiment comparing circle graphs to stacked bar graphs. Eells found participants were slightly faster at interpreting circle graphs than bar graphs. He concluded that the claims of his predecessors were inaccurate—that circle graphs were not inferior to bar graphs. However, it is worth noting that Eells asked for percentage judgments, which could have been more difficult with bar graphs than with circle graphs. It is possible that different processes are involved in judging lengths of bars in bar graphs and angles in circle graphs and that therefore these two data formats support different types of tasks.

Von Huhn (1927) later identified problems with Eells's (1926) method, including the absence of a scale and labels on graphs. Von Huhn went on to measure the accuracy of ratio judgments in a similar fashion to Eells and found that stacked bars led to better performance than circles. However, Eells asked participants for percentage judgments and von Huhn asked for ratio judgments, so the differences could be attributed to the task rather than the type of graph.

The conflicting evidence for performance with bars and with circles prompted Croxton and Stryker (1927) to investigate the accuracy of percentage judgments between bars and circles when varying the number of sections. When participants judged percentages they found that more errors were made for bars than for circles. One important contribution of that study is the conclusion that no general statement could be made as to which is more desirable—circles or bars, for representing component parts—although it appears that their results slightly favor circles. Another important contribution of their study was the attention paid to the rotation of the angle for the circle graph (centered at 1:30 and 6:00, shown in Figure 1). Participants made more errors when the angle was centered at 6:00 than at 1:30, yet even at the 6:00 angle, participants were more accurate than with the equivalent bar graph.

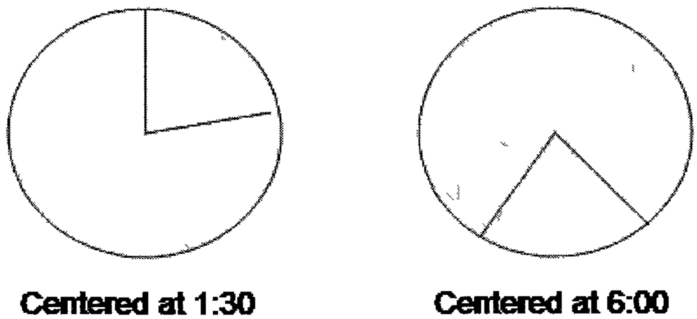


Figure 1. Example of circles shown centered at 1:30 and 6:00.

This finding added a layer of complexity to the question of how best to display data – showing that angle of rotation has an effect on performance.

A further advancement in the quest to display data better was made a few years later when Croxton and Stein (1932) examined graph types based on squares and cubes. For example, participants were shown cue cards with pairs of circles, bars, squares or cubes, and were asked to estimate the size of the smaller one in relation to the larger one. Pairs were arranged either side-by-side (aligned on the bottom), one above the other, or center aligned. The results of that experiment were inconclusive although it did show that there was no difference between bottom- or center alignment. The study was useful because it raised questions regarding how to arrange items to be compared.

These early studies share some commonalities. First, the superiority of one graph type over another was measured primarily in terms of errors. All of the studies asked participants to provide an estimate rather than to choose between estimates provided in, say, a multiple choice format. In those studies, errors could have been due to the necessity to be overly precise. In an effort to ensure that task performance was due to the experimental manipulations and not the response option format, a multiple choice paradigm was adopted throughout this thesis.

These early studies examined error when estimating percentages or ratios. Little attention was paid to the strategies used to interpret the graphs, although Eells' (1926) first study indirectly attempted to evaluate how people judge percentages by asking them about the method they used. The problem with that study, however, was that participants were provided with methodological options from which to select, including areas of sectors, central angles, arcs on circumferences, and subtending chords. That method assumes participants knew and can accurately report which method they used. This thesis investigated how people encode and process information by varying visual dimensions. This was done to determine the source of errors, which in turn could inform the method used to complete the task.

More recently, many authors have continued to explore the best methods for displaying data (Bertin, 1983; Tufte, 1983; and Tukey, 1990). Unfortunately, these authors report little empirical evidence for many of their claims. For example, Tufte (1983) has produced a theory of "data graphics" proposing that the ratio of data to ink on a graph should be maximized. By maximizing the amount of ink dedicated to the data rather than to other graph attributes such as grid lines and icons, for example, the graph becomes easier to interpret. This has become known as the data-to-ink principle. He provides no empirical evidence for his position, as is also true of Tukey's (1990) later work on displaying data. Tukey suggests future directions for graph research as well as for the methods that should be used to display data.

Although the debate regarding the best methods to display data still continues, researchers who have explored the use of bar graphs versus line graphs have generally agreed that bar graphs best represent value identification and comparisons, whilst line graphs provide pattern recognition and trend predictions (e.g., Pinker, 1990; Zacks & Tversky, 1999).

The question of how people interpret and comprehend graphical information is still the focus for many authors (e.g., Carpenter & Shah, 1998; Kosslyn, 1989; Lohse, 1993; Pinker, 1990; Winn, 1994; Zhang, 1996; Zhang & Norman, 1994), as is the question of how graphical elements are perceived (Cleveland & McGill; 1984; Hollands, 2003; Hollands & Spence, 1998; 2001; Lohse, 1993; Simkin & Hastie, 1987). This thesis explored how the perception of visual dimensions may affect interpretation in an effort to better understand graph complexity. The approach of examining errors occurring on only one graph type and exploring only one type of graph reading task can lead to a better understanding of how graphical information is processed. This is because errors can be attributed to perceptual processes only. This approach can lead to a better understanding of the method used to interpret the graph.

Graph Comprehension Process

This Chapter describes the contributions made by graph comprehension authors described earlier. For the purposes of this thesis, ideas that are testable and can be falsified will be referred to as 'theory" (Popper, 1959; Cleveland & McGill, 1984), regardless of whether the authors referred to their ideas as such. With that qualifier in mind, Pinker (1990) developed what I will call a framework, because, although it has many testable segments, his model is not testable or falsifiable as a whole. It is, however, broad enough to accommodate an analysis of other authors' ideas to explore how each contributes to the understanding of graph comprehension and the reasons errors occur. Consequently, Pinker's (1990) framework will be discussed.

The primary objective of this Chapter is to distinguish between the concepts of graph perception, graph interpretation and graph comprehension. To achieve this, Pinker's (1990) model of graph comprehension is discussed with respect to the stages and the processes that guide graph perception, graph interpretation and graph comprehension.

The first section provides an overview of 'graph perception' with reference to a range of models that have been proposed. It aims to show the relationship of these to Pinker's model. Graph perception refers to the physical aspects of the graph that the reader is seeing and the processes involved in encoding the information (e.g., Cleveland & McGill, 1984; Simkin & Hastie, 1987). Next, a section on the 'matching' process is outlined. It describes one of the processes proposed to account for how graph perception leads to graph interpretation. A section on 'graph interpretation' then illustrates that the concept entails extracting meaning from the graph (Kosslyn, 1989; Loshe, 1993; Pinker, 1990; Winn, 1994). That section describes the way the term is used throughout this thesis. The final section on 'graph comprehension' shows that when some meaning is extracted, an iterative cycle is said to ensue, which eventually leads to comprehension, which entails understanding (Pinker, 1990). Evidence for this iterative cycle of graph comprehension has been demonstrated in graph studies exploring eye fixation patterns (Carpenter & Shah, 1998). This thesis is mainly concerned with the perceptual processes that lead to graph interpretation and although the matching process and graph comprehension are discussed briefly, these were not specifically investigated.

Graph Perception

This section discusses how initial perception is argued to affect later interpretation. Graph perception begins with what Pinker (1990) calls the "visual array"—the point at which information is received through the retina, in a two-dimensional form. This initial process of perception is also referred to as the "perceptual image [comprising] basic level graphic constituents" (Kosslyn, 1989) or "visual primitives" (Lohse, 1993; Marr & Nishihara, 1978), and "visual system variables" (Bertin, 1983). Winn (1994) and Kosslyn (1989) argue that the visual array is processed pre-attentively, and that these pre-attentive processes are not open to the influence of a person's prior knowledge. In that pre-attentive view, visual information is first subject to processes Kosslyn refers to as discrimination, distortion, organization and prioritization. Discrimination means that variables can be distinguished from one another. Distortion refers to the way the variables are perceived, which concerns the errors that could occur. For example, the size or length of graphic elements, such as bars, could be underestimated. Prioritization refers to dimensions that take precedence over others, noting that some dimensions are noticed before others. Finally, organization refers to the way objects are grouped together. Winn

(1994) refers to these same processes as discrimination, configuration and detection. Although Winn (1994) describes how people interpret graphical representations, his use of the term configuration means the same as organization, and detection is the same as Kosslyn's notion of prioritization. Both researchers argue that the initial pre-attentive processes are necessarily performed in a bottom-up manner and are hence data-driven. That is, the visual information comes from the graph rather than from information a graph viewer may have stored in memory. This is consistent with Pinker's notion of the visual array. Thus, if a person is unable to detect symbols, discriminate between the symbols, or organize symbols, errors will occur. If indeed graph readers are able to perceive some information pre-attentively, it should be processed automatically with little or no cognitive effort. That is, these pre-attentive, automatic visual processes are said to be responsible for the initial encoding of graphical information (Kosslyn, 1989; Winn, 1994). This is what Pinker refers to as 'default encoding'. Despite these differences in terminology, researchers generally agree on the components of the visual array, its function as well as the initial perceptual processes.

Pinker refers to the second stage in graph perception as the 'visual description'. He refers to the 'visual encoding process' as the mechanism by which the visual description is created from the visual array. The visual description stage serves two functions. First, it provides a structural description of the graph and the visual constraints for the type of graph being viewed. The structural description specifies the relations among the parts, such as bars of varying heights and the visual constraints that specify how the scale on the y-axis, for example, should be used for accurate interpretation. Second, it is the stage at which meaning is ascribed to the various parts such as bars representing values of some variable, which is similar to Kosslyn's (1989) 'semantic' level of analysis. However, according to Pinker, the visual encoding process limits the amount of information that a visual description can contain, which is limited to 7 +/- 2 chunks of information (Miller, 1956). People are believed to have a 'default encoding mechanism' that influences the likelihood of certain aspects of a graph being noticed. The pre-attentive processing said to occur in the initial stage of processing mentioned above, is the result of practice with similar graphs. Thus, familiarity with certain types of graphs may activate a default encoding mechanism that will allow the person to automatically encode the information on the graph. Regardless of the effect of practice, errors are thought to occur as a result of capacity limitations in working memory, which may fail to guide perception accurately.

Visual information may become difficult to detect when the number of visual cues increases (Bertin, 1983; Winn, 1994). A visual cue can represent any item of information shown on a graph, such as the number of lines, or the number of bars, or the colors associated with each. Given this, at least one prediction can be made regarding causes of errors. Graphs with more visual cues will result in more errors than graphs with fewer visual cues because a high number of visual cues in a graph may exceed the capacity of working memory. Increasing the number of visual cues beyond working memory capacity will make it harder to discriminate between them and more difficult to organize them. Unfortunately, the work of Pinker (1990), Bertin (1983), or Kosslyn (1989) does not account for how combining visual dimensions erroneously may result in graph interpretation errors when there are only a few visual cues. For example, graphs showing only two lines could be perceived as complex under some circumstances and simple under other conditions. Assuming that some information is processed pre-attentively with little or no cognitive effort, the number of lines on a line graph may not be the only criterion for ensuring accurate interpretation.

If graphs comprising two lines, of which only one line is of interest, result in poor performance under certain circumstances, one interpretation could be that 'visual noise' may negatively influence perception and cause performance errors. If the line of interest is too similar to the distracter line, the two may not be readily discriminable thus leading to errors. If the two lines differ too much, it is possible that the distracter rather than the target line will be noticed, which could also lead to errors. Evidence has been provided showing that some combinations of visual dimensions are processed easily, for example, those that share the same color (Carswell & Wickens, 1990; Treisman, 1988; Wickens & Carswell, 1995). The experiments conducted here investigated graph perceptual complexity at the perceptual level, which entails the "visual array" and "visual description" stages (Pinker, 1990). This was done to understand where, and explain why, errors may occur in initial perceptual encoding.

Researchers generally agree that visual information is processed to allow it to be 'organized' into chunks. These are susceptible to 'distortion' as people have a tendency to over- or underestimate size and length of bars, for example. The inability to 'discriminate' between visual dimensions could lead to faulty encoding and thus to inaccurate grouping of variables belonging to a category, for example. When a graph displays too much information, the encoding process may become impaired due to inaccurate 'prioritization' of visual cues that should take precedence in processing. Thus, varying visual dimensions such as line thickness, -color and -orientation may affect graph complexity by influencing the processes responsible for accurate encoding. That was explored here.

Matching Process

“Matching’ is a term Pinker borrowed from a theory of long term memory (See Andersen and Bower, 1973) to describe how information gleaned from the visual description informs a stage called, ‘instantiated graph schema’. This ‘matching’ process, Pinker says, compares the visual description to every memory schema for the visual scene, in this case a graph. Pinker defines ‘schema’ as a memory representation of knowledge in some domain. “[It] consists of a description, which contains slots or parameters for as yet unknown information. Thus, a schema can specify both the information that must be true of some represented object or a given class, and the sorts of information that will vary from one exemplar to another” (Pinker, 1990; 95). By ‘instantiated’, Pinker means that every person has unique memory representations, against which incoming stimuli are compared. Pinker was not the first to use the term ‘schema’ (e.g., See Bartlett, 1932), or the notion of ‘instantiated schema’, which is similar in nature to the encoding specificity principle (e.g., See Tulving & Thomson, 1973). Several other researchers also rely on the concept of ‘schema’ to explain how graph stimuli are integrated with memory representations (Carpenter & Shah, 1998; Kosslyn, 1989; Lohse, 1993). Although some errors are likely to occur because graph readers rely on memory representations, which are matched by some process with the incoming stimuli, possibly explained by schema (See McVee, Dunsmore & Gavelek, 2005 for a review), that process is not specifically tested here. This ‘matching’ process is the point at which bottom-up processing meets with top-down processing. Top-down processing refers to the information stored in the individual's long-term memory. Both bottom-up and top-down processing are involved in initial graph interpretation. Pinker's use of the terms 'top-down' and 'bottom-up' is consistent with authors such as Delorme, Rousselet, Mace and Fabre-Thorpe (2004), who investigated these two processes to better understand how they interact at a neural level, when people identify animals in a natural scene.

The outcome of the matching process is related to Pinker's (1990) 'conceptual questions' and 'conceptual messages'. The conceptual question simply refers to the kind of information the reader *wants to extract* from a graph. That is, the reader approaches a graph with a particular question in mind, for example, 'is the stock value increasing?'. The conceptual message simply refers to the kind of information the graph actually *conveys* to the reader. Each graph conveys information in a particular way. That is, the conceptual message in a line graph might be to convey a trend. The conceptual message in a bar graph might be to convey a value. The number of errors of interpretation will depend in part on how easily readers can infer the answers to their conceptual questions from the conceptual messages displayed in the graph. If the conceptual message is unclear and does not match with the readers' conceptual questions, more errors of interpretation could occur. However, if the conceptual message is clear and matches the reader's conceptual questions, and both are always the same, effortless processing could occur, which would facilitate performance.

Graph Interpretation

The point at which graph perception ends and graph interpretation begins is vague but both Pinker and Winn, for example, use the term "interpretation" to indicate the beginning of cognitive top-down processing. The term interpretation has been defined as "a mental representation of the meaning or the significance of something." (Webster's Online Dictionary, 2006). For Pinker, when a graph is perceived and its parts have been processed for long enough to be recognized, interpretation begins. For Winn (1994), interpretation is the cognitive exercise of iterative re-identification and re-configuration of the visual symbols such as shapes, lines and colors until they make sense. An example of re-identification in a graph context would be when a reader re-visits a graph to ensure that the symbols associated with a particular variable were accurately perceived, for example, "Xs represent males". Re-configuration is a process by which the reader ensures that, for example, the symbols associated with a particular variable were accurately grouped together – all the Xs were accounted for. Interpretation is therefore the stage at which the reader attempts to gain meaning from the conceptual message in the graph and if the message is not clear, an iterative process ensues. If the conceptual message is clear and is always the same, interpretation should be effortless, although errors due to faulty perception may still occur. When graph inspection time is limited, more errors are expected to occur due to faulty perception. The term 'interpretation' will be used in the remainder of this thesis because the objective was to understand errors due to faulty perceptual processing of graphs displayed very briefly.

Graph Comprehension

Although the graph literature is rather unclear about the distinction between the terms 'interpretation' and 'comprehension', comprehension can be defined as "an ability to understand the meaning or importance of something" (Webster's Online Dictionary, 2006). This definition indicates that *understanding* is a key component of 'comprehension', which is not a component of 'interpretation'. A further distinction is made by Pinker (1990), who states that three criteria must be met for graph comprehension to occur. First, a graph reader must identify titles and labels that tell the reader what the graph is about, which is what Bertin (1983) calls 'external identification'. The second criterion is identifying relevant visual dimensions such as bars and axes (Pinker 1990), which would represent interpretation. This is what Bertin (1983) calls 'internal identification'. Finally, readers use the information to draw conclusions about the data in the graph. This would map onto what Kosslyn (1989) calls the "pragmatic" level of analysis, as the reader would be able to extract useful information from the graph that s/he could act on. No topics are shown for graphs presented in the experiments reported here, making it impossible for participants to take away any 'understanding' in this sense. Thus, comprehension per se was not investigated here.

Summary

The purpose of this Chapter was to describe previous contributions attempting to explain the processes that take place when readers perceive and interpret a graph and some of the reasons errors occur. The Chapter introduced Pinker's (1990) model of graph comprehension and explained the initial perceptual stages called 'visual array' and 'visual description', which rely on bottom-up processing. It then described how the inferential process commences only if the message is not clear, which necessitates a greater reliance on the top-down processing mechanisms. Graph interpretation was shown to occur when the reader begins to gain some meaning from the graph. This point marks the beginning of the iterative process, which I argue is 'interpretation' that leads to comprehension if the iterative process is allowed to continue. This thesis investigation is limited to understanding errors of interpretation due to faulty perceptual processing and does not investigate the process responsible for the ‘matching’ that occurs between bottom-up and top-down processes or graph comprehension.

Graph Complexity

As the title of this thesis indicates, graph complexity is a central theme. This chapter therefore outlines the concept of complexity and describes what it means in this thesis. Specifically, the research question addresses the issues of how the complexity of a graph may lead to errors; how it affects response times and levels of certainty when tasks and visual dimensions vary. The term complexity is defined first and three ways the concept has been used are then described, focusing mainly on perceptual complexity.

As noted earlier, although the term complexity lacks a concrete definition in the graph literature, the word complex is sometimes used to explain why errors in graph interpretation occur. When graph designers attempt to communicate specific information, a graph may look simple even if the information it represents is complex (e.g., Roth, 2004). So, 'conceptual complexity' relates to the degree of difficulty associated with discerning the conceptual content. For example, a graph reader may understand the graph but not the subject matter of the graph. Conceptual complexity may thus not be affected if the nature of the data is understood. Conceptual complexity is used primarily in the context of graph reading expertise and training (e.g., Roth, 2004), a discussion of which is beyond the scope of this thesis. The notion of conceptual complexity will therefore not be addressed further.

In a graph interpretation context, task complexity refers to the accessibility of the conceptual message given the conceptual question. For example, a graph that shows three variables, where at least one is continuous, is considered complex when the variables are plotted from different perspectives (Carpenter & Shah, 1998; Shah & Carpenter, 1995). Throughout the graph literature researchers generally agree that graph readers seek to complete at least three types of tasks, namely read-off tasks, comparison/relational tasks, and predictions/trend tasks.

'Read off' tasks are considered the easiest of graph comprehension tasks. Thus, for some graph reading tasks, the reader's goal is merely to identify a value associated with a particular variable, which requires attention to precise detail. Although many researchers have called this elementary task a variety of names including, a 'substitute for tables' (read off numbers) (Tukey, 1972), 'exact values' (Pinker, 1990), 'information retrieval' (Zhang, 1996), 'low proximity' (Carswell, 1992), 'point reading' (Lohse, 1993), 'numerical' (Hink, et al., 1998), 'local questions' (Ratwani, Trafton & Boehm-Davis, 2003), and 'read off' (Ratwani & Trafton, 2004), what they mean by their terms is consistent.

The second easiest task is "comparisons" or explorations of relations between variables. Common to these is that the reader's task requires attention to some detail to perform relational judgments, which involve a comparison of two or more values. Again, researchers have labeled this type of task in different ways, including: 'propaganda' – show what has already been learned (Tukey, 1972); 'comparisons' (Hink, et al., 1998; Lohse, 1993; Pinker, 1990); 'comparison tasks' (Zhang, 1996); 'medium proximity (Carswell, 1992); 'multiple search' (Ratwani, et al., 2003) and 'integration' (Ratwani & Trafton, 2004).

Finally, there are predictive tasks, in which the reader is asked to predict the future in some way on the basis of the data presented in the graph, or to describe the direction of a trend. Predictive or trend tasks are thought to be the most cognitively taxing because they involve some amount of extrapolation or summarization. They are referred to as ‘analytical’ (Tukey, 1972), ‘rates of change’ (Pinker, 1990), ‘integration’ (Zhang, 1996), ‘high proximity’ (Carswell, 1992), ‘trends’ (Hink, et al., 1998; Lohse, 1993), ‘global’ (Ratwani, et al., 2003), and ‘inference’ tasks (Ratwani & Trafton, 2004). Although most cognitively taxing because elaboration is required beyond information shown by the specific data points, the trend identification task requires the least attention to specific detail. Trend identification tasks were included in all thesis Experiments. These three task types were explored in the Preliminary Studies conducted to select the best graph task for the graph perception experiments.

‘Perceptual complexity’ is the focus of the remainder of this chapter. It refers to the number and characteristics of visual cues present in a graph make individual variables difficult to identify or compare. For example, the number, color and orientation of lines shown in line graphs could make some lines difficult to detect. There is also evidence that the number of variables and the patterns in the data can affect graph perceptual complexity (Carpenter & Shah, 1998). Thus, the key to the perceptual complexity of a graph seems to reside mainly with the number and the nature of visual cues (Bertin, 1983; Pinker, 1990; Tufte, 1983), which are described next.

Variable Characteristics

According to Bertin (1983), graphs display relations among variables, which he calls components. These are usually displayed in one dimension on a plane. Components can display one variable in relation to the plane such as a bar in relation to the scale of the graph or relations among variables such as a bar in relation to another bar. Each component can comprise several categories. For example, the component ‘sex’ has two categories-male and female. The component 'cars' could have hundreds. For Bertin, one form of graph complexity is linked to the number of categories in a component.

Patterns are represented by the slope of lines, which may be oriented in the same or in different directions (Carpenter & Shah, 1998). The purpose of Carpenter and Shah (1998) study was to understand better how people interpret graphical information when the task required participants to extrapolate from the data shown in the graph. They used eye tracking and a think-aloud protocol to determine where participants looked and what they were thinking while interpreting a line graph. The variables portrayed in the graphs were age, amount of television watched, and vocabulary score. If age and amount of television watched were represented by x and y, the z variable would be vocabulary score. Participants were first shown x and y plotted in a graph as either linear or curvilinear, which were then removed from sight. Next, they were shown graphs depicting x and z. They were then asked to view a comparison graph depicting the same data from another perspective, that is, the x and z variables were reversed. They were asked to identify if the graph was the same as or different from the one just seen. They measured the time it took participants to interpret the graph and make a choice. The researchers also found that it took longer to interpret the meaning of a graph displaying an interaction between variables than when there was no interaction. Because readers detect patterns easily, adding parallel lines did not increase the complexity of that task whereas adding interacting lines did. This suggests that graph complexity concerns not just the number of data points (Bertin, 1983), but is affected by the pattern in the data. The present thesis experiments explored if graph complexity varied when the number of variables and the pattern in the data were held constant.

Visual System Variables

Bertin (1983) identified eight visual variables that can be depicted in a graph. Of these, three are discussed here because they are directly relevant to this thesis, namely, size, color and orientation, which are referred to as dimensions by others (e.g., Garner, 1974). Bertin (1983) refers to these dimensions as 'retinal variables' because their effect on perception will vary depending on the organizational properties they support. Each dimension is important because the number of them that are perceived and required for interpretation is claimed to affect graph perceptual complexity (See for example, Kosslyn, 1989; Pinker, 1990; Tufte, 1983).

Size

Information can be displayed by varying the size of a variable within a graph. Size supports selective perception because larger items stand out from smaller items allowing them to be selected immediately. Size also supports ordered perception. That is, larger items would be perceived as preceding smaller ones according to the ordering principle. Quantitative perception is facilitated by size because larger quantities can be represented using larger sizes and smaller quantities by smaller sizes. Many studies have investigated the effect of size on perception particularly in sorting tasks (e.g., Felfoldy & Garner, 1971, Garner, 1974; Garner & Felfoldy, 1970; Hollands, 2003) and in search tasks (e.g., Treisman, 1985; 1988; Treisman & Sato, 1990). These studies indicate that variation in size makes the perception of other dimensions more difficult and that errors could occur because size may dominate all other perceptual processes. Size is the only visual property that supports quantitative perception. Size facilitates the perception of selectivity, order, and quantity (Bertin, 1983). The effect of varying distracter line size (thickness) in a line graph interpretation task is taken up in this thesis.

Color

Color facilitates association and selectivity of variables (Bertin, 1983). Immediate associations should therefore be clear. The ability to select specific categories should also be supported. Errors could occur because the perception of value dominates color perception. When extrapolating Bertin's theory by applying it to line graphs, one could predict that errors could occur if distracter lines were red, for example, and the target lines black, because red might be more salient. However, red distracters could also enhance performance because they would be perceived as a different category than the black targets. Because the purpose of this thesis was to explore errors of interpretation, it was important to ensure that the task was not too easy. For that reason, target lines were always black just in case red is more perceptually salient. This was done to ensure that the target lines did not stand out more than the distracter lines, which were sometimes red. Some studies that have investigated the effect of color variations in information displays (e.g., Wickens & Andre, 1990) and in maps (e.g., Hastie, Hammerle, Kerwin, Croner & Hermann, 1996) have found that color facilitates the perception of grouping and also discrimination. Those findings support Bertin's notion that color is both associative and selective. The effect of colored distracters on line graph interpretation was investigated in this thesis.

Orientation

Orientation refers to the extent to which an object deviates from parallel in relation to other objects; that is, if all lines are parallel, then the lines that deviate from parallel will be perceived as being different from the others (Bertin, 1983; Treisman, 1985; 1988). Orientation thus facilitates selection. It allows selection because the reader can immediately select an item differing in orientation. For example, changing the orientation of a line can easily be detected; it has been found to influence perception in sorting tasks (Felfoldy & Garner, 1971; Garner & Fellfody, 1970; Pomerantz & Garner, 1973) and in search tasks (Treisman, 1988; Treisman & Gormican, 1988). It also allows association by allowing variables that do, or appear to, belong to the same category to be grouped together (Bertin, 1983). The effect of distracter and target line orientation on graph interpretation was investigated here.

Summary

Three types of complexity were discussed in this section. Conceptual complexity concerns the meaning of the data and was therefore not explored here. Task complexity concerns the amount of mental effort the reader must exert to accomplish a goal. The focus of this Chapter was on perceptual complexity, which concerns the characteristics of the variables displayed on the graph, as well as the way the variables are combined on the graph. Line graphs could differ in perceptual complexity when distracter line orientation, size and color vary, whilst holding constant the number and pattern of the lines, the nature of the task, and the stimulus exposure time.

Processing of Visual Dimensions

This Chapter outlines research relevant for understanding perceptual complexity by showing how the processing of visual dimensions is important to interpreting a graph and why errors can occur if a graph is misperceived. The objective of this thesis research was to understand why errors in interpretation occur and thus how they can be avoided, as well as the extent to which they can be attributed to faulty perceptual processing. When attributing errors to faulty perceptual processing it is important to know how quickly visual information can be processed. One way to do this is to limit stimulus exposure time. In order to understand how errors in perceptual processing can be avoided, it is necessary to identify the sources of perceptual errors.

Perceptual processing errors can be understood in at least two ways. The first way is related to dimension separability. Separable dimensions could give rise to errors due to faulty serial and/or parallel processing mechanisms. Thus, the Chapter begins by defining Garner's (1974) concept of 'integrality' to show how it differs from the concept of separability. The second way errors could be understood is with respect to processing limitations within the person and those of the stimulus itself. Garner (1974) refers to these two sources of errors as state- and process limitations. I will show how the sources of perceptual errors can be revealed by applying Garner's ideas to the present thesis experiments. A section entitled, 'processing efficiency' describes the circumstances that could lead to more efficient processing.

Sources of Perceptual Processing Errors

This section explores some reasons errors may occur due to faulty perception by relying on the work of Garner (1974). The concepts of separable- and integral dimensions are discussed as are the notions of state- and process limitations. The ability to distinguish clearly between the two sources of error should lead to a better understanding of perceptual processing of line graphs.

Separable- and Integral Dimensions

Garner (1974) discusses the concepts of separability and integrality but not in the context of line graphs. I apply his ideas to the line graph here in an effort to understand where errors in graph perception may occur. His approach allows a simple, parsimonious interpretation of this experimental data.

Garner (1974) defines dimensional integrality as the extent to which the existence of one dimension depends on the existence of another dimension. For example, color cannot exist without the dimensions of brightness or hue. Integrality is determined by the occurrence of interference or facilitation in specific stimulus combinations. Garner and Felfoldy (1970) asked participants to sort cards on a specific criterion into two piles as quickly and as accurately as possible. The cards each contained a single colored Munsell chip varying on value and chroma. Value refers to the lightness or darkness of the color. Chroma refers to the color saturation. The color chips varied systematically on these two dimensions or they varied in an uncorrelated fashion. For example, the decks were uncorrelated when high and low value and high and low chroma varied independently from card to card. The time it took participants to sort a deck of cards that varied on both dimensions by, say, value, was then compared to the time it took them to sort one of the control decks of cards. The control decks varied on value, holding chroma constant, or on chroma, holding value constant. They found that it took longer to sort the deck with uncorrelated dimensions when sorting on value, compared to the control deck that varied on value but not on chroma. They refer to this effect as 'interference'. Thus, interference occurred when the stimulus set contained two uncorrelated dimensions, which made variation in one dimension more difficult to perceive when both dimensions varied (Felfoldy & Garner, 1971; Garner & Felfoldy, 1970). This difference in performance led the authors to conclude that value and chroma are 'integral' (See also Sheppard, 1991 for a review). Conversely, facilitation occurs when variation in one dimension makes it easier to perceive changes in the second dimension. For example, using the same stimuli, Garner and Felfoldy (1970) found that sorting speed improved when sorting on either value or chroma when the two dimensions were correlated. The reason for this difference, the researchers argue, is tied to the nature of integral dimensions. Garner (1974) argues that truly integral dimensions are not perceived by dimensional structure at all, which is why participants could not sort efficiently on either dimension unless they varied systematically together, because these were perceived as a single dimension. Thus, when two dimensions are integral, they must both be processed.

The term 'separable' means that perception of one dimension does not depend on the existence of another dimension (Garner, 1974). If no interference or facilitation occurs when the stimuli are orthogonal or correlated, then the dimensions are deemed separable. That is, sorting time with decks varying on both dimensions does not differ from those obtained in a control condition, regardless of whether the stimulus manipulations are correlated. Using that same paradigm, Garner and Felfoldy (1970) found that size and orientation were separable dimensions (See also Felfoldy & Garner, 1971; Garner, 1974). In their experiments, size referred to the graphic representation itself, in their case a circle. A line extending through the center of the circle at varying angles represented orientation. When classifying stimuli on size, participants could ignore orientation (Garner & Felfoldy, 1970) and vise versa, which indicates that the dimensions were processed separately. In other card-sorting tasks, color and form as well as size and lightness, that is 'value', have also been found to be separable dimensions (Garner, 1974). Thus, separable dimensions are processed one at a time; the dimension that is not involved in the classification can be ignored.

The distinction between integral and separable dimensions is occasionally more complex than described thus far. Sometimes the integral relation between the dimensions can be asymmetric. By 'asymmetric integrality' Hollands (2003) means that the relationship varies, depending on which stimulus is the basis for classification, which is consistent with Garner's (1974) definition. Following Garner's card sorting paradigm, Hollands (2003) tested the relations between the cognitive dimension of proportion and the perceptual dimension of size to show their 'asymmetric integrality'. Hollands used cards containing stacked bar graphs or pie graphs varying in size. Again, a control deck was used for each type of graph. When participants were required to sort the stacked bar graphs on the proportion of one of the stacks, performance deteriorated when the size of the graph varied compared with a control condition, in which the graph size was held constant. This provided evidence of interference. However, no performance differences were noted between the treatment- and control conditions when participants sorted the stacked bar graphs on graph size. By the same token, repeating the same procedure for pie graphs showed that performance was undifferentiated from the control, indicating that no facilitation or interference occurred. Hollands (2003) argued that, to judge pie graphs, people use angle and area as the dimensions and they can be processed separately. To judge proportions on stacked bar graphs, participants must use length and area. For a rectangle, these two dimensions are related, as the length of a side will determine the area of the stack. However, to judge pie graphs, the area of the proportion of the pie is independent of the angle of orientation. Thus, area and angle can be processed independently of one another. As in Garner's and Hollands' experiments, a control condition was employed to determine if interference or facilitation occurred when line size, line color and line orientation were manipulated in a line graph context. A condition showing a thin black horizontal distracter acted as a control against which performance on other horizontal-distracter conditions could be assessed when these varied in size and/or color. When the task involves identifying the slope magnitude and direction of a target line in the presence of a distracting line, less interference is expected to occur when the distracter line is oriented consistently with the grid lines on the graph. Thus, line graphs displaying a flat thin black distracter line are argued to represent the least distraction of all the conditions and therefore should produce the best possible performance, as in the control decks of cards that varied only on a single dimension. In this way, the experimental paradigm adopted throughout these thesis experiments was similar to Garner's speeded classification task. The paradigm used here also resembles Garner's in the sense that manipulations of color, orientation and size were always applied to the distracter lines in a correlated fashion. Finally, classification was required but only after discriminating the target from the distracter line.

Although the mixed paradigm used in these thesis experiments is similar to Garner's speeded classification paradigm, some differences should be noted. The first is that each dimension was not tested in isolation to establish a control condition for each. Separability of the dimensions was determined by comparing performance on the flat distracter conditions in which the dimensions of color and/or size were varied. Because orientation differed from that of the target, which was always sloping, orientation varied also. Thus, varying the orientation of the target and the color and/or size of the distracter and comparing performance across flat orientation conditions seemed a reasonable extension of Garner's paradigm to a line graph context.

Garner and Felfoldy (1970) note that the concepts of integrality and separability should be distinguished from the mechanisms responsible for processing. Both serial and parallel processing can occur with separable, but not with integral dimensions. In the latter case, one dimension rather than two is perceived. Hence no opportunity for serial processing exists. The dimensions of line size, line color and line orientation were therefore employed here because they are assumed to be separable and therefore can be processed either serially or in parallel. It is, however, important to note that the results demonstrating both serial and parallel processing in the search task context, could very well be the result of other more elaborate processing mechanisms. Wolfe and colleagues (1989) note that attributing search task results to serial and parallel processing mechanisms should be done with caution. They state that "it may be that all searches are parallel and that the differences between searches lie in capacity limits on different parallel processes" [419]. That is to say that results attributed to serial processing may well be attributable to more complex parallel processing mechanisms (e.g. See Townsend,

1971) or perhaps yet undiscovered processes. A serial processing mechanism was offered to explain the results of search tasks and increasing response time curves. Thus, in these thesis experiments, when manipulating the same dimensions and requiring participants to complete a discrimination task prior to classifying the lines on a graph, the terms serial and parallel were used to explain performance results. Serial and parallel processing was inferred here to offer a parsimonious account of the findings by using the same language as those studies upon which this novel paradigm was based.

Evidence of serial processing of separable dimensions has also been shown in a search task paradigm (Treisman & Gelade, 1980; Treisman & Sato, 1990). Serial processing is necessary in visual search tasks when two visual dimensions define a target as well as the distracters in an array (e.g., Treisman & Gelade, 1980). For example, serial processing is expected to occur if the target is defined as a "brown X" and it is presented with distracters that are "X"s and "L"s and are either brown or green. In that case, information would take longer to process than if the distracters comprised only "X"s or only "L"s or were only green or brown because attention cannot be focused on the letter "X" or on the color brown alone. Instead, each dimension must be considered to identify the target. Performance deteriorated as the number of distracters in the stimulus set increased when targets and distracters were distinguished by two or more dimensions. Treisman and Gelade (1980) note that by analyzing response time curves resulting from search tasks in which set size varied, the separability and integrality of dimensions can be established. These results were interpreted to indicate serial processing of the dimensions and evidence of the separability of them. In search tasks, serial processing thus resulted in degraded performance. One would also expect the same to happen in the kind of line graph interpretation task employed here, as it involved attending to only one of two lines on a graph, thus requiring discrimination. Line graphs containing separable dimensions could result in errors of interpretation if these must be processed serially. This could result if attention must be drawn to the characteristics of each line to distinguish the target from the distracter. Assume, for example, that the characteristics of each line must be processed serially, as Treisman and Gelade's (1980) work on target search suggested. Townsend (1971) argues that there are two ways to distinguish between serial and parallel systems. One is to, "use materials and instructions that will encourage some processing orders over others, especially a single order" [162]. Participants in the present thesis experiments were asked to identify the slope magnitude and direction of a thin black target line with direction changes. When the target line is presented on the same graph with a straight thin black distracter line, oriented oppositely to the target, errors of interpretation could result. This is because attention would have to be focused on each line to determine the straightness and then on the orientation of them, for the target line to be processed correctly. Such a condition was expected to give rise to only one order of serial processing.

Researchers attempt to understand attention and the role of serial and parallel processing in search tasks by measuring response times (Treisman & Gelade, 1980; Treisman & Gormican, 1988; Treisman & Paterson, 1984; Treisman & Sato, 1990; Wolfe, Cave & Franzel, 1989) or by limiting exposure times (Treisman & Gelade, 1980; Treisman & Patterson, 1984; Treisman & Schmidt, 1982). Studies exploring perception (e.g., Bornstein, 1992; Zajonc, 1980) suggest that higher level cognitive processes begin to affect perception after approximately 50 ms has elapsed. Recent research exploring neurological responses of participants viewing scatterplots showed that neural activity occurs after approximately 50 ms of stimulus exposure, which the authors suggest is due to perceptual processing (Best, Hunter & Stewart, 2006). They attribute cognitive processing to neural activity occurring at approximately 300 ms. By limiting stimulus exposure time to a level where processes responsible for encoding are thought to be perceptual only, any differences in performance can be attributed to these- rather than cognitive processing mechanisms. An exposure time of 50 ms was therefore employed throughout the present thesis experiments.

To ensure that traces of a previously exposed stimulus cannot be processed beyond a set exposure time, masking techniques are used (Enns & Di Lollo, 2000). Masking refers to times when "two stimuli are presented in close temporal contiguity; with appropriate adjustment of the interval between the two stimuli, the second stimulus (mask) may interfere with perception of the first (target)" (Schwartz & Pritchard, 1981; p. 678). Enns and Di Lollo (2000) distinguish between various types of visual masking techniques, including pattern masking, which is typified by employing a mask whose spatial contours superimpose those of the target. The target pattern is 'interrupted' by the mask that appears in the same spatial location before the target has been fully processed. In letter identification tasks, pattern masking has been found to yield a 50% accuracy rate at a 100 ms Stimulus Onset Asynchrony with high or medium mask intensity (Enns & Di Lollo, 2000). A pattern-masking paradigm was used here.

State- and Process Limitations

State limitations are due to the stimuli being incompletely represented in the person's mind (Garner, 1974). This could happen, for example, if the person is not alert, is unable to perceive the stimuli in full, or the stimulus exposure time is very limited. State limitations are associated with the context of viewing the stimuli. That is, if the viewing state is completely optimal, for example, with unlimited inspection time, perfect lighting, and perfect discriminability between the stimuli, performance would be expected to be perfect. By contrast, Garner (1974) states that process limitations are due to low discriminability between two stimuli and not to the processing constraints of the person. For example, the figure-ground contrast in the stimulus material may be very low, or lines in a graph may be perceived as identical in appearance such as in color or thickness, even if they differ slightly. The present experiments were designed to test the presence of both state- and process limitations.

When two lines in a graph are both the same color and thickness discriminability between them should be low, particularly when participants must distinguish the target line from the distracter line, as required here. When the distracter line is not discriminable from the grid lines on the graph, the target line should stand out because it would be the only sloping line. Thus, if a horizontal (flat) distracter line is the same orientation, thickness and color as the grid lines on the graph, it should be successfully ignored and participants should be able to identify the sloping target line. If time to view the graph is limited, then errors due to state limitations will be evidenced if performance is less than perfect on the flat distracter condition. Instances of low discriminability should affect performance more in conditions in which the distracter is sloping in the opposite direction than the target line. By Garner's definition, errors that can be attributed to low discriminability should be due to process limitations. Hence, errors due to faulty perception can be parsed into state limitations due to limited exposure times and process limitations due to the complexity of the graph stimuli.

Processing Efficiency

This section identifies circumstances that are likely to lead to more efficient processing strategies than serial processing in a line graph context. The concept of dimension configurability is discussed to show that dimensions can sometimes integrate to improve processing. Selective serial processing' and 'stimulus redefinition' are discussed in an effort to explain how serial processing can be enhanced when visual dimensions are varied.

Configural Dimensions

Parallel processing of information is said to occur when two or more visual dimensions are attended to simultaneously (e.g., Treisman & Sato, 1990; Wickens, 1984). When two visual dimensions are correlated or 'configured' they are perceived together as a unified whole (Pomerantz, 1981; Pomerantz & Garner, 1973; Pomerantz & Schwaitzberg, 1975). The notion of 'configurability' is consistent with Townsend's (1971) definition of parallel processing stating that "by parallel processing, we mean that the entities (hereafter called elements) to be processed are worked on all at once, processing beginning on all simultaneously but possibly finishing at different times" [p. 161]. Configural dimensions are defined by one emergent global dimension that may still allow the individual dimensions to be perceived (Carswell, 1992; Carswell & Wickens, 1990; Pomerantz, 1981; Pomerantz & Garner, 1973; Pomerantz & Schwaitzberg, 1975). When wholes are perceived faster than their parts, the effects on performance are referred to as 'configural superiority effects' (Pomerantz, 1981). Carswell (1992) notes that configural dimensions differ from integral dimensions in that they maintain separate perceptual codes. She also notes that they are not totally separable because the stimuli are coded as a whole rather than separately. Thus, despite the varying ways the notion of configurability has been discussed in the literature, to maintain parsimony throughout these thesis experiments, any process responsible for efficient performance will be referred to as 'parallel'.

Pomerantz (1981) reviewed several studies exploring situations under which dimensions configured and affected performance in selective attention tasks. Participants in Pomerantz and Garner's (1973) study, for example, were required to sort decks of cards, containing four sets of parentheses with varying orientations (e.g., ((, ( ), (---, and ---) in each of four quadrants). Participants were instructed to sort each of the correlated and uncorrelated decks in two ways. For the correlated decks, they had to discriminate between pairs of elements in the two top quadrants, for example, ((and ( ), when only these two elements varied. To sort on these, they had to use the first element in the set that was consistent throughout the deck. Then they sorted by the pairs of elements in the bottom quadrant when the second element was consistent throughout. In the uncorrelated decks, all four dimensions varied. Performance was enhanced when the cards were correlated but degraded when they were orthogonal. An additional experiment showed that sorting time improved when parentheses varied inconsistently, or in an uncorrelated fashion. Thus, the requirement to pay attention to a single item of information led to poorer performance when the items configured than the requirement to pay attention to one item when they did not (Pomerantz & Garner, 1973). Those results showed that when lines are sloping in the same direction, they configured and were therefore processed together. When they were not configured, they can be processed independently of one another.

Carswell (1992) reviewed studies investigating information integration and the role selective attention plays in efficient processing. She explored the results of studies showing when and under what conditions graphical displays or the information on the displays configure to enhance performance. She discusses a study by Carswell and Wickens (1990), for example, who investigated the notion of 'configurability' using various graph types. These included variants on whisker graphs, bar graphs and line graphs. The authors referred to the sets of graphs as either homogeneous or heterogeneous. By homogeneous they mean that the graphs contained elements such as lines *sharing* dimensions, such as length, color and orientation. By heterogeneous, they mean that the graphs contained elements such as lines *varying* in length, color and orientation. By employing a similar speeded classification paradigm as Garner and colleagues (e.g., Garner & Felfoldy, 1970; Garner & Pomerantz, 1973), they tested the integrality of 13 graphical displays and concluded that none of the graph stimulus sets contained integral dimensions. Line graphs were one of the two graph types that contained configural dimensions, and two graph types contained separable dimensions. The other 9 graph types did not fit well with either metric. Homogeneous graph types were more likely to configure than heterogeneous graph types. The authors concluded that these results are consistent with Garner's (1974) notion of configurality as an optional strategy for processing separable dimensions. One would predict that simple line graphs comprising a single target line and a single distracter line should configure and be processed in parallel if they also share the same color and orientation. This was tested here.

The above findings are also consistent with studies in which participants completed tasks requiring either information integration or focused attention (Wickens & Andre, 1990). Participants viewed 'air', 'bank' and 'flap' instruments for an aircraft in a dashboard display or a bar graph. Each task required participants to pay attention to one aspect of the display or to all. They demonstrated that when information integration was required, instruments paired on color, for example, resulted in more efficient processing than when they differed in color. The authors also showed that performance was better on tasks requiring focused attention when different colors were used than when they were all the same. In the target line identification task employed here, focused attention was required to identify the target. One would therefore predict that performance should deteriorate because attention would have to be focused on each line to identify the target when two lines share the same color but not the same orientation. The two stimuli should not configure. One would also predict that, if two lines differ in color and are oriented inconsistently, they should be processed more efficiently than when they are the same color. Thus, using distinct colors should lead to enhanced performance, as in Wickens and Andre's (1990) studies. These conditions were tested here.

Wickens and Carswell (1995), however, note that using homogeneous dimensions results in emergent features in integration tasks as well as the parallel processing of all dimensions. They also explain that heterogeneous display types do not lead to emergent features; although these facilitate parallel processing, they should be used with caution in integration tasks. This is because they may degrade performance, as described by

Carswell (1992). The tasks in these thesis experiments involve only focused attention and thus using heterogeneous manipulations is expected to enhance performance.

Selective Serial Processing

Garner (1974) argues that any set of separable dimensions can contain a dimension that influences or guides perception more than the other. This can lead to a selective serial processing strategy. Garner (1974) defines selective serial processing as "a term meaning simply that the organism processes one dimension before the other, but does so selectively to maximize performance" [p. 132]. Referring to typical card sorting studies, Garner (1974) states that, "if selective serial processing is used, there will be a gain, but it will not be a performance level beyond that provided by the better of the two dimensions alone." [p. 133]. For example, Garner and Felfoldy (1970) and Felfoldy and Garner's (1971) experiments were designed to provide evidence that a selective serial processing strategy was adopted to enhance performance on card sorting tasks with separable dimensions. In both sets of studies, they used the card-sorting paradigm discussed earlier. They found that with high stimulus discriminability and specific instructions, they were able to demonstrate the adoption of a selective serial processing strategy. The authors do note that if stimulus discriminability is very evident, it will maximize the opportunity to identify a selective serial processing strategy, but will minimize the opportunity to observe integration (Felfoldy & Garner, 1971). If stimulus discriminability is almost equal, then the opportunity to observe selective serial processing will be minimized and maximized for integration. Thus, in this thesis, selective serial processing would be evidenced by an improvement in performance over any condition in which a serial processing strategy was used, but performance would not exceed that obtained in the control condition.

The dimensions of size and color have been found to increase the discriminability of stimuli in studies involving card sorting (Garner, 1974) and also in search tasks (Treisman & Sato, 1990). High stimulus discriminability really means that the target 'pops out' of an array in search tasks (e.g., Treisman, 1982; 1985; Treisman & Gormican, 1988). Applying Garner's (1974) notion of dimensional dominance to line graphs would lead to two opposite predictions regarding how differences in target-distracter line size or line color might affect performance when the orientation of target- and distracter lines vary. First, using thicker distracters, for example, may interfere with the identification of a thin target line, thereby degrading performance. Second, it is also possible that, because size is selective (Bertin, 1983), varying that dimension in a distracter line could enhance the discriminability of the target (Garner, 1974), thereby allowing it to be easily identified. These two possibilities were tested here.

Stimulus Redefinition

By stimulus redefinition, Garner (1974) means, "an optional processing mechanism that human information processor may use to redefine the stimuli, so that a new dimension, which is more discriminable than either dimension alone, is provided" [133]. He goes on to say that, "in some cases, the experimental subject may change the two dimensions so they are integral" [p. 133]. For example, in a card-sorting task (Garner, 1974), small black dots were presented in the center, at the top or bottom, or on the edges of the card. The distance between each pair of dots differed. The results showed that sorting time on the correlated decks was as good as for either dimension alone. Thus, Garner (1974) reasons that they must have redefined the two separable dimensions as one integral one because that is the only way sorting speed could have been equal across the control conditions. Efficient processing has been found to occur in search tasks when the characteristics of the targets and distracters were held constant throughout an experiment (Cave & Wolfe, 1990; Wolfe, et al., 1989). These studies showed that targets defined by two or more separable dimensions can be processed in parallel (Wolfe, et al., 1989). The authors describe how attention can be focused almost exclusively on the characteristics of the target when the characteristics of the distracters are held constant throughout the experiment. This efficient strategy of focusing attention on only what is relevant for the task at hand seems consistent with Garner's (1974) notion of "stimulus redefinition". In one of the present thesis experiments, the dimensions of all the distracter lines were always red. This should lead to a strategy in which the two lines become integrated and attention can be focused exclusively on the thin black target line, meaning that it would be unnecessary to process both dimensions in a serial fashion. The adoption of this optional strategy should result in the best possible performance as it did in studies involving both card sorting and search tasks. Stimulus redefinition would be evidenced in the present thesis experiments by undifferentiated performance across all conditions compared to the control.

Summary

The objective of this chapter was to review the literature relevant for understanding the source of errors in graph perception and the processes that could be responsible for overcoming errors. It showed that perceptual errors could be understood in terms of separable- and integral dimensions, because they affect perception differently.

Separable dimensions can be processed in parallel or serially, which could lead to poor performance on the graph interpretation tasks employed here. By limiting stimulus exposure time, perceptual processes can be better understood. Errors can also be understood as being due to state- or process limitations, which are useful for attributing the causes to the processing capabilities of the person and those that are due to the discriminability of the two lines within the graph. Literature describing when visual dimensions are processed efficiently, relying on the notion of configurability, selective serial processing and stimulus redefinition was then presented. These three possible processes were proposed to explain how performance on tasks with separable dimensions could be enhanced. As Garner (1974) and Treisman & Gelade (1980) note, separable dimensions are processed serially, which leads to enhanced performance in card sorting tasks as evidenced by efficient sorting speed. In search tasks, which require focused attention, serial processing leads to degraded performance as evidenced by an increase in response time curves as the number of distracters in the stimulus set increases. It was therefore argued that three possible processing strategies can be adopted to alleviate the requirement of serial processing of separable dimensions in the mixed paradigm adopted here, which entails discrimination and then classification.

Summary of Preliminary Studies

The results of two preliminary studies are summarized here. The objective of the first preliminary study was to discern if participants could complete three graph interpretation tasks when graph stimuli were presented for only 50 ms. The results showed that performance was well above chance levels indicating that a 50 ms stimulus exposure time would be adequate for the formal studies. Details of both preliminary studies are presented in Appendices A and B.

The second Preliminary Study had three objectives. The first was to test the degree to which participants' performance may differ as a function of the type of graph reading task. This was done to determine if three different types of tasks differ in complexity, and, if so, which ones should be selected for subsequent experiments using a 50 ms exposure times. The results showed that of the read-off -, comparison- and trend identification tasks, the trend identification task was completed most accurately, without reaching a ceiling effect (~75% mean accuracy). Details of the tasks are discussed together with the studies, in Appendix B. There was, however, an issue of misalignment of participant-researcher mental models regarding slope magnitude definitions. Some of the line slopes defined by the researcher as 'small' were identified as 'sharp' by more than half the participants, indicating that the two definitions were misaligned. The four problematic graph stimuli were therefore excluded from further study, which eliminated the problem. An error analysis was performed in all Experiments to ensure that no further misalignment of mental models occurred. Since the trend identification task was sufficiently difficult to prevent ceiling- or floor effects from occurring, it was chosen for further exploration.

The second objective of the second preliminary study was to test the effects of two different stimulus exposure times on graph interpretation to determine if performance would improve with longer exposure times. The results showed that performance was best with longer exposure times for all three task types. This result also demonstrated that state limitations were responsible for some errors of interpretation, given that errors in the 50 ms exposure time condition could be attributed to limited exposure times, as discussed in the previous chapter (p. 35). Since the level of performance was adequate with a stimulus exposure time of 50 ms, it was chosen for all four formal experiments.

The third objective was to discern the degree to which two different response modes, namely graphical and textual, may affect performance. A multiple-choice format was used for response options. One format showed four option graphs from which participants chose the graph that best represented the information in the graph previously seen. By contrast, the four textual response options contained only alpha-numeric descriptions of the graph. This was done to determine if participants were relying on the iconic trace from the previously viewed stimulus to make their response in the graphic response mode. No differences were found in the two response modes, indicating that either mode would suffice for the present series of experiments. However, to ensure that lingering iconic trace was not responsible for accurate performance, the textual response mode was chosen for the formal experiments.

The experimental manipulations in the preliminary studies were designed to identify variables that could lead to perceived differences in task complexity and hence improve understanding of their effects on graph perception. The trend identification task was selected for the formal experiments as was the 50 ms exposure time condition, with a textual response mode.

Motivation for the Formal Experiments

The purpose of this Chapter is to describe the rationale for the formal experiments and to introduce a model representing the link between different processing strategies and specific performance outcomes. Details of the formal experiments are outlined first and then a brief review of the concepts related to dimension separability is provided. A model of graph interpretation is presented. It shows how applying the concept of dimension separability can guide our understanding of the reasons errors in perception occur and the experimental manipulations that can account for those errors. Five premises are then presented upon which the rationales for the experimental design in this thesis are based.

Overview of the Formal Experiments

Line graphs containing one distracter and one target line were used in all formal experiments in this thesis. As indicated earlier, the stimulus exposure time was limited to 50 ms throughout. The task was always to identify the slope and the magnitude of a target line presented with a distracter line. The target line was always thin and black with direction changes (i.e. crooked) and sloping positively or negatively. The orientation (positive, negative or flat), color (red or black) and size (thin or thick) of the distracter line varied systematically to enable the tracing of errors as a function of different combinations of dimensions. In the ‘consistent’ condition, the distracter and target lines were both sloping either positively or negatively. In the ‘inconsistent’ condition, one of the two lines was sloped negatively and the other positively. In the ‘flat’ condition, the distracter line was horizontal while the target line was sloping positively or negatively.

The flat distracter condition should require the least amount of effort of the conditions tested here because technically, participants should only notice the target line whilst ignoring the horizontal distracter line, as discussed earlier (p. 37). This would be true in the flat thin black condition as the distracter line was the same thickness, color and orientation as the grid lines on the graph. To determine the separability the three dimensions of line orientation, color and size, performance across all the flat distracter conditions was compared. If performance on the flat thin black distracter condition was found to be indistinguishable from performance in which the distracter line color and/or size varied, that would provide evidence for the separability of all three variable combinations, namely orientation and color, orientation and size and color and size. This is because no interference or facilitation should have occurred in conditions in which the distracter line color and/or size differed from the flat thin black condition.

In an effort to distinguish state- from process limitations, the flat distracter condition was employed as a control. By limiting the stimulus exposure time, participants' ability to fully process the information was reduced. This was referred to earlier (p. 36) as the 'state' in which the graph was presented, which is independent from the complexity of the graph itself. Thus, errors in the flat distracter conditions were taken to be due primarily to the limited exposure time rather than to the distracter line manipulations. This is because the flat distracter was expected to represent the least distraction compared to the other distracter orientations. With limited exposure time, errors are expected to occur in all conditions but the flat distracter condition should be representative of the best performance possible. Therefore performance on flat distracter conditions should represent performance due to state limitations. Additional errors occurring in the consistent- and inconsistent conditions could be attributed to process limitations, brought about by the distracter line manipulations rather than by the limited stimulus exposure time.

Rationale for Line Graph Interpretation Model

The finding that the dimensions of color, size and orientation are separable has implications for performance on the line graph interpretation task adopted in this thesis. The hypothesized performance outcomes on line graph interpretation task employed here and the associated processing strategies that follow from separable dimensions are shown in Figure 2, below.

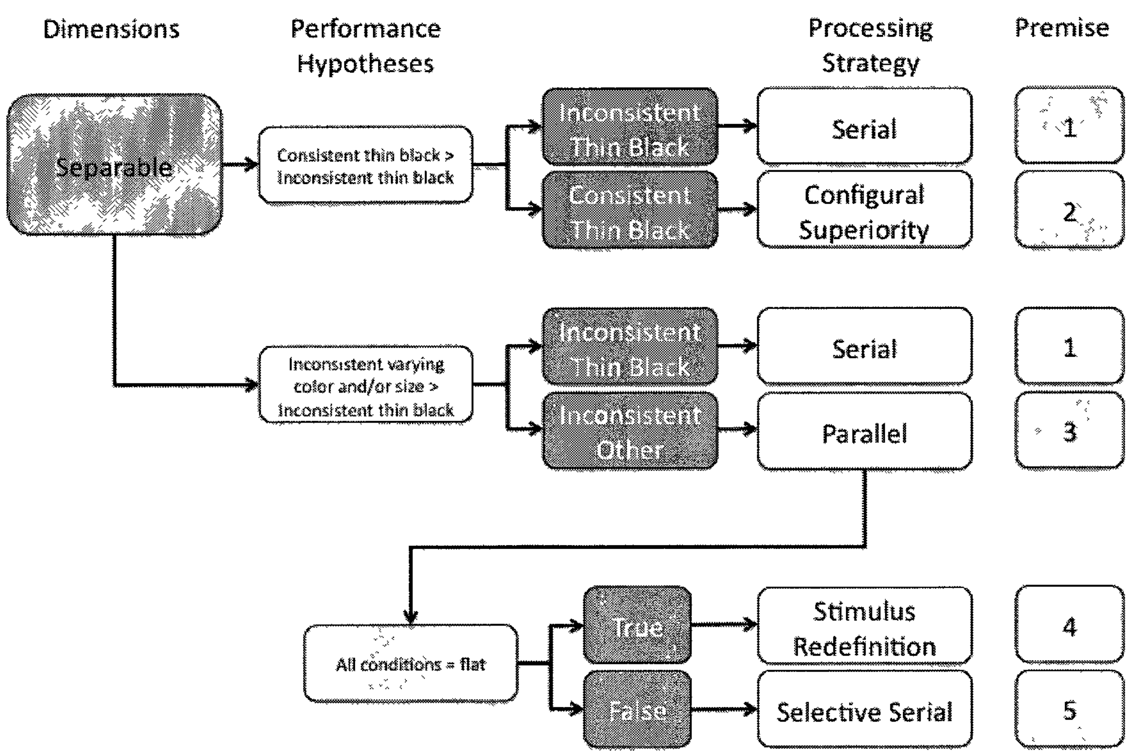


Figure 2. Performance and processing strategies following from separable dimensions

The Model begins with the assumption that the combinations of dimensions tested here are separable. This is indicated in the leftmost column in Figure 2. Next, the possible performance outcomes are presented. The relevant processing strategies are shown with relation to the performance outcomes. The final column shows the relationship of each possible outcome with the premises that are presented next.

Premise 1: Inconsistent thin black → serial processing

When the target line is inconsistently oriented with the distracter line, it should be difficult to discriminate between the two lines, resulting in poor performance. This is because the only dimension separating the target from the distracter is the direction changes (crookedness) in the line. When the target line is not sufficiently distinguishable from the distracter line to allow easy identification, serial processing is expected to occur.

Premise 2: Consistent thin black → parallel processing

When the lines are consistently oriented, they should configure, giving rise to more efficient processing, and hence to better performance, than expected with inconsistently oriented lines. Thus, parallel processing is expected in the consistent conditions due to 'configural superiority' effects (p. 38).

Premises 1 and 2 lead to the prediction that performance would be better in the consistent than in the inconsistent orientations with thin black distracters. This was tested in Experiments 1, 2 and 4.

Premise 3: Inconsistent distracters varying in color or size → parallel processing

A comparison of the inconsistent thin black condition and the inconsistent thin red- and thick black conditions should determine if serial or parallel processing occurred in the thin red- or thick black conditions. Thus, combining premises 1 and 3, leads to two predictions regarding performance when distracter color and size are varied. First, we can infer that parallel processing is used if the red inconsistent distracters improve performance compared to the black inconsistent distracters. This is because serial processing is expected in the thin black inconsistent condition. If performance improves then a different processing strategy could be inferred. If the lines were not processed one at a time, then one possible explanation is that they were processed together, suggesting a parallel processing strategy. This was tested in Experiments 1 and 2. Second, if the inconsistent thick black distracters improve performance compared to the inconsistent thin black distracters, then parallel processing was inferred. Again, because serial processing is expected in the thin black inconsistent condition, any improvement in performance could be attributed to another processing strategy. This was tested in Experiments 1 and 4.

Premise 4: All conditions indistinguishable from the flat → Stimulus Redefinition

If parallel processing is evidenced with thin red or thick black inconsistent distracters then stimulus redefinition would be inferred. Evidence of this would be provided if performance on all conditions is indistinguishable from the flat. This was tested in Experiments 1, 2 and 3.

Premise 5: Not all conditions indistinguishable from the flat → Selective Serial Processing

If performance in inconsistent conditions improve but not all conditions are indistinguishable from performance in the flat condition, then selective serial processing would be inferred. This was tested in all four Experiments.

Experiment 1: Orientation, Color and Size

Since graph perceptual complexity has not previously been tested with line graphs using this limited exposure time paradigm, the purpose of this experiment was to begin to unravel the concept by revealing some sources of graph interpretation errors. Understanding the sources of errors makes it possible to identify whether the dimensions of orientation, size and color are separable in a line graph context. That was done using the flat distracter conditions, as described earlier.

Serial processing will be necessary when the two lines are inconsistently oriented but not when they are consistently oriented. Thus, serial processing was expected to occur in the thin black inconsistent condition, and parallel processing was expected to occur in the thin black consistent condition. Accordingly, Hypothesis 1a predicted that performance would be better in the consistent- than in the inconsistent condition with thin black distracters. Participants should also respond faster (Hypothesis 1b), and be more certain (Hypothesis 1c) in the consistent- than in the inconsistent condition with thin black distracters. This is shown in Figure 3.

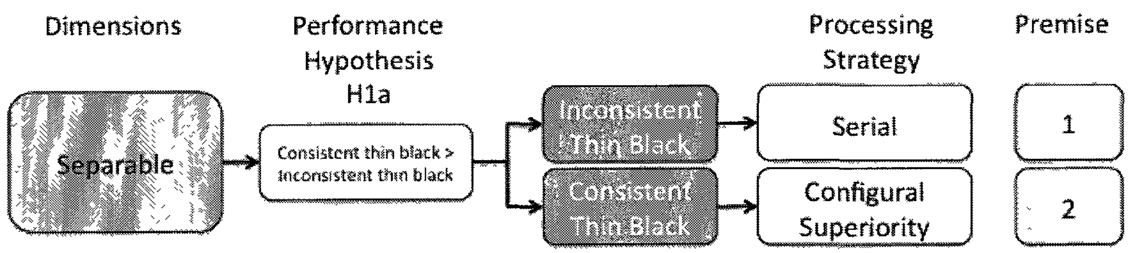


Figure 3. Hypothesis 1a tests performance on thin black consistent against thin black inconsistent distracters

Parallel processing was also expected to occur, in inconsistent conditions, when the target and distracter lines differed on color or in size. To test this, Hypothesis 2a, predicted that performance would be better in the inconsistent condition with thin red distracters than with thin black distracters. This is shown in Figure 4 below. Response times were predicted to be shorter (Hypothesis 2b) and certainty ratings higher (Hypothesis 2c) in the inconsistent condition with thin red distracters than with thin black distracters.

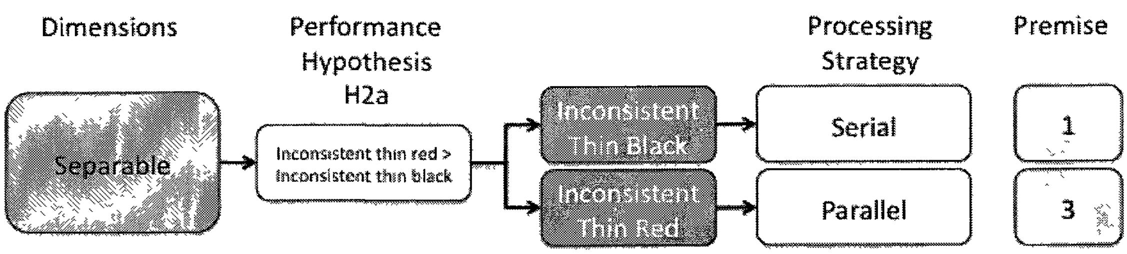


Figure 4. Hypothesis 2a tests performance on thin red inconsistent against thin black inconsistent distracters

Hypothesis 3a predicted that performance on inconsistent conditions with thick black distracters would be better than with thin black distracters, due to parallel processing. This is shown in Figure 5. Hypothesis 3b predicted shorter response times and Hypothesis 3c, higher certainty ratings in the inconsistent conditions with the thick black distracters than with the thin black distracters. The Hypotheses are summarized in Table 1.

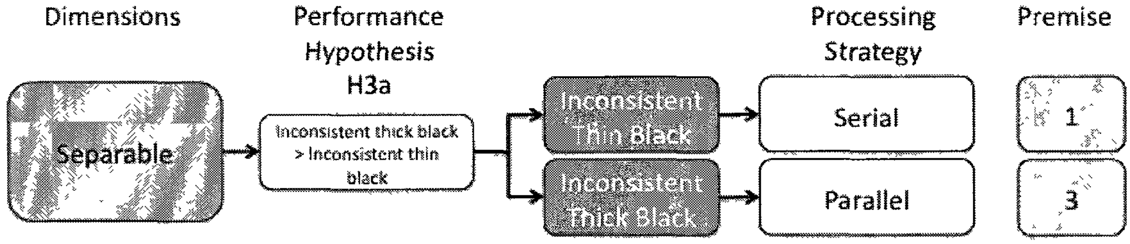


Figure 5. Hypothesis 3a tests performance on thick black inconsistent against thin black inconsistent distracters

Table 1

Hypothesis number, prediction and test used

H #	Prediction	Test Used
1a	Performance will be better in the consistent- than in the inconsistent condition with thin black distracters.	Simple main effects analysis
1b	Response times will be shorter in the consistent- than in the inconsistent condition with thin black distracters.	Simple main effects analysis
1c	Certainty ratings will be higher in the consistent- than in the inconsistent condition with thin black distracters.	Simple main effects analysis
2a	Performance will be better in the inconsistent condition with thin red distracters than with thin black distracters.	Simple main effects analysis
2b	Response times will be shorter in the inconsistent condition with thin red distracters than with thin black distracters.	Simple main effects analysis
2c	Certainty ratings will be higher in the inconsistent condition with thin red distracters than with thin black distracters.	Simple main effects analysis
3a	Performance will be better in the inconsistent condition with thick black distracters than with thin black distracters.	Simple main effects analysis
3b	Response times will be shorter in the inconsistent condition with thick black distracters than with thin black distracters.	Simple main effects analysis
3c	Certainty ratings will be higher in the inconsistent condition with thick black distracters than with thin black distracters.	Simple main effects analysis

Because individual differences in skill levels were expected, short-term visual memory was tested using the Visual Patterns Test (hereafter VPT) (Della Salla, et al., 1997), and field dependence-independence ability was tested using the Group Embedded Figures Test (hereafter EFT) (Witkin, et al., 1971).

Method

Participants

A sample of 10 participants (4 women, 6 men, mean age 21.1 years, range of 18 to 25 years) were recruited from Carleton University. They responded to advertisements posted around campus and were each paid $20.00 for their participation. All were students of Carleton University. All were pre-screened for colorblindness and visual acuity, using a mini eye chart.

Apparatus

Each participant was tested on a Toshiba Satellite laptop with Genuine Intel® Centrino Duo CPU. Screen resolution was set to 1280 by 800 pixels with highest 32bit color. A program created in DirectRT™ was used to present graphs, collect responses, ratings and response times. A stopwatch was also used to time the completion of the Group Embedded Figures Test.

Materials

The materials included informed consent- and debriefing forms, instructions, EFT and VPT Booklets, pens, pencils and erasers. All stimuli were line graphs and comprised two lines: one target and one distracter. The target line showed “small” or “sharp”, “increasing” or “decreasing” slopes. Small changes varied less than 50% between the beginning- and end points and sharp changes varied more than 60% at the beginning- and end points. Examples of line graph stimuli and response options are presented here and all other materials are presented in Appendix C.

Design

This experiment employed a completely within-groups design, shown in Table 2. The stimuli were presented for 50 ms using a backward pattern-masking paradigm. The graph stimuli were completely randomized to avoid serial order effects. Only textual response options were used. Five participants completed the VPT first; the others completed the EFT first.

Table 2

Experimental Design, distracter characteristics (thin and thick line sizes, colored red and black) in consistent (C), inconsistent (I), and flat (F) orientations, all with thin, black targets, with 12 graph stimuli in each condition, all 144 graph stimuli were randomized

		Thin (size)						*Thick (size)*						
		Red			*Black*			*Red*			*Black*			
Test	*P's*	*C*	*I*	*F*	*C*	*I*	*F*	*C*	*I*	*F*	*C*	*I*	*F*	*Test*
		n= 12	n= 12	n= 12	n= 12	n= 12	n= 12	n= 12	n= 12	n= 12	n= 12	n= 12	n= 12	
VPT	P1 … P5													EFT
EFT	P6 … P10													VPT

One hundred and forty-four line graphs were presented, 12 in each condition and of every type, as well as 12 practice graphs. Each graph displayed one target line, oriented positively or negatively, always with one or two direction changes, and one straight distracter line, oriented positively, negatively or flat, as shown in Figure 6 (p. 61). The target line was always presented in black with a standard thickness but distracters were either red or black, thin or thick. Together, targets and distracters were either consistently oriented (both positive or both negative) (n=48), or inconsistently oriented (one positive and the other negative) (n=48), or flat (target positive or negative; distracter always flat) (n=48).

Twelve graphs showed targets with either increasing or decreasing trends; six had sharp- and six had small slopes. The six decreasing trend lines were exact mirror images of the increasing trend lines, yielding 12 target trend lines. These 12 trend lines were reproduced 12 times: once for each distracter orientation (positive, negative or flat), and then again for each distracter line thickness (thin or thick), and again for each distracter color (red or black), yielding the 144 graph stimuli. Participants were asked to choose the textual response option that best described the trend in the graph just presented. An example of a stimulus graph with a sharply decreasing target and increasing thick black distracter is presented in Figure 6 and the textual response options in Figure 7.

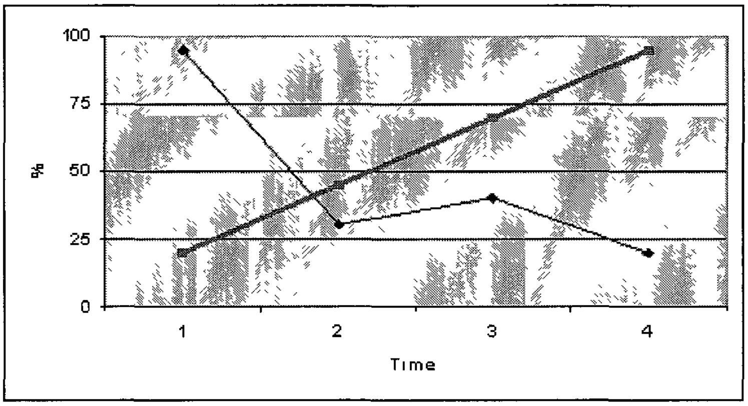


Figure 6. Example of a stimulus graph showing a sharply decreasing target with an increasing (positively sloped) thick red distracter

Small Decrease	**Sharp Increase**
1	**2**
Small Increase	**Sharp Decrease**
3	**4**

Figure 7. Example of textual response options (Correct = 4)

Procedure

Participants were seated and presented with an informed consent form. After reading and signing it, they completed the demographic information sheet and the graph comprehension test, followed by the mini eye test. Next, they were given the Embedded Figures Test or the Visual Pattern Test, depending on the order to which they were assigned. After completing that test, they were instructed to move to the laptop to begin the graph perception study. The onscreen instructions stated,

> "You will see a line graph very briefly. Each graph has two lines: one will have direction changes and one will always be perfectly straight. The one you are to pay attention to is the one with direction changes and not the straight line. The line with direction changes will always be black, whilst the straight ones will be red or black, thick or thin. You are only interested in trend of the crooked black line. You will then be presented with four options and your job is to choose the description of the trend that best describes the crooked black line you saw in the line graph. You will then be asked how certain you are of your choice. When you are done, press the spare bar to move on to the next line graph. You will do this for each graph until you have worked through all the graphs."

Upon seeing the first practice stimulus graph for 50 ms (100 ms SOA and mask for 50 ms), the four textual response options were presented, followed by a text box requesting them to rate their certainty about the correctness of their answer, on a scale from 0 – 100 (0 = not at all certain and 100 = completely certain). When finished, they pressed the 'enter' key followed by the space bar to see the next stimulus. Participants completed the 12 pre-experimental graphs and then all 144 graphs in one randomized block. Participants were encouraged to take a break whenever they desired at any time in the trial. The researcher sat beside the participant throughout. When finished, they completed the VPT or the EFT, whichever they did not complete first. The participant was then thanked for his/her willingness to take part, debriefed, paid and dismissed. The experiment lasted approximately 1.5 hour.

Data Analysis

The percentage of correct responses was calculated first for each condition and for each participant. The same was done for certainty scores. An error analysis was undertaken to identify possible problematic stimuli. Additionally, the average response time was calculated for each participant in each task. Because the dependent measures were the same for all experiments, the data were treated in the same fashion throughout. All experimental data shown include Confidence Intervals calculated using the approach described by Jarmasz and Hollands (2009).

Results

The results section is divided into six sub-sections. The first outlines the results of an error analysis. The second discusses data exploration with regard to assumption testing and outlying scores. The third presents mean performance data and explores the effects of the dimensions of orientation, color and line size (thickness). The fourth part explores the effects of the same independent variables on mean response times. In the fifth section, mean certainty ratings are presented with the same independent variables. The final section briefly addresses the relationship between the two tests (VPT and EFT) and the mean performance, certainty ratings and response times. An alpha level of .05 was used for all statistical tests in these thesis experiments. All descriptive statistics, details of assumption and outlier tests, and error data are included in Appendix C.

Error Analysis

To ensure that no mismatch occurred between the researcher's and participants' mental models, percent correct was calculated for each participant and for each graph.

This was done for each subsequent Experiment as well. A criterion was set and any graph showing a percent correct less than 50% across all distracter conditions would be removed prior to the data analysis. None of the graphs met this criterion in this Experiment or in any of the three remaining Experiments, thus, all the graph stimuli were retained throughout.

Data Exploration

Outliers were defined in two ways: first those scores lying three standard deviations from the mean were considered 'extreme', and second, those scores that might lie closer to the mean but may pull the distribution in one way or another resulting in violations of normality. Thus, it was possible to have an extreme outlier that did not influence the distribution and a not quite so extreme outlier that did have an influence. Thus they could be outlying and influential, outlying and not influential or slightly outlying and influential. If they were outlying and influential, the scores were adjusted. If they were just outlying and did not affect the distribution, they were left alone and included in the analyses. Box plots, stem and leaf plots, frequency, histograms, detrended Q-Q plots and normalized Q-Q plots were used throughout these thesis experiments to identify outliers. Outlying scores were dealt with on a case-by-case basis, generally following the same sequence of testing. When an outlier was identified on performance scores or response times, using one of the above noted methods, its relation to others' scores in that same condition was first examined. If the participant's score differed from those in the same condition, and if it resulted in a violation of normality, the score was adjusted to one standard deviation above or below the mean (depending on whether the score was much higher or lower than the others), unless the adjustment would have resulted in a misrepresentation of the meaning of the outlying score. For example, if many participants show outlying scores on different tasks, this is more likely due to individual differences and task difficulty. If only one participant's score was outlying in any given condition, then that is more likely to be due to some force other than the experiment or to chance. Either way, in that case the score would have been adjusted as described above. That type of adjustment seemed reasonable because it would have been acceptable to replace the score with the mean of the group. Instead, it was replaced by a score that more closely approximated the outlying one by accounting for its place in the distribution – either higher or lower than the mean. In the case of certainty ratings, outlying scores were first compared to the participant's own ratings on other conditions to determine if they were consistent for that person. If the ratings were similar across conditions, the rating was left alone. If it was out of line with that participant's other ratings, the ratings of the group, and it caused a violation of normality, the score was adjusted.

The data for performance, response times and certainty were explored to ensure that assumptions of ANOVA were met and that outlying scores did not adversely affect the results. Beginning with performance data, tests of normality showed a violation in the black thin inconsistent and the red flat thick orientation conditions. An outlier analysis revealed two outlying scores (participants 3 and 10). Both were outlying and influential. They were moved back within one standard deviation higher or lower than the mean for the group. Participant 10's score in the black thin inconsistent orientation was decreased and Participant 3's score in the red thick flat orientation condition was increased.

An analysis of response time data showed that the normality assumption was violated in the condition with red flat thin distracters. Closer inspection showed that four outliers were identified although only one was influential and thus affected normality. Participant 7's response time, in the red flat thin condition was adjusted downward to one standard deviation above the mean.

Finally, certainty data was also explored for assumption violations and outliers. Participant 6 was identified as outlying in 9 conditions. This participant deviated from the group but not from his/her own certainty ratings, which were consistently low. Of the 9 somewhat outlying scores, only 2 were adjusted as they were identified as influential. They were adjusted upward to one standard deviation above the mean but the others were left as they were.

Performance

Figure 8 shows the mean percent correct results for the three orientation conditions, for both color conditions and for both thin and thick distracters. The Figure suggests that performance was more variable across the black thin distracter conditions than the red thin distracter conditions. Within the thin black distracter conditions, performance appeared best for flat distracters, followed by consistently oriented distracters and then for thin inconsistently oriented distracters.

When looking across inconsistent conditions, performance with black thin distracters appeared lower than with the red thin distracters. Similarly, in the same orientation condition, performance with thin black distracters appeared lower than with thick black distracters.

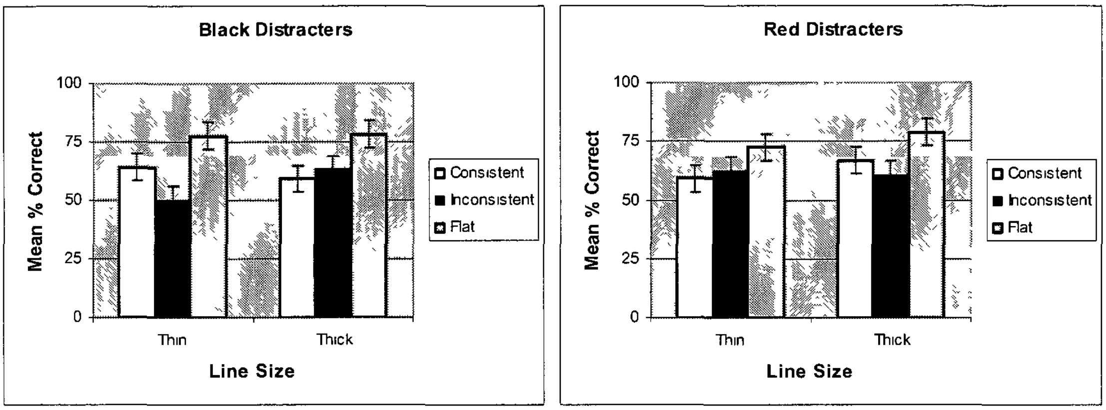


Figure 8. Mean percent correct responses for distracter color (red, black) orientation (consistent, inconsistent, flat) and line size (thin, thick), bars represent *CI*s for 3-way interaction = $M_i \pm 5.76$

To test Hypotheses 1a, 2a and 3a, a (3 x 2 x 2) within groups ANOVA was conducted for mean performance scores on target and distracter orientation (consistent, inconsistent, flat), distracter line size (thin, thick) and color (red, black). The three-way interaction was significant, $F(2, 18) = 3.78, p = .043, \eta_p^2 = .30$, but none of the two way interactions reached significance: orientation by color, $F(2, 18) = 0.96, p = .401, \eta_p^2 = .10$; orientation by size, $F(2, 18) = 0.53, p = .597, \eta_p^2 = .06$; or color by size, $F(2, 9)= 0.05, p = .836, \eta_p^2 = .005$. The main effect for color, $F(1, 9) = 1.42, p = .264, \eta_p^2 = .14$, and size, $F(1, 9)= 3.28, p = .103, \eta_p^2 = .27$, did not reach significance either. However, there was a highly significant main effect of orientation, $F(2, 18) = 20.58, p = .000, \eta_p^2 = .70$, (consistent: $M = 62.29$, $SE = 3.909$; inconsistent: $M = 59.21$, $SE = 3.68$; flat: $M = 76.84$, $SE = 3.01$). To unravel the three-way interaction, simple main effects were calculated.

There was a significant simple effect of orientation in all four distracter conditions namely with thin black distracters, $F(2, 18) = 10.77, p = .001, \eta_p^2 = .54$, with thick black distracters, $F(2, 18) = 6.69, p = .07, \eta_p^2 = .43$, as well as with thin red distracters, $F(2, 18) = 3.88, p = .040, \eta_p^2 = .30$, and with thick red distracters, $F(2, 18) = 10.39, p = .001, \eta_p^2 = .54$.

Comparing the three orientations within the thin black distracter conditions showed that performance on the inconsistent orientation was significantly lower than on the flat orientation, $F(1, 9) = 21.71, p = .001, \eta_p^2 = .71$. The same was true for the consistent orientation, $F(1, 9) = 7.11, p = .026, \eta_p^2 = .44$. Performance on the consistent orientation was marginally higher than the inconsistent orientation, $F(1, 9) = 4.39, p = .066, \eta_p^2 = .33$, thus showing some support for Hypothesis 1a, which stated that performance would be better in the consistent- than in the inconsistent condition with thin black distracters. This finding shows support the notion of dimensional configurality (e.g., Pomerantz, 1981).

Comparing the different orientations within the thick black distracter conditions showed that performance in the inconsistent orientation was significantly lower than in the flat orientation, $F(1, 9) = 11.39, p = .008, \eta_p^2 = .56$, as was performance on the consistent orientation, $F(1, 9) = 13.96, p = .005, \eta_p^2 = .61$. The consistent orientation did not differ significantly from the inconsistent orientation, $F(1, 9) = 0.38, p = .550, \eta_p^2 = .04$. It was thus more difficult to distinguish the target from the distracter line in both consistent and inconsistent orientations than when distracters were flat and both the target and distracters were black, regardless of the thickness of the distracter line.

Comparing the different orientations within the thin red distracter conditions showed that performance did not differ between the inconsistent and flat orientation, $F(1, 9) = 3.12, p = .111, \eta_p^2 = .26$, suggesting that the target may have been highly "discriminable". However, performance was significantly lower in the consistent than in the flat orientation, $F(1, 9) = 14.05, p = .005, \eta_p^2 = .61$. Performance in the consistent orientation however, did not differ significantly from performance in the inconsistent orientation, $F(1, 9) = 0.38$, $p = .555$, $\eta_p^2 = .04$.

Comparing the different orientations within the thick red distracter conditions showed that performance in the inconsistent was significantly lower than the flat orientation, $F(1, 9) = 18.36$, $p = .002$, $\eta_p^2 = .67$, and the same was true for the consistent orientation, $F(1, 9) = 10.35$, $p = .011$, $\eta_p^2 = .53$. Consistent orientations did not differ significantly from the inconsistent orientations, $F(1, 9) = 2.00$, $p = .191$, $\eta_p^2 = .18$.

Looking now at differences for color within orientation and size, the three-way interaction was further explored to test Hypotheses 2a and 3a and to identify if differences exist between only the flat distracter conditions. The only significant differences for color in simple simple effects were found in the inconsistent orientations with thin distracters, $F(1, 9) = 5.96$, $p = .041$, $\eta_p^2 = .39$, showing that performance was better with red than with black distracters. This supports Hypothesis 2a, which stated that performance would be better on inconsistent conditions with thin red distracters than with thin black distracters. This finding shows support for Garner's (1974) theory regarding the discriminability of the color dimension and shows support for the notion of selective serial processing. This finding also indicates that flat orientation conditions did not differ from each other despite distracter color and size manipulations.

To test Hypotheses 3a, asserting that performance in inconsistent conditions with thick black distracters would be better than with thin black distracters, simple simple main effects of line thickness at each color and each orientation confirmed this, $F(1, 9) = 7.53$, $p = .023$, $\eta_p^2 = .46$. This supports Hypothesis 3a and provides further evidence for the discriminability of dimensions (Garner, 1974) and selective serial processing. No other differences were found, which also indicates that the flat distracter conditions did not differ despite variations in distracter color and size. A simple simple main effects contrast compared inconsistent thick red to inconsistent thin black distracters. The difference was only marginally significant, $F(1, 9) = 3.90, p = 0.08$.

Response Times

From Figure 9, it can be seen that mean response times for red and black distracters did not vary as much in the consistent orientation as they did in the inconsistent- and the flat orientations.

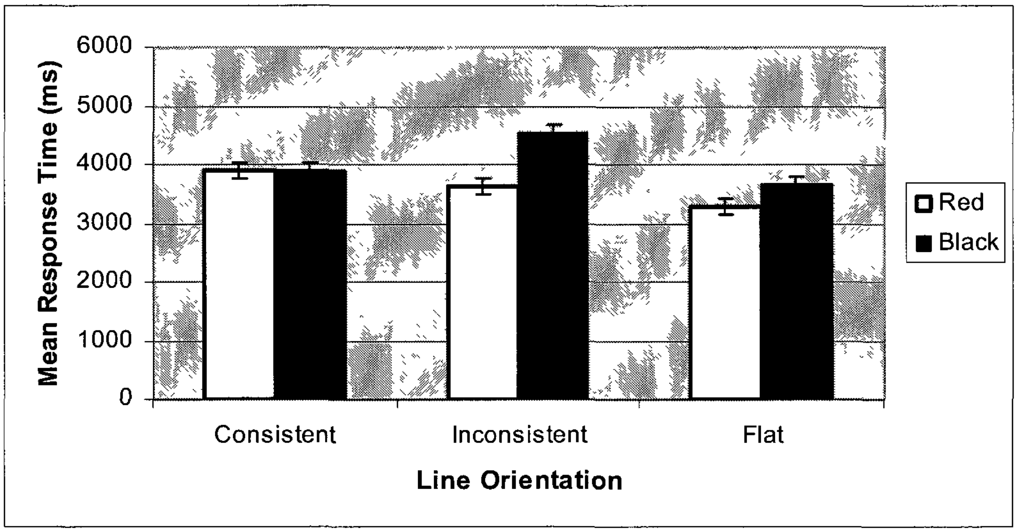


Figure 9. Mean response time (ms) for distracter color (red, black) and orientation (consistent, inconsistent, flat), bars represent *CIs* for 2-way interaction = $M_i \pm 136.26$

To test Hypotheses 1b, 2b and 3b, a (3 x 2 x 2) within groups ANOVA was conducted for response times on target and distracter orientation (consistent, inconsistent, flat), distracter size (thin, thick) and color (red, black). The three-way interaction was not significant, $F(2, 18) = 0.32, p = .730, \eta_p^2 = .03$. However, the two-way interaction of orientation and color was significant, $F(2, 18) = 5.79, p = .011, \eta_p^2 = .39$. As Figure 9 shows, response times for black distracters were considerably longer than response times for graphs with red distracters only when orientation was inconsistent.

Tests of simple main effects of orientation within black distracters showed that responses to inconsistent orientations took longer than to flat orientations, $F(1, 9) = 15.17$, $p = .004$, $\eta_p^2 = .63$, and longer than consistent orientations, $F(1, 9) = 6.24$, $p = .034$, $\eta_p^2 = .41$. However, responses to consistent orientations did not take longer than they did for flat orientations, $F(1, 9) = 1.00$, $p = .344$, $\eta_p^2 = .10$.

Tests of simple main effects showed that within red distracter conditions, consistent distracter orientations took longer than flat ones, $F(1, 9) = 6.85$, $p = .028$, $\eta_p^2 = .43$. However, inconsistent and flat orientations did not differ, $F(1, 9) = 1.85$, $p = .207$, $\eta_p^2 = .17$. Similarly, comparing consistent and inconsistent orientations revealed no differences, $F(1, 9) = 1.30$, $p = .284$, $\eta_p^2 = .13$.

The two-way interaction of orientation and line size did not reach significance, $F(2, 18) = 2.66$, $p = .097$, $\eta_p^2 = .23$, but the interaction of color and size was marginally significant, $F(1, 9) = 4.87$, $p = .055$, $\eta_p^2 = .35$. These results are shown in Figure 10.

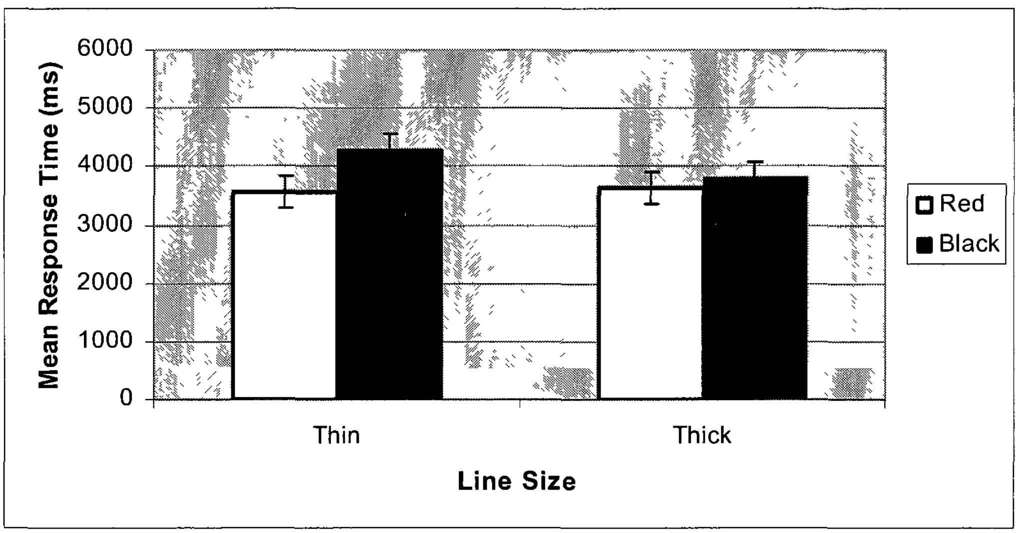


Figure 10. Mean response time (ms) for distracter color (red, black) and line size (thin, thick), bars represent *CIs* for 2-way interaction = $M_i \pm 276.31$

Tests of simple main effects were conducted for size within red and black distracter conditions. The results showed that there was a difference between response times to thin and thick distracters in the black distracter condition, $F(1, 9) = 7.87$, $p = .021$, $\eta_p^2 = .47$, showing that responses to thin black distracters took longer than to thick black distracters. However, distracter thickness within the red distracter condition did not contribute to response time differences, $F(1, 9) = 0.19$, $p = .677$, $\eta_p^2 = .02$.

Looking at simple main effects of color within size showed that it took significantly longer to respond to black thin distracters than to red thin distracters, $F(1, 9) = 6.51$, $p = .031$, $\eta_p^2 = .42$, but not when they were thick, $F(1, 9) = 3.74$, $p = .085$, $\eta_p^2 = .29$.

There was a main effect for orientation, $F(2, 18) = 4.91$, $p = .020$, $\eta_p^2 = .35$. Pairwise LSD comparisons showed that responses in flat distracter conditions were significantly faster than in the inconsistent conditions ($p = .006$) and were only marginally faster than in the consistent conditions ($p = .061$) (see Figure 9, p. 66). There was also a main effect of color, $F(1, 9) = 7.10$, $p = .026$, $\eta_p^2 = .44$, which showed that participants responded faster when distracters were red than when they were black. The main effect of size was only marginally significant, $F(1, 9) = 3.97$, $p = 0.077$, $\eta_p^2 = .31$, and was thus not explored further.

As the three-way interaction was not significant, tests of the specific hypotheses were not completed during the initial exploration of the ANOVA findings. Tests of the specific hypotheses were therefore conducted as planned comparisons. Hypothesis 1b stated that responses to consistent- would be faster than for inconsistent conditions with thin black distracters. The results of these simple simple comparisons showed that responses for consistent thin black distracters were faster than those for inconsistent black distracters, $F(1, 9) = 7.17$, $p = .025$, $\eta_p^2 = .44$, showing support for Hypothesis 1b and therefore the configurability of dimensions (e.g., Pomerantz, 1981).

Hypothesis 2b stated that responses would be faster in inconsistent conditions with thin red distracters than with thin black distracters. Results of simple simple comparisons showed that there were significant differences between red and black thin distracters, in inconsistent conditions, $F(1, 9) = 6.86$, $p = .028$, $\eta_p^2 = .43$, with black taking longer. This finding shows support for Hypothesis 2b and therefore the discriminability of dimensions (Garner, 1974).

Finally, Hypothesis 3b stated that responses in inconsistent conditions would be faster with thick black distracters than with thin black distracters. Tests of simple simple comparisons showed that the difference was marginally significant, $F(1, 9) = 4.67$, $p = .059$, $\eta_p^2 = .34$. Responses were marginally faster to thick black compared to thin black inconsistently oriented distracters, thus showing some support for Hypothesis 3b.

Certainty

Certainty was lowest and most variable in the thin black inconsistent condition, with less variability in the other two distracter conditions. Flat distracter conditions appear to result in the highest certainty. This is shown in Figure 11.

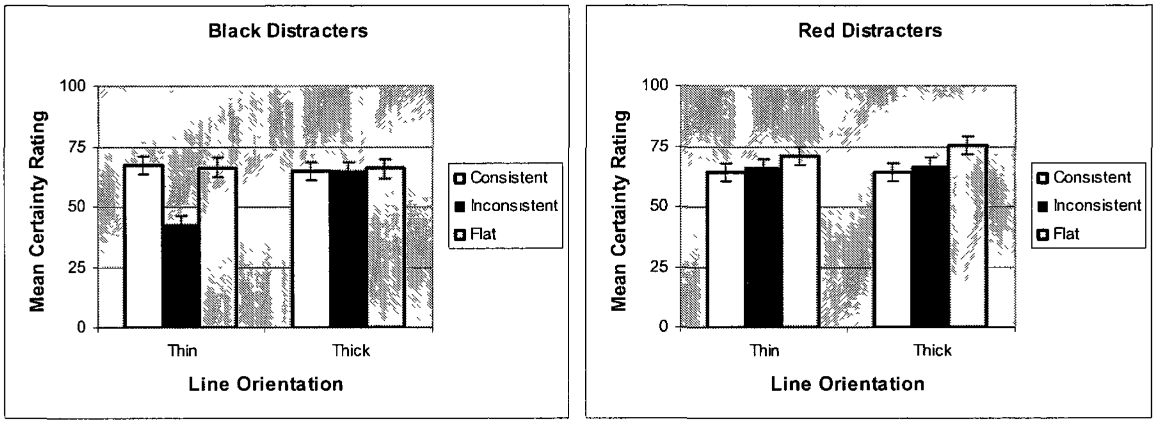


Figure 11. Mean certainty rating for distracter color (red, black), orientation (consistent, inconsistent, flat) and line size (thin or thick), bars represent *CI*s for 3-way interaction = $M_i \pm 3.76$

To test the Hypotheses 1c, 2c and 3c, a (3 x 2 x 2) within groups ANOVA was conducted for certainty ratings on target and distracter orientation (consistent, inconsistent, flat), distracter size (thin, thick) and color (red, black). The three-way interaction was significant, $F(2, 18) = 16.44, p = .000, \eta_p^2 = .65$, and so were the two-way interactions of orientation and size, $F(2, 18) = 8.3, p = .003, \eta_p^2 = .48$, as well as orientation and color, $F(2, 18) = 9.37, p = .002, \eta_p^2 = .51$. The interaction of color and size did not reach significance, $F(1, 9) = 3.60, p = .090, \eta_p^2 = .29$. The main effect of size was highly significant, $F(1, 9) = 86.06, p = .000, \eta_p^2 = .91$, showing that certainty was higher when distracters were thick ($M = 66.96$, $SE = 4.59$) than when they were thin ($M = 62.90$, $SE = 4.80$). The main effect of orientation was also highly significant, $F(2, 18) = 14.46, p = .000, \eta_p^2 = .61$. Pairwise LSD comparisons showed that certainty was rated higher when they were oriented consistently ($M = 65.10$, $SE = 4.86$) than when they were oriented inconsistently ($M = 60.10$, $SE = 5.08$) ($p = .010$), but the consistent orientations were rated significantly lower than the flat ($M = 69.59$, $SE = 4.42$) ($p = .035$). Certainty was rated significantly higher for flat distracter orientations than for inconsistent orientations ($p = .001$). A main effect of color, $F(1, 9) = 14.81, p = .004, \eta_p^2 = .62$, showed that certainty on graphs with red distracters was rated higher than on graphs with black distracters.

Drilling down into the three-way interaction, tests of simple simple effects were conducted. Tests exploring orientation within size and within color showed that certainty differed significantly within thick red, $F(2, 18) = 7.24, p = .005, \eta_p^2 = .45$, and thin black, $F(2, 18) = 49.07, p = .000, \eta_p^2 = .85$; but not within thick black, $F(2, 18) = 0.07, p = .935, \eta_p^2 = .007$, or thin red distracters, $F(2, 18) = 2.04, p = .159, \eta_p^2 = .18$.

Simple simple comparisons between conditions with thin black distracters showed that certainty for inconsistent conditions was lower than for flat conditions, $F(1, 9) = 59.68, p = .000, \eta_p^2 = .87$. However certainty did not differ between consistent and flat conditions, $F(1, 9) = 0.09, p = .768, \eta_p^2 = .01$. Certainty regarding inconsistent conditions was lower than consistent conditions, $F(1, 9) = 84.44, p = .000, \eta_p^2 = .90$. This supported hypothesis 1c, which stated that certainty ratings would be higher in the consistent- than in the inconsistent conditions with thin black distracters.

Within thick red distracter orientations, significant differences were found for certainty between consistent and flat conditions, $F(1, 9) = 19.72, p = .002, \eta_p^2 = .69$, and between inconsistent and flat conditions, $F(1, 9) = 6.16, p = .035, \eta_p^2 = .41$, but not between consistent and inconsistent conditions, $F(1, 9) = 0.61, p = .456, \eta_p^2 = .06$. This result shows that certainty ratings were similar for consistent- and inconsistent conditions, but both were lower than the flat condition.

Simple simple comparisons of distracter color within each orientation and size showed that distracter color was responsible for significant differences in certainty ratings between inconsistent thin distracter conditions, $F(1, 9) = 45.59$, $p = .000$, $\eta_p^2 = .84$. Certainty was higher when distracters were red rather than black. Hypothesis 2c, which stated that certainty ratings would be higher in inconsistent conditions with thin red distracters than with thin black distracters, was thus supported. In the flat thick conditions, $F(1, 9) = 10.98$, $p = .009$, $\eta_p^2 = .55$, the red was also rated higher in certainty than the black. No other significant differences were found in consistent thin, $F(1, 9) = 0.82$, $p = .387$, $\eta_p^2 = .08$, consistent thick, $F(1, 9) = 0.02$, $p = .891$, $\eta_p^2 = .002$, inconsistent thick, $F(1, 9) = 1.04$, $p = .333$, $\eta_p^2 = .10$, or flat thin conditions, $F(1, 9) = 1.48$, $p = .254$, $\eta_p^2 = .14$.

To test Hypothesis 3c, which stated that certainty ratings would be higher in the inconsistent condition than with thin black distracters with thick black distracters, simple simple comparisons were conducted between line thicknesses within black distracters for each orientation. The results showed that there was a highly significant difference between certainty ratings for thin and thick distracters in the inconsistent conditions, $F(1, 9) = 66.04$, $p = .000$, $\eta_p^2 = .88$. This finding supports Hypothesis 3c and is further evidence for the discriminability of dimensions (Garner, 1974). No other differences were found between thin and thick distracters in consistent conditions, $F(1, 9) = 0.80$, $p = .393$, $\eta_p^2 = .08$, or in flat conditions, $F(1, 9) = 0.6$, $p = .817$, $\eta_p^2 = .006$. A summary of the hypothesis testing results is shown in Table 3.

Summary of Hypotheses

Table 3

Hypothesis number, predictions, test used and result

H #	Prediction	Test Used	Result
1a	Performance will be better in the consistent- than in the inconsistent condition with thin black distracters.	Simple simple comparisons	Supported (marginal)
1b	Response times will be shorter in the consistent- than in the inconsistent condition with thin black distracters.	Simple simple comparisons	Supported
1c	Certainty ratings will be higher in the consistent- than in the inconsistent condition with thin black distracters.	Simple simple comparisons	Supported
2a	Performance will be better in the inconsistent condition with thin red distracters than with thin black distracters.	Simple simple comparisons	Supported
2b	Response times will be shorter in the inconsistent condition with thin red distracters than with thin black distracters.	Simple simple comparisons	Supported
2c	Certainty ratings will be higher in the inconsistent condition with thin red distracters than with thin black distracters.	Simple simple comparisons	Supported
3a	Performance will be better in the inconsistent condition with thick black distracters than with thin black distracters.	Simple simple comparisons	Supported
3b	Response times will be shorter in the inconsistent condition with thick black distracters than with thin black distracters.	Simple simple comparisons	Supported (marginal)
3c	Certainty ratings will be higher in the inconsistent condition with thick black distracters than with thin black distracters.	Simple simple comparisons	Supported

Individual Differences

The two tests administered to determine if differences in short term visual memory (VPT) and ability in field dependence-independence (EFT) could account for differences in performance, response times and rated certainty, yielded little explanatory power. Pearson Product Moment Correlation Coefficients were calculated and the results showed that these tests did not consistently correlate with any of the dependent variables, but they did correlate with each other in 3/5 comparisons, conducted in the four Experiments and the second Preliminary Study. The correlation tables are shown in Appendices BB-F. These tests were administered in all four Experiments as well as Preliminary Study 2 and out of the 36 possible correlations (the total of all conditions), the VPT accounted for a significant relation between it and performance, 0/36 times, between it and response times, 6/36 times, and between it and certainty, 0/36 times. The EFT accounted for significant relations between it and performance 5/36 times, between it and response times 7/36 times and finally, between it and certainty 2/36 times. Thus, none of the correlation analyses showed that these tests systematically accounted for a significant amount of variance in either of the dependent variables. Therefore, they are not discussed further in this thesis.

Discussion

When exploring the three-way interaction, no differences were identified between the four flat distracter conditions (thin black, thick black, thin red, and thick red) despite variations in size and color. One possible way to interpret this result is that combinations of distracter line orientation, color and size are separable dimensions when presented in simple line graphs.

The remainder of the discussion addresses three important topics. First, evidence of serial- and parallel processing is discussed. Second, evidence for state- and process limitations is presented and third, possible strategies that can account for performance enhancement in inconsistent conditions are outlined.

Serial and Parallel Processing

Taking the thin black distracters first, Hypothesis 1a stated that performance would be better in consistent- than in inconsistent conditions. It was marginally supported. Hypothesis 1b stated that response times would be shorter in the consistent- than in the inconsistent condition. It was supported. Hypothesis 1c stating that certainty ratings would be higher in the consistent- than in the inconsistent condition, was also supported. Performance in the consistent condition represents parallel processing because the two lines were same color and oriented in the same direction. In turn, it supports the notion of dimension configurability (e.g., Carswell, 1992; Pomerantz, 1981; Pomerantz & Garner, 1973). Performance in the inconsistent condition may be attributed to a serial processing mechanism because the lines could not easily be discriminated from one another, which forced participants to pay attention to each line. The finding that response times were shorter for consistent than for inconsistent thin black distracters is similar to the result obtained by Carpenter and Shah (1998), who found that graphs showing non-interacting lines were interpreted faster than interacting lines. The higher certainty ratings evidenced in the consistent than in the inconsistent distracter conditions indicates that participants knew when their answers were correct. In the consistent condition with thin black distracters, performance, response times and certainty ratings indicate that the target line was easier to interpret than in the inconsistent condition.

Hypothesis 2a predicted that performance in the inconsistent condition would be better with thin red- than with thin black distracters. That was supported. Now, if serial processing had been applied, performance would have been very similar in the inconsistent thin red and thin black distracter conditions. Since that was not the case, the results suggest that a different processing mechanism was used. It is possible that a parallel processing strategy was used with inconsistent thin red distracters. Hypothesis 2b stated that response times would be shorter in the inconsistent condition with thin red- than with thin black distracters. It was also supported. In accordance with Hypothesis 2c, which was also supported, participants responded faster in inconsistently oriented thin red distracter conditions than in thin black distracter conditions, indicating that the correct trend was more readily identified when distracter lines were red. Finally, the higher certainty ratings in the thin red- than in the thin black inconsistent conditions indicate that participants were aware of the correctness of their responses.

For black distracters, Hypothesis 3a predicted that performance on inconsistent conditions would be better with thick than with thin distracters. It was also supported. Hypothesis 3b stating that response times in the inconsistent thick black distracter condition would be shorter than with thin black distracters, was marginally supported. Finally, as predicted by Hypothesis 3c, certainty ratings were higher in the inconsistent thick black- than in the thin black distracters condition. Since performance was better in the inconsistent thick black- than in the inconsistent thin black distracter condition, this indicates that a parallel processing mechanism was applied. Response time measures indicate that it was easier to identify the correct trend when distracter lines varied from the target on size or on both size and color. Performance, response time and certainty ratings all indicate that the target line was easier to interpret when the distracter line differed in size than when both were the same size. This was true even when both lines were the same color.

Performance was also marginally better with thick red- than with thin black distracters also indicating a parallel processing mechanism. Taken together, these results indicate that manipulation of distracter line color or size enhanced the discriminability of targets and distracters, in inconsistent conditions, which therefore resulted in parallel processing.

State Limitations and Process Limitations

Performance across the four flat distracter conditions did not differ, suggesting that when the orientation was flat, the other dimensions did not matter. Furthermore, performance in any of the flat distracter conditions was not surpassed by any of the other conditions, suggesting that the flat condition represented the best possible performance under these circumstances. However, the one exception was that, in the thin red inconsistent condition, performance did not differ from performance in the thin red flat condition. Apparently, manipulation of distracter line color alone rendered the target line very discriminable. Still, the overall results from the flat distracter conditions reinforce the contention that errors in the flat conditions were due to exposure time limitations only. They also suggest that performance was as good as it could be in the flat distracter conditions, indicating that any errors were due to state limitations, as described earlier (p. 36).

Looking only at performance in the thin black distracter conditions, it was argued earlier that the flat distracter condition would be the least perceptually complex because the distracter would more readily blend into the background than in the consistent- and in the inconsistent distracter orientations. This is because the distracter line was the same orientation, color and size as the grid lines on the graph. Performance was indeed found to be best in that particular flat distracter condition compared to the consistent and inconsistent conditions. This is further evidence that the flat distracter condition represented the best possible performance, thereby providing evidence for state limitations. It is therefore safe to conclude that any additional errors identified in the comparison of the flat conditions to other orientation conditions must be due to stimulus characteristics rather than to the limited exposure time alone. That is, they would be due to process limitations.

Strategy Adoption

Focusing only on performance in inconsistent conditions, as the above results showed, the high discriminability of red distracters facilitated accurate target selection, particularly when they were thin. This supports Bertin (1983), who would predict that color and orientation should affect perceptual complexity the most because they are both associative and selective. However, performance with thick black inconsistent distracters was less than the best possible. It differed from performance on the flat condition, whereas thin red inconsistent distracters did not. This shows that when only one dimension was manipulated, color enhanced performance more than size in the inconsistent conditions. However, when both color and size were manipulated, as in the thick red inconsistent conditions, performance was still less than the best possible. This shows that manipulating color alone resulted in the best performance in inconsistent conditions. Overall, however, performance was better in inconsistent conditions when distracter color and/or size was different from the target line than when they were both the same color and size.

With a thick red and a thick black distracter line, performance did not differ between the consistent and inconsistent conditions, which still suggests that a parallel processing strategy was adopted. So, performance was better than expected in the thick black and thick red conditions than if serial processing had occurred. Since performance was poorer than on the flat condition, it suggests a selective serial processing strategy.

Stimulus redefinition would have been shown if performance across all distracter conditions had been indistinguishable from performance on the flat condition. This result would have provided evidence that the two stimulus lines were processed as one integral dimension. This clearly did not happen. The stimulus redefinition strategy may be possible if the targets and distracters differed consistently on only one dimension throughout an experiment. That was tested in Experiment 3.

In summary, performance was enhanced in inconsistent conditions when distracter color or size varied from the targets but not when they shared both color and size. One possibility is that participants adopted selective serial processing. Variations in distracter color or size, or both of these, increased the discriminability between targets and distracters, thus facilitating the perception and processing of separable dimensions. This was shown in the comparisons of inconsistent and consistent conditions in which no differences were identified. Serial processing only appears to be necessary when distracter and target lines are inconsistently oriented and share both color and size. However, parallel processing appears to be employed when distracter lines are flat or when they are oriented consistently with target lines. Parallel processing seems to explain performance when target and distracter lines are inconsistently oriented but can be differentiated by color, or size, or both of these.

It is premature to draw any firm conclusions because participants were exposed to all combinations of distracter orientations, sizes and colors. Experiment 2 was therefore conducted to test the impact of variations in color and orientation on parallel and serial processing mechanisms and the strategies adopted for efficient performance.

Experiment 2: Orientation and Color

The purpose of this experiment was to test the effects of line orientation and color on graph perception and interpretation, and to unravel the three-way interaction of color, size and orientation found in Experiment 1. The colors tested were again red and black, and all target and distracter lines were thin. As in Experiment 1, the thin black distracter conditions were employed to establish the conditions under which serial- and parallel processing mechanisms could be observed. Hypothesis 1a predicted that performance would be better in the consistent- than in the inconsistent condition with black distracters. Response times would also be shorter (Hypothesis 1b) and certainty ratings would be higher (Hypothesis 1c) in the consistent- than in the inconsistent condition with black distracters.

Consistent with Garner's (1974), the target should be "discriminable" when inconsistently oriented target and distracter lines differ in color. That is, participants should be easily able to select the black target lines when the distracter lines are red. Hypothesis 2a therefore predicted that performance would be better in the inconsistent condition with red distracters than with black distracters. Likewise, response times would be shorter (Hypothesis 2b) and certainty ratings higher (Hypothesis 2c) in the inconsistent condition with red distracters than for thin black distracters. The Hypotheses are outlined in Table 4. Flat distracter conditions were again employed.

Table 4

Hypothesis number, predictions and test used

H #	Prediction	Test Used
1a	Performance will be better in the consistent- than in the inconsistent condition with black distracters.	Simple main effects analysis
1b	Response times will be shorter in the consistent- than in the inconsistent condition with black distracters.	Simple main effects analysis
1c	Certainty ratings will be higher in the consistent- than in the inconsistent condition with black distracters.	Simple main effects analysis
2a	Performance will be better in the inconsistent condition with red distracters than with black distracters.	Simple main effects analysis
2b	Response times will be shorter in the inconsistent condition with red distracters than with black distracters.	Simple main effects analysis
2c	Certainty ratings will be higher in the inconsistent condition with red distracters than with black distracters.	Simple main effects analysis

Method

Participants

A sample of 16 new participants were recruited from Carleton University but two participants indicated that they did not follow the instructions and thus their data was eliminated leaving data for 14 participants to be analysed (4 women, 10 men, mean age of 24.93 years, ranging from 17 to 40 years). All responded to advertisements posted around campus and each was paid $10.00 for participation. All were Carleton University students. All were pre-screened for colorblindness and for visual acuity using a mini eye chart.

Apparatus and Materials

The apparatus and materials were the same as in Experiment 1, with the exception that the graph stimuli included only the line graphs with thin distracters, used in Experiment 1.

Design

This experiment employed a completely within-groups design, shown in Table 5, and the stimuli were presented the same as they were in Experiment 1. Seven participants completed the VPT first and the EFT after the experimental trials, whilst the other seven completed the EFT first and the VPT last.

Table 5

Experimental design, red and black distracters in consistent (C), inconsistent (I), and flat (F) orientations, all with thin black targets, with 12 graph stimuli in each condition

		Thin (size)						
			Red			*Black*		
Test	*P's*	*C*	*I*	*F*	*C*	*I*	*F*	*Test*
		n= 12	n= 12	n= 12	n= 12	n= 12	n= 12	
VPT	P1 ... P7							EFT
EFT	P8 ... P14							VPT

Seventy-two line graphs were presented, preceded by 12 practice line graphs. The stimulus graphs and textual options were the same as in Experiment 1, excluding the 72 stimuli with thick distracters.

Procedure

The procedure was exactly the same as for Experiment 1. The experiment lasted approximately 1 hour.

Results

The results are divided into four sections: data exploration, performance, certainty, and response times. All descriptive statistics, details of assumption and outlier tests, and error data are included in Appendix D.

Data Exploration

The data for performance, response times and certainty were explored to ensure that assumptions of ANOVA were met and that outlying scores did not adversely affect the results. Beginning with performance data, tests of normality showed that normality was violated but no outlying participant scores were identified. Mauchley's test of Sphericity was performed and showed a violation for orientation, $W = 0.58$, $\chi^2 (2) = 6.58$, $p = .037$, thus a Huynh-Feldt Epsilon correction was applied (0.76) and the degrees of freedom in both the numerator and denominator of the F ratio were re-calculated and used where differences in degrees of freedom resulted.

An analysis of response time data revealed that the assumption of normality was violated in three conditions. Three participants' response times were identified as outlying and influential. Therefore, the three response times were adjusted. A re-analysis of the data showed that after moving the values to one standard deviation above or below the mean, the data was normally distributed. Finally, Mauchley's test of

Sphericity was performed and showed a violation for orientation, $W = 0.57$, $\chi^2(2) = 6.81$, $p = .033$, thus a Huynh-Feldt Epsilon correction was applied (0.76) and the degrees of freedom in both the numerator and denominator of the F ratio were re-calculated.

An analysis was also conducted for certainty ratings and tests of normality showed the assumption to be satisfied, although one participant's ratings appeared to be outlying in three of the 6 conditions, it was retained because it fell within normal limits in the other conditions and did not cause a violation of the normality assumption or sphericity.

Performance

The results of correct performance data are shown in Figure 12 for red and black distracter colors in consistent, inconsistent and flat conditions. It appears that black distracters resulted in slightly better performance than red distracters when target and distracters were consistently oriented. However, red distracters appear to have resulted in better performance than black distracters when distracters were inconsistently oriented. There appears to be no difference between the red and black flat distracter conditions.

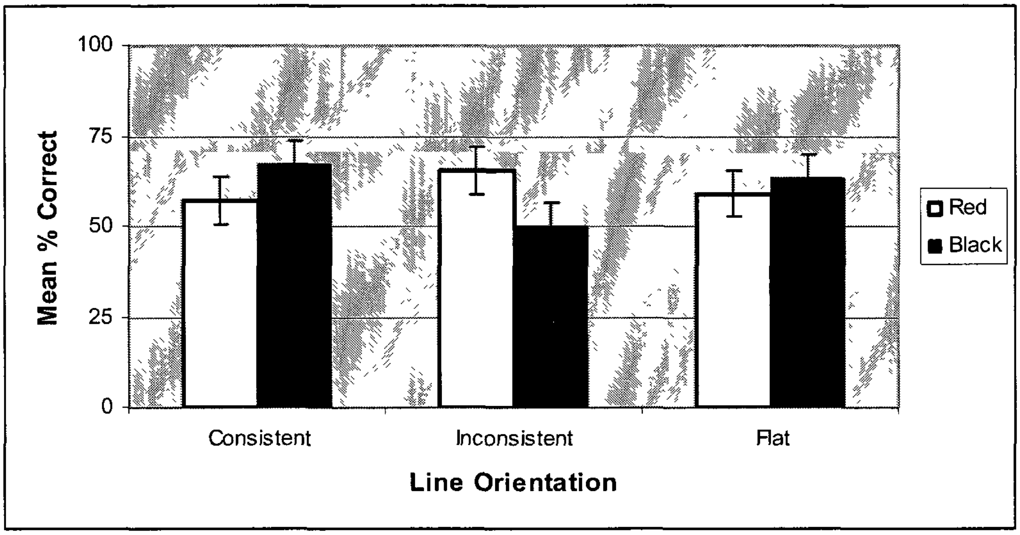


Figure 12: Mean percent correct responses for distracter color (red, black) and orientation (consistent, inconsistent, flat), bars represent *CI*s for 2-way interaction = $M_i \pm$ 6.44

To test the performance Hypotheses (1a and 2a), a (2 x 3) ANOVA was conducted for performance on distracter color (black, red) and target and distracter orientation (consistent, inconsistent, flat). The interaction of color and orientation was highly significant, $F(2, 26) = 9.13, p = .001, \eta_p^2 = .41$.

Exploring simple main effects within color and between orientations showed that there was an effect of orientation for black distracters, $F(2, 26) = 6.51, p = .005, \eta_p^2 = .33$, but not for red ones, $F(2, 26) = 2.25, p = .126, \eta_p^2 = .15$. Simple comparisons were therefore made within orientation for the black distracters only. The results showed that there was a significant difference between inconsistent and flat orientations, $F(1, 13) = 12.69, p = .003, \eta_p^2 = .49$. Similarly, there was a significant difference between the consistent and inconsistent orientations, $F(1, 13) = 9.89, p = .008, \eta_p^2 = .43$, which confirms support for Hypothesis 1a. However, there was no significant difference between the consistent and flat orientations, $F(1, 13) = 0.56, p = .468, \eta_p^2 = .04$.

Tests of simple main effects showed that there was a difference between distracter-line colors in the consistent orientations, $F(1, 13) = 11.01, p = .006, \eta_p^2 = .46$; black distracters led to better performance. In the inconsistent orientation red distracters led to better performance than black distracters, $F(1, 13) = 12.63, p = .004, \eta_p^2 = .49$. This supports Hypothesis 2a. No differences were found in the flat orientation, $F(1, 13) = 0.65, p = .433, \eta_p^2 = .05$. No main effect for color, $F(1, 13) = 0.03, p = .864, \eta_p^2 = .07$, or for orientation, $F(1.53, 19.84) = 0.95, p = .379, \eta_p^2 = .002$, was found.

Response Times

Figure 13 shows the results for the analysis of mean response times for both red and black distracters in consistent, inconsistent and flat conditions. Response times do not appear to differ between red and black distracters in the consistent condition. However, it appears that participants responded faster to red distracters than to black distracters in the inconsistent condition. Response times appear to be slightly shorter in the flat condition regardless of line color.

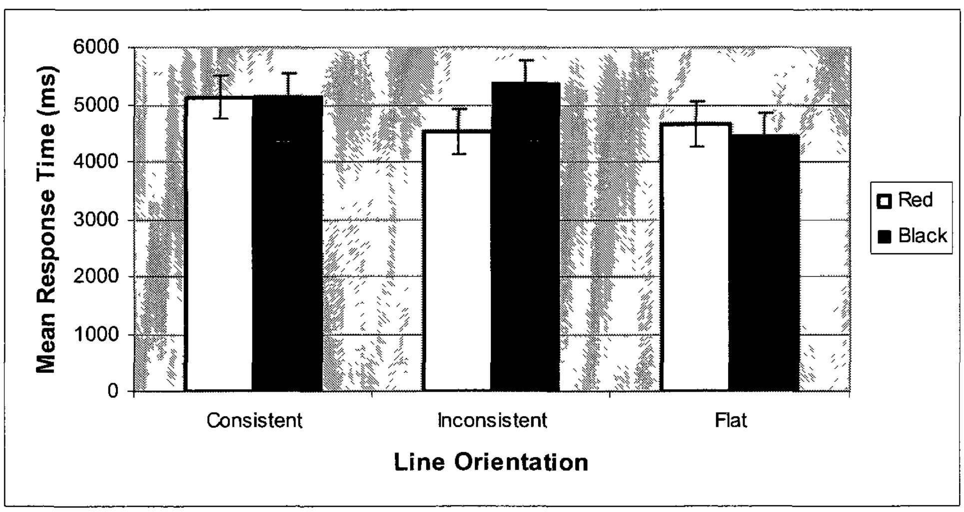


Figure 13. Mean response time (ms) for distracter color (red, black) and orientation (consistent, inconsistent, flat), bars represent *CI*s for 2-way interaction = $M_i \pm 389.93$

To test Hypotheses 1b and 2b, a (2 x 3) ANOVA was conducted for response times on distracter color (black, red) and target and distracter orientation (consistent, inconsistent, flat). The interaction was significant, $F(2, 26) = 4.40$, $p = .023$, $\eta_p^2 = .25$. As Figure 13 shows, response times within the black distracter conditions appear to be more variable than the red distracter conditions. This was confirmed by tests of simple main effects showing an effect of orientation, $F(2, 26) = 4.93$, $p = .015$, $\eta_p^2 = .28$. Simple comparisons showed that response times were longer with consistent than with flat orientations, $F(1, 13) = 9.16$, $p = .010$, $\eta_p^2 = .41$, and also longer than with inconsistent and flat orientations, $F(1, 13) = 9.33$, $p = .009$, $\eta_p^2 = .42$. Response times did not differ between consistent and inconsistent orientations, $F(1, 13) = 0.43$, $p = .523$, $\eta_p^2 = .03$, refuting Hypothesis 1b, which stated that response times in the consistent- would be shorter with black distracters than in the inconsistent condition. It is thus evident that participants responded fastest to flat black distracters.

A simple main effects analysis conducted on red distracters showed no significant differences within the three orientations, $F(2, 26) = 1.61$, $p = .219$, $\eta_p^2 = .11$.

Exploring response times within each orientation revealed that response times did not differ between black and red distracters in the consistent orientations, $F(1, 13) = 0.00$, $p = .980$, $\eta_p^2 = .0001$, and nor was there a difference between colors in the flat orientations, $F(1, 13) = 0.72$, $p = .412$, $\eta_p^2 = .05$. There was however a significant difference in response times associated with black and red distracters in the inconsistent orientations, $F(1, 13) = 11.75$, $p = .004$, $\eta_p^2 = .47$. When distracters were black and oriented inconsistently with the target line, responses took longer than when they were red. This supports Hypothesis 2b, predicting that response times in the inconsistent condition would be shorter with red distracters than when they were black. No main effect was found for orientation, $F(1.51, 19.65) = 2.35$, $p = .132$, $\eta_p^2 = .15$, or for color, $F(1, 13) = 2.21$, $p = .161$, $\eta_p^2 = .15$.

Certainty

Figure 14 shows the results for the analysis of mean certainty ratings for both red and black distracters in consistent, inconsistent and flat conditions. Certain ratings do not appear to differ between red and black distracters in the consistent conditions.

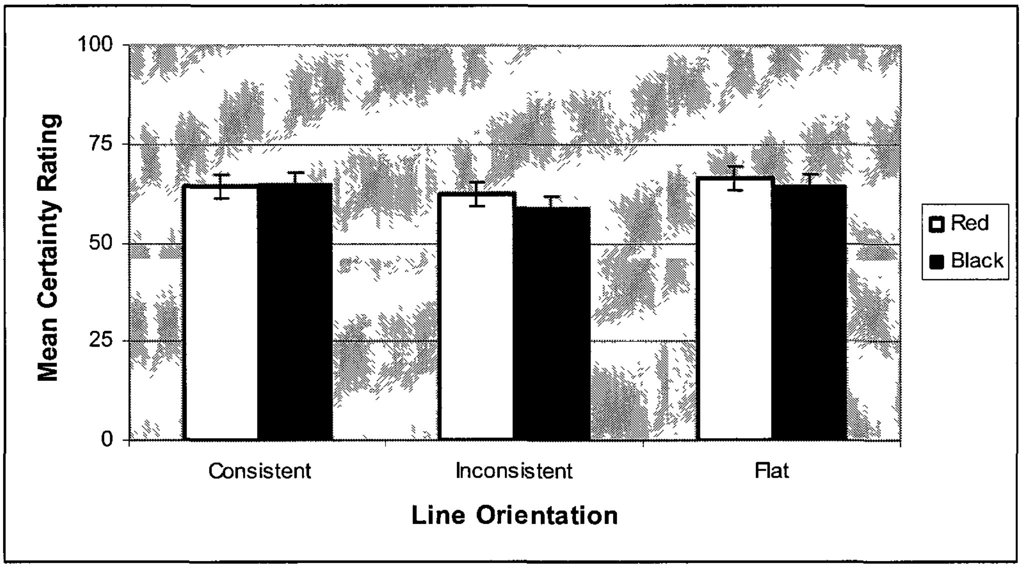


Figure 14. Mean certainty rating for distracter color (red, black) and orientation (consistent, inconsistent, flat), bars represent *CI*s for 2-way interaction = $M_i \pm 2.90$

To test Hypotheses 1c and 2c, a (2 x 3) ANOVA was conducted for certainty on distracter color (black, red) and target and distracter orientation (consistent, inconsistent, flat). The interaction was not significant, $F(2, 26) = 0.88$, $p = .426$, $\eta_p^2 = .06$, and there was no main effect found for distracter color, $F(1, 13) = 1.53$, $p = .238$, $\eta_p^2 = .11$. There was, however, a significant main effect for orientation, $F(2, 26) = 6.18$, $p = .006$, $\eta_p^2 = .32$. Simple main effects analyses showed that certainty ratings were higher with consistent- (M = 64.45, SE = 6.28) than with inconsistent (M = 60.63, SE = 6.63) orientations, $F(1, 13) = 8.00$, $p = .014$, $\eta_p^2 = .38$, but they did not differ from flat orientations (M = 65.44, SE = 6.23), $F(1, 13) = 0.65$, $p = .435$, $\eta_p^2 = .05$. However, certainty ratings were lower with inconsistent orientations than with flat orientations, $F(1, 13) = 7.86$, $p = .015$, $\eta_p^2 = .38$, regardless of distracter line color. Thus, inconsistent orientations brought about the lowest certainty ratings. Simple comparisons were then conducted as planned comparisons. A simple comparison between orientations within black distracters was conducted to test Hypothesis 1c. The results showed that the difference was significant, $F(1, 13) = 6.69$, $p = .023$, $\eta_p^2 = .34$, thus showing support for Hypothesis 1c, which stated that, certainty ratings would be higher in the consistent- than the inconsistent condition with black distracters.

The simple main effect analysis of color within the inconsistent orientation was conducted to test Hypothesis 2c. The results showed that certainty ratings did not differ significantly between red and black distracters in the inconsistent condition, $F(1, 13) = 2.73$, $p = .122$, $\eta_p^2 = .17$. This refutes Hypothesis 2c, which stated that certainty ratings would be higher in the inconsistent condition with red distracters than with black distracters.

Summary of Hypotheses

Table 6

Hypothesis number, prediction, test used and result

H #	Prediction	Test Used	Result
1a	Performance will be better in the consistent- than in the inconsistent condition with black distracters.	Simple main effects analysis	Supported
1b	Response times will be shorter in the consistent- than in the inconsistent condition with black distracters.	Simple main effects analysis	Refuted
1c	Certainty ratings will be higher in the consistent- than in the inconsistent condition with black distracters.	Simple main effects analysis	Supported
2a	Performance will be better in the inconsistent condition with red distracters than with black distracters.	Simple main effects analysis	Supported
2b	Response times will be shorter in the inconsistent condition with red distracters than with black distracters.	Simple main effects analysis	Supported
2c	Certainty ratings will be higher in the inconsistent condition with red distracters than with black distracters.	Simple main effects analysis	Refuted

Discussion

The two flat distracter conditions (black, red) compared first revealed no differences; variations in the color of the line had no impact on performance in the flat condition. This reinforces the observation made in Experiment 1 that orientation and color are separable dimensions.

Evidence of serial and parallel processing is discussed first. Second, evidence for state- and process limitations is presented and finally, possible strategies that can account for performance enhancement in inconsistent conditions are outlined.

Serial and Parallel Processing

Hypothesis 1a stated that performance would be better in consistent- than inconsistent conditions with black distracters. It was supported. Performance in the black consistent condition represents parallel processing due to 'configural superiority effects' (e.g., Pomerantz, 1981), suggesting that participants paid attention to both lines simultaneously. The poor performance in the black inconsistent condition is a sign that the lines could not be readily distinguished from one another. Instead, participants paid attention to each line, suggesting that they adopted a serial processing strategy. This result therefore suggests that both parallel and serial processing occurred. However, contrary to the prediction of Hypothesis 1b, response times did not differ between the consistent- and inconsistent conditions with black distracters. It was therefore refuted. Hypothesis 1c predicting that certainty ratings would be higher in the consistent- than in the inconsistent condition with black distracters was supported. Hence, differences in the processing mechanism employed had no effect on response times, but it did affect certainty in the predicted direction, suggesting that participants knew when they were correct.

Hypothesis 2a, predicting that performance would be better in the inconsistent condition with red- than with black distracters, was also supported, thereby providing support for the discriminability of dimensions (Garner, 1974). It also provides further evidence of parallel processing. If the lines had been processed serially, performance in the inconsistent condition with the red distracters would have been undifferentiated from performance with the black distracters. Hypothesis 2b predicted that response times would be shorter in the inconsistent condition with red- than with black distracters. It was also supported. However, Hypothesis 2c, which predicted that certainty ratings would be higher in the red inconsistent than in the black inconsistent conditions, was refuted, as these did not differ. Apparently, participants failed to realize that they were more often wrong on graphs with black- than with red distracters.

Performance in the consistent condition was better with black- than with red distracters. This suggests that the target and distracter did not configure when they differed in color; thus they appear to have been processed serially. Participants were unable to predict the color of the next distracter, which may have made it difficult to adopt a consistent strategy for identifying the target lines. One way to test this would be to present graphs with only one distracter color instead of two. That was done in Experiments 3 and 4.

State Limitations and Process Limitations

Performance across in the two flat distracter conditions did not differ, nor was that level of performance surpassed in any other condition. As in Experiment 1, this suggests that errors in the flat conditions were due to exposure time limitations only, hence representing state limitations.

Looking only at black distracters, performance on the flat- was better than in the inconsistent conditions but did not differ from the consistent condition. Performance was also better in the consistent- than in the inconsistent conditions. Evidence for the distinction between state- and process limitations was provided by performance results from the black distracter conditions. The black inconsistent distracter line was responsible for errors additional to those committed in flat- and consistent conditions, which may be attributed to process limitations. The most plausible explanation for this is that the stimuli were quite difficult to discriminate and process accurately.

Looking only at performance with red distracters, however, neither the consistent nor the inconsistent conditions differed from performance on the flat condition. Thus, in the presence of red distracters, there was no evidence for process limitations; the manipulation of the distracter color apparently rendered the target very 'discriminable'. Thus, both state- and process limitations must have accounted for the additional errors in the black inconsistent condition, but only state limitations accounted for errors when distracters were red.

Strategy Adoption

Performance was better in the inconsistent condition with red distracters than with black distracters, suggesting that the black target must have been highly discriminable. Performance with the red distracters was undifferentiated from the best possible performance in the flat distracter condition. This allowed participants easily to select it, suggesting that a selective serial processing strategy was adopted when distracters were red. However, because performance with red distracter lines was undifferentiated from the flat, an alternative explanation is that a stimulus redefinition strategy may have been adopted. That is, in the presence of a red distracter line, a black target can be processed unitarily. However, because performance on consistent conditions was better with black- than with red distracters, this suggests that the black target could not always be redefined. This indicates that stimulus redefinition did not occur. It is thus most likely that selective serial processing was adopted for target identification with red distracters. The above findings indicate that one of at least two possible processing strategies may have been adopted for target identification when distracter orientation and color were varied. Experiments 3 and 4 both explored perceptual complexity by varying only size and orientation while holding distracter color constant.

Experiment 3: Orientation and Size

The purpose of this Experiment was to explore how graph interpretation is affected when varying the size and orientation of distracter lines while holding target and distracter color constant. The target line was always thin and black and the distracter lines were always red, thick and thin.

Recall that size, represented here by the thickness of distracter lines, is claimed to be the most perceptible of all dimensions (Bertin, 1983; Felfoldy & Garner, 1973, Garner, 1974; Treisman & Sato, 1990), which means that thicker lines should be processed easily because they should stand out more than the thin black target lines. Thus, increasing distracter line size relative to the target line should make the distracter stand out very prominently, thereby decreasing the perceptibility of the target lines. With red, thick distracter lines, attention could be drawn to these first, when target lines are thin and black possibly leading to faulty target identification. The same should be less likely to occur with thin red distracter lines. Hypothesis 1a therefore stated that performance would be better in the inconsistent condition with thin- than with thick distracters. Response times were predicted to be shorter (Hypothesis 1b), and certainty ratings higher (Hypothesis 1c) in inconsistent conditions with thin- than with thick distracters. These hypotheses are summarized in Table 7.

Table 7

Hypothesis number, prediction and test used

H #	Prediction	Test Used
1a	Performance will be better in inconsistent conditions with thin than with thick distracters.	Simple main effects analysis
1b	Response times will be shorter in inconsistent conditions with thin than with thick distracters	Simple main effects analysis
1c	Certainty ratings will be higher in inconsistent conditions with thin than with thick distracters.	Simple main effects analysis

Method

Participants

A sample of 12 new participants, comprising 6 women and 6 men (mean age 24.36 years, range of 21 to 29 years), were recruited from Carleton University and through personal contacts. They responded to advertisements posted around campus and were each paid $10.00 for their participation. All were pre-screened for colorblindness and tested for visual acuity using a mini eye chart.

The apparatus and materials was the same as in Experiment 1, with the exception that the 72 graph stimuli were only those line graphs with red, thin or thick distracters. As before, line graphs were presented in random order. The design was exactly the same as in Experiment 2.

Results

The results section has four sub-sections: data exploration, performance, certainty, and response times. All descriptive statistics, details of assumption and outlier tests, and error data are included in Appendix E.

Data Exploration

An analysis was undertaken for performance scores to identify threats to the assumptions underlying the ANOVA test. Tests of normality showed that it may have been violated in two conditions. An outlier analysis identified only Participant 12 as outlying. However, since the score deviated from the other scores in that condition, it was representative of that participant's scores on the other conditions and was therefore not altered. Sphericity was not violated.

An analysis for response times revealed no violations of normality in the dataset. Only one participant was identified as an outlier, but because the score did not affect the normality of the distribution, it was not altered. Sphericity was not violated.

Finally, the same analysis was undertaken for certainty ratings. Participant 4 was identified as an outlier in three conditions and the scores were adjusted as before for that participant, to one standard deviation above or below the mean. After this adjustment no other outliers were identified. Tests of normality did however indicate that the assumption was slightly violated still, but due to the small sample size and the nature of the scores, no further adjustments were made. Sphericity was not violated.

Performance

Performance results for orientation and distracter size are shown in Figure 15. There appears to be very little difference between the orientations and between thin and thick distracter sizes.

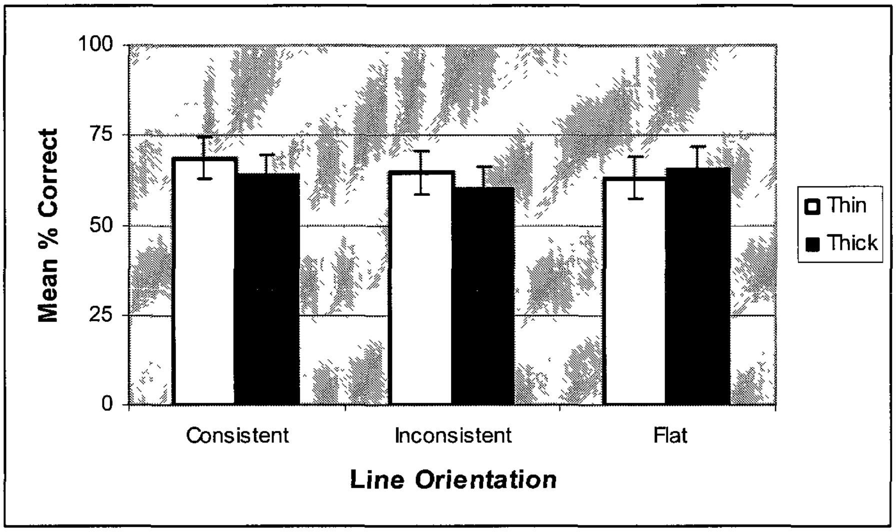


Figure 15. Mean percent correct responses for distracter line size (thin and thick) and orientation (consistent, inconsistent, flat), bars represent *CI*s for 2-way interaction = $M_i \pm$ 5.88

To test Hypothesis 1a, a (2 x 3) ANOVA was conducted for performance on distracter line size (thin, thick) and target and distracter orientation (consistent, inconsistent, flat). The interaction of orientation and size was not significant, $F(2, 22) = 1.11$, $p = .348$, $\eta_p^2 = .09$, and neither were the main effects of orientation, $F(2, 22) = 0.52$, $p = .604$, $\eta_p^2 = .04$, or of size, $F(1, 11) = 0.84$, $p = .379$, $\eta_p^2 = .07$. To test Hypothesis 1a, stating that performance will be better in inconsistent conditions with thin- than with thick distracters, a simple main effects analysis was conducted. The difference between the conditions was not significant, $F(1, 11) = 0.65$, $p = .438$, $\eta_p^2 = .06$. Thus, Hypothesis 1a was refuted. To test for dimension separability, a simple main effects analysis was conducted on the two flat distracter conditions. No differences were identified $F(1, 11) = 0.60$, $p = .457$, $\eta_p^2 = .05$, indicating that no interference occurred.

Response Times

Response time results for orientation and distracter size are shown in Figure 16. The consistent distracter conditions appear to have taken longer than the inconsistent and the flat orientations, with the flat distracter conditions taking the least amount of time.

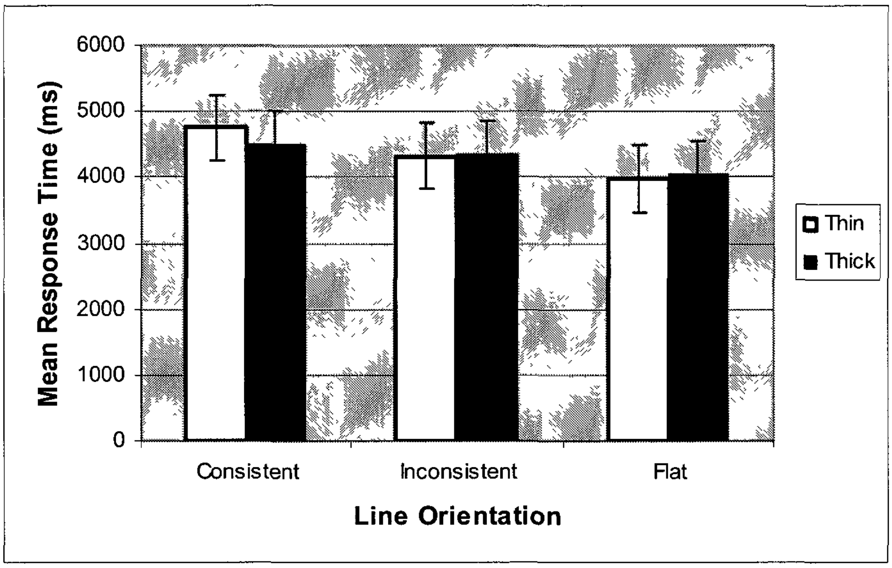


Figure 16. Mean response time (ms) for distracter line size (thin and thick) and orientation (consistent, inconsistent, flat), bars represent *CI*s for 2-way interaction = $M_i \pm 503.18$

To test Hypothesis 1b, a (2 x 3) ANOVA was conducted for response times on distracter size (thin, thick) and target and distracter orientation (consistent, inconsistent, flat). The interaction of orientation and size was not significant, $F(2, 22) = 0.25$, $p = .781$, $\eta_p^2 = .02$, and neither was the main effect of size, $F(1, 11) = 0.07$, $p = .791$, $\eta_p^2 = .006$, but the main effect of orientation was significant, $F(2, 22) = 4.62$, $p = .021$, $\eta_p^2 = .30$. Pairwise LSD comparisons revealed that graphs with flat oriented distracter lines were responded to faster than both the consistent- ($p = .010$) and the inconsistent orientations

(p = .056 – marginal), but these did not differ from each other (p = .262). To test Hypothesis 1b, stating that response times will be shorter in inconsistent conditions with thin- than with thick distracters, a simple main effects analysis was conducted. The difference was not significant, $F(1, 11) = 0.009, p = .927, \eta_p^2 = .001$. Thus, Hypothesis 1b was refuted.

Certainty

Figure 17 shows the mean certainty ratings for distracter line orientation and line size. There appears to be very little difference between the conditions.

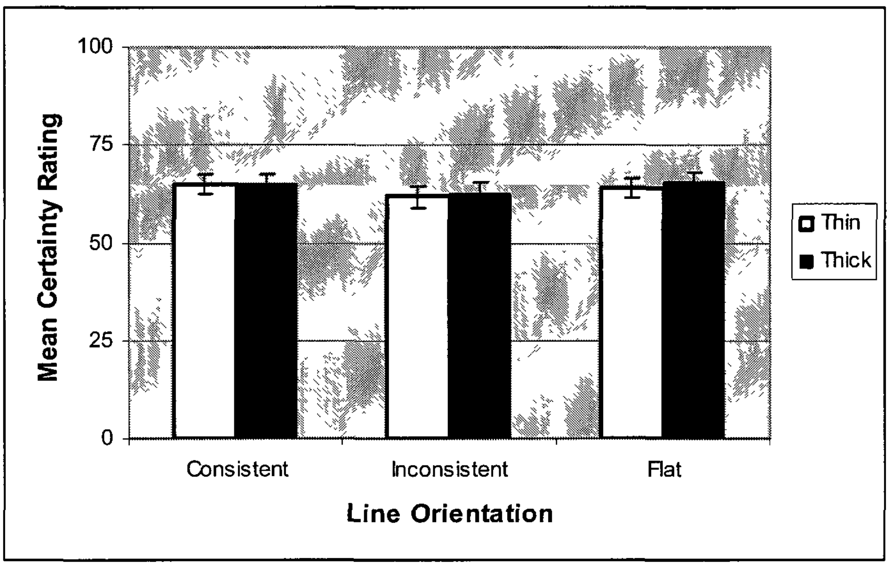


Figure 17. Mean certainty rating for distracter line size (thin and thick) and orientation (consistent, inconsistent, flat), bars represent *CI*s for 2-way interaction = $M_i \pm 2.74$

To test Hypothesis 1c, a (2 x 3) ANOVA was conducted for certainty on distracter line size (thin, thick) and target and distracter orientation (consistent, inconsistent, flat). The results showed that the interaction of orientation and size was not significant, $F(2, 22) = 0.19, p = .832, \eta_p^2 = .02$, and neither were the main effects of orientation, $F(2, 22) = 1.74, p = .199, \eta_p^2 = .14$, or size, $F(1, 11) = 0.29, p = .604, \eta_p^2 = .03$. To test Hypothesis

1c, stating that certainty will be higher in inconsistent conditions with thin- than with thick distracters, a simple main effects analysis was conducted. The difference was not significant, $F(1, 11) = 0.36, p = .561, \eta_p^2 = .03$. Thus, Hypothesis 1c was also refuted.

Summary of Hypotheses

Table 8

Hypothesis number, prediction, test used and result

H #	Prediction	Test Used	Result
1a	Performance will be better in inconsistent conditions with thin than with thick distracters.	Simple main effects analysis	Refuted
1b	Response times will be shorter in inconsistent conditions with thin than with thick distracters	Simple main effects analysis	Refuted
1c	Certainty ratings will be higher in inconsistent conditions with thin than with thick distracters.	Simple main effects analysis	Refuted

Discussion

The experimental conditions explored here still made it possible to identify errors attributable to state- or process limitations. As the target and distracter lines in this Experiment always differed on color, serial processing was not expected to occur at all. Hence, parallel processing was assumed throughout.

Contrary to the predictions of Hypothesis 1a, stating that performance would be better in the inconsistent condition with thin- than with thick distracters, no statistically traceable differences were identified. Hypothesis 1a was therefore refuted. Hypothesis 1b, stating that response times in inconsistent conditions with thin distracters would be shorter than with thick distracters was also refuted. Responses, however, were faster on the flat- than on the consistent- and inconsistent distracter conditions. Hypothesis 1c, which stated that certainty would be higher in inconsistent conditions with thin- than with thick distracters, was refuted as certainty ratings were also undifferentiated throughout. These results differed from those found in Experiments 1 and 2 and the implications for the model and line graph processing will be taken up in the general discussion.

Interestingly, performance on consistent- and inconsistent conditions was indistinguishable from the flat condition regardless of the size of the distracter. Thus, since no apparent interference was observed when targets and distracters varied on color, size and orientation, this again points to the separability of these dimension combinations.

State Limitations and Process Limitations

Performance across the two flat distracter conditions did not differ, and nor were performance differences identified in any of the three orientation conditions. The flat conditions represent optimal performance with any errors due to state limitations because of constraints on stimulus processing. No additional errors were found in the consistent or inconsistent orientation conditions. Thus, there was no evidence for errors that occurred because of stimulus characteristics. The results therefore suggest that state limitations were solely responsible for errors of interpretation with only red distracters.

Strategy Adoption

Performance was unaffected by variations in line size or orientation, thus, separability was assumed. Recall that Garner (1974) argues that people can opt for the most efficient processing strategy when dimensions are separable. The best possible performance is generally considered a standard against which variations in performance can be compared. In this Experiment, the best performance was represented by the flat conditions in which state limitations were responsible for errors. Stimulus redefinition may have occurred because performance across all orientation conditions was indistinguishable from performance on the flat condition. Participants knew that all distracter lines would be red. This knowledge may have led them to redefine the separable dimensions to reflect one integral dimension whereby attention was directed exclusively to the black target lines while successfully ignoring the red distracter line, regardless of whether it was thick or thin. The above results therefore demonstrate that circumstances under which stimulus redefinition may be adopted for interpreting simple line graphs can be tested experimentally.

Experiment 4: Orientation and Size

As in Experiment 3, the purpose of this Experiment was to identify sources of graph interpretation errors by exploring the effects of varying distracter line size and orientation, whilst holding target and distracter color constant. This time, only black distracters were used.

Hypothesis 1a predicted that performance would be better in the consistent- than in the inconsistent condition with thin distracters. Response times should be shorter (Hypothesis 1b) and certainty ratings higher (Hypothesis 1c) in the consistent- than the inconsistent condition with thin distracters.

The target should be "discriminable" when inconsistently oriented target and distracter lines differ in size (Garner, 1974). That is, in inconsistent conditions participants should be easily able to select the black target lines when the distracter lines are thick. Accordingly, Hypothesis 2a predicted that performance in inconsistent orientations should be better with the thick- than with the thin distracters. It was also predicted that response times should be shorter (Hypothesis 2b) and certainty ratings higher (Hypothesis 2c) in the inconsistent condition with thick than with thin distracters. These Hypotheses are shown in Table 9.

Flat distracter conditions were again employed to test for the separability of dimensions and also to observe instances in which errors could be attributed to state- or process limitations.

Table 9

Hypothesis number, predictions and test used

H #	Prediction	Test Used
1a	Performance will be better in the consistent- than in the inconsistent condition with thin distracters.	Simple main effects analysis
1b	Response times will be shorter in the consistent- than in the inconsistent condition with thin distracters.	Simple main effects analysis
1c	Certainty ratings will be higher in the consistent- than in the inconsistent condition with thin distracters.	Simple main effects analysis
2a	Performance will be better in the inconsistent condition with thick distracters than with thin distracters.	Simple main effects analysis
2b	Response times will be shorter in the inconsistent condition with thick distracters than with thin distracters.	Simple main effects analysis
2c	Certainty will be higher in the inconsistent condition with thick distracters than with thin distracters.	Simple main effects analysis

Method

Everything in this experiment was exactly the same as it was in Experiment 3, with the exception that black distracters were used instead of the red ones.

Participants

A sample of 12 new participants, comprising 5 women and 7 men (mean age 22.83 years, range of 17 to 49 years), were recruited from Carleton University and through personal contacts. They responded to advertisements posted around campus and were each paid $10.00 for their participation. All were students of Carleton University.

All were pre-screened for colorblindness and tested for visual acuity using a mini eye chart.

Results

The results are divided into four sections: data exploration, performance, response times and certainty. All descriptive statistics, details of assumption and outlier tests, and error data are included in Appendix F.

Data Exploration

An analysis was undertaken on performance data to identify threats to the assumptions of ANOVA. Tests of normality showed that normality was violated in one condition, however no cases were identified as outlying thus the data were not adjusted. Sphericity was not violated.

An analysis was done for response times. Tests of normality showed that the distributions for average response times did not deviate from normal and thus the data were not adjusted. Sphericity was not violated.

An analysis was undertaken for certainty and tests of normality showed that normality was violated in one condition. Only one case was identified as outlying and because this case had consistently low ratings, the data were not adjusted. Sphericity was violated for size by orientation, $W = .35$, $\chi^2(2) = 10.57$, $p = .005$. A Huynh-Feldt Epsilon correction was applied, if the correction resulted in different degrees of freedom than when Sphericity was assumed.

Performance

The mean percent correct for line orientation and size conditions are shown in Figure 18. There appears to be very little difference in performance for thin and thick conditions. Differences do appear to exist between the three orientation conditions; performance appears higher in the flat than in both the consistent and inconsistent conditions.

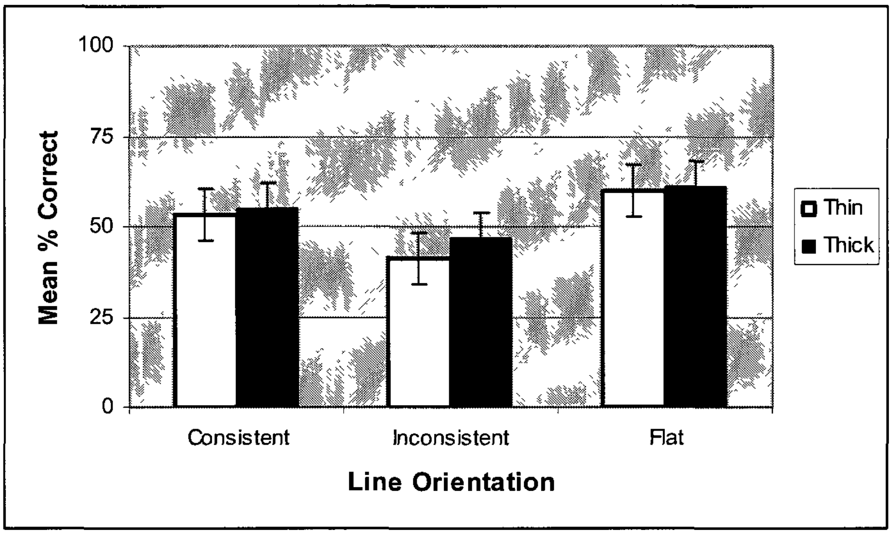


Figure 18. Mean percent correct responses for distracter line size (thin and thick) and orientation (consistent, inconsistent, flat), bars represent *CI*s for 2-way interaction = $M_i \pm$ 7.11

To test Hypotheses 1a and 2a, a (2 x 3) ANOVA was conducted for performance on distracter line size (thin, thick) and target and distracter orientation (consistent, inconsistent, flat). The interaction of size and orientation was not significant, $F(2, 22) = 0.25$, $p = .784$, $\eta_p^2 = .02$, and neither was the main effect of size, $F(1, 11) = 0.90$, $p = .362$, $\eta_p^2 = .08$. There was however a main effect of orientation, $F(2, 22) = 3.96$, $p = .034$, $\eta_p^2 = .26$. Pairwise LSD comparisons showed that performance was better with flat- than with inconsistently oriented distracters ($p = .006$). The difference between consistent-and inconsistent distracter orientations was marginally significant (p = .08) but there was no significant difference between consistent- and flat orientations (p = .409). To test Hypothesis 1a, which stated that performance would be better in the consistent- than the inconsistent condition with thin distracters, a simple main effects analysis was conducted. No differences were identified $F(1, 11) = 3.13$, $p = .105$, $\eta_p^2 = .22$. Hypothesis 2a stated that performance would be better in the inconsistent condition with thick distracters than with thin distracters. A simple main effects analysis showed that the conditions did not differ, $F(1, 11) = 1.07$, $p = .323$, $\eta_p^2 = .09$, thus Hypothesis 2a was refuted. Regardless of line thickness, inconsistent orientations resulted in the lowest performance, thereby refuting Hypothesis 2a, but showing partial support for Hypothesis 1a. Finally, a simple main effects analysis showed that flat conditions did not differ, $F(1, 11) = 0.10$, $p = .755$, $\eta_p^2 = .01$.

Response Times

Response time data for line orientation and line size are shown in Figure 19. Mean response times appear quite consistent with the consistent condition taking slightly less time than the inconsistent and the flat conditions.

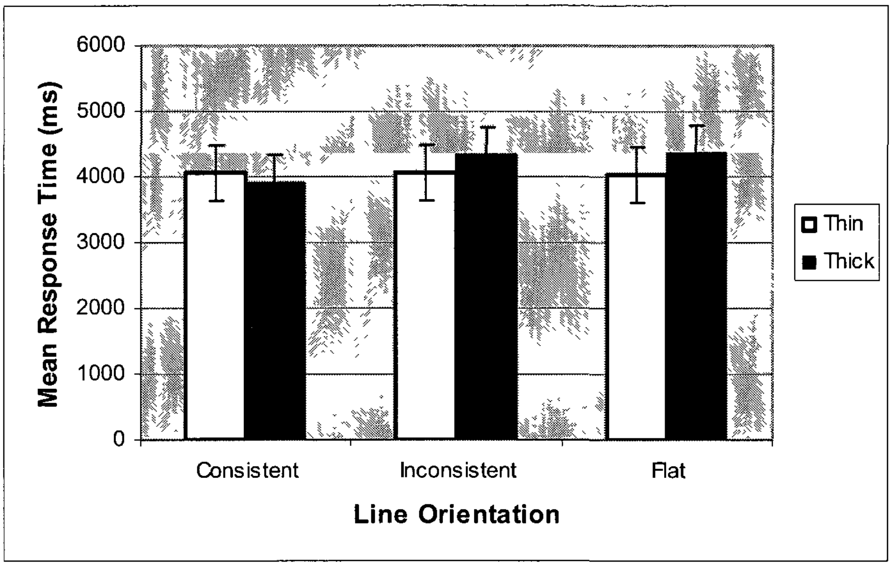


Figure 19. Mean response time (ms) for distracter line size (thin and thick) and orientation (consistent, inconsistent, flat), bars represent *CI*s for 2-way interaction = $M_i \pm$ 416.90

To test Hypotheses 1b and 2b, a (2 x 3) ANOVA was conducted for response times on distracter line size (thin, thick) and target and distracter orientation (consistent, inconsistent, flat). The interaction was not significant, $F(2, 22) = 0.90, p = .422, \eta_p^2 = .08$, and neither was the main effect of size, $F(1, 11) = 1.04, p = .330, \eta_p^2 = .09$, or orientation, $F(2, 22) = 1.04, p = .369, \eta_p^2 = .09$. Hypothesis 1b stated that response times would be shorter in the consistent- than the inconsistent condition with thin distracters. The difference was not significant, $F(1, 11) = 0.00, p = .996, \eta_p^2 = .000$, thus, it was not supported. Hypothesis 2b, which stated that response times would be shorter in the inconsistent condition with thick distracters than with thin distracters, was not supported either, $F(1, 11) = 0.65, p = .436, \eta_p^2 = .06$.

Certainty

The mean certainty data for line orientation and line size are shown in Figure 20. Certainty ratings appear quite consistent across conditions.

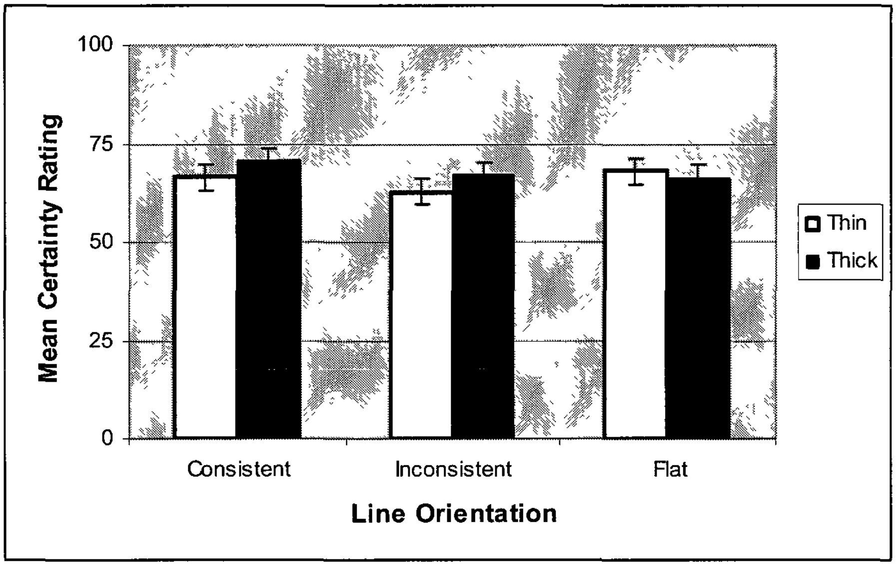


Figure 20. Mean certainty rating for distracter line size (thin and thick) and orientation (consistent, inconsistent, flat), bars represent *CI*s for 2-way interaction = $M_i \pm 3.33$

To test Hypotheses 1c and 2c, a (2 x 3) ANOVA was conducted for rated certainty on distracter line size (thin or thick) and target and distracter orientation (consistent, inconsistent or flat). The interaction between size and orientation was not significant, $F(2, 22) = 2.27, p = .127, \eta_p^2 = .17$, and neither was the main effect of size, $F(1, 11) = 1.75, p = .213, \eta_p^2 = .14$, or orientation, $F(2, 22) = 1.77, p = .194, \eta_p^2 = .14$. Hypothesis 1c stated that certainty would be higher in the consistent- than the inconsistent condition with thin distracters. Tests of simple main effects showed that the difference was not significant, $F(1, 11) = 2.37, p = .152, \eta_p^2 = .18$. It was therefore refuted. Hypothesis 2c stated that rated certainty would be higher in the inconsistent condition with thick distracters than with thin distracters. A test of simple main effects showed that certainty ratings did not differ significantly either, $F(1, 11) = 3.33$, $p = .095$, $\eta_p^2 = .23$. Thus, Hypothesis 2c was not supported.

Summary of Hypotheses

Table 10

Hypothesis number, prediction, test used and result

H #	Prediction	Test Used	Result
1a	Performance will be better in the consistent- than in the inconsistent condition with thin distracters.	Simple main effects analysis	Partially Supported
1b	Response times will be shorter in the consistent- than in the inconsistent condition with thin distracters.	Simple main effects analysis	Refuted
1c	Certainty ratings will be higher in the consistent- than in the inconsistent condition with thin distracters.	Simple main effects analysis	Refuted
2a	Performance will be better in the inconsistent condition with thick distracters than with thin distracters.	Simple main effects analysis	Refuted
2b	Response times will be shorter in the inconsistent condition with thick distracters than with thin distracters.	Simple main effects analysis	Refuted
2c	Certainty will be higher in the inconsistent condition with thick distracters than with thin distracters.	Simple main effects analysis	Refuted

Discussion

This section begins by outlining instances of serial and parallel processing strategies and how errors of interpretation can be understood. Errors due to state and process limitations are then discussed. Finally, evidence for the adoption of processing strategies is outlined.

Serial and Parallel Processing

Hypothesis 1a stated that performance would be better in the consistent- than in the inconsistent condition with thin distracters. The finding that performance was marginally better in the predicted direction offers some support for Hypothesis 1a. The consistent condition resulted in better performance than the inconsistent condition but no differences were identified between the thin and thick conditions. This suggests that both distracter line sizes were processed serially. Thus, contrary to the predictions of Hypothesis 1b, response times were not shorter in the consistent- than in the inconsistent condition with thin distracters. Hypothesis 1c stated that certainty would be higher in the consistent- than in the inconsistent condition with thin distracters. It was also refuted. The poor performance in the inconsistent condition compared to the consistent condition suggests that the lines could not readily be distinguished from one another and processed together. Serial processing was thus most likely adopted.

Hypothesis 2a predicted that performance in the inconsistent condition would be better with thick- than with thin distracters. No such differences were found, thus refuting Hypothesis 2a. Response times were predicted to be shorter in the inconsistent condition with thick distracters than with thin distracters, according to Hypothesis 2b. It was also refuted. The same was true for the certainty ratings, which, according to Hypothesis 2c, predicted that certainty would be higher in the inconsistent condition with thick distracters than with thin distracters. Apparently, participants paid attention to each inconsistently oriented line, regardless of whether line size made distracters more discriminable. Evidently, targets were not more 'discriminable' in the thick- than in the thin inconsistent conditions. Thus, there was no evidence for parallel processing with thick inconsistent distracters, suggesting instead that a serial processing strategy was adopted. Serial processing appears to have occurred for both thin- and thick inconsistent distracter lines, which contradicts the claim that performance should be better with thick- than with thin distracter lines. It therefore appears that participants were unable to utilize a more efficient processing strategy than serial processing in the inconsistent condition regardless of distracter line size variations. By contrast, performance in the flat- and the consistent conditions was significantly better than in the inconsistent condition regardless of distracter line size variations. There were no differences between the flat and the consistent conditions, which also indicates parallel processing.

State Limitations and Process Limitations

Performance in the two flat distracter conditions did not differ and were not surpassed by performance in any other condition. Collapsing across line thickness, performance did not differ between the flat- and the consistent condition. It was also better in the flat- than in the inconsistent condition. Performance was marginally better in the consistent condition than in the inconsistent condition. These results suggest that the flat- and consistent conditions represented state limitations only. By contrast, errors in the inconsistent condition can be attributed to state- as well as process limitations.

Strategy Adoption

Target and distracters did not consistently vary in size or in color. The stimulus redefinition strategy was therefore not expected to be an option for more efficient processing. Performance with thin and thick distracters was indistinguishable in inconsistent conditions. Thick distracters should have resulted in better performance than thin distracters if a selective serial processing strategy had been adopted. Thus, no optimizing strategy appears to have been adopted for enhanced performance in the inconsistent condition. These results also run contrary to those found in Experiments 1 and 2 and the implications for the model and graph processing will be taken up in the General Discussion.

The above results show that performance did not differ systematically within the flat orientation conditions despite differences in distracter line size. This suggests that variations in distracter line size did not create interference when target orientation varied. The results therefore indicate that line size and line orientation are separable dimensions. This is consistent with findings of Experiments 1, 2 and 3.

General Discussion

This chapter summarizes the main findings of the four formal experiments in this thesis and presents the contributions these findings make to better understanding perceived line graph complexity. This is followed by an outline of the limitations of the above experiments. The final section describes future research.

Summary of the Main Findings and Contributions

The goal of this thesis was to improve our understanding of the sources of perceptual errors in graph interpretation. Perceptual processes were explored by limiting stimulus exposure time and varying certain aspects of distracter lines in simple line graphs.

Accuracy of Perceptual Processing

Throughout all four experiments, performance was well above chance levels in all target- and distracter line orientations. Limiting stimulus exposure time to 50 ms reduced the opportunity for iterative information processing and prevented participants from relying on top-down processes, suggesting instead that accurate interpretation relied mainly on bottom-up processing. Carpenter and Shah (1998) demonstrated that graph comprehension is an iterative process that entails inspecting the pattern on the graph and the axes until the meaning of the graph has been fully comprehended. The present findings showed clearly that interpretation occurs even with very limited opportunities for iterating the graph information processing activity. This observation that line graphs can be interpreted very quickly suggests that automatic processing occurs in some form, which is consistent with Pinker's (1990) notion of a 'default encoding mechanism'. Thus, this manipulation of the stimulus exposure time facilitated a better understanding of perceptual processing of line graph information. That has not been shown before. It allowed an exploration of the separability of line orientation, color and size, which enabled clarification of the reasons processing errors occurred. The approach also revealed some strategies that participants probably employed to overcome processing errors.

Separability of Dimensions

The main objective of the above experiments was to identify ways in which errors of interpretation can be explained by perceptual processing mechanisms. Two sources of evidence emerged to suggest that the combinations of distracter line size, orientation, and color are separable dimensions. One came from comparisons of the flat distracter line conditions in each of the four Experiments. These comparisons revealed no statistically traceable performance differences whatsoever: variations in distracter line combinations of color or size did not interfere with interpretation in the presence of a horizontal distracter line. This result provides some evidence that the combinations of the three dimensions tested here are indeed separable in the context of the kinds of simple line graphs employed throughout this thesis. The second source of evidence came from performance differences that could be attributed to participants adopting different task-completion strategies, namely configurability of dimensions, selective serial processing and stimulus redefinition. These can only be adopted if dimensions are separable. The specific findings are discussed in a later section. Suffice it to say here that both sources of evidence suggest that combinations of distracter line orientation, color, and size, are separable in the context of simple line graphs with only two lines.

The literature exploring dimension separability with graphs has focused on graph types where the height of bars, the size of a graph, or the area and the angle are the defining stimulus characteristics of the variables in the graph (e.g., Hollands, 2003). For example, as was discussed earlier (p. 31), length, area and size judgments can be measured with stacked bar graphs for which angle judgments are not possible, just like length judgments are not possible for pie charts. Hollands' (2003) experiments, employing these two types of graphs, lead to two important findings. First, Hollands' results showed that the same manipulations performed on different graph types can lead to differences in performance. Second, they showed that at least some dimensions are separable for some, but not for all, graph types, providing evidence for the asymmetric integrality of the dimensions. Hollands attributed these performance differences to the type of dimensions being judged. He found that bars were judged by length and area but pies were judged by area and angle. Area and angle were shown to be separable dimensions.

The above thesis findings demonstrate that it is possible to apply Garner's theory of dimension separability to interpreting trends in line graphs as well as to classifying graphical images by sizes, lengths, areas and angles. This is new knowledge. Thus, even graph types that do not lend themselves easily to judgments of areas and angles can also demonstrate dimension separability.

Line Graph Interpretation Model

Separable dimensions can lead to specific performance outcomes that can be understood in terms of serial and parallel processing mechanisms. The relevant section of the model presented earlier (p. 51) and the associated premises are shown in Figure 21.

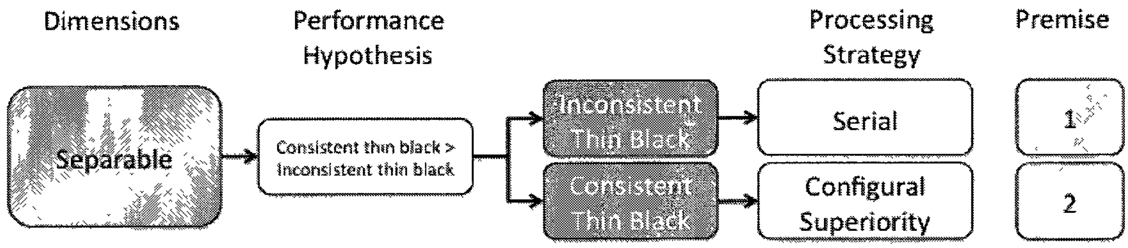


Figure 21. Performance comparisons in thin black distracter conditions and processing strategies reflected by premises 1 and 2

As the Figure shows, when two lines are both thin, black, and inconsistently oriented, it should be difficult and hence take longer than 50 ms to identify the slope direction and magnitude of the target line than when both lines are oriented consistently. Here the only distinction between the two lines is the target line has direction changes and the distracter line is straight. When inconsistently oriented, the two lines would have to be attended to one at a time to complete the task correctly. Poor performance compared to the consistent orientation would be evidence that 50 ms exposure times were insufficient to allow iterative processing. In accordance with Premise 1 presented earlier (p. 52), this poorer performance would allow one to infer that serial processing would most likely be responsible for inaccurate performance. Performance should be better with consistently oriented lines because they would configure and be processed simultaneously with 50 ms exposure times rendering iterative processing unnecessary for accurate interpretation to occur. According to Premise 2, parallel processing due to configural superiority effects is represented by consistent thin black orientations. If performance is undifferentiated between the inconsistent and consistent orientations, then parallel processing would be assumed for both orientation conditions and there would be no evidence that a different processing mechanism, namely serial processing, was adopted. The performance difference between the consistent and inconsistent conditions was significant in the predicted direction in Experiment 2 and marginally significant in Experiment 1. These results therefore support Premises 1 and 2, thereby providing some evidence that the two processing mechanisms can be experimentally distinguished from one another. It will be recalled that Experiment 1 contained both red and black as well as thick and thin distracters. In Experiment 2, all distracters were thin, varying only in color. It is therefore safe to conclude that evidence of serial processing occurred in inconsistent conditions with thin black distracters. Parallel processing was inferred from performance in the conditions with thin black and consistently oriented distracters. In Experiment 4, where all distracters were black, varying in size only, performance differences were not observed between thin and thick distracters regardless of orientation condition. However, despite the lack of variation in distracter size, consistent orientations resulted in better performance than inconsistent orientations, also showing support for the distinction between serial and parallel processing. One possible reason for this result is that participants did not notice the variation in distracter line size.

The next segment in the model, shown in Figure 22, displays the implications for performance in inconsistent conditions when both targets and distracters differ in color and/or size as well as when they do not differ on these dimensions. If performance with thin black distracters is statistically indistinguishable from performance with thin or thick red or thick black distracters in inconsistent conditions, this would provide evidence that the same processing strategy was likely used in all conditions, in this case, serial processing. If performance on the thin red, thick red or thick black inconsistent conditions is statistically distinguishable from the thin black inconsistent condition, then this would indicate that a different kind of processing occurred. This is the assumption underlying Premise 3. Thus, if performance in the thin red, thick red or thick black inconsistent conditions is better than in the thin black condition, this would indicate parallel processing. This is shown in Figure 22.

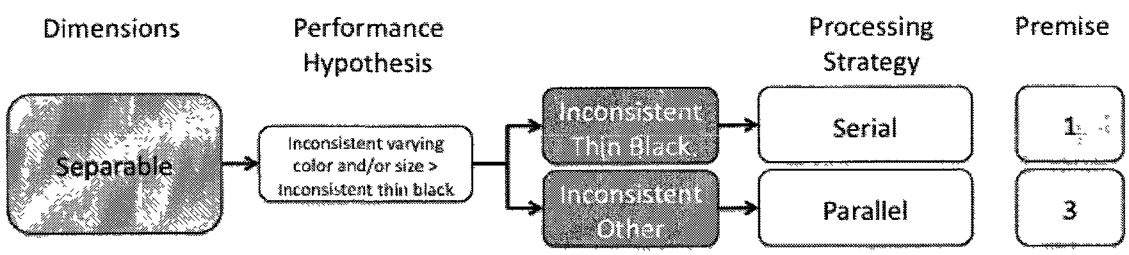


Figure 22. Performance comparisons in inconsistent conditions and processing strategies reflected by premises 1 and 3

It will be recalled that the flat distracter conditions was the benchmark for the best possible performance, under these limited exposure time conditions, and that any errors occurring in the presence of a flat distracter line could be attributed to state limitations. This is because performance on any condition did not surpass the flat condition, thus it was taken to represent the best possible performance with limited exposure times. Any additional errors occurring in the other conditions could then be attributed to information processing limitations, due to the perceptual complexity of the graphs. Thus, by limiting the stimulus exposure time, it was possible to explore variations in perceptual complexity and distinguish state- from process limitations. This exploration allowed us to identify which manipulations were responsible for rendering even a simple line graph perceptually complex. If no statistically reliable difference in performance were to be found between all conditions including the flat, this would represent the best possible outcome, and one would have been able to infer that a stimulus redefinition strategy would have been applied. This is the assumption underlying Premise 4. However, if performance improved in inconsistent conditions compared to those in which serial processing was expected, and other conditions were found to differ from the flat as well, then performance would not be the best it could have been in all cases. In that case, one would infer that a selective serial processing strategy had been applied. As shown in Figure 23, this is the assumption underlying Premise 5. Thus, Premises 4 and 5 represent inferences regarding the strategies adopted when parallel processing was observed in inconsistent conditions.

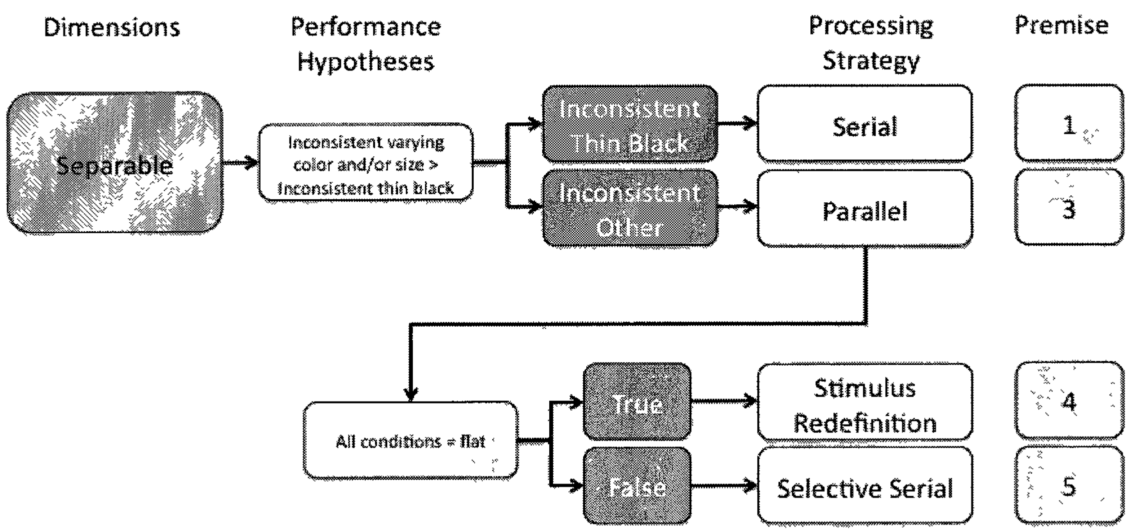


Figure 23. Performance comparisons to flat distracter conditions and processing strategies reflected in premises 4 and 5

The results from these formal experiments offered support for the five premises in explaining the relations between the performance hypotheses and the processing strategies adopted.

Perceptibility of Color. In Experiment 1, the distracter lines could be thin or thick, red or black. In addition to performance being better in the thin red than in the thin black inconsistent condition, it was also statistically indistinguishable from performance in the flat red condition. This indicates that processing was the best it could have been. Performance in the thick red inconsistent condition was also marginally better than performance in the thin black inconsistent condition, but was not as good as in the thick red flat condition. This indicates that performance on inconsistent conditions was enhanced most when targets and distracters differed only in color.

In Experiment 2, the distracters were all thin but could be either black or red. Performance was again better in the thin red than in the thin black inconsistent distracter condition, and it was also statistically indistinguishable from the thin red flat condition. These findings suggest that parallel processing can adequately account for accurate performance when distracter color differs from that of the target. Evidently, a distracter line that differs in color from the target line increased the discriminability of the two lines, enabling participants to selectively attend to the target, as was the case with red distracter lines. An increase in the discriminability of the lines was also observed with thick red inconsistently oriented distracters, compared to the thin black inconsistently oriented distracters, as in Experiment 1, although performance was poorer than the flat condition. However, the target was not always as readily discriminable, since performance differed between the flat and the other conditions in some instances in Experiments 1 and 2. These findings indicate that selective serial processing occurred in inconsistent conditions when the distracters color was unpredictable. The observation that performance was better than expected with thin black distracters, but not always as good as the best possible performance, supports this interpretation. Feature Integration Theory of Attention (Treisman & Gelade, 1980) holds that the dimensions such as color, size and orientation would have to be processed serially for the target to be accurately identified if targets and distracters vary throughout a stimulus set. This is because they are not sufficiently distinguishable to be attended to simultaneously. If the distracters do stand out, then the parallel process could explain efficient performance. The results of the above comparisons from Experiments 1 and 2 supports the interpretation offered by Feature Integration Theory as well.

In Experiment 3 in which all distracter lines were red but could be either thin or thick, no differences could be attributed to the size variations. Performance across all conditions was undifferentiated from the flat conditions, indicating that it was the best it could be under these limited exposure time conditions. This finding lends support to Bertin's (1983) claims that color and orientation are selective and should therefore be highly perceptible. Furthermore, a stimulus redefinition strategy (Garner, 1974) can explain the reason for the parallel processing in Experiment 3. Apparently, participants were able to focus exclusively on the target and ignore the distracters when the color of both targets and distracters were predictable. Guided Search Theory (Wolfe, Cave & Franzel, 1989) holds that the parallel process guides the serial process and thus, attention can be weighted more heavily on the characteristics of the target than on the characteristics of the distracters when they differ consistently and are predictable. In the present thesis, the undifferentiated performance between all conditions in Experiment 3 suggests that attention was guided to the dimensions relevant to the target, in that case, the color black. The data from Experiment 3 support this interpretation and also suggests that line graph processing can be explored using Guided Search Theory, which has not yet been applied in a line graph context. In the trend identification task employed in these thesis experiments the manipulation of distracter line color made the targets easier to identify, thereby enhancing performance.

Perceptibility of Size. In Experiment 1, it was found that thick black inconsistent distracter conditions resulted in better performance than thin black inconsistent distracter conditions. This finding provides evidence of parallel processing because distracter size improved processing above what would be expected if processing was serial. It supports Bertin's claim regarding the high perceptibility of size and is also evidence of a selective serial processing strategy. However, in Experiment 3, as noted above, line size variations had no effect on efficient processing when all distracters were red. The same was also true in Experiment 4 when all distracters were black. Variations in size variations had no effect on performance. It was apparently difficult to detect the differences in inconsistent conditions when all distracters were black. The results from Experiment 4 contradicted the prediction that performance in the thick black inconsistent conditions should be better than in the thin black inconsistent conditions. The results from Experiment 3 contradicted the prediction that thin red inconsistent conditions would be better than thick red inconsistent conditions. Consequently, these findings suggest two things regarding the effects of manipulating distracter size in stimulus sets that hold distracter color constant. First, if the distracters are a consistently different color from the targets, line size will not affect performance because color will guide processing. Second, if the distracters are consistently the same color as the target, line size manipulations are not salient enough to guide processing. The results of Experiment 4 suggest that it was not possible to adopt a performance-enhancing strategy beyond that of serial processing, when all distracters were black. Thus, when target and distracter color was held constant throughout an experiment, line size manipulations had no effect on performance as evidenced in Experiments 3 and 4. Line size manipulations did, however, have an effect in inconsistent conditions when the stimulus set also contained distracters that varied on color from targets, as in Experiment 1.

Taking a common sense approach, one could reason that the adoption of different strategies is dependent entirely upon participants noticing the experimental manipulations. The results from Experiment 1 suggest that when manipulating color and size within a stimulus set, color as well as size differences are noticed. However, when color is held constant, as in Experiments 3 and 4, size differences are not as noticeable; they do not enhance or degrade performance in inconsistent conditions. This could be because participants may become aware that other less noticeable things, like line size, may also differ when color does as well, but when color does not vary, they may forget to pay attention to the less noticeable manipulations. It would therefore appear that differences in target and distracter color are more important than size for enhancing the visual discriminability of lines in a line graph.

Carpenter & Shah (1998) developed a model to describe the processes involved in line graph interpretation. They showed that adding lines to a graph did not affect task performance if the additional lines were consistently oriented with the existing lines.

Performance, however, was adversely affected when the additional lines were inconsistently oriented with the existing lines. They did not test the effects of varying line color or line size on the interpretation tasks. The above thesis findings suggest that processing efficiency can be improved if inconsistent target and distracter lines differ in color. In some cases, processing efficiency also increases when the two lines differ in size.

In summary, these thesis results suggest that a selective serial processing strategy can be adopted when the color of the distracter lines could be either black or red, rendering distracter color unpredictable. Similarly, the results demonstrate that a stimulus redefinition strategy can be adopted when the color of targets and distracters consistently differ from each other, making it possible to focus attention exclusively on the target characteristics. However, when the color of the targets and distracters do not differ from each other, no performance enhancing strategy will likely be adopted, even if line size differs as thick black distracters did not enhance performance compared to the thin black distracters in inconsistent conditions. These interpretations provide a parsimonious account of the circumstances under which possible performance enhancing strategies can be adopted in the context of line graph interpretation tasks. This is new to the literature.

The objective of this research was to improve our understanding of how information is processed in simple line graphs to explain differences in accuracy on trend identification judgments. This research complements the work of Simkin and Hastie (1987), whose research objective was to explain the processes responsible for accurate proportion judgments in stacked bar graphs and pie charts. They showed stacked bar graphs or pie charts for 500 ms or for an unlimited time and asked participants to judge proportions and estimate magnitudes. By measuring errors and response times, they found that processes such as scanning and anchoring were responsible for accurate proportion judgments. Although these processes can help to explain proportion judgments about stacked bar graphs, the processes they describe cannot be applied in a line graph context. This is because the task of judging a proportion is very different from that required in a trend identification task. In the present thesis, these experiments showed that the processing strategies involved in line graph interpretation can be better understood by limiting the stimulus exposure times enabling the allocation of errors to serial and parallel processing mechanisms. It has not been possible before to predict performance outcomes on trend identification tasks using simple line graphs. This thesis makes these kinds of predictions possible – which is a contribution to the literature.

Limitations

There are a few limitations that should be noted regarding these thesis experiments. First, the above series of experiments employed small samples, which raises two sample-related issues. Most participants were recruited from Carleton University, which represents a homogeneous population and all were paid for their participation. Payment may have been a motivator attracting a particular subgroup of the population. Thus, these experimental results may not be generalizable beyond a Canadian University population receiving payment for participation. Second, outlying scores, response time data and certainty ratings did pose some threats to the assumptions underlying the ANOVA statistic. Although small adjustments to a specific datum eliminated the threats to assumption violations, dealing with outliers always poses a challenge. Here, this challenge was addressed by moving outlying scores to one standard deviation higher or lower than the group mean, depending on its place in the distribution. Although this modification seemed reasonable, the results do reflect those data analysis decisions. Hence, different results could have been obtained if the original data have been retained.

The issue of serial and parallel processing of line graph stimuli warrants further investigation. Although the results of this unique paradigm suggest evidence of each process, the results are far from conclusive. There are many more complex parallel processing paradigms which could have been employed to interpret these thesis results. However, to maintain consistency between the terminology used by Garner (1974) and Treisman and Gelade (1980), the terms serial and parallel processing were used throughout this thesis.

Finally, these thesis results may not be generalizable beyond the simple line graph reading task employed here or beyond simple line graph reading tasks completed during very brief exposure times. Similarly, it is unclear if these results could be generalized to experiments employing a similar research paradigm and to card sorting- or search task paradigms.

Future Research

Findings in this thesis suggested that at least some graphical information can be interpreted very quickly. However, it would also be of interest to investigate how quickly it can be interpreted because this would help to establish the limits of line graph perceptual processing. This could be done by using three or four different exposure times and shorten these even below the 50 ms used here. As there were a few problems identified with terminology describing the slope of the lines (See Appendix B, Preliminary Study 2), it would be worthwhile to investigate how participants define small and sharp slopes.

The issue of how well people can complete graph interpretation tasks on interfaces of varying sizes is a very timely and interesting question that was not addressed by the present thesis. It would be interesting to conduct the same experiments using different graph display sizes that would fit onto a small mobile device. It is difficult to read the information on a graph when presented on a small device. This series of studies can help to determine the amount of information required for simple interpretation to occur without requiring graph readers to enlarge the image. One could also include samples of people with varying graph reading experience to determine the effect of graph reading experience on graph perception and interpretation.

The experimental paradigm employed in the present series of studies used controlled exposure times but it is possible that similar results would materialize if other paradigms were used. One possibility would be to employ a card-sorting paradigm asking participants to sort graphs as quickly as possible, based on the magnitude or the direction of the slopes. This would be comparable to Hollands' (2003) studies and could answer questions aimed at assessing asymmetric integrality of the visual dimension of graph size on the cognitive dimension of trend judgment in line graphs, varying the perceptual dimensions of line size, color and orientation. Asymmetric integrality could not be assessed here. Employing this paradigm would also make the results more comparable to Garner's card sorting paradigm, which would allow a more conventional exploration of dimension separability of line size, color and orientation in a line graph context. Another possibility is a forced choice paradigm, for example, which could address the question of which target distracter combinations are easiest to identify. This could be done by testing two graphs at a time, asking participants to choose a graph with a particular slope, or with particular values. These two possible paradigms could be employed in a two-part repeated-measures study. These next steps could help to increase our understanding of graph perceptual complexity by comparing the results obtained across different experimental paradigms. It would also be interesting to test the effects of differing target and distracter line colors or differing line textures on graph interpretation. This would address the question of which line colors or characteristics enhance discriminability most in a line graph context. As an additional manipulation, one could also add more distracter lines to better understand the effect of varying line colors on more complex graphs. This would be comparable to the search task paradigm and could provide further evidence of serial and parallel processing mechanisms if response time curves were generated.

Conclusion

The main objective of this thesis was to understand why errors of graph interpretation occur and how they can be avoided. The above experiments allow six conclusions to be drawn with respect to the stated objective. First, and perhaps most importantly, these experiments established conditions under which serial and parallel processing could possibly be identified. This led to the finding that color is more perceptible than size in a line graph context. Second, the limited exposure time paradigm was effective for understanding line graph perceptual processing better. The above experiments showed that information was processed rather accurately and very quickly. Third, the results show some support for the notion that orientation, color and size are separable dimensions in a line graph context, suggesting that Garner's theory can be applied in a novel way to understand graph perceptual complexity. Fourth, it was possible to vary perceptual complexity even when the pattern and the amount of data were held constant across the graph stimuli. Fifth, the simple line graphs allowed us to establish the source of errors observed and partition errors due to state- and process limitations, which appeared to depend on the variations in the experimental stimuli. Finally, it appears that graph perceptual complexity affects errors of interpretation, making it possible to answer the research question posed at the beginning of this thesis, namely that separable dimensions do matter!

References

Andersen, J. R. & Bower, G. H. (1973). *Human Associative Memory*. Washington DC: V. H. Winston & Sons.

Bartlett, F. C. (1932). *Remembering.* Cambridge: University Press.

Bertin, J. (1983). *Semiology of Graphics: Diagrams, Networks, Maps*. Madison: University of Wisconsin Press.

Best, L., Hunter, A. C. , & Stewart, B. M. (2006). Perceiving relationships: A physiological examination of the perception of scatterplotts. In D. Barker-Plummer et al. (Eds.), *Diagrams*, pp. 244-257.

Bornstein, R. (1992). Subliminal mere exposure effect. In F. Bornstein & T. Pittman, (Eds.), *Perception Without Awareness: Cognitive, Clinical and Social Perspectives*. NY: The Guilford Press, 191-210.

Brinton, W. C. (1915). Joint committee on standards for graphic presentation. *Publications of the American Statistical Association. 14*, 112, 790-797.

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied, 4*, 2, 75-100.

Carswell, M. C. (1992). Reading graphs: Interactions of processing requirements and stimulus structure. In B. Burns (Ed). *Precepts, Concepts and Categories* (pp. 605-645). Elesvier: Amsterdam.

Carswell, M. C., & Wickens, C.D. (1987). Information integration and the object display and interaction of task demands and display superiority. *Ergonomics*, *30*, 3, p. 511-527.

Carswell, M. C., & Wickens, C.D. (1990). The perceptual interaction of graphical attributes: Configurality, stimulus homogeneity, and object integration. *Perception & Psychophysics, 47*, 157-168.

Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual research. *Cognitive Psychology, 22*, 225-271.

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, education, and application to the development of graphical methods. *Journal of the American Statistical Association, 79*, 387, 531-554.

Croxton, F. E., & Stein, H. (1932). Graphic comparisons by bars, squares, circles and cubes. *Journal of the American Statistical Association, 27*, 177, 54-60.

Croxton, F. E., & Stryker, R. E. (1927). Bar charts versus circle diagrams, *Journal of the American Statistical Association, 22*, 160, 473-482.

Culbertson, H. M., & Powers, R. D. (1959). A study of graph comprehension difficulties. *AV Communication Review, 7*, 97-110.

Della Salla, S., Gray, C., Baddley, A., & Wilson, L. (1987). *Visual Pattern Test (VPT).* Pearson.

Delorme, A., Guillaume, A, Rousselet, M., J.-M., Fabre-Thorpe, M. (2004). Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. *Cognitive Brain Research, 19*, 103-113.

Eells, W. C. (1926). The relative merits of circles and bars for representing component parts. *Journal of the American Statistical Association. 21*, 154, 119-132.

Enns, J., T. & Di Lollo, V. (1997). Object substitution: A new form of masking in unattended visual stimuli. *Psychological Science, 8,* 2, 135-139.

Enns, & Di Lollo, (2000). What's new in visual masking? *Trends in cognitive Science, 4,* 9, 345-352.

Feinberg, S. E. (1979). Graphical methods in statistics. *The American Statistician, 33*, 4, 165-178.

Felfoldy, G. L., & Garner, W. R. (1971). The effects of speeded classification of implicit and explicit instructions regarding redundant dimensions. *Perception and Psychophysics*, *9*, 3(A), 289-292.

Garner, W. R. (1974). *The Processing of Information and Structure.* Potomac: Lawrence Erlbaum Associates.

Garner, W. R., & Felfoldy, G. L. (1970). Integrity of stimulus dimensions in various types of information processing. *Cognitive Psychology*, *1*, 225-241.

Gattis, M., & Holyoak, K. J. (1996). Mapping conceptual to spatial relations in visual reasoning. *Journal of Experimental Psychology: Learning Memory and Cognition, 22*, 1, 231-239.

Hastie, R., Hammerele, O., Kerwin, J., Croner, C. M., & Hermann, D. J. (1996). Human performance reading statistical maps. *The Journal of Environmental Psychology: Applied, 2*, 1, 3-16.

Hink, J. K., Euastace, J. K., & Wogalter, M. S. (1998). Do graphs enable the extraction of quantitative information better than pure graphs? *International Journal of Industrial Ergonomics,* 22, 439-447.

Hollands, J. G. (2003). The classification of graphical elements. *Canadian Journal of Experimental Psychology*, *57*, 1, 38-47.

Hollands, J. G., & Spence, I. (2001). The discrimination of graphical elements. *Applied Cognitive Psychology, 15*, 413-431.

Hollands, J. G., & Spence, I. (1998). Judging proportion with graphs: The summation model. *Applied Cognitive Psychology, 12*, 173-190.

Jarmasz, J. & Hollands, J. G. (2009). Confidence intervals in repeated-measures designs: The number of observations principle. *Canadian Journal of Experimental Psychology, 63*, 2, 124-138.

Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology, 3*, 185-226.

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

McClure, G. B. (1998). *The Effects of Spatial and Verbal Abilities on Graph Use.* Unpublished doctoral dissertation, University of New Mexico.

Marr, D. & Nishihara, H. K. (1978). Representations and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London, Series B, Biological Sciences*, *200*, 1140, 269-294.

McVee, M. B., Dunsmore, K., & Gavelek, J., R. (2005). Schema theory revisited. *Review of Educational Research*, *75*,4, 531-566.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 2, 81-97.

Murphy, E. D. (2000). *Cognitive Issues in Autonomous Spacecraft Control Operations: An Investigation of Software-Mediated Decision Making in a Scaled Environment.* Unpublished doctoral dissertation, University of Maryland.

Navon, D. (1977). Forest before the trees: The precedence of global features in visual perception. *Cognitive Psychology. 9*, 353-383.

Pinker, S. (1990). A theory of graph comprehension. In R. Freedle (Ed). *Artificial Intelligence and the Future of Testing* (pp. 73-126). New Jersey: Lawrence Erlbaum Associates.

Pomerantz, J. R. (1981). Perceptual organization in information processing. In M. Kubovy and J. R. Pomerantz (Eds.). *Perceptual Organization* (pp. 144-180). Hillsdale, New Jersey: Erlbaum.

Pomerantz, J. R., & Garner, W. R. (1973). Stimulus configuration in selective attention tasks. *Perception & Psychophysics, 14*, 3, 565-569.

Pomerantz, J. R., & Schwaitzberg, S. D. (1975). Grouping by proximity: Selective attention measures. *Perception & Psychophysics, 18*, 5, 355-361.

Popper, K. (1959). *The Logic of Scientific Discovery.* London: Routledge Classics.

Ratwani, R. M., & Trafton, J. G. (2004). Making graphical inferences: A hierarchical framework. In K. Forbuus, D. Genter, & K. Regier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, Chicago, IL: Earlbaum.

Ratwani, R. M., Trafton, J. G., & Boehm-Davis, D. A. (2003). Thinking graphically: Extracting local and global information. In R. Alterman & D. Kirsch (Eds.). *25th Annual Meeting of the Cognitive Science Society.* Boston, MA: Erlbaum.

Roth, W. M. (2004). Emergence of graphic practices in scientific research. *Journal of Cognitive Culture, 4,* 595-627.

Schwartz, M., & Pritchard, W., S. (1981). AERs and detection in tasks yielding U-shaped backward masking functions. *Psychophysiology, 18,* 678-685.

Shah, P., & Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology: General, 1*, 24, 43-61.

Sheppard, R. N. (1991). Integrity versus separability of stimulus dimensions: From an early convergence of evidence to a proposed theoretical basis. In G.R. Lochead and J. R. Pomerantz (Eds). *The Perception of Structure,* Washington DC: APA.

Simkin, D., & Hastie, R. (1987). An information-processing analysis of graph perception. *Journal of the American Statistical Association*, *82*, 398, 454-465.

Treisman A. (1982). Perceptual grouping and attention in visual search for features and for objectives. *Journal of Experimental Psychology*, *8*, 2, 194-214.

Treisman, A. (1985). Pre-attentive processing in vision. *Computer Vision, Graphics, and Image Processing*, *31*, 156-177.

Treisman A. (1988). Features and objectives: The fourteenth Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology, 40*, 2, 201-237.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetrics. *Psychological Review, 95*, 1, 15-48.

Treisman A., & Patterson, R. (1984). Emergent features, attention, and object perception. *Journal of Experimental Psychology*, *10*, 1, 12-31.

Treisman, A., & Sato, S. (1990). Conjunction Search Revisited. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 3, 459.

Treisman, A., & Schmidt, H. (1982). Illusionary conjuctions in the perception of objects. *Cognitive Psychology*, *14*, 107-141.

Townsend, J. T. (1971). A note on the identifiability of parallel and serial processes. *Perception & Psychophysics, 10*, 3, 161-163.

Tufte, E. R. (1983). *The Visual Display Of Quantitative Information.* Chesire: Graphics Press.

Tukey, J. W. (1972). Some graphic and semi graphic displays. In T. A. Bancroft (Ed). *Statistical Papers in Honour of George W. Snedecor*, (pp. 293-316). Ames: Iowa State University Press.

Tukey, J. W. (1990). Data-based graphics: Visual display in the decades to come. *Statistical Science, 5*, 3, 327-339.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity principle and retrieval processes in episodic memory. *Psychological Review, 80*, 5, 352-373.

Van Rullen, R & Thorpe, S., J. (2001). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience*, *13,* 4, 454-461.

von Huhn, R. (1927). Further studies in the graphic use of circles and bars: A discussion of the Eell's Experiment. *The Journal of American Statistical Association, 22*, 157, 31-39.

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

Webster's Online Dictionary (2006). Retrieved October 31, 2010 from: http://www.websters-online-dictionary.org/

Wickens, C. D. (1984). *Engineering Psychology and Human Performance.* Columbus: Charles E. Merrill Publishing.

Wickens, C. D., & Andre, A. D. (1990). Proximity compatibility and information display: Effects of color, space, and objectness on information integration. *Human Factors, 32*, 1, 61-77.

Wickens, C. D., & Carswell, M. C. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors, 37*, 3, 473-494.

Winn, W. (1994). Contributions of perceptual and cognitive processes to the comprehension of graphics. In Schnotz, W., Kulhavy, R. W. (1994). *Comprehension of Graphics* (pp. 1-27). The Netherlands: Elsvier Science.

Witkin, H., Oltman, P., Raskin, E. & Karp, S. (1971). A Manual for the Group Embedded Figures Test. Palo Alto, CA.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided Search: An alternative to feature integration model for visual research. *Journal of Experimental Psychology, 15*, 3, 419-433.

Zacks, J., & Tversky, B. (1999). Bars and lines: A study of graphic communication. *Memory & Cognition, 27*, 6, 1073-1079.

Zhang, J. (1996). A representation analysis of relational information displays. *International Journal of Human-Computer Studies, 45*, 59-74.

Zhang, J., & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science, 18*, 87-122.

Zajonc, R. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35(2), 151-175.

Appendix A

Preliminary Study 1

If initial processing occurs from the bottom up and symbols are detected, organized and configured automatically, people should be able to respond accurately to some types of questions about graphs presented very briefly. By showing graph stimuli very briefly and administering tasks that require attention to detail as well as to global aspects, it should be possible to demonstrate that complex tasks require more top-down processing than simple ones. Generally, researchers agree that there are three types of graph reading tasks, which vary in difficulty due to the level of attention to detail required for each (p. 22).

Previous research has found that line graphs are best suited for trend identification tasks and bar graphs are best suited for tasks that require comparison or value identification (Pinker, 1990; Zacks & Tversky, 1999). Both types were therefore employed in this experiment. The preliminary study tested initial processing to determine if a 50 ms exposure time would suffice for participants to interpret the graphs correctly. For tasks that require people to identify values, performance has been found to be better on graphs with grid lines than on graphs without grid lines (Culbertson & Powers, 1959; Lohse, 1993). All graphs in this thesis had grid lines and that issue was not explored further. There is also some agreement that variable labels are more easily interpreted than legends or keys (Carpenter & Shah, 1998; Culbertson & Powers, 1959; Lohse, 1993). Thus, all graphs in this thesis had labels instead of legends. Gattis and Holyoak (1996) found that axis labeling improved performance. In this thesis, all axes were labeled and that issue was not explored further. Three task types were presented in this preliminary experiment: trend identification-, comparison-, and read-off tasks. All involved extracting information and interpreting the meaning of the graph.

To isolate the perceptual stages, with the objective of focusing this investigation on bottom-up perceptual processing, participants' ability to invoke top-down processing was limited in the following five ways: 1) participants were aware of the graphs they were to interpret, limiting the reliance on graph schema to interpret the conceptual message; 2) the nature of the tasks was known and explained in advance to limit the number of possible conceptual questions; 3) exposure times were limited to eliminate the iterative graph interpretation cycle; 4) only bar- and line graphs were used to ensure familiarity with the graph types; and 5) bar graphs were employed to signal read-off- and comparison messages and line graphs were employed to signal trend identification messages. Control of these extraneous variables ensured that performance errors were due to faulty perception and not to cognitive graph interpretation issues.

The prediction task involved line graphs only. This brief exposure time task has not been tested with line graphs before. Therefore, if initial processing is bottom-up (Kosslyn, 1989; Pinker, 1990; Winn, 1994), participants should select the correct option more often than if a top-down process is necessary.

The comparison task required comparisons between two variables presented in a bar graph. Again, if the graph is initially processed in a bottom-up fashion, the bars should separate from the background (Pinker, 1990), leaving only the two target bars to be considered for the comparison judgment. If the bars do not separate from the background, the task should not be completed successfully, suggesting that more top-down processing was required. Furthermore, if the items to be compared are the same color, they should be perceived together (Bertin, 1983; Carswell & Wickens, 1987; Wickens & Carswell, 1995) thus enhancing performance. Finally, the read-off task required identification of values associated with the variables, also in a bar graph. For this type of task, bar graphs should signal length and value relations, according to Pinker (1990). If the relation of interest, namely 'value', is signaled, performance should be better than chance. If performance is less accurate than predicted by chance at limited exposure times, then it is more likely that value identification tasks requires more top-down processing. A value identification task with graphs presented very briefly has not been tested before.

Method

Participants

Eleven participants were recruited from the general population, but the first two assisted in clarifying instructions and the procedure, so their data were excluded. Nine participants took part (5 women and 4 men, mean age of 32.6 years, range of 19 to 46 years). Each was screened to ensure that their eye sight was 20/20 or corrected to 20/20 and that they had some experience with reading both bar- and line graphs. A simple eye-chart reading test and a short graph interpretation test were administered. Participants were tested in individual sessions lasting approximately 30 minutes and were offered $5.00 for their participation.

Location

The experiment took place at the HOTLab of the Social Science Research Building , at Carleton University, or at another suitable location with comparable conditions.

Apparatus

Each participant was tested on a Hewlett Packard laptop workstation with Pentium 4 CPU, 512 Mbytes of RAM, with screen resolution set to 1024 by 768 pixels. A program created in DirectRT™, was used to present graphs, collect responses, ratings and response times.

Materials

An informed consent- and a debriefing form were prepared. A mini eye chart to test eye-sight, a pre-qualifying questionnaire, and a graph comprehension test were administered. Pens were provided. Ten graphs were used for each task, totaling 30 graphs (10 line - trend identification, 10 bar - comparison and 10 bar - read-off), with a total of four response option graphs shown in a multiple-choice paradigm for each. A pop-up text box was used to collect certainty ratings which ranged from 0 = 'not certain at all' to 100 = 'completely certain'. Instructions were presented on the screen. Examples of graph stimuli and response options are presented on pages 149-152. See Appendix AA for all other materials.

Design

This preliminary study employed a completely within-group design. The stimuli were presented for 50 ms each. The three tasks were counterbalanced to avoid order effects. Three participants received them in the order: trend identification (trend), comparison (compare), read-off (read); another three received them in the order, compare, read, trend; and the final three received them in the order, read, trend, compare.

For the trend identification task, the 10 line graphs were presented, preceded by 5 practice graphs. Each graph displayed only one line, with one or two direction changes. The line on the graphs was created in MS Excel and was presented in black with a standard thickness. The first two pilot participants were asked to "predict what would happen next". That instruction was not clear so it was amended to, "describe the trend" for other the nine participants. An example of a stimulus graph is presented in Figure A1a.

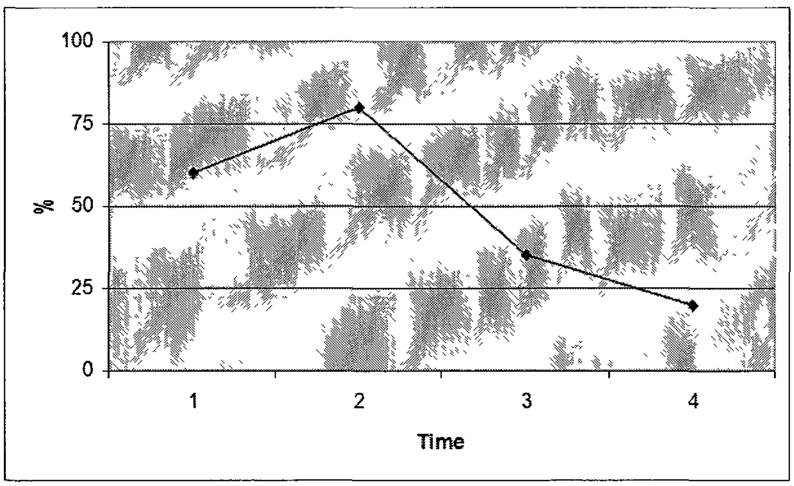


Figure A1a. Example of a stimulus graph for trend identification task.

Participants were provided with four option graphs from which to choose the one that best described the trend shown in the previous graph, as shown in Figure A1b.

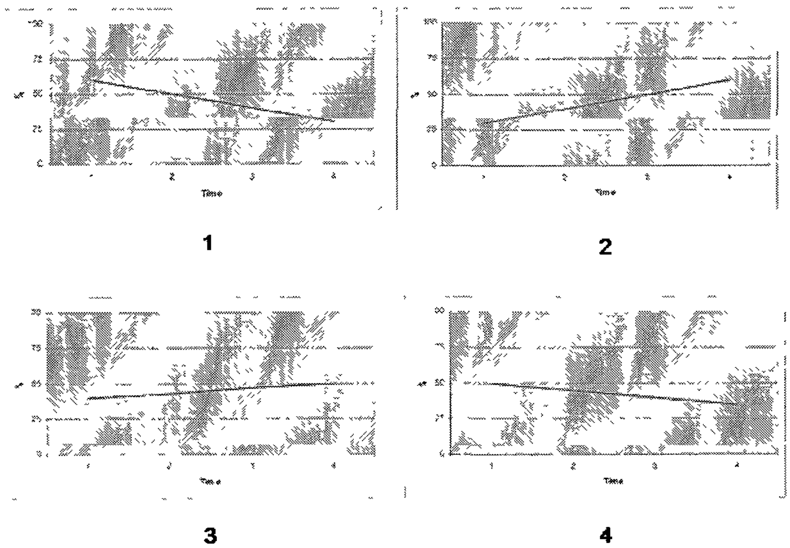


Figure A1b. Example of option graphs for the trend identification task (correct = 1)

In the comparison task, 10 bar graphs were used. Bar graphs were displayed with an 'A' and a 'B' variable, both in black. Participants were asked to choose the option graph that best described the relation between them. An example of a stimulus graph is presented in Figure A2a. The option graphs are shown in Figure A2b. Two of the option graphs showed 'A' greater than 'B' (2 & 3) and the other two showed 'B' greater than 'A' (1 & 4).

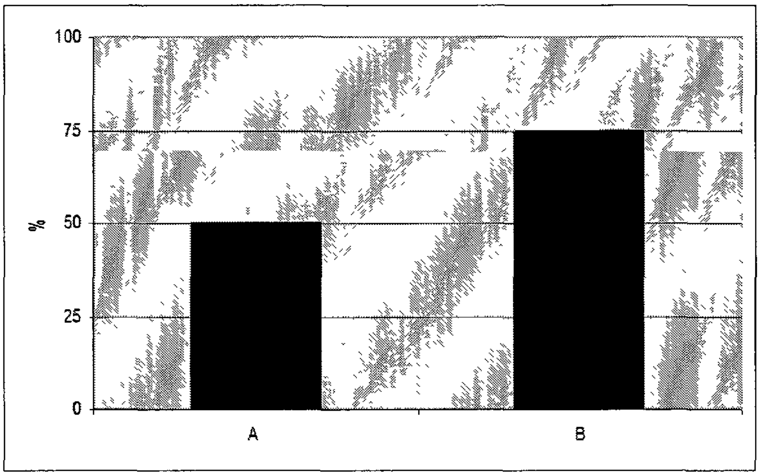


Figure A2a. Example of a stimulus graph for the comparison task

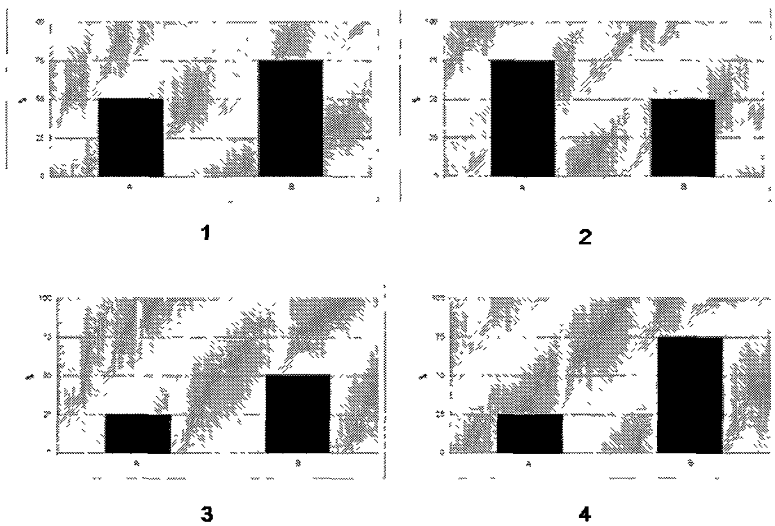


Figure A2b. Example of option graphs for the comparison task (correct = 1)

The read-off task entailed estimating the value of 'A', when the variable 'B' was held constant at a value of 50% on a scale of 0 to 100%. An example of the stimulus graph is shown in Figure A3a, and the option graphs are in Figure A3b.

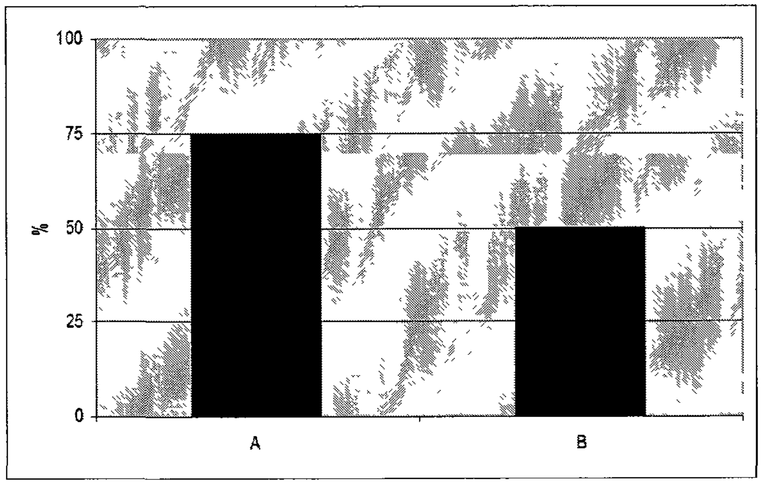


Figure A3a. Example of stimulus graph for the read-off task

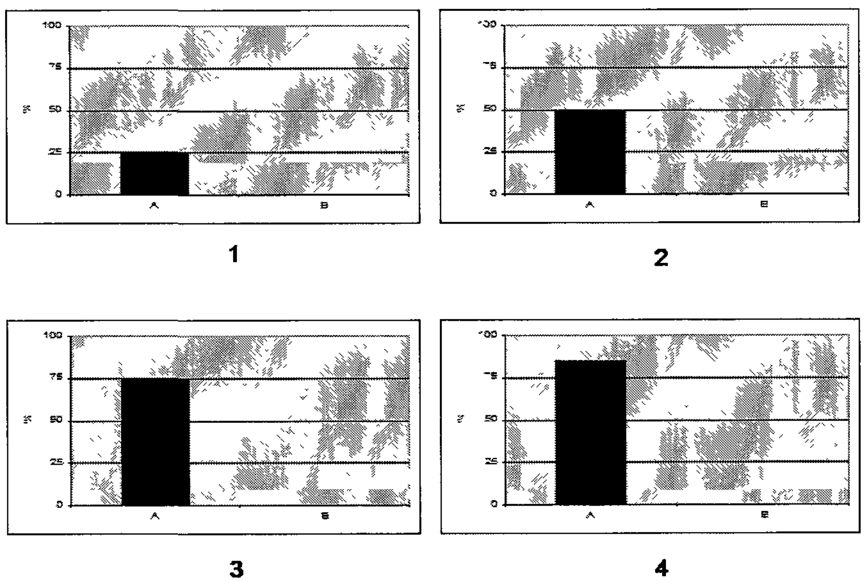


Figure A3b. Example of option graphs for the read-off task (correct = 3)

Procedure

Participants were escorted into a quiet room, seated at a table and asked to complete the informed consent form. If they still wished to take part, they then received the pre-qualifying questions, the mini eye-test as well as the graph interpretation test.

They were then seated in front of a computer displaying the experimental instructions. After they read the instructions, they were given an opportunity to ask questions. Five practice graphs preceded the stimulus graphs for each task.

Participants selected the graph deemed correct by entering the relevant number. The option graphs were exposed for an unlimited time. They were then presented with a text box and told to enter a value ranging from 0 - 100 to indicate their level of certainty that their response was correct. They pressed the space bar to continue to the next graph. The 10 line graphs were presented in a different random order for each participant.

For the comparison task, participants were shown a stimulus graph and asked to choose the option graph that most closely represented the relation of 'A' to 'B'. Exposure time, sequence of events, and number of option graphs were the same as before.

For the read-off task, the procedure was the same, however the option graphs only displayed the 'A' variable, and participants were asked to choose the graph representing the value of 'A' in the stimulus graph.

Upon completing the experiment, participants were thanked for their willingness to take part, paid and debriefed before being excused.

Data analysis

The percentage of correct responses was calculated first for each task and for each participant. The same was done for certainty scores. To the extent that the percentage score for accuracy exceeded the chance level of 25%, it was concluded that the task was doable at a 50 ms exposure time. An error analysis was subsequently undertaken for the graphs in all tasks to determine the adequacy of the stimuli and the response options. It was decided that if more than 50% of the participants erred on any one graph that the stimuli and response options would be scrutinized and modified, if necessary. Additionally, the average response time was calculated for each participant across each task.

Results

The purpose of this study was to determine if participants could accurately complete the three types of tasks when graphs were presented at 50 ms exposure times. The results are presented first by performance, then by response times and finally by the certainty ratings.

Performance

The mean percent correct for the nine participants, displayed by task type (trend identification, comparison and read-off) are shown in Figure A4.

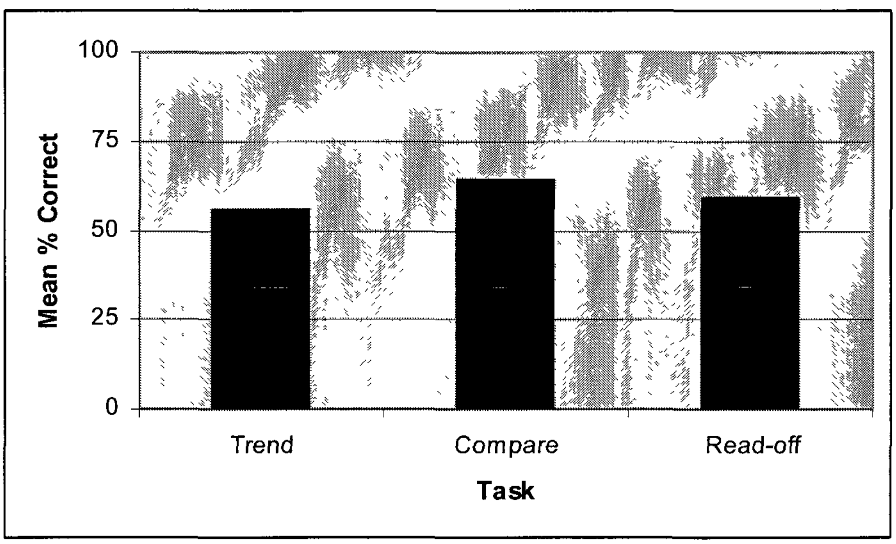


Figure A4. Mean percent correct for each task (trend, compare, read-off)

The results showed that participants were able to identify the correct trend 55.6% of the time, on average (*Mdn* = 60, *SD* = 13.3, *Range* = 30% - 70%). For the comparison task, participants chose the correct response 64% of the time, on average (*Mdn* = 60, *SD* = 10, *Range* = 60% - 80%). Finally, the number of correct responses in the read-off task was 59% on average (*Mdn* = 60, *SD* = 19, *Range* = 30% - 80%). This task was more difficult for some participants than for others, judging from the variability in the scores. Participants number 2 and number 8 performed just better than chance level, whilst all others performed above the 25% level. Most participants were thus able to complete the tasks successfully at better than chance level (25%). Therefore, it is reasonable to conclude that they could successfully complete the tasks when graphs were presented at 50 ms.

Response Times

The results showed that response times varied by task and that, on average, responses took longer on the trend identification task than on the others, as shown in Figure A5. Response times on the comparison and read-off tasks were quite short with the exception of participant number 4, who took considerably longer on average to respond on those two tasks. This could be due to the participant being overly cautious about every response and for that reason the data for that participant were retained. However, no such pattern was evident for any other participant.

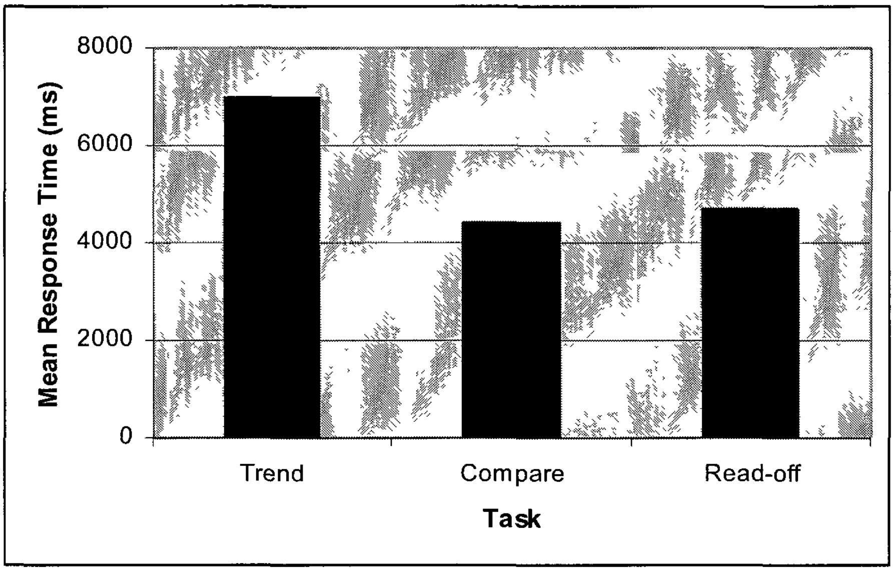


Figure A5. Mean response times for each task (trend, compare, read-off)

Mean response times calculated, for the trend identification task showed that it took approximately 7 seconds, on average, for participants to respond (*Mdn* = 7.8 s, *SD* = 2.5, *Range* = 3.9 s – 11.7 s). Participants took approximately 4.4 seconds, on average, to respond in the comparison task (*Mdn* = 3.9 s, *SD* = 2.2, *Range* = 2.5 s – 10 s). It took approximately 4.7 seconds, on average, for participants to respond on the read-off task (*Mdn* = 3.67 s, *SD* = 4.3, *Range* = 1.7 s – 16.1 s). These results showed that participants responded very quickly. However, the range in scores can be accounted for by participant 4, who took much longer than others and two participants who responded much faster than others.

Certainty

The results showed that the mean certainty was quite high across participants for each task. These results are shown in Figure A6.

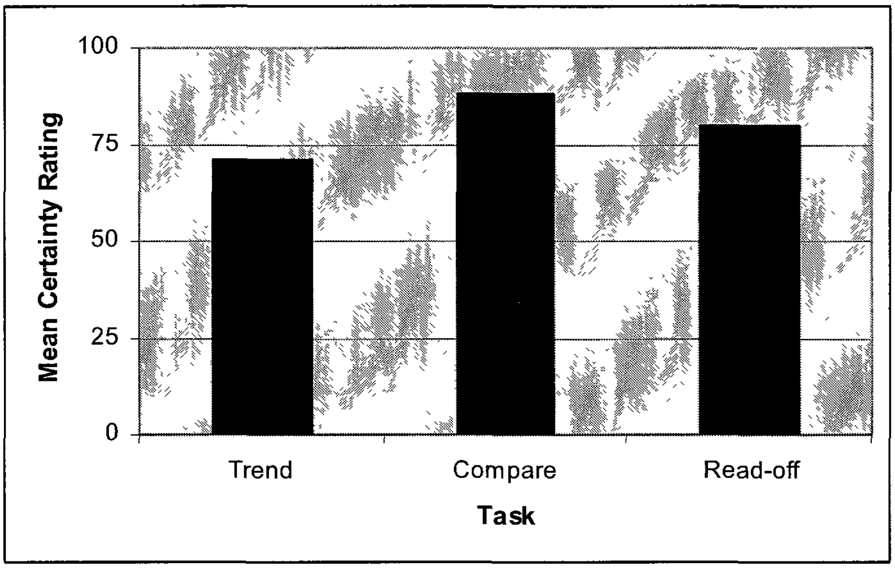


Figure A6. Mean certainty rating for each task (trend, compare, read)

Mean certainty scores of 71.3% were obtained for the trend identification task (*Mdn* = 72.5, *SD* = 13.58, *Range* = 52% - 90%). As the Figure shows, most participants' were relatively certain of their responses. Mean certainty for the comparison task was 88% (*Mdn* = 90, *SD* = 8.2, *Range* = 75% - 99%). Finally, for the read-off task, on average, participants were 80% certain of their responses (*Mdn* = 83, *SD* = 15.6, *Range* = 43% - 100%). Overall, participants were quite certain about their performance indicating that they found the tasks easy, even at the 50 ms exposure time.

Discussion

The results show that participants could complete the three types of tasks when stimulus graphs were presented at 50 ms, as all the mean values showed that the percent of correct responses exceeded the chance level of 25%. The lowest score on any task was 30%. It occurred twice on the read-off task and once on the trend identification task.

The standard deviations show that participants' scores varied similarly suggesting that participants were reasonably consistent in their performance. It is worth noting that participants' performance was better on the comparison and read-off tasks than it was on the trend identification task.

Modifications to Preliminary Study

Based on the results of this experiment some 9 methodological changes were implemented to the tasks, the design and the stimuli and response options.

Tasks. First, all bar graphs were amended to include a third variable ('A', 'B' 'C') to ensure participants did not habituate to always looking in the same location. Second, for both the comparison and read-off tasks, participants were prompted to attend to one of the three variables prior to viewing each graph. The amendment to the comparison task prompted participants to "Compare 'A' to 'C' for example, rather than always comparing 'A' to 'B'. The read-off task also had only two variables, and participants were instructed always to look for the value of 'A'. This task was amended so that participants were prompted for the value of 'A', 'B', or 'C', at random, with each equally represented.

Design. Third, to ensure that the graph stimuli could not be processed beyond the 50 ms exposure time, a visual mask presented for 50 ms, with an SOA of 100 ms followed each stimulus graph in all subsequent experiments. Fourth, a textual mode was added as another response option. Fifth, a 300 ms exposure time condition was added, to compare with responses at the 50 ms exposure time. Sixth, two visual tests were added as a way to measure individual differences. The Group Embedded Figures Test (Witkin, Oltman, Raskin, & Karp, 1971) was employed to measure how well people can separate figure from background. The Visual Patterns Test (Della Salla, Gray, Baddley & Wilson,

1997) was administered to measure people's ability to remember patterns. Seventh, the instructions were amended for all the tasks to account for the changes noted above. The modifications to the tasks and the design are shown in Table A1.

Table A1

Modifications to Experiment 1

Modifications	Preliminary Experiment	Experiment 1
Tasks		
Read-off	Value of A	Value of A, B, or C
Comparison	Compare A to B	Compare A to B, A to C or B to C
Design		
Mask	No mask	Mask (50 ms) with SOA of 100 ms
Response Options	Graphical	Graphical or Textual
Exposure Time	50 ms	50 ms and 300 ms
Tests for Individual Differences	None	Visual Pattern Test Group Embedded Figures Test
Instructions	For 50 ms only	For 50 ms and 300 ms + changes

Stimuli and response options. Eighth, the graphs for the comparison and read-off tasks were amended. An error analysis was undertaken for the graphs in the trend identification task found that three graphs were problematic as more than 50% of participants erred on them. Finally, careful inspection of the graphs revealed that two of the response options were quite similar, which appeared to cause interpretation problems for the participants. Thus, the stimulus graphs were retained and the relevant response option graphs were amended to reduce those problems.

Appendix AA

This appendix contains materials for the preliminary study. It includes: the pre-qualifying questionnaire, graph comprehension test and that task instructions.

Pre-qualifying Questionnaire
(Same for all studies)

This information is completely confidential
(Please, do not put your name on this form!)

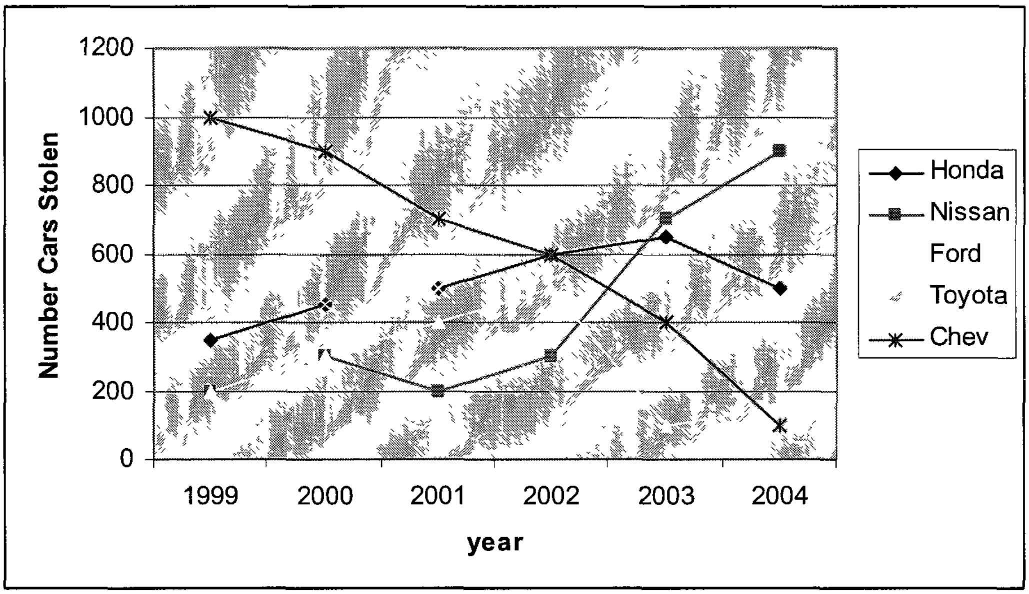
Age: (check one)	☐ 19-30	☐ 31-40	☐ 41-50
Sex: (circle one)		Male	Female
Do you suffer from colour blindness?		Yes	No
Do you typically encounter graphs in your everyday life?		Yes	No

If so, for what do you use them?

____
____
____
____
____

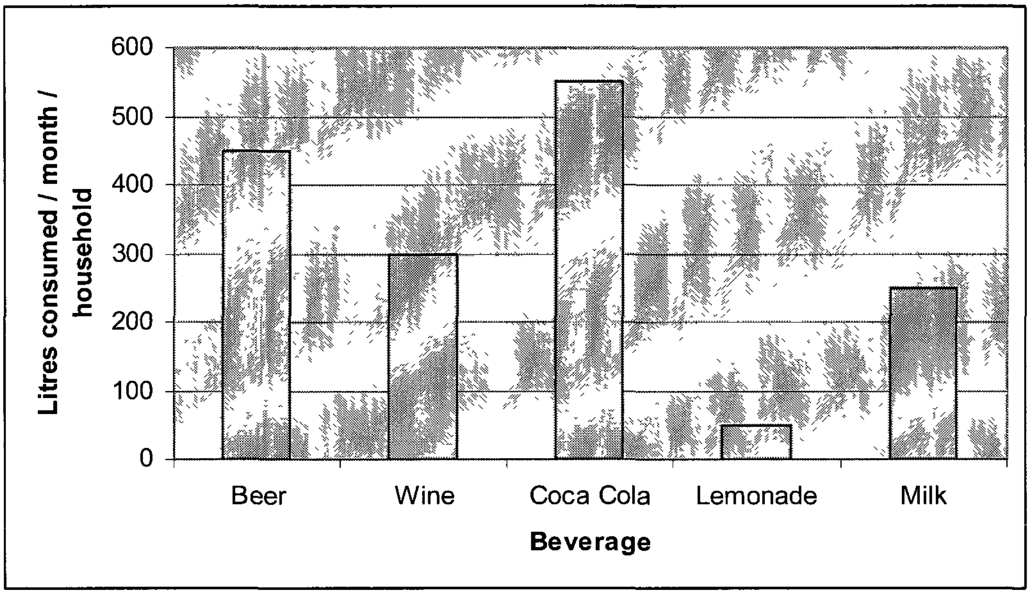
Graph Comprehension Test
(Same for all studies)

Please look at the following graph and answer the following questions:



What is the most stolen car in 2004? ____________________
What make of car has dropped most in popularity with thieves? ____________
In 2001, what was the most stolen car? ______________________________
If you were afraid to have your car stolen, which make would you most avoid today?
____

Please look at the following graph and answer the following questions:



How many more beer are consumed per month than milk? _______________

What is the most popular beverage consumed per household? _______________

How much wine is consumed every month per household? _______________

What is the second most popular beverage? _______________

Instructions
(Presented on screen)

[Always presented first]

In the first section, you will be shown a series of graphs for very brief periods of time. You will be asked to answer three different kinds of questions regarding the variables, called simply 'A' and 'B'. Instructions will be given throughout with more detail as you come to each task.

When you are ready to do the first task, press the space bar.

[Participant receives one of the six block orders]

[Instructions presented when they receive the 'predict' block]

Predict Instructions [amended to trend identification]

For this task you will see a line graph very briefly. Your job is to predict what the variable will do next. You will then be presented with four option graphs. To answer the question you will choose which of the four option graphs you think best represents what 'A' will do, based on the line graph that was just presented to you. You will then be asked how certain you are of your choice. You will enter your response on a scale from 0 to 100. A rating of 0 means you are just guessing and a rating of 100 means you are certain that you are correct. When you are done, press the continue button to move on to the next line graph. You will do this for each graph until you have worked through all the graphs.

Do you have any questions?

When you are ready to begin, press the space bar.

[participant receives 5 practice graphs]

When you are ready to continue to the test graphs, press the space bar

[participant receives 10 test graphs]

You have finished this task. When you are ready to move onto the next task, press the space bar.

[Instructions presented when they receive the 'compare' block]

Compare Instructions

For this task, you will see a bar graph very briefly. You will be presented with four option graphs. Your job is to compare two variables, called 'A' and 'B'. To answer the questions you will choose which of the four bar graphs best represents the relation between 'A' and 'B' that you were previously shown. You will then be asked how certain you are of your choice. When you are done, you will press the continue button to move on to the next bar graph until you have worked through all the graphs. Do you have any questions?

Do you have any questions?

When you are ready to begin, press the space bar.

[participant receives 5 practice graphs]

When you are ready to continue to the test graphs, press the space bar

[participant receives 10 test graphs]

You have finished this task. When you are ready to move onto the next task, press the space bar.

[Instructions presented when they receive the 'read off' block]

Read Off Instructions

For this task, you will see a bar graph very briefly. You will be asked to estimate the value of variable 'A'. The variables are labeled 'A' and 'B'. In this task, variable 'B' will always represent a value of 50% but 'A' will vary from graph to graph. In the four option graphs, only variable 'A' will be shown. After you make your choice, you will then be asked how certain you are of your choice. When you are done, you will press the continue button to move on to the next bar graph until you have worked through all the graphs. Do you have any questions?

Do you have any questions?

When you are ready to begin, press the space bar.

[participant receives 5 practice graphs]

When you are ready to continue to the test graphs, press the space bar

[participant receives 10 test graphs]

You have finished this task. When you are ready to move onto the next task, press the space bar.

or

Thank you for participating in this study! Your willingness to help us is greatly appreciated. Please exit the room and let the experimenter know you have finished.

Appendix B

Preliminary Study 2

This Experiment had three objectives, all are related to task complexity and with the aim of isolating perceptual complexity. The first was to test the degree to which participants' performance may differ as a function of the type of graph reading task. This was done to determine if tasks differ in complexity, and which could be done most successfully at very brief exposure times. The second objective was to test the effects of different exposure times on graph comprehension. This was done to determine if performance would improve with longer exposure times, which could also account for differences in task complexity. The third objective was to discern the degree to which different response modes may affect performance. This objective was met by manipulating response modes to understand if they contributed systematically to task complexity. The experimental manipulations were aimed at satisfying the three stated objectives by identifying variables that could lead to increased task complexity and hence improve understanding of their effects on graph perception. Armed with such understanding, the effects of these variables could be controlled in subsequent experiments, all of which investigated perceptual complexity.

As discussed earlier (p. 22), graph comprehension tasks can be divided into three types: trend identification tasks, comparison tasks, and read-off tasks. It has been found to be easier to extract global level information than to extract precise information, which takes more effort and time (e.g., Lohse, 1989; Navon, 1977; Ratwani, et al., 2003). In Lohse's studies, participants had unlimited time to answer questions using tables, bar-and line graphs. He found that extracting precise information in comparison tasks took longer than general trend information, but both took more time than the read-off tasks. These results showed that with unlimited inspection time, the read-off task was the easiest and completed fastest, however, when limiting exposure time, Navon (1977) found that precise information was harder to detect. In those studies, participants could recognize a large 'H', 'S' or rectangle but did not notice the small letters or rectangles that defined each, without also attending to the global feature. Trends are believed to be identifiable at a glance because the message the viewer is looking for stands out, unlike other tasks, which require more processing and hence more time to complete. Thus, the trend identification task should be perceived as the simplest of the three types. It should lead to better performance, shorter response times and higher certainty than the other two tasks. To test this, Hypothesis 1a predicted that trend identification task performance would be higher and Hypothesis 1b that response times would be shorter than both the comparison and the read-off tasks. Hypothesis 1c predicted that the level of certainty for the trend identification task would be higher than for both the comparison- and the read-off tasks. These and all following hypotheses are summarized in Table B1 (p. 170) together with the tests employed to test their effects.

According to Pinker (1990) we all have a 'graph schema' or pictorial representation against which incoming graph stimuli are compared. If, during graph perception, incoming stimuli (bottom-up) are immediately compared to existing representations (top-down), then perceptual processing should lead to some level of comprehension, even when graph stimuli are shown very briefly. To test this, graph stimuli were presented at two different stimulus exposure times. According to Pinker

(1990), graph comprehension is a cyclical process where information from the graph is compared to existing graph schema and then to the conceptual question. This iterative cycle continues until the graphical information is comprehended (discussed earlier on p. 18). Thus, it is reasonable to assume that the task would be perceived as more complex when less time is allowed for iteration to occur as in the 50 ms condition. Hypothesis 2a predicted that performance would be better and Hypothesis 2b predicted that response times would be shorter at 300 ms than at 50 ms stimulus exposure times. Hypothesis 2c predicted that reported certainty would be higher at 300 ms than at 50 ms.

Finally, it is possible that the type of response options could affect performance Participants could be relying solely on memory of the stimulus to select the correct response if the response options correspond to the stimuli presented (i.e., graphical stimulus and graphical response option). However, when the response mode differs from the stimulus mode participants cannot rely exclusively on memory of the stimulus (i.e., graphical stimulus and textual response option). Thus, it was desirable to discern how those two response modes would affect performance. To test that, a graphical and a textual response mode was employed. Accordingly Hypothesis 3a predicted that graphical response options would lead to better performance and Hypothesis 3b that response times would be shorter for graphical than for textual response options, regardless of stimulus exposure time and of the graph comprehension task. Finally, Hypothesis 3c predicted that certainty would be higher for graphical than textual response options.

Table B1

Hypothesis number, predictions and test used

H #	Prediction	Test Used
1a	Performance on trend identification task will be better than both the comparison and read-off tasks.	Main effects analysis & Simple main effects analysis
1b	Response times will be shorter for the trend identification task than both the comparison and read-off tasks.	Main effects analysis & Simple main effects analysis
1c	Certainty ratings for the trend identification task will be higher than both the comparison and read-off tasks.	Main effects analysis & Simple main effects analysis
2a	Performance at 300ms exposure time will be better than performance at 50ms exposure time.	Main effects analysis
2b	Response times will be shorter at 300 ms exposure times than at 50 ms exposure time.	Main effects analysis
2c	Certainty ratings will be higher for 300ms exposure time than at 50ms exposure time.	Main effects analysis
3a	Performance will be better for graphical response options than for textual response options.	Main effects analysis
3b	Response times will be shorter for graphical response options than textual response options.	Main effects analysis
3c	Certainty will be higher for graphical response options than for textual response options.	Main effects analysis

To test short-term visual memory, the Visual Patterns Test (hereafter VPT) (Della Salla, et al., 1997) was used. To test participants' ability to separate figure from background, or field dependence-independence ability, the Group Embedded Figures Test (hereafter EFT) (Witkin, et al., 1971) was administered. Both tests were administered to help identify individual differences in skill levels.

Method

Participants

A total of 33 participants were recruited from Carleton University and through personal contacts. The first 6 participants' data were excluded because their data files were incomplete due to a programming error. Three were recruited through the SONA system and the other three were paid for their participation. A sample of 27 new participants were recruited without the use of the SONA system. The last three participants' data were not necessary to fill the experimental conditions, so their data were excluded. The data for the first 24 participants, comprising 13 women and 11 men (mean age 23.13 years and a range of 18 to 45 years) were included for analysis. All responded to advertisements posted around campus and were each paid $15.00 for participation. All were tested for visual acuity using a mini eye chart and were pre-screened for colorblindness.

Apparatus

Each participant was tested on a Toshiba Satellite laptop with Genuine Intel® Centrino Duo CPU. Screen resolution was set to 1280 by 800 pixels with highest 32bit colour. A program created in DirectRT™ was used to present graphs, collect responses, ratings and response times. A stopwatch was also used to time the completion of the Group Embedded Figures Test.

Materials

The materials were the same as before with the addition of new consent- and debriefing forms, amended instructions, a Group Embedded Figures Test Booklet, and

Visual Pattern Test Booklet, pens, pencils and erasers (See Appendix BB for materials, where permitted by copyright law). For each task, the stimulus materials consisted of five practice graphs, and 15 test graphs. Bar graphs were used in the comparison- and read-off tasks, each showing three variables. The trend identification task employed line graphs, showing "small" or "sharp", "increasing" or "decreasing" slopes. Small changes varied less than 50% between the beginning - and end points and sharp changes varied more than 60% at the beginning- and end points.

Design

The experiment employed a mixed design. The stimuli were presented for 50ms in the first, and for 300 ms in the second condition. In both conditions, a backward masking technique was employed to cancel further processing of the stimulus information. The mask was presented for 50 ms and 300 ms, respectively, with a 100 ms delay (stimulus onset asynchrony, SOA) (Enns & Dillolo, 2000). The design is shown in Table B2.

Table B2

The Experimental Design

Test	Ps	50 ms Condition			Test	300 ms Condition		
		Trend n=15	Comp n=15	Read n=15		Trend n=15	Comp n=15	Read n=15
				Graphical Options				
VPT/ EFT	P1 … P6				EFT/ VPT			
EFT/ VPT	P7 … P12				VPT/ EFT			
				Textual Options				
VPT/ EFT	P13 … P18				EFT/ VPT			
EFT/ VPT	P19 … P24				VPT/ EFT			

The 50 ms condition was always presented first. The three task types were counterbalanced (six orders) within each response mode condition to avoid serial order effects, but they were presented in the same order in both exposure time conditions. One half of the participants completed the VPT at the start of the experiment and the EFT just prior to the 300 ms exposure time condition. This was reversed for the other half. Each task comprised 5 practice graphs, followed by 15 test graphs. Test graphs were presented in random order within each task type. The first graph stimulus was presented for 50 ms, with an SOA of 100 ms, followed by the mask for 50 ms, a blank screen for 1000 ms, and then the four response options. Participants were instructed to select the graph deemed correct by typing the relevant number into the computer. Response options were shown for an unlimited time. Twelve participants were given graphical response options as in Preliminary Study 1, and another 12 were given textual response options, as shown in Figure B1.

For the trend identification task, all graphs displayed a single line, 14 with one or two direction changes resulting in a crooked line, and one had a perfectly straight decreasing trend line. Thus, seven line graphs were mirror opposites and one was added to keep the number of graphs consistent. Lines were presented in black of standard thickness. Participants were asked to choose the response option that best indicated the trend in the graph just seen. The procedure was the same as in Preliminary Study 1 (p. 145) with the addition of five graphs (two with mirror opposites and one straight line).

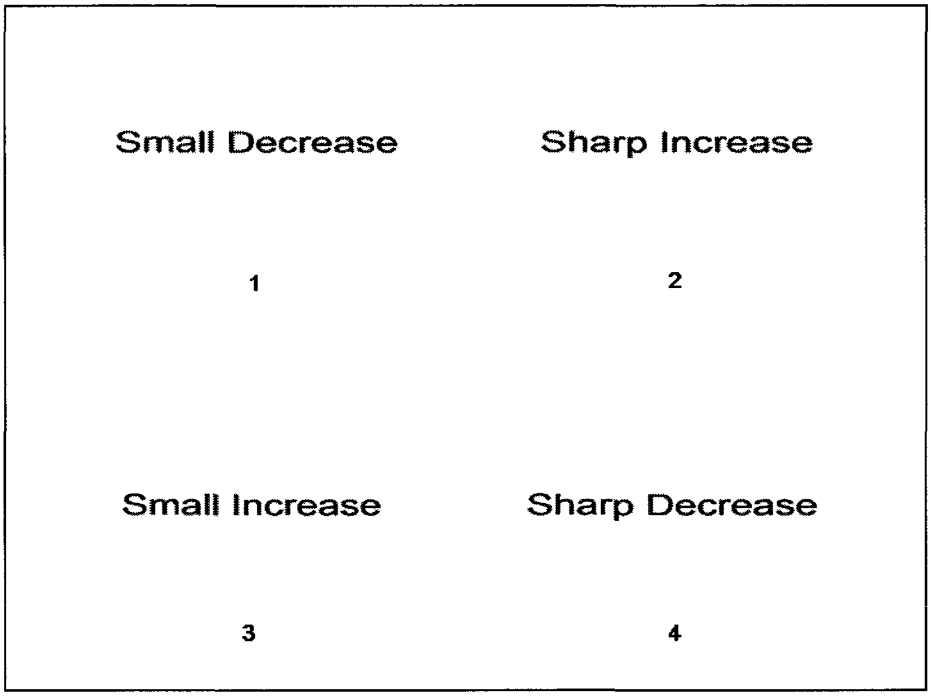


Figure B1. Example of textual response options for trend identification task

In the comparison task, 15 bar graphs were shown following the same procedure as the trend identification task. Bar graphs were displayed with an ‘A’, ‘B’ and ‘C’ variable, all in black. Participants were asked to “compare A and C” (or A & B, or B & C), for example, and to choose the option graph or textual description that best described the magnitude of the difference between the two target variables. An example of a comparison task stimulus graph is shown in Figure B2, with both graphical and textual response options shown in Figures B3a and B3b (pp. 176 & 177). One response option graph showed the correct relation (#1), one showed the variables in the opposite relation (#3), and two showed the specified variables varying randomly.

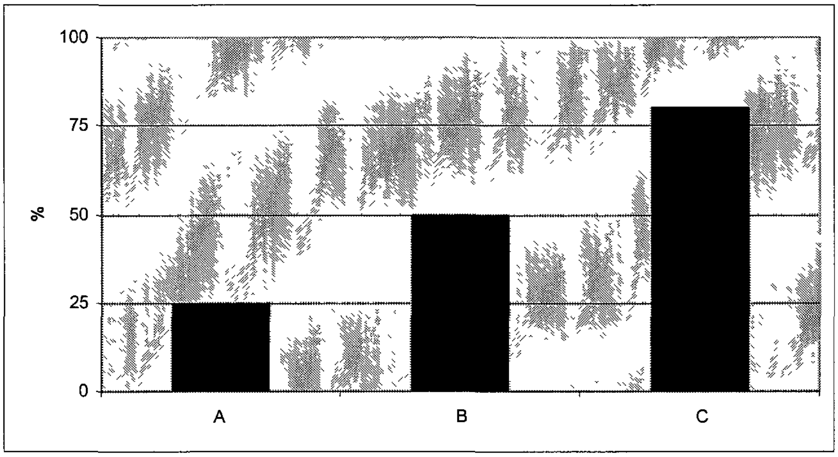


Figure B2. Example of comparison stimulus graph (Compare A to C)

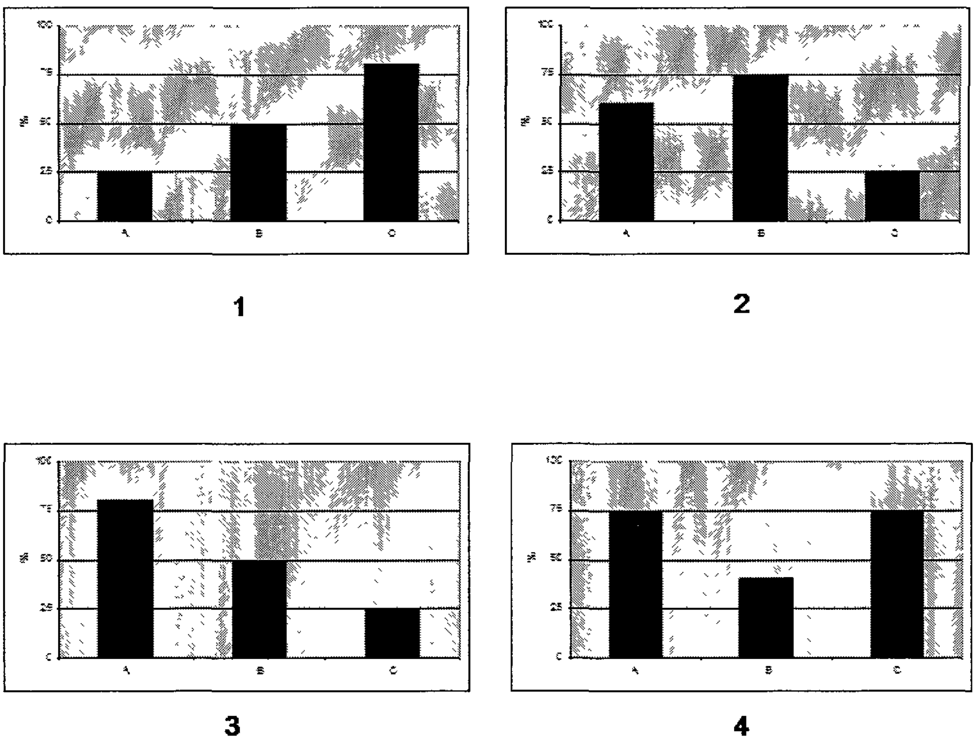


Figure B3a. Example of graphical response options for the comparison task (Correct = 1)

$A < C$	$C = A$
1	2
$A > C$	$C > A$
3	4

Figure B3b. Example of textual response options for the comparison task (Correct = 1)

The read-off task entailed estimating the value of a particular variable 'A', 'B' or 'C' in the initial screen. Participants were asked to focus on the value of one of the variables, "value of A", or B or C, and when ready, to press the space bar to see the stimulus graph. The stimulus graphs were the same as in the comparison task. The stimulus graph is shown in Figure B4 in which the value of B is 75%, and the graphical and textual options are shown in Figures B5a and B5b (pp. 178 & 179) respectively.

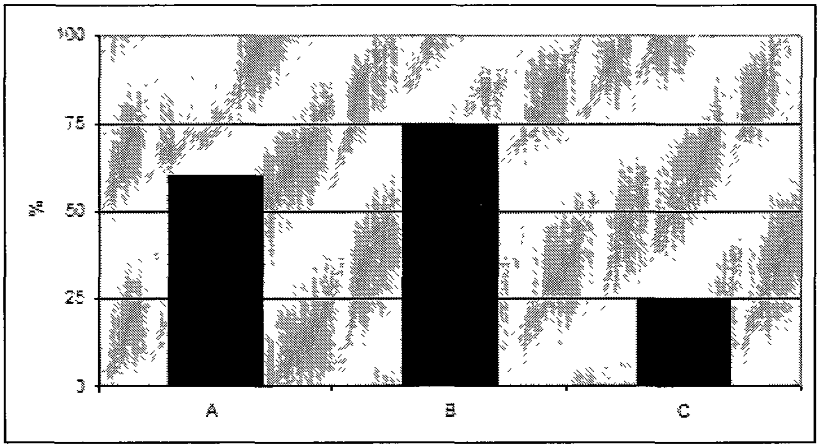


Figure B4. Example of read-off stimulus graph (Value of B)

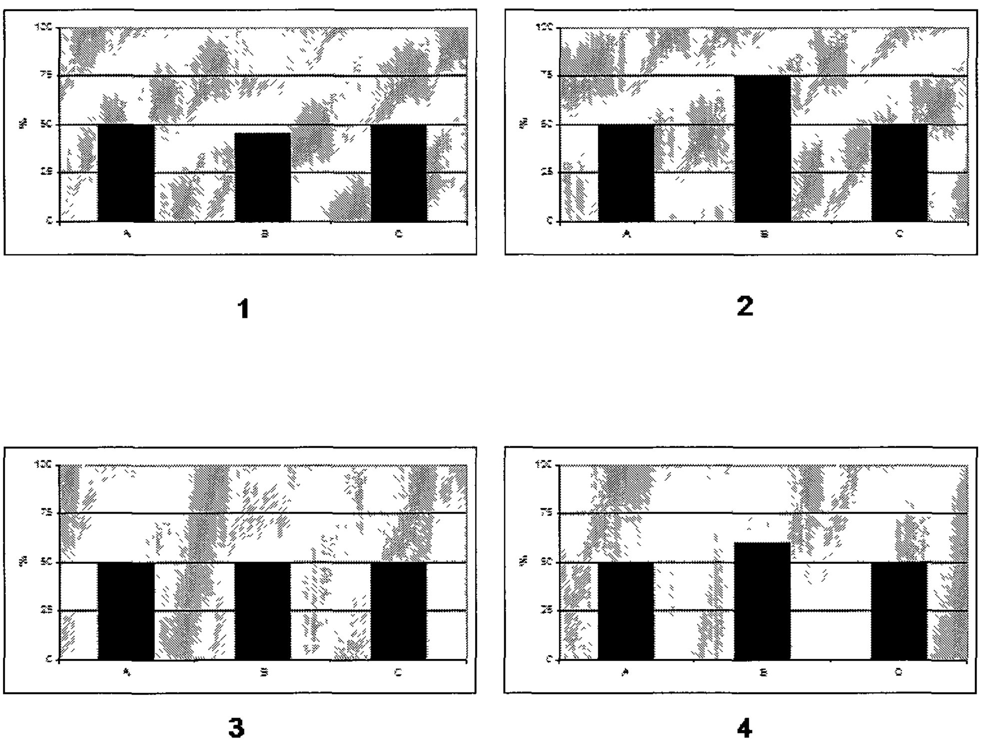


Figure B5a. Example of graphical response options for read-off task (Correct =2)

B = 45%	B = 75%
1	2
B = 50%	B = 60%
3	4

Figure B5b. Example of textual response options for read-off task (Correct = 2)

Procedure

Participants were seated and presented with an informed consent form. After reading and signing it, they completed the demographic information sheet and the graph comprehension test, followed by the mini eye test. Next, they were given the Embedded Figures Test or the Visual Pattern Test, depending on the order to which they were assigned. After completing that test, they were instructed to move to the laptop to begin the graph perception study. Upon seeing the first practice stimulus graph for 50 ms (100 ms SOA and mask for 50 ms), the four textual or graphical response options were presented, followed by a text box requesting them to rate their certainty about the correctness of their answer, on a scale from 0 – 100 (0 = not at all certain and 100 = completely certain). When finished, they pressed the 'enter' key followed by the space bar to see the next stimulus. After completing the five practice graphs, participants were asked if they had any questions. They then continued to the test graphs. In each task condition, the 15 graphs were presented in random order. The researcher sat beside the participant throughout. At the end of the trend identification task, participants were asked how they had estimated the trend. Their responses were recorded for later analysis.

For the comparison task, exposure time, sequence of events, and number of option graphs were the same as in the trend identification task, except that participants were told which variables to compare prior to seeing the stimulus graph. For the value identification task, participants were also prompted to look for the value of one specific variable and to choose the response option graph representing that value.

Upon completing the 50 ms condition, they were offered a short break. After the break, participants completed the remaining test, either EFT or VPT. Then they completed the 300 ms condition with tasks presented in exactly the same order as in the 50 ms condition. The stimulus graphs were randomized again within each task. When finished, the participant was thanked for his/her willingness to take part, debriefed and paid and dismissed. The experiment lasted approximately 1.5 hours.

Data analysis

As in the preliminary experiment, the percentage of correct responses was calculated first for each task and for each participant. The same was done for certainty scores. An error analysis was undertaken, as before, to identify possible problematic stimuli. Additionally, the average response time was calculated for each participant in each task. Because the dependent measures were the same for all experiments, the data were treated in the same fashion throughout.

Results

The results are divided into six sub-sections. The first presents an error analysis to identify and possibly remove potentially problematic stimuli. The second section presents the results of data exploration that ensured that assumptions of ANOVA were met and any outlying scores were identified and dealt with. The third section presents performance data on all three tasks types, exploring the effects of stimulus exposure times, tasks, and response modes. Mean response time data are presented for stimulus exposure times, tasks and response modes in the fourth sub-section. In the fifth sub-section, an analysis of participants' mean certainty ratings is presented. The final section briefly addresses the relationship between the two tests (VPT and EFT) and the mean performance, certainty ratings and response times. An alpha level of .05 was used for all statistical tests in these thesis experiments. All descriptive statistics as well as tests for violations of assumptions are included in Appendix BB.

Error Analysis

Participant responses were recorded for each graph in each of the three task types, for both exposure time conditions and in both response modes. The percentage of incorrect responses was first calculated for each condition and every task. As performance was predicted to be higher than chance levels, even in the 50 ms exposure condition, the mean percentage of incorrect responses on each task should be less than 75%. The mean percent of incorrect responses for both response modes, three tasks and both exposure time conditions is shown in Figure B6.

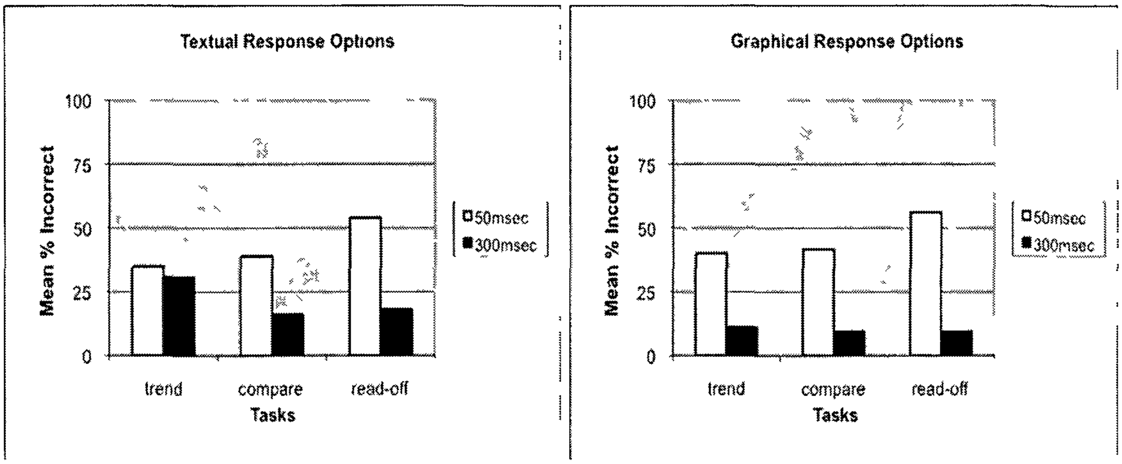


Figure B6: Mean incorrect responses in percent for response modes, tasks, and exposure time condition

As the Figure shows, the mean percentage of incorrect responses was higher in the 50 ms than the 300 ms conditions in both response modes. The read-off task led to the lowest performance in the 50 ms condition in both response modes. The number of errors declined quite dramatically across all tasks from the 50- to 300 ms conditions with both response-options. However, in the textual response condition, the decline was minimal in the trend identification task. The trend identification and comparison tasks were completed approximately equally well in the 50 ms condition regardless of response option, but the results look different in the 300 ms conditions because the error rate for the comparison tasks was considerably lower in the 300 ms condition compared to the trend tasks. The assumption that performance should be higher in the 300 ms exposure time condition clearly held for the comparison and the read-off tasks in both the graphical and textual response modes, but not for the trend identification task in the textual response mode. In order to identify any outliers, a detailed analysis was conducted on each of the 15 stimulus graphs in all three task types.

The analysis showed that error rates exceeded the 50% criterion in both exposure time conditions for four graphs in the trend identification task only. Therefore, these four graphs were removed prior to analyzing participants' performance, response times and certainty ratings. In all four graphs, the difference between the first and last value was smaller than 50%, which was defined earlier as a 'small' slope. However, more than half of the participants chose to call that a 'sharp' slope in the textual response-mode suggesting a possible misalignment of the researcher's and participants' mental models. As these exceeded the 50% error rate criterion, it justified excluding these four graphs from further analyses. Certainty ratings and response times associated with correct and incorrect responses were further examined prior to removing those four graph stimuli for the trend identification task only, to determine if a misalignment of mental models occurred. The data for two participants were missing from these two analyses because they did not have incorrect responses for some graphs in the textual condition.

To determine if participants completing the trend identification tasks realized when they had selected an incorrect response, a 2 x (2 x 2) ANOVA was conducted on mean certainty ratings for response mode (graphical, textual), accuracy (correct, incorrect responses) and exposure time (50 ms, 300 ms). It revealed an interaction of response mode and accuracy, $F(1, 20) = 14.88, p = .001, \eta_p^2 = .43$. Analyses of simple main effects showed that certainty was significantly higher for correct than incorrect responses in the textual response mode, $F(1, 20) = 17.67, p = .000, \eta_p^2 = .47$, but not in the graphical response mode, $F(1, 20) = 1.25, p = .277, \eta_p^2 = .06$. Apparently, participants did not always realize when they had made an incorrect response in the graphical response mode, suggesting that the textual mode was easier to understand than the graphical mode. This finding supports the assertion that there was a misalignment between the researcher's and the participants' definition for a "small" slope in the textual response mode, because participants were likely aware that they did not know how to define the slope for some of the graph stimuli. In the graphical response mode, however, it is possible that because participants had only to choose a graph with a similar slope, they were unaware that they were incorrect because they didn't have to indicate if the slope was 'small' or 'sharp'.

Participants were more certain in the 300 ms than the 50 ms condition, as evidenced by a significant main effect for exposure time, $F(1, 20) = 10.07, p = .005, \eta_p^2 = .34$, and they were more certain when their responses were correct than when they were incorrect as was shown by a significant main effect for accuracy, $F(1, 20) = 5.53, p = .029, \eta_p^2 = .22$. Additionally, certainty was highest in the graphical response mode compared to the textual, as demonstrated by a significant main effect of mode, $F(1, 20) = 8.01, p = .01, \eta_p^2 = .29$. No other effects were significant.

Response times were explored in the same fashion by comparing the mean response times for correct and incorrect responses. The result of the 2 x (2 x 2) ANOVA conducted for response mode (graphical, textual), accuracy (correct, incorrect responses) and exposure time (50 ms, 300 ms) showed that correct responses were offered faster than incorrect ones as evidenced by a main effect of accuracy, $F(1, 20) = 16.14, p = .001, \eta_p^2 = .45$. No other effects were significant. Thus, both certainty ratings and response times aligned reasonably well with performance. Participants were generally less certain and took longer to respond when they had selected an incorrect than a correct response and less certain in the textual than in the graphical response mode.

Data Exploration

An analysis was conducted for performance scores across response modes to determine if the assumptions of ANOVA were met. Outliers were defined in two ways: first those scores lying three standard deviations from the mean were considered 'extreme', and second, those scores that might lie closer to the mean but may pull the distribution in one way or another resulting in violations of normality. Thus, it was possible to have an extreme outlier that did not influence the distribution and a not quite so extreme outlier that did have an influence. Thus they could be outlying and influential or just outlying. If they were outlying and influential, the scores were adjusted. If they were just outlying and did not affect the distribution, they were left alone and included in the analyses. (For a complete description of how outlying scores were dealt with, refer to Experiment 1, results section on data exploration).

For the current experiment, the test of normality showed that the performance data were normally distributed across tasks in the 50 ms condition, but not in the 300 ms condition. These violations seemed acceptable given that different participants were identified as outliers on different tasks, suggesting that individual differences were responsible for the violation. The data were not adjusted for that reason. Mauchley's test of Sphericity was performed and showed no violation for task or time by task.

To ensure that the assumptions of the ANOVA were met, an analysis of response time data was conducted across response modes. Tests of normality showed that the assumption was violated in some conditions and thus outlying cases were explored. One outlying participant was identified in the 50 ms trend condition, one in the 50 ms comparison condition, one in the 300 ms comparison condition; three in the 50 ms read-off condition and one in the 300 ms comparison condition. Thus, the eight outlying response times were moved within one standard deviation of the mean. Tests of normality were re-calculated and after adjusting the 8 response times, all were normally distributed except the read-off task in the 50 ms condition. It was left as is because the tests showed only a slight violation. Mauchley's test of Sphericity was performed and showed no violation for task.

An analysis on certainty ratings was conducted for tasks and time conditions across response modes to ensure that the ANOVA assumptions were met. Tests of normality showed that the data were normally distributed across all tasks times and response modes, except in the 300 ms read-off condition, which showed a violation of normality. No outliers were identified and normality was not violated in any other condition, thus the data were retained without adjustment. Mauchley's test of Sphericity was performed and showed a violation for task, $W = 0.50$, $\chi^2\,(2) = 14.75$, $p = .001$, and time by task, $W = 0.45$, $\chi^2\,(2) = 16.76$, $p = .000$. Thus, a Huynh-Feldt Epsilon correction was considered to interpret the results of the three-way interaction; the corrected degrees of freedom had no effect on the F-value for the other ANOVA sources, thus they were not altered.

Since the experimental design employed repeated measures, the assumption of independence of residuals was not considered as the assumption was knowingly violated by choosing this design. This is true for all experiments, thus it will not be discussed again.

Performance

As a result of the error analysis, the four problematic graphs were removed, leaving a sample of 11 stimuli in the trend identification task. Figure B7 shows that performance on all tasks was at well above chance level (25%) and participants performed better in the 300 ms than in the 50 ms condition. There appears to be a difference in performance on tasks in the 50 ms condition but not in the 300 ms condition.

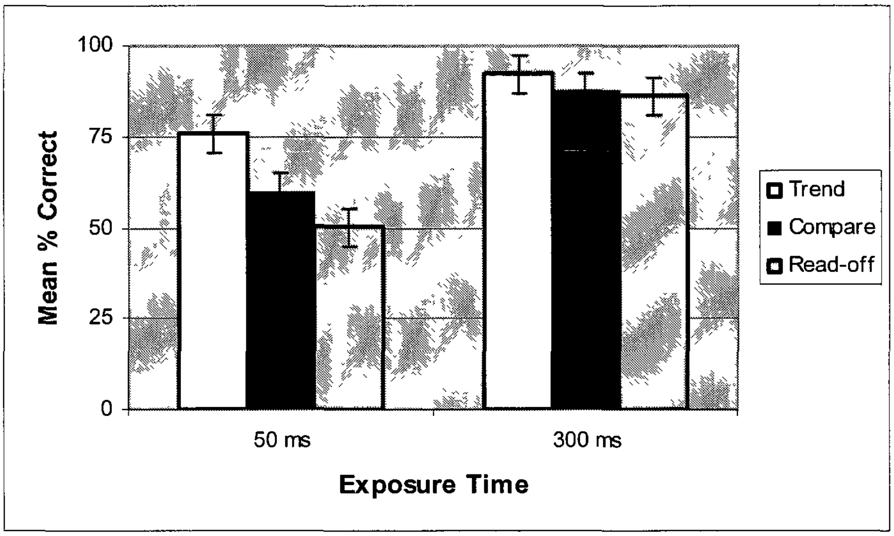


Figure B7. Mean percent correct responses on the three tasks (trend, compare and read-off), at both stimulus exposure times (50ms and 300ms), collapsed across response modes, bars show *CIs* for 2-way interaction = $\pm$ 5.19

To determine the locus of these suspected differences, a 2 x (2 x 3) ANOVA was conducted for response mode (textual, graphical), exposure time (50 ms, 300 ms), and task (trend identification, comparison, read-off) on the mean performance scores.

The three-way interaction of task by exposure time by response mode was not significant, $F(2, 44) = 0.73, p = .49, \eta_p^2 = .03$, and nor were the two-way interactions of response mode by task, $F(2, 44) = 0.37, p = .693, \eta_p^2 = .02$, or response mode by exposure time, $F(1, 22) = 2.05, p = .166, \eta_p^2 = .09$. The two-way interaction of task by exposure time was however significant, $F(2, 44) = 7.53, p = .002, \eta_p^2 = .26$.

Exploring the locus of the differences between tasks at each exposure time condition, analysis of simple main effects revealed a significant difference between tasks in the 50 ms condition, $F(2, 44) = 12.58, p = .000$, but not in the 300ms condition, $F(2, 44) = 1.39, p = .259$. Simple comparisons of tasks within the 50 ms condition showed that performance on the trend identification task was significantly better than on both the read-off task, $F(1, 22) = 27.90, p = .000$, and the comparison task, $F(1, 22) = 8.50, p = .008$. The comparison- and read-off tasks did not differ, $F(1, 22) = 3.53, p = .074$. Applying a Bonferonni correction to control for familywise error yielded an alpha of 0.05/3 = 0.017; both simple comparisons between trend identification task and comparison and read-off tasks remained significant.

To test Hypothesis 1a, which stated that performance would be better on the trend identification task than on the other two tasks, the main effect for task, $F(2, 44) = 9.20, p = .000, \eta_p^2 = .26$ was explored. Pairwise LSD comparisons showed that performance was better on the trend identification task than both the comparison- ($p = .020$) and the read-off task ($p = .000$). Hypothesis 1a was therefore supported. Performance did not differ between the comparison and read-off tasks ($p = .163$).

To test Hypothesis 2a, which predicted that performance would be better in the 300 ms- than the 50 ms exposure time condition, the main effect, $F(1, 22) = 97.99, p = .000, \eta_p^2 = .82$, was explored. Performance was better in the 300 ms exposure time condition ($M = 88.46, SE = 1.89$) than the 50 ms condition ($M = 61.89, SE = 3.61$), thus supporting Hypothesis 2a.

Finally, Hypothesis 3a stated that performance would be higher on the graphical- than the textual response mode. The main effect of response mode was not significant, $F(1, 22) = 1.19$, $p = .286$, $\eta_p^2 = .05$, thus refuting Hypothesis 3a.

Response Times

The mean response time was calculated for each participant for each task, each exposure time condition, and each response mode. Figure B8 suggests that the mean response times varied more for tasks in the graphical than in the textual response mode. They were also lower in the trend identification and read-off tasks in the graphical than in the textual response mode.

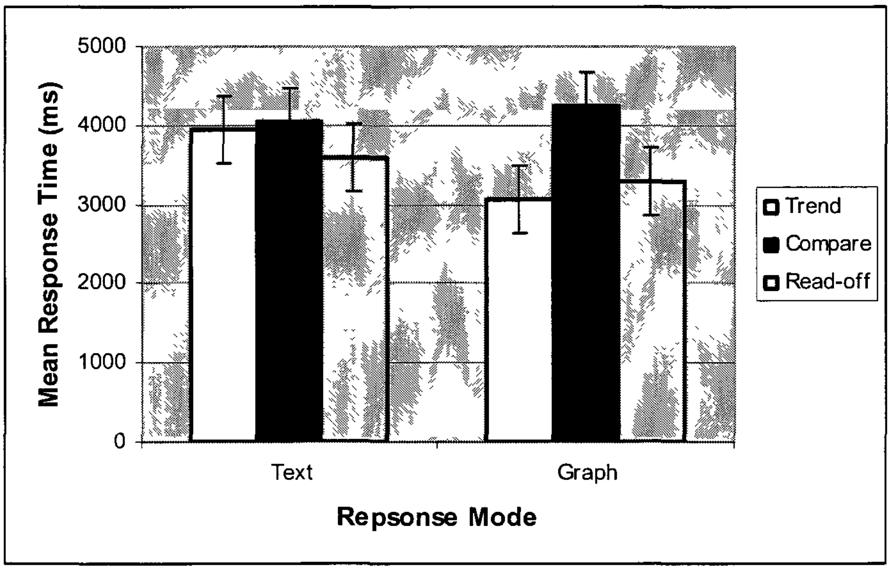


Figure B8. Mean response time for three tasks (trend, compare and read-off), and at both response modes (Text and Graph), collapsed across exposure time conditions, bars show *CIs* for 2-way interaction = $\pm$ 427.15

A 2 x (2 x 3) ANOVA was conducted for the mean response times for response mode (graphical, textual), exposure times (50 ms, 300 ms), and task (trend identification, comparison, read off). The three-way interaction did not reach significance, $F(2, 44) =$

0.84, $p = .438$, $\eta_p^2 = .04$, and neither did the two-way interactions of response mode and exposure time, $F(1, 22) = 0.81$, $p = .379$, $\eta_p^2 = .04$, or exposure time and task, $F(2, 44) = 1.51$, $p = .233$, $\eta_p^2 = .04$. The interaction of response mode and task, however, was marginally significant, $F(2, 44) = 3.18$, $p = .051$, $\eta_p^2 = .13$. Simple main effects analyses revealed a difference in the graphical response mode, $F(2, 44) = 8.86$, $p = .001$, but none in the textual mode, $F(2, 44) = 1.36$, $p = .267$.

Simple simple comparisons conducted within the graphical- and textual response modes showed that the comparison task took significantly longer than the trend identification task, in the graphical mode, $F(1, 24) = 14.61$, $p = .001$, but not in the textual response mode, $F(1, 22) = 0.16$, $p = .691$. The trend and read-off task did not differ significantly the graphical response mode, $F(1, 22) = 0.88$, $p = .359$, or in the textual response mode, $F(1, 22) = 1.89$, $p = .183$. Participants responded faster in the graphical response mode when completing the trend identification tasks and read-off tasks, but not when completing the comparison tasks. Finally, simple simple comparisons of the comparison task and the read-off task showed that the tasks differed in the graphical response mode, $F(1, 22) = 8.50$, $p = .008$, but not in the textual response mode, $F(1, 22) = 2.13$, $p = .258$. Taken together, these results show that participants responded faster in the graphical response mode when completing the trend or read-off tasks but not when completing the comparison task. This finding suggests that in the graphical condition, participants may be relying on memory. This will be taken up in the discussion section. This partially supports Hypothesis 3b, claiming that response times would be shorter for graphical- than for textual response options.

There was a significant main effect of exposure time, $F(1, 22) = 47.85, p = .000$, $\eta_p^2 = .69$, indicating that responses took longer when stimuli was were exposed for 50 ms (M= 4287.29, SE = 218.68) than for 300 ms (M = 3101.35, SE = 112.19), supporting Hypothesis 2b. There was also a significant main effect for task, $F(2, 44) = 7.04, p = .002$, $\eta_p^2 = .24$. Pairwise LSD comparisons indicated that response times were longer for the comparison task than for both the read-off task (p = .005) and the trend identification task (p = .007), but there were no differences between the trend identification and read-off tasks (p = .759). Hypothesis 1b predicting that response times would be shorter for the trend identification task than for both the comparison and read-off tasks, was thus only partially supported.

Certainty

Figure B9 shows that mean certainty ratings were higher in the 300 ms- than in the 50 ms condition. The Figure also shows that certainty was higher in the textual- than in the graphical response mode.

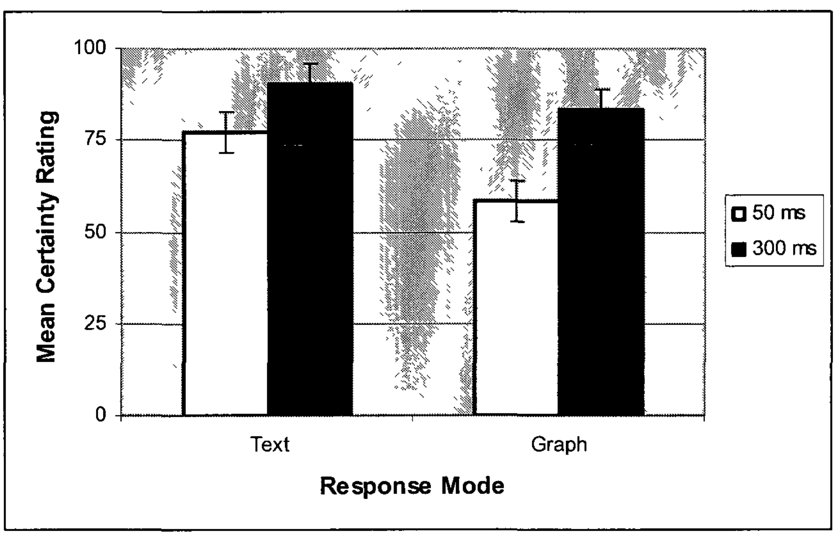


Figure B9. Mean certainty ratings on response mode (textual, graphical), and exposure time condition (50 ms, 300 ms), bars show *CIs* for 2-way interaction = $\pm$ 5.61

A 2 x (2 x 3) ANOVA was conducted for certainty ratings for response mode (graphical, textual), exposure times (50 ms, 300 ms), and task (trend identification, comparison, read off). It showed that the three-way interaction was not significant, $F(1.29, 28.39) = 2.87, p = .093, \eta_p^2 = .12$, indicating that mean certainty ratings did not differ systematically at each level of each independent variable. However, the two-way interaction of response mode and exposure time was significant, $F(1, 22) = 4.64, p = .043, \eta_p^2 = .17$, indicating that certainty ratings differed in the two exposure time conditions as a function of response mode. The main effect of exposure time conditions was also significant, $F(1, 22) = 48.45, p = .000, \eta_p^2 = .69$. As stated earlier, certainty was lower, on average, in the 50ms- than the 300ms condition. This supports Hypothesis 2c, which predicted that certainty would be higher in the 300ms condition than the 50ms condition. Comparison of the mean certainty ratings for the textual and the graphical response mode was also significant, $F(1, 22)= 7.31, p = .013, \eta_p^2 =.25$, showing that certainty was higher in the textual- than the graphical response mode. This contradicts Hypothesis 3c, which stated that participants should be more certain about their performance in the graphical mode than the textual one.

There were no significant two-way interactions involving tasks, namely, task by response mode, $F(2, 44) = 1.21, p = .307, \eta_p^2 = .05$, and task by exposure time, $F(2, 44) = 0.79, p = .460, \eta_p^2 = .04$. There was no main effect for task either, $F(2, 44) = 0.10, p = .908, \eta_p^2 = .004$. Differences in certainty could thus not be accounted for by systematic differences in tasks. Hypothesis 1c stating that participants would be more certain about their performance on the trend identification task than on the other two tasks was thus also refuted.

Individual Differences

Two tests were administered to determine if differences in short term visual memory (VPT) and ability in field dependence-independence (EFT) could account for differences in performance, response times and rated certainty. For a description of the results of the Pearson Product Moment Correlation Coefficients, see Experiment 1.

Summary of Hypotheses

Nine hypotheses were tested in this experiment. The specific hypotheses, the results of each test and the results are reported in Table B3.

Table B3

Hypothesis number, prediction, test used and result

H #	Prediction	Test Used	Result
1a	Performance on trend identification task will be higher than both the comparison and read-off tasks.	Simple main effect	Supported
1b	Response times will be shorter for the trend identification task than both the comparison and read-off tasks.	Pairwise comparisons	Partially Supported
1c	Certainty ratings for the trend identification task will be higher than both the comparison and read-off tasks.	Simple main effect	Refuted
2a	Performance at 300 ms exposure time will be better than performance at 50 ms exposure time.	Main effect of exposure time	Supported
2b	Response times will be shorter at 300 ms exposure times than at 50 ms exposure time.	Main effect of exposure time	Supported
2c	Certainty ratings will be higher for 300 ms exposure time than at 50 ms exposure time.	Main effect of exposure time	Supported
3a	Performance will be better for graphical response options than for textual response options.	Main effect of response mode	Refuted
3b	Response times will be shorter for graphical response options than textual response options.	Main effect of response mode	Partially Supported
3c	Certainty will be higher for graphical response options than for textual response options.	Main effect of response mode	Refuted

Discussion

When considering the demands of the three task types on graph perception and interpretation, task complexity appears to depend on three criteria, namely attention to detail, stimulus exposure time, and stimulus response compatibility. Each of these is discussed.

Attention to Detail

Global information is said to be more readily identified than detailed information (Navon, 1977). This would lead one to predict that the trend identification task should be the easiest of the three types of task tested here because it only involves attending to one item of information. Hypothesis 1a predicting that performance would be best on the trend identification task compared to the comparison and read-off tasks was borne out by the results. One would also assume that, the trend identification task should be completed faster than the other two tasks (as per Hypothesis 1b). This was only partially supported by the results, as response times were not shorter for the trend identification task than for the read-off task. Finally, the trend identification task should lead to higher certainty levels (as per Hypothesis 1c). That was not the case. Therefore, Hypothesis 1c was refuted.

One way to interpret these results is that the three types of task differed in the amount of information requiring attention. The line graphs for example, contained only one line, thus only one item required attention; participants could have ignored everything else. One possible strategy is to attend only to the slope of the line in the trend identification task. This is not possible in the comparison task, in which participants must locate and attend to two elements. The same is true in the read-off task, in which they had to identify both the value and the location of the target element and then remember the value when confronted with four possible responses. Thus, in terms of the response requiring most attention to precision, the read-off task was most demanding and hence the most complex task, and the trend identification task was the least demanding. That interpretation is reflected quite clearly in the results.

The read-off task required only attention to one bar, comparing it to the ordinate, whilst the comparison task involved comparing two bars on the graph. It should have taken longer to complete, but it did not. One possible explanation for these results is that different strategies were used to complete each of the two tasks but the different strategies took the same amount of time.

The level of certainty was similar across all tasks, which refuted Hypothesis 1c. The results indicate that variations in certainty levels could not be attributed to difficulty of the tasks. Participants were not always aware when their responses were incorrect, thus their certainty ratings could not have been based on their actual performance. Instead, they could be a reflection of the perceived ease of the task; apparently this did not differ between the three task types.

Stimulus Exposure Time

As Pinker (1990) contends, if the message the graph creator intended to convey matches the graph readers' conceptual question, the information should be processed automatically by the default encoding mechanism. However, if the task is too complex to rely on automatic processing, the information will take longer to process and errors will occur. In the above experiment, all tasks at the 50 ms stimulus exposure time were completed at higher than chance level on average, showing that some information is likely to be processed automatically. This supports Pinker's (1990) framework and demonstrates that the tasks were not too complex to be completed with some success. It also follows that the longer the inspection time, the better performance will be. Opportunity for iteration is one possibility, as Pinker's (1990) argues.

Performance on graph stimuli presented at 300 ms was predicted to be better (Hypothesis 2a), faster (Hypothesis 2b) and participants should be more certain (Hypothesis 2c). All of these hypotheses were supported. Thus, longer exposure times lead to better and faster performance as well as to higher certainty ratings. Task complexity apparently decreased with longer exposure times, also showing support for Pinker's (1990) theory. As a result of these findings, only the 50 ms stimulus exposure time was retained in future experiments. Since the 300 ms- was clearly easier than the 50 ms condition, there was no reason to continue testing that exposure time.

Stimulus-Response Compatibility

Recall that graphical representations of the responses were predicted to lead to enhanced performance because of the direct mapping between the stimuli and the responses. The conversion of graphical stimuli into text was predicted to require more cognitive processing thereby increasing the complexity of the task and leading to degraded performance. Accordingly, Hypothesis 3a predicted that performance would be better on the graphical- than on the textual response mode. It was refuted because there were no differences in performance between the two response modes. However, responses were faster in the trend identification task and the read-off task than in the comparison task in the graphical response mode. Hypothesis 3b, which stated that participants would be faster at responding to options presented in graphical than in textual formats, was therefore only partially supported.

It was also hypothesized that participants should be more certain about the correctness of their responses in the graphical- than in the textual response mode. However, the reverse was found: certainty ratings were significantly higher in the textual response mode. It may have been harder to picture the trend line represented in the graph option than to identify the magnitude and direction of the slope in the trend identification task. Alternatively, the textual response options may have been perceived as being visually simpler than the graphical response options in the comparison and read-off tasks, which led to higher certainty ratings.

The results showed that response times were shorter in the graphical response mode for the trend identification task and the read-off task, despite the fact that their performance was no better than in the textual response mode. Although it took participants longer to respond to textual options, their higher levels of certainty indicate that they found the textual response mode easier than the graphical response mode, irrespective of their level of performance. The predictions that graphical response options would lead to improved performance, thus resulting in faster responses and higher certainty, were not borne out by these results. Thus, only the textual response mode was retained in subsequent experiments because, first, there were no differences in performance identified between the graphical and textual options. And second, because response times were shorter and certainty was higher for the textual options, indicating that participants found the textual easier than the graphical response mode.

Results from this experiment showed that performance was affected by the amount of attention to detail required to complete the task as well as the duration of stimulus exposure. Task complexity was not affected by the two response options offered here. Thus, this exploration of some variables that could account for differences in 'task' complexity led to methodological decisions allowing a more focused exploration of perceptual complexity in the subsequent experiments.

Methodological Modifications

As a result of the above findings, the following five methodological changes were made to the Experiments 1 through 4, in which a repeated measures design was used. First and foremost, only line graphs were used because line graphs allow the dimensions of orientation, size and color to be manipulated, whereas this is not possible with bar graphs. Furthermore, it was clear that the trend identification task was completed above chance levels at the 50 ms exposure time, which resulted in better performance than the other two tasks. When exploring perceptual processes in graph interpretation, it was necessary to select a specific task that was doable in 50 ms when stimuli were in their simplest form but also one which participants would still be able to complete if was made more complex with the addition of extra information. Thus, it made sense methodologically to select line graphs over bar graphs to explore the processes involved in graph perception further.

Since line graphs were selected for subsequent experiments, the following changes were to overcome the differences in the researcher's and participants' mental models noted in the above experiment. The second methodological change was to replace the four problematic trend identification graphs with ones that ensured the steepness of slopes were obvious. As before (with a small increase/decrease represented by less than 50% difference in magnitude between first and last point and a sharp represented by greater than 60% difference), and the lines were evidently crooked (with at least one very clear direction change). All line graphs were designed to ensure that the patterns differed from one another. All displayed four data points. Third, the number of line graphs was reduced from 15 to 12 because it was a more manageable number for the factorial design employed in the next study, and it allowed the graphs to be completely balanced for direction, magnitude of slope and mirror opposite. Each graph was reproduced 12 times to account for all levels of independent variables, which yielded 144 graphs. The VTP and the EFT were counterbalanced and administered at the beginning and end of the single trial, depending upon condition. Fourth, a distracter line was added to all line graphs to increase the complexity of the task. Finally, the thickness (hereinafter "size") and color of the distracter lines were varied, but the target was always black and of a standard thickness. These changes helped to ensure that there was less discrepancy regarding the 'correct' answer and accounted for conflicting mental models. These changes were also instrumental in ensuring that amount of attention to detail, stimulus exposure time, and response mode variables were not responsible for differences in performance, response times and certainty when next exploring graph perceptual complexity.

Limitations

The textual options employed in this Experiment for the comparison task could have been considered too easy because so little text was displayed. One way to make the options appear more difficult would be to spell out the relation between the variables (A is greater than B) in sentences rather than using symbols (e.g., A > B, etc…) or to perhaps show the values associated with each (A=50, B=90, etc…). This would be worth exploring. Furthermore, four options were selected to be consistent across all experimental tasks, also displaying four options. It was very difficult to create a fourth 'wrong' option, given that there were only 3 variables. For example, the choices for the comparison of A to B were: A < B, A > B, A = B, and B < A. There is a possibility that participants might have noticed that $A > B$ is equivalent to $B < A$. This could have been reflected in their responses if they did not select either of the options and thus, their chance of being correct could have increased to 50% from 25%. Future studies could avoid this problem by assigning values to the variables and letting participants decide whether they were greater than, less than, or equal to the other variable. Finally, there could be a possible interpretation issue regarding the tasks and the graph stimuli used for each. Because the comparison- and read-off tasks employed bar graphs and the trend identification task employed line graphs, the criticism of a possible confound between task and graph could be raised. Although it is possible to raise this concern, the results of this Experiment were not interpreted to address the question of which graph type is better for which graph reading task or vise versa. The question this Experiment intended to answer was whether there was a difference between graph reading tasks, and which can be done most efficiently. Here graph task and graph type are considered simultaneously and not separated for the purpose of interpretation. Thus, the possible criticism regarding a confound between graph task and graph type can be dispelled.

Appendix BB

This Appendix contains all the materials used in Preliminary Study 2, including: consent and debriefing forms, Visual Pattern Test (Della Salla, Gray, Baddley & Wilson, 1997), the Group Embedded Figures Test (Witkin, Oltman, Raskin, & Karp, 1971), the experimental task instructions and notice of recruitment.

All performance, response time and certainty data are also provided in this Appendix.

Performance data includes: descriptive statistics (Table BB1), tests of normality (Table BB2), ANOVA table (Table BB3).

Response time data includes: descriptive statistics (Table BB4), tests of normality (Table BB5 & BB6), ANOVA table (Table BB7).

Certainty data includes: descriptive statistics (Table BB8), tests of normality (Table BB9), ANOVA table (Table BB10).

Data used for the analysis of errors is presented in Tables BB11 and BB12, followed by descriptive statistics for the error analysis for certainty ratings (Table BB13) and response times (Table BB14). ANOVA tables analyzing certainty (Table BB15) as well as response times (Table BB16) for correct and incorrect responses, response mode, and exposure time.

Correlation tables are shown for relations between task and response time conditions and both the Visual Pattern Test and the Group Embedded Figures Test using: performance scores (Table BB17), response times (Table BB18) and certainty ratings (Table BB19).

Consent
(Amended for length of study and remuneration for all other studies)

The purpose of this consent form is to inform you as to what the purpose of the experiment is and what you will be required to do so that you have adequate information to decide whether or not you wish to participate.

If you have questions regarding this study, please contact:
Cathy Dudek (Principal Investigator)
Telephone: (613) 520-2600 Ex-6628, e-mail:
Dr. G. Lindgaard (Thesis Supervisor)
Telephone: (613) 520-2600 Ex-2255, e-mail: gitte_lindgaard@carleton.ca

If you have ethical concerns regarding this study, please contact:
Dr. Avi Parush (Chair of the Carleton University Research Ethics Committee for Psychological Research)
Telephone: (613) 520-2600 Ex-6026, e-mail: avi_parush@carleton.ca
Dr. Anne Bowker (Departmental Chair)
Telephone: (613) 520-2600 Ex-2648, e-mail: psychchair@carleton.ca

Purpose

The purpose of this experiment is to learn more about graph perception and comprehension.

Task Requirements

You will be required to view a number of different graphs and then answer questions about them. After you do that you will be asked to rate how certain you are about your answer.

Duration and Locale

Each session should take no longer than two hours. The experiment will take place in the HOT Lab research facility located in 210 SSRB or at similar other locations that share the same characteristics.

Consideration

You will receive $15.00 for taking part in this experiment, regardless of whether you complete the experiment or not.

Potential Risk/Discomfort

There is the possibility that some people may experience some discomfort because the graphs are presented for very brief periods of time. To ensure that any discomfort is kept to a minimum, we encourage you to take as many breaks as you feel are necessary throughout the experiment. You control the pace at which the graphs are presented. If you feel like taking a break, you are welcome and encouraged to do so at any time.

Confidentiality

All data that is collected will be held completely confidential. The data will only be made available to those people involved with this research. Data will be coded for identification purposes.

Right to Withdraw

You have the right to withdraw at any time, without any explanation as to the reason for withdrawing and the right to not answer any questions that you don't want to.

Signatures

I have read and understand the above terms of the web site study and I understand the conditions of my participation. My signature indicates that I agree to participate in this experiment.

Participant's Name: ______________________________

Participant's Signature: ___________________________

Researcher's Name: ______________________________

Researcher's Signature: ___________________________

Debriefing
(Amended for all studies)

You were a participant in an experiment conducted by Cathy Dudek, Supervised by Dr. Gitte Lindgaard, of the Department of Psychology.

This study was designed to test people's ability to interpret graphical information quickly. You answered questions about graphs that you saw once for 50 milliseconds and then you answered the same questions again about graphs that you saw for 300 milliseconds. These two answers were then compared to see if you could answer questions about a graph as accurately at 50 msecs as you can at 300 msecs. Your ratings will be combined with others who did the same task you did to arrive at a mean score for each exposure time. We were looking specifically for how much agreement there is between your first and second answers. Further, we were looking at how certain you were regarding your answers. Those ratings were also compared between your first and second answers. We are also interested in whether or not there are differences in performance, certainty and response times for people who were presented with option graphs compared to those who received written descriptions of the graphs.

The tests of spatial ability and visual memory will be used to understand what role these abilities play in assisting people with graph comprehension tasks.

It is important to note that all data used in the graphs presented during the study and in the preliminary graph comprehension test are completely ficticious and do not represent real statistics.

Thank you for your willingness to help us to learn more about graph perception and comprehension. If you have any further questions regarding this study please contact:

If you have questions regarding this study, please contact:
Cathy Dudek (Principal Investigator)
Telephone: (613) 520-2600 Ex-6628, e-mail: cathy@interactingwithcomputers.com
Dr. G. Lindgaard (Thesis Supervisor)
Telephone: (613) 520-2600 Ex-2255, e-mail: gitte_lindgaard@carleton.ca

If you have ethical concerns regarding this study, please contact:
Dr. Avi Parush (Chair of the Carleton University Research Ethics Committee for Psychological Research)
Telephone: (613) 520-2600 Ex-6026, e-mail: avi_parush@carleton.ca
Dr. Anne Bowker (Departmental Chair)
Telephone: (613) 520-2600 Ex-2648, e-mail: psychchair@carleton.ca

Visual Pattern Test
(Same for all studies)

The Visual Pattern Test was administered according to the instructions given by (Della Salla, Gray, Baddley & Wilson, 1997) in the test package. *Permission to provide a sample is required.*

Group Embedded Figures Test
(Same for all studies)

The Group Embedded Figures Test was administered according to the instructions given by (Witkin, Oltman, Raskin, & Karp, 1971) in the test package. *Permission to provide a same is required.*

Instructions

Introductory Instructions

In the first part of the study there are 3 different
sections. In each section, you will be shown a series
of graphs for very brief periods of time.
Each section requires you to complete a diferent task:
indicate trends, compare between variables and
identify values.
Instructions will be given and you will get
the opportunity to practice at the beginning of
each section.
When you are ready to do the first task,
press the space bar.

Trend Identification Instructions (graphical)

To begin, you will see a line graph very briefly.
You will then be presented with four option graphs.
Your job is to choose the trend that is most similar
to the one you saw in the line graph.
You will then be asked how certain you are of your choice.
When you are done, press the spare bar to move
on to the next line graph. You will do this for each
graph until you have worked through all the graphs.
Do you have any questions?
When you are ready to begin, press the space bar.

Trend Identification Instructions (Textual)

To begin, you will see a line graph very briefly.
You will then be presented with four options.
Your job is to choose the trend that describes
the one you saw in the line graph.
You will then be asked how certain you are of your choice.
When you are done, press the spare bar to move
on to the next line graph. You will do this for each
graph until you have worked through all the graphs.
Do you have any questions?
When you are ready to begin, press the space bar.

Value Identification Instructions (graphical)

Before being shown a bar graph very briefly, you will be
prompted to look for a certain variable. Each graph
will have three columns,labelled 'A', 'B' and 'C':
in that order. For example, you will see, "value of B".
You are to look for that value on the graph presented.
You will then be presented with four option graphs. Your
job is to choose which one shows the value you were asked
to look for, in this case ‘B'.
You will then be asked how certain you are. When
you are done, you will press the space bar to
move on to the next bar graph until you have
worked through all the graphs.
Do you have any questions?
When you are ready to begin, press the space bar.

Value Identification Instructions (Textual)

Before being shown a bar graph very briefly, you will be
prompted to look for a certain variable. Each graph
will have three columns,labelled 'A', 'B' and 'C':
in that order. For example, you will see, "value of B".
You are to look for that value on the graph presented.
You will then be presented with four options. Your
job is to choose which one shows the value you were asked
to look for, in this case ‘B'.
You will then be asked how certain you are. When
you are done, you will press the space bar to
move on to the next bar graph until you have
worked through all the graphs.
Do you have any questions?
When you are ready to begin, press the space bar.

Comparison Instructions (Graphical)

Before being shown a bar graph very briefly, you will be prompted to look for the relation between specific variables. Each graph will have three columns,labelled 'A', 'B' and 'C': in that order. For example, you will see, "compare A and B". You are to look for those variables on the graph presented.
You will then be presented with four option graphs. Your job is to choose which graph shows the values you were asked to look for, in this case 'A' and 'B'.
You will then be asked how certain you are of your choice.
When you are done, press the space bar to move on to the next graph.
Do you have any questions?
When you are ready to begin, press the space bar.

Comparison Instructions (Textual)

Before being shown a bar graph very briefly, you will be prompted to look for the relation between specific variables. Each graph will have three columns,labelled 'A', 'B' and 'C': in that order. For example, you will see, "compare A and B". You are to look for those variables on the graph presented.
You will then be presented with four options describing the relation between the variables. Your job is to choose the one that describes the relation between'A' and 'B'. You will then be asked how certain you are of your choice. When you are done, press the space bar to move on to the next graph.
Do you have any questions?
When you are ready to begin, press the space bar.

Performance Data

Table BB1

Descriptive statistics for performance on textual and graphical response options for task and exposure time

Task and Exposure Time	N	Range	Minimum	Maximum	Mean	Std. Deviation
			Textual			
Trend 50 ms	12	82	18	100	75.92	23.97
Trend 300 ms	12	46	54	100	85.67	16.66
Compare 50 ms	12	73	20	93	61.08	28.48
Compare 300 ms	12	73	27	100	83.83	19.67
Read 50 ms	12	86	7	93	46.08	29.10
Read 300 ms	12	53	47	100	81.75	18.60
			Graphical			
Trend 50 ms	12	55	45	100	75.83	15.60
Trend 300 ms	12	9	91	100	98.50	3.50
Compare 50 ms	12	54	33	87	58.42	17.87
Compare 300 ms	12	33	67	100	90.50	10.32
Read 50 ms	12	67	13	80	54.00	19.40
Read 300 ms	12	27	73	100	90.50	9.20

Table BB2

Tests of normality for performance on tasks and exposure times, collapsed across response mode

Tests of Normality						
	Kolmogorov-Smirnov			Shapiro-Wilk		
Task and Exposure Time	Statistic	df	Sig.	Statistic	df	Sig.
Trend 50 ms	.149	24	.179	.903	24	.026*
Trend 300 ms	.388	24	.000***	.660	24	.000***
Compare 50 ms	.171	24	.069	.921	24	.063
Compare 300 ms	.228	24	.002**	.715	24	.000***
Read 50 ms	.130	24	.200*	.954	24	.338
Read 300 ms	.232	24	.002**	.812	24	.000***

*Indicates a lower bound of the true significance

Table BB3

Analysis of Variance for performance on task, exposure time and response mode

Source	*df*	*F*	η_p^2	*p*
		Between subjects		
Response Mode (A)	1	1.19	.05	.286
S within – group error	22	(935.08)		
		Within subjects		
Task (B)	2	9.20^{**}	.294	.001
B X A	2	0.37	.02	.693
B X A within-group error	44	(341.05)		
Time (C)	1	97.99^{***}	.81	.000
C X A	1	2.05	.085	.166
C X A within-group error	22	(259.35)		
B X C	2	7.53^{**}	.255	.002
B X C X A	2	0.73	.032	.485
B X C X A within-group error	44	(158.21)		

Response Time Data

Table BB4

Descriptive statistics for response times (after adjustments) on textual and graphical response options for task and exposure time

Descriptive Statistics						
Task and Exposure Time	N	Range	Minimum	Maximum	Mean	Std. Deviation
Textual						
Trend 50 ms	12	4291.37	2738.18	7029.55	4551.85	1412.35
Trend 300 ms	12	2169.81	2288.55	4458.36	3312.73	645.44
Compare 50 ms	12	7605.82	1172.13	8777.95	4777.38	2180.93
Compare 300 ms	12	3620.73	1824.67	5445.40	3337.52	1130.67
Read 50 ms	12	2653.03	3213.87	5866.90	4251.79	1062.17
Read 300 ms	12	1810.53	1962.07	3772.60	2911.28	649.27
Graphical						
Trend 50 ms	12	3666.54	2235.46	5902.00	3310.85	1003.82
Trend 300 ms	12	1812.09	1910.36	3722.45	2800.08	633.67
Compare 50 ms	12	5328.07	2963.93	8292.00	4824.64	1518.08
Compare 300 ms	12	1799.67	2987.33	4787.00	3665.28	644.71
Read 50 ms	12	3607.60	2874.20	6481.80	4007.24	1077.52
Read 300 ms	12	755.27	2186.93	2942.20	2581.22	262.70

Table BB5

Tests of normality for response times for task and exposure time, collapsed across response mode

	Tests of Normality					
	Kolmogorov-Smirnov			Shapiro-Wilk		
Task and Exposure Time	Statistic	df	Sig.	Statistic	df	Sig.
Trend 50 ms	.169	24	.076	.881	24	.009**
Trend 300 ms	.130	24	.200*	.955	24	.345
Compare 50 ms	.233	24	.002**	.651	24	.000***
Compare 300 ms	.199	24	.015*	.923	24	.068
Read 50 ms	.205	24	.011*	.803	24	.000***
Read 300 ms	.177	24	.050	.866	24	.004**

*Indicates a lower bound of the true significance

Table BB6

Tests of normality for response times (after adjustments) times for task and exposure time, collapsed across response mode

	Tests of Normality					
	Kolmogorov-Smirnov			Shapiro-Wilk		
Task and Exposure Time	Statistic	df	Sig.	Statistic	df	Sig.
Trend 50 ms	.123	24	.200*	.913	24	.042*
Trend 300 ms	.080	24	.200*	.979	24	.877
Compare 50 ms	.122	24	.200*	.960	24	.433
Compare 300 ms	.182	24	.039*	.946	24	.221
Read 50 ms	.179	24	.046*	.887	24	.012*
Read 300 ms	.109	24	.200*	.944	24	.202

*Indicates a lower bound of the true significance

Table B7

Analysis of Variance for response times on task, exposure time and response mode

Source	*df*	*F*	η_p^2	*p*
Between subjects				
Response Mode (A)	1	1.16	.050	.293
S within – group error	22			
Within subjects				
Task (B)	2	7.04	.242**	.002
B X A	2	3.18	.126	.051
B X A within-group error	44	(1072117.4)		
Time (C)	1	47.85	.685	.000
C X A	1	0.81	.035*	.379
C X A within-group error	22	(1058147.6)		
B X C	2	1.51	.064	.233
B X C X A	2	0.84	.036*	.433
B X C X A within-group error	44	(592121.29)		

Certainty Data

Table BB8

Descriptive statistics for certainty on textual and graphical response options for task and exposure time

Descriptive Statistics						
Task and Exposure Time	N	Range	Minimum	Maximum	Mean	Std. Deviation
Textual						
Trend 50 ms	12	49	51	100	74.75	17.34
Trend 300 ms	12	30	70	100	91.17	11.16
Compare 50 ms	12	46	54	100	79.50	15.74
Compare 300 ms	12	25	75	100	91.08	8.23
Read 50 ms	12	48	52	100	76.67	15.70
Read 300 ms	12	26	74	100	88.67	9.55
Graphical						
Trend 50 ms	12	70	16	86	61.92	23.55
Trend 300 ms	12	44	53	97	79.92	14.92
Compare 50 ms	12	63	19	82	55.75	21.76
Compare 300 ms	12	32	65	97	81.50	10.17
Read 50 ms	12	75	17	92	56.92	22.49
Read 300 ms	12	30	70	100	89.00	10.07

Table BB9

Tests of normality for certainty for task and exposure time collapsed across response mode

Tests of Normality						
	Kolmogorov-Smirnov			Shapiro-Wilk		
Task and Exposure Time	Statistic	df	Sig.	Statistic	df	Sig.
Trend 50 ms	.138	24	.200*	.914	24	.043*
Trend 300 ms	.165	24	.088	.885	24	.011*
Compare 50 ms	.102	24	.200*	.941	24	.171
Compare 300 ms	.148	24	.186	.937	24	.136
Read 50 ms	.132	24	.200*	.962	24	.486
Read 300 ms	.198	24	.016*	.907	24	.030*

*Indicates a lower bound of the true significance

Table BB10

Analysis of Variance for certainty on task, exposure time and response mode

Source	*df*	*F*	η_p^2	*p*
		Between subjects		
Response Mode (A)	1	7.31*	.249	.013
S within – group error	22			
		Within subjects		
Task (B)	2	0.96	.004	.908
B X A	2	1.21	.052	.307
B X A within-group error	44	(124.05)		
Time (C)	1	48.45***	.688	.000
C X A	1	4.64*	.174	.043
C X A within-group error	22	(276.91)		
B X C	2	.79	.035	.460
B X C X A	2	2.87	.115	.068
B X C X A within-group error	44	(93.42)		

Error Data

Table BB11

Graph number and the number of participants who erred on each of the tasks, in each exposure time, in the textual response mode (n=12)

Graph #	Trend 50 ms	Trend 300 ms	Compare 50 ms	Compare 300 ms	Read 50 ms	Read 300 ms
1	12	9	5	1	6	3
2	7	8	4	1	6	1
3	6	10	4	2	6	3
4	6	9	3	1	7	3
5	4	1	10	6	7	2
6	2	2	3	1	9	1
7	4	3	3	1	8	3
8	2	3	3	4	5	3
9	2	1	6	2	6	3
10	2	2	2	1	8	2
11	4	1	8	3	7	4
12	3	0	5	0	4	1
13	3	2	6	2	3	0
14	4	2	5	2	8	1
15	2	2	3	1	7	2

Table BB12

Graph number and the number of participants who erred on each of the tasks, in each exposure time, in the graphical response mode (n=12)

Graph #	Trend 50 ms	Trend 300 ms	Compare 50 ms	Compare 300 ms	Read 50 ms	Read 300 ms
1	10	5	5	1	4	1
2	11	4	3	0	5	2
3	10	3	2	0	5	1
4	9	5	3	1	7	3
5	2	1	6	2	9	0
6	4	0	4	3	6	2
7	2	0	5	2	6	1
8	2	0	4	1	6	0
9	2	0	6	0	5	0
10	2	0	7	2	2	0
11	4	0	8	2	4	1
12	6	1	4	0	8	4
13	2	0	4	1	6	0
14	5	1	10	0	5	2
15	1	0	5	2	5	0

Table BB13

Descriptive statistics for certainty of correct and incorrect responses in each exposure time condition

Descriptive Statistics						
Accuracy and Exposure Time	N	Range	Minimum	Maximum	Mean	Std. Deviation
Textual						
Correct 50 ms	10	74.38	16.25	90.63	60.57	25.79
Incorrect 50 ms	10	69.28	16.43	85.71	53.85	23.38
Correct 300 ms	10	42.63	54.64	97.27	78.26	14.09
Incorrect 300 ms	10	89.50	7.5	97.00	60.70	27.26
Graphical						
Correct 50 ms	12	72.5	27.50	100	72.43	21.81
Incorrect 50 ms	12	51.5	48.5	100	71.84	16.94
Correct 300 ms	12	31.11	68.89	100	92.88	10.12
Incorrect 300 ms	12	36.43	63.57	100	87.58	14.03

Table BB14

Descriptive statistics for response times for correct and incorrect responses for each exposure time condition

Descriptive Statistics						
Accuracy and Exposure Time	N	Range	Minimum	Maximum	Mean	Std. Deviation
Textual						
Correct 50 ms	10	2468.85	2202.40	4671.25	2972.51	719.89
Incorrect 50 ms	10	5501.09	2497.20	7998.29	4242.09	1588.38
Correct 300 ms	10	1719.34	1994.23	3713.57	2933.78	649.68
Incorrect 300 ms	10	4910.00	1790.00	6700.00	4118.05	1915.68
Graphical						
Correct 50 ms	12	6304.07	2682.50	8986.57	4897.10	1832.43
Incorrect 50 ms	12	8000.25	2324.00	10324.25	5217.37	2254.87
Correct 300 ms	12	2169.81	2288.55	4458.36	3168.20	639.50
Incorrect 300 ms	12	8343.75	2127.00	10515.75	4457.92	2243.69

Table BB15

Analysis of Variance for certainty for correct and incorrect responses, exposure times and response modes

Source	*df*	*F*	η_p^2	*p*
		Between subjects		
Response Mode (A)	1	8.01*	.286	.010
S within – group error	20	(6944.53)		
		Within subjects		
Accuracy (B)	1	14.88**	.427	.001
B X A	1	5.53*	.217	.029
B X A within-group Error	20	(83.41)		
Time (C)	1	10.07**	.335	.005
C X A	1	0.37	.018	.550
C X A within-group Error	20	(499.61)		
B X C	1	3.182	.137	.090
B X C X A	1	0.495	.024*	.490
B X C X A within-group error	20	(103.48)		

Table BB16

Analysis of Variance for response times for correct and incorrect responses, exposure times and response modes

Source	*df*	*F*	η_p^2	*p*
Between subjects				
Response Mode (A)	1	3.64	.154	.071
S within – group error	20	(4527895.24)		
Within subjects				
Accuracy (B)	1	16.14^{**}	.447	.001
B X A	1	0.70	.034	.414
B X A within-group Error	20	(1394980.51)		
Time (C)	1	2.89	.105	.126
C X A	1	2.22	.100	.152
C X A within-group Error	20	(3319154.64)		
B X C	1	0.66	.032	.425
B X C X A	1	0.95	.045	.343
B X C X A within-group error	20	(1608415.02)		

Correlations

Table BB17

Correlations between performance scores on the three different tasks, at 50- and 300 ms exposure times (collapsed across response mode) and Visual Patterns Test (VPT) and Group Embedded Figures Test (EFT)

	EFT	Trend 50	Trend 300	Comp 50	Comp 300	Read 50	Read 300
Performance (N=24)							
VPT	.618**	.153	.183	.339	.251	.219	-.071
EFT	-	.514*	.311	.215	.181	.324	.001

* indicates p < .05, indicates p < .01

Table BB18

Correlations between response times, on the three different tasks, at 50- and 300 ms exposure times (collapsed across response mode) and Visual Patterns Test (VPT) and Group Embedded Figures Test (EFT)

	EFT	Trend 50	Trend 300	Comp 50	Comp 300	Read 50	Read 300
Response Times (N=24)							
VPT	.618**	-.245	-.064	-.064	-.228	.095	-.107
EFT	-	-.320	-.122	.115	-.086	.203	.110

* indicates p < .05, indicates p < .01

Table BB19

Correlations between certainty ratings on the three different tasks, at 50- and 300 ms exposure times (collapsed across response mode) and Visual Patterns Test (VPT) and Group Embedded Figures Test (EFT)

	EFT	Trend 50	Trend 300	Comp 50	Comp 300	Read 50	Read 300
				Certainty (N=24)			
VPT	.618**	.305	.022	.115	.211	-.041	.255
EFT	-	.182	.067	-.037	.023	-.136	.004

* indicates p < .05, indicates p < .01

Appendix C

This Appendix contains the experimental task instructions for Experiment 1 and all other materials used were provided before.

All performance, response time and certainty data are also provided in this Appendix.

Performance data includes: descriptive statistics (Table C1), tests of normality (Table C2 & C3), ANOVA table (Table C4).

Response time data includes: descriptive statistics (Table C5), tests of normality (Table C6 & C7), ANOVA table (Table C8).

Certainty data includes: descriptive statistics (Table C9), tests of normality (Table C10 & C11), ANOVA table (Table C12).

Data used for the analysis of errors is presented in Tables C13, C14 and C15 and information for the *t*-test for illusory conjunctions is presented in Table C16.

Correlation tables are shown for relations between task and response time conditions and both the Visual Pattern Test and the Group Embedded Figures Test using: performance scores (Table C17), response times (Table C18) and certainty ratings (Table C19).

Instructions
(Same for subsequent experiments but description of distracters was amended to include, red or black, thick or thin)

Introductory Instructions

You will see a line graph very briefly. Each graph has two lines: one will have direction changes and one will always be perfectly straight. The one you are to pay attention to is the one with direction changes and not the straight line. The line with direction changes will always be black, whilst the straight ones will be red or black. You are only interested in trend of the crooked black line. You will then be presented with four options and your job is to choose the description of the trend that best describes the crooked black line you saw in the line graph. You will then be asked how certain you are of your choice. When you are done, press the spare bar to move on to the next line graph. You will do this for each graph until you have worked through all the graphs.
Do you have any questions?
When you are ready to begin, press the space bar.

[practice graphs completed]

Thanks!
You're done with the practice section.

Do you have any questions?

If you are ready to continue to the test graphs, press the space bar.

Performance Data

Table C1

Descriptive statistics for performance (after adjustments) on color, orientation and size of distracter lines

Distracter Characteristics	N	Range	Minimum	Maximum	Mean	Std. Deviation
Red Distracters (*N*=10)						
Consistent thin	10	50.00	41.67	91.67	59.17	14.41
Consistent thick	10	33.34	50.00	83.34	66.67	9.62
Inconsistent thin	10	41.67	41.67	83.34	62.50	15.34
Inconsistent thick	10	50.00	33.33	83.33	60.83	17.59
Flat thin	10	41.67	50.00	91.67	72.50	14.19
Flat thick	10	34.86	56.81	91.67	79.02	10.19
Black Distracters (*N*=10)						
Consistent thin	10	41.67	41.67	83.34	64.17	15.74
Consistent thick	10	66.66	33.34	100.00	59.17	20.95
Inconsistent thin	10	34.91	33.33	68.24	50.16	9.93
Inconsistent thick	10	50.00	41.67	91.67	63.34	15.81
Flat thin	10	50.00	41.67	91.67	77.50	15.74
Flat thick	10	41.66	58.34	100.00	78.34	12.55

Table C2

Tests of normality for performance on color, orientation and size of distracter lines

Tests of Normality						
	Kolmogorov-Smirnov			Shapiro-Wilk		
Distracter Characteristics	Statistic	df	Sig.	Statistic	df	Sig.
Red Distracters						
Consistent thin	.238	10	.115	.862	10	.080
Consistent thick	.200	10	.200*	.953	10	.703
Inconsistent thin	.113	10	.200*	.928	10	.426
Inconsistent thick	.190	10	.200*	.928	10	.424
Flat thin	.170	10	.200*	.939	10	.541
Flat thick	.383	10	.000	.675	10	.000
Black Distracters						
Consistent thin	.254	10	.066	.867	10	.093
Consistent thick	.269	10	.039	.894	10	.187
Inconsistent thin	.263	10	.048	.811	10	.020
Inconsistent thick	.184	10	.200*	.942	10	.573
Flat thin	.244	10	.092	.841	10	.045
Flat thick	.155	10	.200*	.969	10	.886

*Indicates a lower bound of the true significance

Table C3

Tests of normality for performance (after adjustments) on color, orientation and size of distracter lines

Tests of Normality						
	Kolmogorov-Smirnov			Shapiro-Wilk		
Distracter Characteristics	Statistic	df	Sig.	Statistic	df	Sig.
Red Distracters						
Consistent thin	.238	10	.115	.862	10	.080
Consistent thick	.200	10	.200*	.953	10	.703
Inconsistent thin	.113	10	.200*	.928	10	.426
Inconsistent thick	.190	10	.200*	.928	10	.424
Flat thin	.170	10	.200*	.939	10	.541
Flat thick	.247	10	.086	.878	10	.123
Black Distracters						
Consistent thin	.254	10	.066	.867	10	.093
Consistent thick	.269	10	.039	.894	10	.187
Inconsistent thin	.206	10	.200*	.952	10	.697
Inconsistent thick	.184	10	.200*	.942	10	.573
Flat thin	.244	10	.092	.841	10	.045
Flat thick	.155	10	.200*	.969	10	.886

*Indicates a lower bound of the true significance

Table C4

Analysis of Variance for performance on distracter color, orientation and size

Source	*df*	*F*	η_p^2	*p*
Within Subjects				
Orientation (A)	2	20.58^{***}	.695	.000
A within-group error	18	(172.34)		
Size (B)	1	3.28	.267	.103
B within-group Error	9	(115.79)		
Color (C)	1	1.42	.136	.264
C within-group error	9	(37.82)		
A X B	2	0.53	.055	.597
A X B within-group error	18	(95.64)		
A X C	2	0.96	.096	.401
A X C within-group error	18	(130.48)		
B X C	1	.05	.005	.836
B X C within-group error	9	(202.87)		
A X B X C	2	3.78^{*}	.295	.043
A X B X C within-group error	18	(133.97)		

Response Time Data

Table C5

Descriptive statistics for response times (after adjustments) on distracter color, orientation and size

Descriptive Statistics						
Distracter Characteristics	N	Range	Minimum	Maximum	Mean	Std. Deviation
Red Distracters						
Consistent thin	10	3090.00	2822.42	5912.42	3725.57	928.09
Consistent thick	10	3151.33	2853.92	6005.25	4062.42	958.87
Inconsistent thin	10	4337.92	2456.50	6794.42	3816.20	1353.52
Inconsistent thick	10	2791.92	2148.83	4940.75	3423.08	886.42
Flat thin	10	1175.08	2819.00	3994.08	3170.81	382.49
Flat thick	10	2650.67	2462.58	5113.25	3420.58	987.23
Black Distracters						
Consistent thin	10	3503.41	3072.67	6576.08	4137.59	1020.90
Consistent thick	10	1750.83	2849.42	4600.25	3666.85	532.98
Inconsistent thin	10	3491.66	3049.92	6541.58	4971.46	1113.77
Inconsistent thick	10	3360.25	2512.33	5872.58	4119.11	1173.62
Flat thin	10	3341.09	2238.08	5579.17	3697.42	884.03
Flat thick	10	2453.00	2445.33	4898.33	3596.89	724.27

*Indicates a lower bound of the true significance

Table C6

Tests of normality for response times on distracter color, orientation and size

Tests of Normality						
	Kolmogorov-Smirnov			Shapiro-Wilk		
Distracter Characteristics	Statistic	df	Sig.	Statistic	df	Sig.
Red Distracters						
Consistent thin	.169	10	.200*	.859	10	.074
Consistent thick	.197	10	.200*	.938	10	.532
Inconsistent thin	.169	10	.200*	.885	10	.148
Inconsistent thick	.177	10	.200*	.952	10	.689
Flat thin	.278	10	.027	.672	10	.000
Flat thick	.255	10	.064	.863	10	.082
Black Distracters						
Consistent thin	.217	10	.198	.853	10	.063
Consistent thick	.130	10	.200*	.980	10	.964
Inconsistent thin	.237	10	.119	.908	10	.270
Inconsistent thick	.159	10	.200*	.941	10	.567
Flat thin	.181	10	.200*	.948	10	.648
Flat thick	.213	10	.200*	.957	10	.747

*Indicates a lower bound of the true significance

Table C7

Tests of normality for response times (after adjustments) on distracter color, orientation and size

Tests of Normality						
	Kolmogorov-Smirnov			Shapiro-Wilk		
Distracter Characteristics	Statistic	df	Sig.	Statistic	df	Sig.
Red Distracters						
Consistent thin	.169	10	.200*	.859	10	.074
Consistent thick	.197	10	.200*	.938	10	.532
Inconsistent thin	.169	10	.200*	.885	10	.148
Inconsistent thick	.177	10	.200*	.952	10	.689
Flat thin	.213	10	.200*	.859	10	.075
Flat thick	.255	10	.064	.863	10	.082
Black Distracters						
Consistent thin	.217	10	.198	.853	10	.063
Consistent thick	.130	10	.200*	.980	10	.964
Inconsistent thin	.237	10	.119	.908	10	.270
Inconsistent thick	.159	10	.200*	.941	10	.567
Flat thin	.181	10	.200*	.948	10	.648
Flat thick	.213	10	.200*	.957	10	.747

*Indicates a lower bound of the true significance

Table C8

Analysis of Variance for response times on distracter color, orientation and size

Source	*df*	*F*	η_p^2	*p*
Within subjects				
Orientation (A)	2	4.90^{*}	.353	.020
A within-group error	18	(801152.08)		
Size (B)	1	3.97	.306	.077
B within-group Error	9	(317659.37)		
Color (C)	1	7.10^{*}	.441	.026
C within-group error	9	(775301.58)		
A X B	2	2.66	.228	.097
A X B within-group error	18	(510582.68)		
A X C	2	5.79^{*}	.391	.011
A X C within-group error	18	(371329.89)		
B X C	1	4.87	.351	.055
B X C within-group error	9	(447648.21)		
A X B X C	2	0.32	.034	.730
A X B X C within-group error	18	(444966.43)		

Certainty Data

Table C9

Descriptive statistics for certainty (after adjustments) on distracter color, orientation and size

Distracter Characteristics	N	Range	Minimum	Maximum	Mean	Std. Deviation
Red Distracters						
Consistent thin	10	54.59	22.08	76.67	64.01	17.04
Consistent thick	10	52.50	34.17	86.67	64.43	14.28
Inconsistent thin	10	60.17	26.25	86.42	66.18	18.78
Inconsistent thick	10	60.67	25.83	86.50	66.83	17.82
Flat thin	10	35.44	50.31	85.75	70.90	11.369
Flat thick	10	32.88	55.45	88.33	75.36	10.12
Black Distracters						
Consistent thin	10	67.41	27.92	95.33	67.16	17.61
Consistent thick	10	56.25	31.25	87.50	64.82	16.69
Inconsistent thin	10	47.83	15.17	63.00	42.82	12.94
Inconsistent thick	10	53.67	25.83	79.50	64.60	17.21
Flat thin	10	70.33	24.17	94.50	66.35	20.10
Flat thick	10	59.83	23.75	83.58	65.75	17.47

Table C10

Tests of normality for certainty on distracter color, orientation and size

Tests of Normality						
	Kolmogorov-Smirnov			Shapiro-Wilk		
Distracter Characteristics	Statistic	df	Sig.	Statistic	df	Sig.
Red Distracters						
Consistent thin	.229	10	.147	.769	10	.006
Consistent thick	.208	10	.200*	.918	10	.344
Inconsistent thin	.246	10	.089	.865	10	.087
Inconsistent thick	.207	10	.200*	.872	10	.107
Flat thin	.265	10	.045	.812	10	.020
Flat thick	.316	10	.006	.753	10	.004
Black Distracters						
Consistent thin	.200	10	.200*	.914	10	.310
Consistent thick	.186	10	.200*	.953	10	.705
Inconsistent thin	.179	10	.200*	.944	10	.594
Inconsistent thick	.235	10	.124	.831	10	.034
Flat thin	.208	10	.200*	.924	10	.387
Flat thick	.221	10	.182	.846	10	.052

*Indicates a lower bound of the true significance

Table C11

Tests of normality for certainty (after adjustments) on distracter color, orientation and size

Tests of Normality						
	Kolmogorov-Smirnov			Shapiro-Wilk		
Distracter Characteristics	Statistic	df	Sig.	Statistic	df	Sig.
Red Distracters						
Consistent thin	.229	10	.147	.769	10	.006
Consistent thick	.208	10	.200*	.918	10	.344
Inconsistent thin	.246	10	.089	.865	10	.087
Inconsistent thick	.207	10	.200*	.872	10	.107
Flat thin	.155	10	.200*	.949	10	.657
Flat thick	.221	10	.184	.927	10	.417
Black Distracters						
Consistent thin	.200	10	.200*	.914	10	.310
Consistent thick	.186	10	.200*	.953	10	.705
Inconsistent thin	.179	10	.200*	.944	10	.594
Inconsistent thick	.235	10	.124	.831	10	.034
Flat thin	.208	10	.200*	.924	10	.387
Flat thick	.221	10	.182	.846	10	.052

*Indicates a lower bound of the true significance

Table C12

Analysis of Variance for certainty on distracter color, orientation and size

Source	*df*	*F*	η_p^2	*p*
Within subjects				
Orientation (A)	2	14.46^{***}	.613	.000
A within-group error	18	(62.26)		
Size (B)	1	86.06^{***}	.905	.000
B within-group Error	9	(5.75)		
Color (C)	1	14.81^{**}	.622	.004
C within-group error	9	(73.79)		
A X B	2	8.30^{**}	.479	.003
A X B within-group error	18	(48.74)		
A X C	2	9.37^{**}	.510	.002
A X C within-group error	18	(57.45)		
B X C	1	3.60	.286	.090
B X C within-group error	9	(41.02)		
A X B X C	2	16.44^{***}	.646	.000
A X B X C within-group error	18	(32.00)		

Error Analysis

Table C13

Number of errors of magnitude (small vs. sharp slope) on each graph for distracter size, orientation and color

Distracter Size:		Thin						Thick					
Distracter Orientation:		Consistent		Inconsistent		Flat		Consistent		Inconsistent		Flat	
Target Orientation:		Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
Graph	Mag												
Black Distracters (N = 10)													
1	Sharp	1	0	1	0	1	3	0	0	4	1	1	3
2	Small	8	6	4	5	3	4	8	8	6	7	3	4
3	Sharp	2	1	2	1	1	4	0	2	2	2	1	4
4	Small	4	2	3	2	2	1	4	1	0	2	2	1
5	Sharp	2	3	3	2	1	1	1	2	2	1	1	1
6	Small	2	2	0	2	2	1	1	3	0	2	2	0
Red Distracters (N = 10)													
1	Sharp	0	0	3	1	2	2	2	2	0	1	4	4
2	Small	7	6	8	8	4	6	6	8	9	9	5	5
3	Sharp	2	2	1	0	5	1	0	1	2	4	3	1
4	Small	3	0	2	3	0	2	1	0	1	2	1	3
5	Sharp	2	1	1	1	0	1	2	1	3	1	0	1
6	Small	1	3	1	3	1	0	3	2	1	5	0	0

Table C14

Number of errors of direction (increase vs. decrease)) on each graph for distracter size, orientation and color

Distracter Size:		Thin						Thick					
Distracter Orientation:		Consistent		Inconsistent		Flat		Consistent		Inconsistent		Flat	
Target Orientation:		Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
Graph	Mag												
		Black Distracters (N = 10)											
1	Sharp	1	2	0	1	0	0	1	1	1	0	0	0
2	Small	0	0	3	0	2	1	0	0	0	0	0	0
3	Sharp	1	0	1	1	1	1	1	0	1	0	0	0
4	Small	0	0	4	2	2	0	0	2	2	2	1	0
5	Sharp	0	1	1	1	0	1	0	0	1	0	1	0
6	Small	2	3	1	2	1	0	0	1	4	0	0	0
		Red Distracters (N = 10)											
1	Sharp	2	2	0	0	0	0	1	2	0	0	0	0
2	Small	0	0	0	0	1	0	0	0	0	0	0	0
3	Sharp	0	0	1	0	0	0	2	0	0	0	1	0
4	Small	2	1	1	2	0	0	1	0	1	0	0	1
5	Sharp	0	0	0	1	1	0	0	1	0	1	2	0
6	Small	3	0	5	2	0	1	1	1	5	0	1	0

Table C15

Number of errors of both (magnitude & direction) on each graph for distracter size, orientation and color

Distracter Size:		Thin						Thick					
Distracter Orientation:		Consistent		Inconsistent		Flat		Consistent		Inconsistent		Flat	
Target Orientation:		Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
Graph	Mag												
		Black Distracters (N = 10)											
1	Sharp	1	0	0	1	1	0	1	3	0	0	1	1
2	Small	0	1	1	0	0	0	1	0	1	0	0	0
3	Sharp	0	0	1	1	2	0	0	0	1	0	1	1
4	Small	0	3	1	1	0	0	0	2	2	1	0	0
5	Sharp	0	1	1	0	0	0	0	0	0	1	0	0
6	Small	0	0	1	2	0	1	2	0	0	1	0	0
		Red Distracters (N = 10)											
1	Sharp	2	2	0	0	1	0	3	2	0	0	1	0
2	Small	0	0	0	1	0	1	0	0	0	0	0	1
3	Sharp	0	1	0	0	0	0	0	0	0	0	0	0
4	Small	0	0	0	0	1	0	0	1	0	0	1	0
5	Sharp	0	0	0	1	0	0	0	0	0	1	0	0
6	Small	1	2	0	0	0	0	1	0	0	0	0	0

Table C16

T-test for illusory conjunctions (testing for errors of direction between red inconsistent and black inconsistent orientations) (n=12)

Mean	Std. Deviation	Std. Error Mean	Lower CI	Upper CI	*t*	*df*	p
-.50000	1.83402	.52944	-1.66528	.66528	-.944	11	.365

Correlations

Table C17

Correlations between performance on distracter conditions (e.g., red consistent thin distracters), Visual Patterns Test (VPT) and Group Embedded Figures Test (EFT)

	EFT	Red con thin	Red con thick	Red incon thin	Red incon thick	Red flat thin	Red flat thick	Black con thin	Black con thick	Black incon thin	Black incon thick	Black flat thin	Black flat thick
						Performance (N=10)							
VPT	.148	-.282	.346	.347	.273	-.385	.444	.372	.092	.038	.164	.474	.112
EFT	-	-.021	.079	.131	.034	.224	.188	-.214	.034	.394	.013	-.042	.184

Table C18

Correlations between response times on distracter conditions (e.g., red consistent thin distracters), Visual Patterns Test (VPT) and Group Embedded Figures Test (EFT)

	EFT	Red con thin	Red con thick	Red incon thin	Red incon thick	Red flat thin	Red flat thick	Black con thin	Blac kcon thick	Black incon thin	Black incon thick	Black flat thin	Black flat thick
						Response Times (N=10)							
VPT	.148	.133	.318	-.032	.367	.339	.071	.197	.316	.120	.711*	.307	-.059
EFT	-	.064	.104	-.217	-.140	-.079	-.370	-.096	-.167	-.616	-.206	.113	-.196

* indicates p < .05

Table C19

Correlations between certainty ratings on distracter conditions (e.g., red consistent thin distracters), Visual Patterns Test (VPT) and Group Embedded Figures Test (EFT)

	EFT	Red con thin	Red con thick	Red incon thin	Red incon thick	Red flat thin	Red flat thick	Black con thin	Black con thick	Black incon thin	Black incon thick	Black flat thin	Black flat thick
							Certainty (*N*=10)						
VPT	.148	-.314	-.342	-.167	-.418	-.466	-.008	-.216	-.328	-.389	-.179	-.174	-.202
EFT	-	-.087	-.034	.085	.054	.011	.136	-.056	-.123	.240	.165	.051	.076

Appendix D

This Appendix contains Experiment 2 data as all other materials used were provided before.

All performance, response time and certainty data are provided in this Appendix.

Performance data includes: descriptive statistics (Table D1), tests of normality (Table D2), ANOVA table (Table D3).

Response time data includes: descriptive statistics (Table D4), tests of normality (Table D5 & D6), ANOVA table (Table D7).

Certainty data includes: descriptive statistics (Table D8), tests of normality (Table D9), ANOVA table (Table D10).

Data used for the analysis of errors is presented in Tables D11, D12 and D13 and information for the *t*-test for illusory conjunctions is presented in Table D14.

Correlation tables are shown for relations between task and response time conditions and both the Visual Pattern Test and the Group Embedded Figures Test using: performance scores (Table D15), response times (Table D16) and certainty ratings (Table D17).

Performance Data

Table D1

Descriptive statistics for performance on distracter color and orientation

Descriptive Statistics					
Distracter Characteristics	N	Minimum	Maximum	Mean	Std. Deviation
Consistent Black	14	33.33	91.67	67.26	18.04
Consistent Red	14	25.00	75.00	57.15	14.93
Inconsistent Black	14	16.67	66.67	50.00	14.25
Inconsistent Red	14	33.34	91.67	65.48	18.45
Flat Black	14	33.33	83.34	63.10	15.58
Flat Red	14	25.00	83.34	58.93	19.46

Table D2

Tests of normality for performance on distracter color and orientation

Tests of Normality						
	Kolmogorov-Smirnov			Shapiro-Wilk		
Distracter Characteristics	Statistic	df	Sig.	Statistic	df	Sig.
Consistent Black	.237	14	.031	.887	14	.072
Consistent Red	.167	14	.200*	.926	14	.268
Inconsistent Black	.149	14	.200*	.915	14	.186
Inconsistent Red	.240	14	.028	.918	14	.204
Flat Black	.234	14	.037	.880	14	.058
Flat Red	.181	14	.200*	.910	14	.158

*Indicates a lower bound of the true significance

Table D3

Analysis of Variance for performance on distracter color and orientation

Source	*df*	*F*	η_p^2	*p*
Within Subjects				
Orientation (A)	2	.95	.002	.398
A within-group error	26	(149.51)		
Color (B)	1	.031	.068	.864
B within-group error	13	(108.36)		
A X B	2	9.13^{**}	.413	.001
A X B within-group error	26	(137.54)		

Response Time Data

Table D4

Descriptive statistics for response times (after adjustments) on distracter color and orientation

Descriptive Statistics					
Distracter Characteristics	N	Minimum	Maximum	Mean	Std. Deviation
Consistent Black	14	2906.75	8065.42	5137.56	1576.44
Consistent Red	14	2766.92	8547.67	5129.89	1782.52
Inconsistent Black	14	2802.50	8637.76	5380.19	1623.87
Inconsistent Red	14	2771.58	7990.50	4522.15	1463.17
Flat Black	14	3038.92	7273.58	4457.71	1333.52
Flat Red	14	2557.33	8184.82	4662.64	1727.69

Table D5

Tests of normality for response times on distracter color and orientation

Tests of Normality						
	Kolmogorov-Smirnov			Shapiro-Wilk		
Distracter Characteristics	Statistic	df	Sig.	Statistic	df	Sig.
Consistent Black	.232	14	.040	.787	14	.003
Consistent Red	.232	14	.039	.746	14	.001
Inconsistent Black	.158	14	.200*	.940	14	.417
Inconsistent Red	.169	14	.200*	.916	14	.190
Flat Black	.190	14	.185	.891	14	.083
Flat Red	.264	14	.009	.718	14	.001

*Indicates a lower bound of the true significance

Table D6

Tests of normality response times (after adjustments) on distracter color and orientation

Tests of Normality

Distracter Characteristics	Kolmogorov-Smirnov Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
Consistent Black	.154	14	.200*	.934	14	.345
Consistent Red	.143	14	.200*	.965	14	.808
Inconsistent Black	.158	14	.200*	.940	14	.417
Inconsistent Red	.169	14	.200*	.916	14	.190
Flat Black	.190	14	.185	.891	14	.083
Flat Red	.160	14	.200*	.928	14	.283

*Indicates a lower bound of the true significance

Table D7

Analysis of Variance for response times on distracter color and orientation

Source	*df*	*F*	η_p^2	*p*
Within Subjects				
Orientation (A)	2	2.35	.152	.116
A within-group error	26	(1023969.7)		
Color (B)	1	2.21	.145	.161
B within-group error	13	(461183.21)		
A X B	2	4.40*	.252	.023
A X B within-group error	26	(503554.00)		

Certainty Data

Table D8

Descriptive statistics for certainty on distracter color and orientation

Descriptive Statistics					
Distracter Characteristics	N	Minimum	Maximum	Mean	Std. Deviation
Consistent Black	14	10.00	100.00	64.69	23.60
Consistent Red	14	10.00	96.00	64.21	23.88
Inconsistent Black	14	10.00	95.00	59.00	25.07
Inconsistent Red	14	10.00	98.75	62.25	25.05
Flat Black	14	10.00	96.67	64.55	23.76
Flat Red	14	10.00	97.92	66.32	23.77

Table D9

Tests of normality for certainty on distracter color and orientation

Tests of Normality						
	Kolmogorov-Smirnov			Shapiro-Wilk		
Distracter Characteristics	Statistic	df	Sig.	Statistic	df	Sig.
Consistent Black	.102	14	.200*	.963	14	.765
Consistent Red	.161	14	.200*	.933	14	.333
Inconsistent Black	.145	14	.200*	.964	14	.780
Inconsistent Red	.146	14	.200*	.960	14	.726
Flat Black	.196	14	.150	.927	14	.276
Flat Red	.158	14	.200*	.939	14	.401

*Indicates a lower bound of the true significance

Table D10

Analysis of Variance for certainty on distracter color and orientation

Source	*df*	*F*	η_p^2	*p*
Within Subjects				
Orientation (A)	2	6.18**	.322	.006
A within-group error	26	(29.28)		
Color (B)	1	1.53	.105	.238
B within-group error	13	(31.40)		
A X B	2	.88	.064	.365
A X B within-group error	26	(27.93)		

Error Analysis

Table D11

Number of errors of magnitude (small vs. sharp slope) on distracter color and orientation

Distracter Color:		Black						Red					
Distracter Orientation:		Consistent		Inconsistent		Flat		Consistent		Inconsistent		Flat	
Target Orientation:		Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
Graph	Mag												
1	Sharp	2	0	5	4	5	3	0	0	1	1	0	1
2	Small	8	12	9	6	6	5	10	8	9	7	10	7
3	Sharp	2	1	2	5	4	9	1	13	5	0	3	8
4	Small	8	5	4	0	3	9	5	5	1	2	2	1
5	Sharp	4	2	2	2	6	1	4	0	4	4	7	0
6	Small	4	5	3	7	4	1	1	3	3	8	3	5

Table D12

Number of errors of direction (increase vs. decrease) on distracter color and orientation

Distracter Color:		Black						Red					
Distracter Orientation:		Consistent		Inconsistent		Flat		Consistent		Inconsistent		Flat	
Target Orientation:		Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
Graph	Mag												
1	Sharp	0	1	1	0	0	1	3	2	6	3	7	6
2	Small	0	1	1	0	2	2	0	2	1	2	0	0
3	Sharp	0	0	4	3	1	1	2	0	0	1	0	1
4	Small	1	2	0	4	1	1	4	3	2	3	2	1
5	Sharp	0	0	5	0	1	1	1	0	0	0	1	4
6	Small	2	3	2	3	3	2	3	2	2	2	2	2

Table D13

Number of errors of both (magnitude & direction) on distracter color and orientation

Distracter Color:		Black						Red					
Distracter Orientation:		Consistent		Inconsistent		Flat		Consistent		Inconsistent		Flat	
Target Orientation:		Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
Graph	Mag												
1	Sharp	1	2	2	2	1	1	2	3	0	1	1	0
2	Small	1	0	3	2	0	0	0	1	0	1	1	1
3	Sharp	1	2	2	2	1	1	2	0	0	2	1	2
4	Small	1	1	5	5	0	4	1	1	0	2	0	2
5	Sharp	0	0	3	1	1	1	1	2	0	1	1	0
6	Small	0	0	0	1	1	1	2	0	0	1	1	0

Table D14

T-test for illusory conjunctions (testing for errors of direction between red inconsistent and black inconsistent orientations)(n=12)

Mean	Std. Deviation	Std. Error Mean	Lower CI	Upper CI	*t*	*df*	p
.08333	2.84312	.82074	-1.72310	1.88977	.102	11	.921

Table D15

Correlations between performance on distracter conditions (e.g., red consistent distracters), Visual Patterns Test (VPT) and Group Embedded Figures Test (EFT)

	EFT	Black con	Red con	Black incon	Red incon	Black flat	Red flat
	Performance (*N*=14)						
VPT	.537*	.466	.462	.180	.529	.503	.109
EFT	1	.067	.426	.584*	.582*	.785**	.444

* indicates p < .05

Table D16

Correlations between response times on distracter conditions (e.g., red consistent distracters), Visual Patterns Test (VPT) and Group Embedded Figures Test (EFT)

	EFT	Black con	Red con	Black incon	Red incon	Black flat	Red flat
	Response Times (*N*=14)						
VPT	.537*	-.589*	-.513	-.222	-.427	-.535*	-.289
EFT	1	-.838**	-.659*	-.791**	-.818**	-.864**	-.761**

* indicates p < .05, ** indicates, p < .001

Table D17

Correlations between certainty ratings on distracter conditions (e.g., red consistent distracters), Visual Patterns Test (VPT) and Group Embedded Figures Test (EFT)

	EFT	Black con	Red con	Black incon	Red incon	Black flat	Red flat
	Certainty (*N*=14)						
VPT	.537*	.043	-.021	.189	.091	.019	-.016
EFT	1	.005	-.024	.081	-.015	.064	.026

* indicates p < .05

Appendix E

This Appendix contains Experiment 3 data as all other materials used were provided before.

All performance, response time and certainty data are provided in this Appendix.

Performance data includes: descriptive statistics (Table E1), tests of normality (Table E2), ANOVA table (Table E3).

Response time data includes: descriptive statistics (Table E4), tests of normality (Table E5), ANOVA table (Table E6).

Certainty data includes: descriptive statistics (Table E7), tests of normality (Table E8 & E9), ANOVA table (Table E10).

Data used for the analysis of errors is presented in Tables E11, E12 and E13 and information for the *t*-test for illusory conjunctions is presented in Table E14.

Correlation tables are shown for relations between task and response time conditions and both the Visual Pattern Test and the Group Embedded Figures Test using: performance scores (Table E15), response times (Table E16) and certainty ratings (Table E17).

Performance Data

Table E1

Descriptive statistics for performance on distracter size and orientation

Descriptive Statistics					
Distracter Characteristics	N	Minimum	Maximum	Mean	Std. Deviation
Consistent thin	12	41.67	83.34	68.75	13.82
Consistent thick	12	25.00	83.34	63.89	16.41
Inconsistent thin	12	41.67	91.67	64.59	16.71
Inconsistent thick	12	33.34	91.67	60.42	16.71
Flat thin	12	41.67	83.33	63.20	16.07
Flat thick	12	33.34	91.67	65.98	14.41

Table E2

Tests of normality for performance on distracter size and orientation

Tests of Normality						
	Kolmogorov-Smirnov			Shapiro-Wilk		
Distracter Characteristics	Statistic	df	Sig.	Statistic	df	Sig.
Consistent thin	.258	12	.027	.888	12	.109
Consistent thick	.234	12	.069	.896	12	.142
Inconsistent thin	.229	12	.082	.922	12	.306
Inconsistent thick	.146	12	.200*	.967	12	.875
Flat thin	.185	12	.200*	.880	12	.087
Flat thick	.269	12	.016	.905	12	.183

*Indicates a lower bound of the true significance

Table E3

Analysis of Variance for performance on distracter orientation and size

Source	*df*	*F*	η_p^2	*p*
Within Subjects				
Orientation (A)	2	0.52	.044	.604
A within-group error	22	(169.83)		
Size (B)	1	0.84	.071	.348
B within-group error	11	(92.83)		
A X B	2	1.11	.091	.348
A X B within-group error	22	(96.52)		

Response Time Data

Table E4

Descriptive statistics for response times on distracter size and orientation

Descriptive Statistics					
Distracter Characteristics	N	Minimum	Maximum	Mean	Std. Deviation
Consistent thin	12	3008.08	7126.92	4753.35	1184.86
Consistent thick	12	2489.50	5903.00	4505.01	1168.76
Inconsistent thin	12	2436.75	7458.25	4320.70	1372.27
Inconsistent thick	12	2549.00	6717.17	4351.83	1297.14
Flat thin	12	2410.67	5760.75	3976.38	1083.28
Flat thick	12	2736.67	5943.17	4040.75	984.26

Table E5

Tests of normality for response times on distracter size and orientation

Tests of Normality						
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
Distracter Characteristics	Statistic	df	Sig.	Statistic	df	Sig.
Consistent thin	.145	12	.200*	.971	12	.924
Consistent thick	.210	12	.150	.916	12	.258
Inconsistent thin	.170	12	.200*	.925	12	.334
Inconsistent thick	.178	12	.200*	.930	12	.375
Flat thin	.129	12	.200*	.946	12	.583
Flat thick	.153	12	.200*	.955	12	.713

*Indicates a lower bound of the true significance

Table E6

Analysis of Variance for response times on distracter size and orientation

Source	*df*	*F*	η_p^2	*p*
Within Subjects				
Orientation (A)	2	4.62^{*}	.295	.021
A within-group error	22	(500879.82)		
Size (B)	1	.07	.0066	.791
B within-group error	11	(630855.27)		
A X B	2	0.25	.022	.781
A X B within-group error	22	(706335.65)		

Certainty Data

Table E7

Descriptive statistics for certainty (after adjustments) on distracter size and orientation

Descriptive Statistics					
Distracter Characteristics	N	Minimum	Maximum	Mean	Std. Deviation
Consistent thin	12	24.33	85.42	64.94	19.12
Consistent thick	12	33.00	85.00	64.84	18.48
Inconsistent thin	12	31.00	76.00	61.87	17.89
Inconsistent thick	12	35.83	78.33	62.56	15.93
Flat thin	12	32.17	86.67	63.97	19.33
Flat thick	12	25.83	84.92	65.48	20.90

Table E8

Tests of normality for certainty on distracter size and orientation

Tests of Normality						
	Kolmogorov-Smirnov			Shapiro-Wilk		
Distracter Characteristics	Statistic	df	Sig.	Statistic	df	Sig.
Consistent thin	.319	12	.001	.782	12	.006
Consistent thick	.345	12	.000	.800	12	.009
Inconsistent thin	.295	12	.005	.742	12	.002
Inconsistent thick	.319	12	.001	.747	12	.003
Flat thin	.306	12	.003	.827	12	.019
Flat thick	.274	12	.013	.781	12	.006

*Indicates a lower bound of the true significance

Table E9

Test of normality for certainty (after adjustments) on distracter size and orientation

Tests of Normality

	Kolmogorov-Smirnov			Shapiro-Wilk		
Distracter Characteristics	Statistic	df	Sig.	Statistic	df	Sig.
Consistent thin	.286	12	.008	.823	12	.017
Consistent thick	.318	12	.001	.816	12	.014
Inconsistent thin	.259	12	.026	.743	12	.002
Inconsistent thick	.288	12	.007	.757	12	.003
Flat thin	.270	12	.016	.837	12	.026
Flat thick	.255	12	.030	.800	12	.009

*Indicates a lower bound of the true significance

Table E10

Analysis of Variance for certainty on distracter size and orientation

Source	*df*	*F*	η_p^2	*p*
Within Subjects				
Orientation (A)	2	1.74	.136	.199
A within-group error	22	(30.88)		
Size (B)	1	0.29	.025	.604
B within-group error	11	(31.02)		
A X B	2	0.19	.0166	.832
A X B within-group error	22	(20.99)		

Error Analysis

Table E11

Number of errors of magnitude (small vs. sharp slope) on distracter size and orientation

Distracter Size:		Thin						Thick					
Distracter Orientation:		Consistent		Inconsistent		Flat		Consistent		Inconsistent		Flat	
Target Orientation:		Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
Graph	Mag												
1	Sharp	1	0	2	2	3	5	0	1	3	1	1	1
2	Small	8	6	8	5	4	5	5	6	9	8	5	4
3	Sharp	1	1	4	3	3	2	2	0	4	3	3	3
4	Small	0	0	5	3	1	2	2	2	5	5	1	3
5	Sharp	2	2	2	3	1	1	2	3	1	0	2	2
6	Small	3	5	0	0	0	2	1	3	3	6	1	0

Table E12

Number of errors of direction (increase vs. decrease) on distracter size and orientation

Distracter Size:		Thin						Thick					
Distracter Orientation:		Consistent		Inconsistent		Flat		Consistent		Inconsistent		Flat	
Target Orientation:		Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
Graph	Mag												
1	Sharp	0	2	1	0	1	0	2	2	0	0	0	0
2	Small	0	0	1	3	1	1	0	0	0	0	0	0
3	Sharp	0	0	0	0	0	0	0	0	0	2	0	1
4	Small	1	3	0	1	1	0	3	1	1	0	0	1
5	Sharp	0	1	1	0	1	0	0	0	0	1	0	1
6	Small	3	1	2	4	1	1	4	0	0	1	1	2

Table E13

Number of errors of both (magnitude & direction) on distracter size and orientation

Distracter Size:		Thin						Thick					
Distracter Orientation:		Consistent		Inconsistent		Flat		Consistent		Inconsistent		Flat	
Target Orientation:		Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
Graph	Mag												
1	Sharp	2	1	0	0	1	0	2	2	0	0	0	0
2	Small	0	0	1	0	1	1	0	2	0	1	0	0
3	Sharp	0	0	0	1	0	0	1	1	0	0	0	1
4	Small	2	0	0	1	1	0	1	0	1	2	0	1
5	Sharp	0	0	0	1	1	0	1	1	1	0	0	1
6	Small	0	0	0	0	1	1	1	2	1	0	1	2

Table E14

T-test for illusory conjunctions (testing for errors of direction on red inconsistent compared to black inconsistent orientations)

Mean	Std. Deviation	Std. Error Mean	Lower CI	Upper CI	*t*	*df*	p
.66667	1.55700	.44947	-.32260	1.65594	1.483	11	.166

Correlations

Table E15

Correlations between performance on distracter conditions (e.g., thin consistent distracters), Visual Patterns Test (VPT) and Group Embedded Figures Test (EFT)

	EFT	Thin con	Thick con	Thin incon	Thick incon	Thin flat	Thick flat
Performance (N=12)							
VPT	.558	.046	-.207	.394	.470	.119	.162
EFT	1	.440	.197	.509	.594*	.164	.463

* indicates p < .05

Table E16

Correlations between response times on distracter conditions (e.g., thin consistent distracters), Visual Patterns Test (VPT) and Group Embedded Figures Test (EFT)

	EFT	Thin con	Thick con	Thin incon	Thick incon	Thin flat	Thick flat
Response Times (N=12)							
VPT	.558	-.294	-.719**	-.549	-.584*	-.683*	-.708*
EFT	1	-.348	-.331	-.255	-.304	-.535	-.618*

* indicates p < .05, indicates p < .01

Table E17

Correlations between certainty ratings on distracter conditions (e.g., thin consistent distracters), Visual Patterns Test (VPT) and Group Embedded Figures Test (EFT)

	EFT	Thin con	Thick con	Thin incon	Thick incon	Thin flat	Thick flat
				Certainty (N=12)			
VPT	.558	.412	.232	.417	.370	.458	.305
EFT	1	.551	.337	.561	.515	.621*	.461

* indicates p < .05

Appendix F

This Appendix contains Experiment 4 data as all other materials used were provided before.

All performance, response time and certainty data are provided in this Appendix.

Performance data includes: descriptive statistics (Table F1), tests of normality (Table F2), ANOVA table (Table F3).

Response time data includes: descriptive statistics (Table F4), tests of normality (Table F5), ANOVA table (Table F6).

Certainty data includes: descriptive statistics (Table F7), tests of normality (Table F8), ANOVA table (Table F9).

Data used for the analysis of errors is presented in Tables F10, F11 and F12.

Correlation tables are shown for relations between task and response time conditions and both the Visual Pattern Test and the Group Embedded Figures Test using: performance scores (Table F13), response times (Table F14) and certainty ratings (Table F15).

Performance Data

Table F1

Descriptive statistics for performance on distracter size and orientation

Distracter Characteristics	N	Minimum	Maximum	Mean	Std. Deviation
Consistent thin	12	8.34	100.00	53.48	24.99
Consistent thick	12	25.00	100.00	54.86	20.55
Inconsistent thin	12	.00	91.67	40.97	30.25
Inconsistent thick	12	.00	91.67	46.53	29.61
Flat thin	12	.00	100.00	59.73	30.32
Flat thick	12	.00	91.67	61.11	24.70

Table F2

Tests of normality for performance on distracter size and orientation

	Kolmogorov-Smirnov			Shapiro-Wilk		
Distracter Characteristics	Statistic	df	Sig.	Statistic	df	Sig.
Consistent thin	.195	12	.200*	.951	12	.650
Consistent thick	.260	12	.024	.906	12	.188
Inconsistent thin	.157	12	.200*	.939	12	.488
Inconsistent thick	.120	12	.200*	.954	12	.699
Flat thin	.198	12	.200*	.933	12	.413
Flat thick	.160	12	.200*	.898	12	.147

*Indicates a lower bound of the true significance

Table F3

Analysis of Variance for performance on distracter size and orientation

Source	*df*	*F*	η_p^2	*p*
Within Subjects				
Orientation (A)	2	3.96*	.264	.034
A within-group error	22	(429.27)		
Size (B)	1	.90	.075	.362
B within-group error	11	(153.60)		
A X B	2	0.25	.021	.784
A X B within-group error	22	(140.99)		

Response Time Data

Table F4

Descriptive statistics for response times on distracter size and orientation

Descriptive Statistics					
Distracter Characteristics	N	Minimum	Maximum	Mean	Std. Deviation
Consistent thin	12	2910.00	5210.50	4060.35	895.95
Consistent thick	12	2510.42	5415.08	3909.32	1094.89
Inconsistent thin	12	2976.58	6156.83	4058.92	894.49
Inconsistent thick	12	2756.33	6511.00	4331.97	1246.54
Flat thin	12	2470.67	5642.33	4022.59	1071.98
Flat thick	12	3391.58	6388.58	4370.94	824.66

Table F5

Tests of normality for response times on distracter size and orientation

Tests of Normality						
	Kolmogorov-Smirnov			Shapiro-Wilk		
Distracter Characteristics	Statistic	df	Sig.	Statistic	df	Sig.
Consistent thin	.188	12	.200*	.881	12	.089
Consistent thick	.152	12	.200*	.905	12	.184
Inconsistent thin	.171	12	.200*	.919	12	.276
Inconsistent thick	.161	12	.200*	.938	12	.475
Flat thin	.120	12	.200*	.950	12	.643
Flat thick	.217	12	.123	.884	12	.098

*Indicates a lower bound of the true significance

Table F6

Analysis of Variance for response times on distracter size and orientation

Source	*df*	*F*	η_p^2	*p*
Within Subjects				
Orientation (A)	2	1.04	.086	.369
A within-group error	22	(342351.32)		
Size (B)	1	1.04	.086	.330
B within-group error	11	(425100.44)		
A X B	2	0.90	.075	.422
A X B within-group error	22	(484871.60)		

Certainty Data

Table F7

Descriptive statistics for certainty

Descriptive Statistics					
Distracter Characteristics	N	Minimum	Maximum	Mean	Std. Deviation
Consistent thin	12	25.08	97.25	66.56	18.78
Consistent thick	12	27.50	96.67	70.47	19.01
Inconsistent thin	12	23.33	86.58	62.71	16.60
Inconsistent thick	12	27.50	89.17	67.02	15.76
Flat thin	12	24.17	95.50	68.12	18.51
Flat thick	12	30.00	96.25	66.31	17.78

Table F8

Tests of normality for certainty on distracter size and orientation

Tests of Normality						
	Kolmogorov-Smirnov			Shapiro-Wilk		
Distracter Characteristics	Statistic	df	Sig.	Statistic	df	Sig.
Consistent thin	.139	12	.200*	.963	12	.832
Consistent thick	.237	12	.062	.917	12	.264
Inconsistent thin	.170	12	.200*	.931	12	.390
Inconsistent thick	.188	12	.200*	.892	12	.124
Flat thin	.255	12	.031	.908	12	.200
Flat thick	.147	12	.200*	.967	12	.875

*Indicates a lower bound of the true significance

Table F9

Analysis of Variance for certainty on distracter size and orientation

Source	*df*	*F*	η_p^2	*p*
Within Subjects				
Orientation (A)	2	1.77	.138	.194
A within-group error	22	(46.56)		
Size (B)	1	1.75	.137	.213
B within-group error	11	(47.14)		
A X B	2	2.27	.170	.127
A X B within-group error	22	(31.04)		

Error Analysis

Table F10

Number of errors of magnitude (small vs. sharp slope) on distracter size and orientation

Distracter Size:		Thin						Thick					
Distracter Orientation:		Consistent		Inconsistent		Flat		Consistent		Inconsistent		Flat	
Target Orientation:		Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
Graph	Mag												
1	Sharp	3	1	4	0	3	5	2	1	7	2	4	1
2	Small	9	7	3	6	3	3	8	8	5	8	3	3
3	Sharp	3	1	2	2	5	3	2	2	4	1	5	4
4	Small	7	6	1	2	2	4	3	5	4	3	2	0
5	Sharp	3	2	1	1	0	3	2	2	3	2	0	3
6	Small	5	5	6	2	0	0	3	5	3	3	0	1

Table F11

Number of errors of direction (increase vs. decrease) on distracter size and orientation

Distracter Size:		Thin						Thick					
Distracter Orientation:		Consistent		Inconsistent		Flat		Consistent		Inconsistent		Flat	
Target Orientation:		Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
Graph	Mag												
1	Sharp	1	0	0	1	1	0	1	5	0	2	0	0
2	Small	1	1	4	1	2	2	2	0	1	1	1	1
3	Sharp	0	0	1	0	0	0	1	0	3	1	1	2
4	Small	1	1	1	1	3	3	2	2	2	1	2	1
5	Sharp	0	0	4	3	2	0	1	1	3	2	3	0
6	Small	2	1	0	1	4	2	1	0	1	0	2	4

Table F12

Number of errors of both (magnitude & direction) on distracter size and orientation

Distracter Size:		Thin						Thick					
Distracter Orientation:		Consistent		Inconsistent		Flat		Consistent		Inconsistent		Flat	
Target Orientation:		Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
Graph	Mag												
1	Sharp	1	2	1	3	1	1	1	0	0	1	2	2
2	Small	0	1	1	1	0	0	0	1	2	1	0	1
3	Sharp	0	0	3	2	1	2	0	2	0	1	1	2
4	Small	1	1	6	7	1	0	1	1	4	3	0	1
5	Sharp	0	0	1	1	2	2	0	0	2	3	1	0
6	Small	0	1	0	2	0	1	1	2	1	5	0	0

Correlations

Table F13

Correlations between performance on distracter conditions (e.g., thin consistent distracters), Visual Patterns Test (VPT) and Group Embedded Figures Test (EFT)

	EFT	Thin con	Thick con	Thin incon	Thick incon	Thin flat	Thick flat
	Performance (N=12)						
VPT	.694*	.025	.155	.414	.179	.392	.069
EFT	1	.262	.380	.661*	.370	.271	.021

* indicates p < .05

Table F14

Correlations between response times on distracter conditions (e.g., thin consistent distracters), Visual Patterns Test (VPT) and Group Embedded Figures Test (EFT)

	EFT	Thin con	Thick con	Thin incon	Thick incon	Thin flat	Thick flat
	Response Times (N=12)						
VPT	.694*	-.342	.072	.271	-.495	-.051	.267
EFT	1	-.350	-.153	.283	-.491	-.085	.166

* indicates p < .05

Table F15
Correlations between response times on distracter conditions (e.g., thin consistent distracters), Visual Patterns Test (VPT) and Group Embedded Figures Test (EFT)

	EFT	Thin con	Thick con	Thin incon	Thick incon	Thin flat	Thick flat
	Certainty (N=12)						
VPT	.694*	.293	.455	.323	.485	.457	.337
EFT	1	.457	.519	.532	.614*	.458	.535

* indicates p < .05