

**EXPERT AND NOVICE STRATEGIES IN THE EXPLORATION OF
DATA IN LINE GRAPHS**

by

Bruce Hidekazu Tsuji
MA (University of Waterloo, 1979), BSc (Trend University, 1977)

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

©2011 Bruce Hidekazu Tsuji



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 978-0-494-81558-8
Our file *Notre référence*
ISBN: 978-0-494-81558-8

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

Abstract

In three experiments domain experts from business and psychology were compared with domain novices in the strategies they utilized to understand the data in line graphs. Experiment 1 compared experts and novices in their ratings of the visual and semantic complexity of a series of graphs. This resulted in a set of graphs distributed along a complexity continuum. Experiment 2 required experts and novices to visually describe and then attempt to explain graphs that were drawn from their own- (business or psychology) or from the other domain. The results indicated an expertise effect which contradicts previous research in graph exploration (Shah & Carpenter, 1995). The nature of the expert/novice differences included the strategies with which experts and novices approached the graphs: their visual descriptions, the completeness of their explanations, and the order in which they reported different elements of the graphs. Relying on the results of Experiment 2, Experiment 3 investigated the degree to which it might be possible to teach novices expert strategies by giving them a very brief instructional session. Explanations generated before- and after instruction were compared to assess the effects of the instruction. The results showed that novices adopted at least some of the expert strategies uncovered in Experiment 2. The results are discussed in relation to three theories, namely Pinker's (1990), graph comprehension model, Lemaire and Siegler's (1995) development of strategy use, and Feldon, Timmerman, Stowe, and Showman's (2010) instructional methods for teaching science, technology, engineering, and mathematics.

Acknowledgements

First, to my wife and long-suffering partner Susan Mary Woollam. Thank you for your support and patience—I really didn't think it would take this long! Second, to Gitte Lindgaard for your intellectual guidance and many many re-reads, re-edits, and re-treads. Third, I would like to thank the members of my examination committee; Thomas Hewett, Tony Bailetti, Jo-Anne Lefevre, and Avi Parush. Fourth, I pay tribute to my venerable Japanese ancestors and especially my mother Aiko Margaret, and my father James Hideaki for life itself and for getting me my first microscope at the age of five (amongst many other things). Fifth, I want to acknowledge and thank those who prodded me, pushed me, or pointed the way. Some of those people include Robert Scott, Carolyn McMillan, Ian McMillan, The Microbe Hunters, James Loughton, Joachim Voss, Doug Lowe, Alan Worthington, Keith Crockatt, Lawrence Weissmann, Tom Fennell, Carlyle T. Smith, Paul Wong, Sandra Thwaites, Phil Merikle, Lyle MacDonald, Phil Bryden, Neil Charness, Herbert Simon, Fran Allard, Dick Dillon, Jo Wood, Lise Paquet, John Clegg, Paul Guild, John Tyson, Artemis March, Coralie Lalonde, Brad Furlott, Joan Norris, Randy Norris, Robert Biddle, Tim Pynchyl, Anne Bowker, Jing Liu, Kate Oakley, Lena and Larry Jones, Suzanne White, Bob Minns, James O'Regan, Lois Frankel, Steven Murphy, Richard Hoptroff, June Drayton-Callender, Etelle Bourassa, and Janet Tsuji. (Apparently, it takes a village to raise a PhD.) Sixth, I would like to thank my cohort, Cathy Dudek, Judy Brown, and stats buddies Michelle Gauthier and Cathy MacMillan for their friendship and support. Seventh, to my son Thomas for the motivation to be a good example, not a bad one. Finally, (and again) Susan. This giraffe's for you Cubby Sue! Love you.

Table of Contents

Expert and Novice Strategies in the Exploration of Data in Line Graphs.....	i
<i>Abstract</i>	<i>ii</i>
<i>Acknowledgements</i>	<i>iii</i>
<i>Table of Contents</i>	<i>iv</i>
<i>List of Figures</i>	<i>vii</i>
<i>List of Tables</i>	<i>viii</i>
<i>List of Appendices</i>	<i>ix</i>
Expert and Novice Strategies in the Exploration of Data in Line Graphs	1
<i>Outline of the thesis</i>	5
<i>Graphs Defined</i>	6
<i>Graphical Task Analysis</i>	7
<i>Graph Exploration Theoretical Framework</i>	9
<i>Top-Down Influences On Graph Comprehension</i>	18
<i>Graph Comprehension Theories Reconciled</i>	22
<i>Expertise</i>	23
<i>Expert Problem-Solving Strategies</i>	28
<i>A Model Of Strategy And Expertise</i>	36
<i>Summary of Preliminary Studies</i>	37
Experiment 1: Judging Visual versus Semantic Graph Complexity.....	42
<i>Method</i>	43
<i>Participants</i>	43
<i>Apparatus and Materials</i>	43
<i>Design and Procedure</i>	44
<i>Results and Discussion</i>	46
<i>Complexity Ratings and Response Times</i>	46
<i>Summary</i>	48
<i>Graph Stimulus Selection</i>	49
Experiment 2: Graph Exploration By Domain Experts and Novices	50
<i>Method</i>	52
<i>Participants</i>	52

Expert/Novice Strategies in Graph Exploration

<i>Apparatus and Materials</i>	53
<i>Design and Procedure</i>	54
<i>Data Analysis</i>	55
Results	55
<i>Visual Descriptions of Graphs</i>	56
<i>Semantic Explanations of Graphs</i>	56
<i>Serial Order in Explanations of Graphs</i>	60
<i>Proportion of Explanations</i>	63
<i>Familiar and Unfamiliar Domain Explanations</i>	64
<i>Nominal, Ordinal, and Metric Conceptual Messages</i>	65
<i>Time Measures</i>	66
<i>Visual Patterns Test</i>	69
<i>Results Summary</i>	69
<i>Discussion</i>	71
Experiment 3: Graph Exploration Instruction For Novices	73
<i>Method</i>	74
<i>Participants</i>	74
<i>Apparatus and Materials</i>	74
<i>Design and Procedure</i>	75
<i>Results and Discussion</i>	76
<i>Visual Descriptions of Graphs</i>	76
<i>Semantic Explanations of Graphs</i>	76
<i>Serial Order of Themes</i>	79
<i>Nominal, Ordinal, and Metric Conceptual Messages</i>	81
<i>Time Measures</i>	81
<i>Summary</i>	81
General Discussion	82
<i>Future Research</i>	86
<i>Contributions of the Research</i>	88
Conclusions	89
References	91

Appendices 106

List of Figures

Figure 1. Pinker’s (1990) graph comprehension model 15

Figure 2. Preliminary Study 1, example graph stimulus 38

Figure 3. Preliminary Study 2, example graph stimulus 39

Figure 4. Experiment 1 trial procedure. On Visual Criterion trials the mask was displayed at 150 ms SOA. On Semantic Criterion trials the mask was displayed at 3000 ms SOA. 46

Figure 5. Experiment 1, expertise x criterion for mean complexity ratings. Error bars indicate 95% confidence interval in this and all subsequent graphs. 47

Figure 6. Experiment 1 response times. 48

Figure 7. Example graph stimuli: business domain (top) and psychology domain (bottom)..... 53

Figure 8. Word count for visual description task for novices, business-, and psychology experts 56

Figure 9. Mean Completeness Scores for Expertise and Graph Domain 60

Figure 10 Proportion of explanations provided by novices, business-, and psychology experts. 64

Figure 11. . Business- and psychology expert mean completeness scores as a function of same- versus different graph domain. 65

Figure 12. Mean conceptual messages generated by novices, business, and psychology experts. 66

Figure 13. Mean silent time by Expertise (novices, business-, and psychology experts) x Domain (business, psychology). 68

Figure 14. A hybrid graph comprehension model 84

List of Tables

Table 1 Graphical task analysis: What people do with graphs and ask of graphs	9
Table 2 Preliminary Study 3, Mean attributes of the simple/complex, visual/semantic graphs selected from Preliminary Study 2.....	41
Table 3 Examples and Mean Proportion of Themes Uttered by Novices, Business Experts, and Psychology Experts.....	58
Table 4 Mean Frequency of Step 1 Themes by Novices, Business-, and Psychology Experts	63
Table 5 Experiment 2 Summary of Results	69
Table 6 Experiment 2 Hypotheses, Results, and Conclusions.....	70
Table 7 Experiment 3 before- and after-instruction means. Experiment 2 novice and expert means are included for comparison purposes.	77
Table 8 Most common themes before- and after instruction by Step. Each theme listed was the most common observed for each step. Values in parentheses are proportion of instances each theme was observed in each step.....	80
Table 9 Mean proportion of initial (Step 1) graph explanation themes before- and after- instruction	80

List of Appendices

Appendix A. Preliminary Study 1	106
Appendix B. Preliminary Study 2	120
Appendix C. Preliminary Study 3	132
Appendix D. Experiment 1 Instructions, Stimuli, Descriptive Statistics, ANOVA Tables.....	140
Appendix E. Experiment 2 Instructions, Stimuli, Descriptive Statistics, ANOVA Tables	150
Appendix F. Experiment 3 Instructions, Stimuli, Descriptive Statistics, ANOVA Tables	176

Expert and Novice Strategies in the Exploration of Data in Line Graphs

Graphs are everywhere. They are found in textbooks, television, newspapers, magazines, journal articles in the sciences, social sciences, arts, and humanities, financial reports, government reports, on the internet, before the law courts, and in theses. Roth, Bowen, and McGinn (1999) found 1.5 graphs per page in five highly ranked ecology research journals and a similar number in six high school textbooks on biology. Before the advent of the internet, Tufte (1983) estimated that between 900 billion and 2 trillion statistical graphs were printed annually. Certainly the growth of the internet combined with the development of tools to assist in graph creation (see Heer, Bostock, & Ogvietzky, 2010; Kimball, 1996; Neil, 2009) implies that the number must be much larger now. Whenever a politician, CEO, marketer, researcher, or author wants to communicate or persuade an audience of something (Huff, 1954) graphs are often utilized. Whether the graph aids or hinders that communication (Larkin & Simon, 1987) is immaterial for the moment; the point is that graphs are a key artifact in our communications repertoire.

A special case of graph communication occurs when the communicant is oneself. This is sometimes known as the *analytical* (Tufte, 1983) or *exploratory* (Behrens, 1997; du Toit, Steyn, & Stumpf, 1986; Tukey, 1977) use of graphs where the purpose is interpretation or understanding the underlying data by an individual who is often the graph's creator. While a common application of graph exploration is in scientific reasoning, graphs intended for exploratory purposes are found in many other domains including virtually all aspects of business (e.g., Bertin, 1983; Stutely, 1993) and psychology (Cleveland, 1984, 1987; Postigo & Pozo, 2004).

This thesis is concerned with the use of graphs by novices and experts in familiar and unfamiliar domains for analytical or exploratory purposes. In particular, the pragmatic goals of the thesis are to:

1. Identify performance and strategy differences (if any) between novices and experts in their exploration of simple and more complex graphs drawn from familiar and unfamiliar domains.
2. Determine if brief instructions about graph exploration strategies might make novices perform more like experts when they are analyzing graphs.

The theoretical goal of the thesis is to contribute to the theory of graph comprehension.

Confronted with a graph that requires domain experts and novices to utilize complex inferential processes, a number of interesting psychological questions arise: Do domain experts apply qualitatively different strategies than novices? For example, do domain experts compare a particular graph to archetypal graphs in Long Term Memory (LTM) representing analogous domains (e.g., Gick & Holyoak, 1980, 1983), while novices avoid such LTM comparisons altogether? Do novices focus on the graph's surface or syntactic structure at the expense of an analysis of the deeper structural or semantic components (e.g., Chi, Feltovich, & Glaser, 1981)? For example, Preece and Janvier's (1993) teenaged novices often described the purely physical, graphical content of graphs without integrating it into the context. Alternatively, do domain experts simply execute their strategies more efficiently (Lemaire & Siegler, 1995) than novices? Alternatively, it may be that domain experts are able to quickly recognize patterns in graphs in ways that may be similar to how expert chess players recognize a board as a Queen's Gambit (e.g., Newell and Simon, 1972). Or, drawing from a common pool of strategies, are experts more adept at selecting the most effective strategy for the circumstance at hand? For example,

perhaps domain experts run “mental simulations” of situations that might lead to the data depicted in a particular graph that novices might not think of doing without prompting (e.g., Schraagen, 1993). More generally, what differences do experts and novices exhibit in graph exploration? Perhaps the difference is simply that experts discriminate visual features of the graphs better than novices in the same way that chicken sexers see the differences between males and females in day-old chicks (Biederman & Shiffrar, 1987).

Relatively few studies have addressed the issue of expertise in graph exploration. One instance is the ethnographic research of Roth and Bowen (2003) who examined how domain experts in biology, physics, and forest sciences interpreted familiar and unfamiliar graphs. Roth and Bowen found that experts had significant difficulty interpreting graphs taken from undergraduate textbooks from their respective domains but they had little difficulty with the familiar graphs taken from their own personal research. Roth and Bowen ascribed these differences to the semiotic properties of the personally familiar graphs in which those graphs became “transparent” and came to signify a great deal of domain- and contextual knowledge. Instead of referring to the content of their graphs, Roth and Bowen’s expert participants often focused on the experiential context of the studies from which the graph was created. These issues are discussed in more depth in the section devoted to Expertise.

Basing their research in a different domain, Trafton and colleagues (Trafton, Marshall, Mintz, & Trickett, 2002) described how expert meteorologists create spatial transformations of meteorological data when the information requested of them is not explicitly present on complex graphs. For example, in determining the air pressure over Pittsburgh, Trafton et al.’s eye movement data suggested that their participants were identifying nearby isobars, calculating the

distance between them, and then using the proportional distance to calculate the atmospheric pressure.

However, neither the Roth and Bowen nor the Trafton et al. studies directly compared experts against novices in their respective domains. Thus, it is difficult to know if the strategies inferred by these authors were attributable to expertise per se or were idiosyncratic to the domains selected. For example, it is difficult to imagine how one would perform the meteorological tasks set forth by Trafton et al. without the sort of spatial transformations they describe, although it is entirely possible that other strategies, less accessible to observers, are employed. Calculating the closest isobar to the city of Pittsburgh is not a necessary or sufficient attribute of expertise!

Freedman and Shah (2001, cited in Freedman & Shah, 2002) conducted one of the few studies in graph comprehension to explicitly compare experts with novices. Freedman and Shah used cognitive studies of aging; their experts were psychology graduate students and their novices, undergraduates. They also manipulated the domain by utilizing graphs concerned with non-cognitive aging data. Freedman and Shah reported that their novices tended to describe main effects while their experts were more likely to describe the underlying mathematical functions in the graph stimuli. However, the domain manipulation had no effect. Freedman and Shah interpreted these results as supporting the notion that novices attend to lower-level perceptual features of a graph whereas experts enrich and elaborate the visual features of a graph with their domain knowledge. In other words, experts might infer the underlying mathematical function from the lines describing a set of data. However, Freedman and Shah were unable to explain the fact that their experts performed equally well on cognitive and non-cognitive graphs. Their graph stimuli were extremely simple, for example, memory for digit span increasing from

4 to 7 between the ages of 2 and 20 years (Freedman & Shah, 2001, cited in by Freedman & Shah, 2002). It is therefore possible that their findings represented a ceiling effect, hence failing to show any performance differences between domains. Thus it is unclear whether domain-specific knowledge was really exerting an effect. This thesis systematically varied graph complexity to avoid ceiling- or floor effects in the graph exploration tasks employed. Another possible explanation for the lack of participant-group differences in Freedman and Shah's research is the dilution of experimental power because their 'experts' were all graduate students. This thesis attempted to maximize differences in expertise by recruiting faculty members as well as senior graduate students as experts and first-year, first semester undergraduate student novices in Experiment 2. Like Freedman and Shah, the present thesis experiments manipulated the domain to be either consistent or inconsistent with an individual participant's expertise.

While performance differences between experts and novices have been demonstrated in different domains the strategies employed have received somewhat less attention. In particular the focus of this thesis is to contrast expert and novice graph exploration strategies.

Outline of the Thesis

The next section explores the topic of graphs in detail including a definition of graphs, a task analysis of graph usage, and a review of theories of graph comprehension. Following that discussion is a focus on the concept of expertise in general, and problem-solving strategies in particular. Next, three Preliminary Studies are summarized as an introduction to Experiment 1, which identified a set of simple and complex graphs that differed on visual and semantic dimensions. Experiment 2 utilized those graphs to examine the strategies of experts and novices while graph domain and graph complexity were manipulated. Experiment 3 tested the notion, arising from Experiment 2, that graph expertise might be facilitated through some simple

instructions. Finally, the General Discussion summarizes the empirical results of the thesis in the context of the theories of graph comprehension and expertise introduced here and it outlines future work before presenting the concluding remarks.

Graphs Defined

Pinker (1990) defined graphs as a set of visual objects whose component dimensions correlate with values on associated mathematical scales. More precisely, he conceived of graphs as a set of n -tuples consisting of visual objects whose physical dimensions correlate with a set of mathematical scales. These are the abscissa and ordinate of conventional line graphs. A somewhat more accessible definition asserts that graphs "...visually display[s] measured quantities by means of combined use of points, lines, a coordinate system, numbers, symbols, words, shading, and color" (Tufté, 1983, p. 9).

Throughout this thesis, the parts of a graph will be identified according to Kosslyn's (1989) classification in which a typical graph contains four constituent parts. The *background* is the backdrop against which the other parts of the graph are displayed. The *framework* includes the axes and (optionally) tick marks or pips and/or grid lines. The *specifier* is the "working" portion of the graph and usually consists of visual objects such as lines, points, curves, bars, or 'slices'. Sometimes the specifier may include pictograms such as barrels of oil or cars or sheep. Finally, the *labels* consist of words, letters, or numbers that identify the amounts or categories depicted in the graph; these often include a legend. In the present experiments, the background and framework of the graph stimuli were held constant. However, the complexity of the specifiers and domain consistency between the labels and the native expertise of participants were manipulated.

While these broad definitions include a variety of artifacts such as maps, diagrams, and other kinds of visualizations (see Heer, Bostock, & Ogvietzky, 2010; Kosslyn, 1993; Neil, 2009), the focus of this thesis was exclusively on line graphs, the most common manifestation of data visualizations (Stutely, 1993).

Graphical Task Analysis

To provide a context for an evaluation of graph comprehension theories, it is useful to position those theories within a larger task analysis framework since the nature of tasks play an important role in determining comprehension. This preliminary step aims to understand what people *do* with, and *ask* of, graphs. In his seminal work Bertin (1981, 1983) proposed three fundamental uses of graphical representations:

1. **Storage:** This is the economical (in terms of human cognitive resources) representation of data in a form that avoids the need to memorize long lists of numbers.
2. **Communication:** This is by far the most common use of graphs; its most common manifestation is in newspapers and the internet, often with pictograms representing the amounts to be communicated. For example, ingots may represent units of international gold reserves. This communication function is the one considered by most authors in the area of graph comprehension (Lohse, 1993; Pinker, 1990; Shah & Hoeffner, 2002).
3. **Exploration:** This refers to the ability of graphs to augment cognition and to assist in reasoning about the underlying data. This function of graphs is often neglected in the graph comprehension literature (Roth & Bowen, 2003).

The exploration role of graphs is central to exploratory data analysis (Chambers, Cleveland, Kleiner, & Tukey, 1983; Kimball, 1996; Tukey, 1977) as opposed to the more common data analysis in which a formal hypothesis is either confirmed or disconfirmed (Haig, 2005).

While storage, communication, and exploration describe what people *do* with graphs, it is also reasonable to describe what people *ask* of graphs. Ratwani and Trafton (2004) distinguish amongst three types of questions that can be asked of graphs. Fact extraction is the direct extraction of a single piece of data, for example, “What was the temperature on June 22?” might be asked of a graph displaying daily temperatures. Integration questions include fact extractions accompanied by some kind of mental operation, for example, “Was the average temperature in June increasing or decreasing?” Finally, inference questions involve fact extraction and integration as well as extrapolating beyond the information in the graph, for example, “What will be the temperature tomorrow?”

The taxonomies of Bertin (1981, 1983) and Ratwani and Trafton (2004) are complementary and together, they provide a reasonably complete picture of graph behaviour as shown in Table 1. One way of interpreting this Table is to consider the different types of questions that a user of a graph might ask depending on whether they are engaged in storage, communication, or exploration. For example, for storage the only relevant question is fact extraction. However, when someone is communicating with graphs, then fact extraction and integration come into play. Exploration involves all three—fact extraction, integration, and inference. Since the focus of this thesis is on graph exploration, Table 1 points to the need to consider all three components. All of these task-types will therefore be explored to learn how experts differ from novices in their graph exploration strategies when faced with familiar and unfamiliar domain graph stimuli.

Table 1

Graphical task analysis: What people do with graphs and ask of graphs

Use / Ask	Fact Extraction	Integration	Inference
Exploration	√	√	√
Communication	√	√	
Storage	√		

Graph Exploration Theoretical Framework

This section reviews theories and models of graph comprehension. These theories have evolved from an emphasis on the perceptual qualities of graphs towards some of the higher-level cognitive attributes of the graph user. That evolution was accompanied by the use of stimulus material that started as simple two- and three-point line graphs to complex representations with five or six variables plotted simultaneously.

At the same time, the tasks employed in graph research have evolved from simple fact extractions such as “What was the price of coal in July?” where the participant identifies the month of July on a histogram and “reads off” the height of its corresponding bar against a scale on the ordinate, to include the integration of many complex variables involving inferences about, for example, future weather conditions.

As well, graph research has progressed from a purely nomothetic viewpoint and has come to embrace a more idiographic approach (Newell & Simon, 1972). Empirical studies of graphs have progressed from testing undergraduate students exclusively through to PhDs and domain

experts with decades of experience in their respective fields. This thesis employed both domain experts and domain novices.

The theories of Kosslyn (1989), Pinker (1990), Freedman and Shah (2002), and Trafton, Marshall, Mintz, and Trickett (2002) have substantially furthered our understanding of graph comprehension. However, their elucidation of the structure and process nature of graph cognition does not include the strategies, plans or heuristics that people may use when interpreting graphs varying in complexity. In the following section, graph theories are evaluated from three perspectives: theoretical, graph behaviour, and expertise with a view to expanding our understanding of strategies in graph exploration.

Kosslyn (1989) was among the first to apply an information processing approach to graph comprehension. He followed a “modal” model (e.g., Baddeley, 1976) wherein an image of a graph fed into a limited-capacity Short-Term Memory (STM) store. STM also had inputs from Long Term Memory (LTM) from which knowledge was used to help discriminate between (for example) line- and bar graphs. Within this framework, Kosslyn identified syntactic, semantic, and pragmatic levels of graph analysis. These are similar to their equivalents in linguistics where the syntactic level is concerned with the physical properties of the lines and objects contained in graphs. The semantic level is concerned with the meanings of the lines and objects in graphs and what they signify. Finally, the pragmatic level mediates between the information contained in a graph and the individual user’s needs.

Kosslyn used Goodman’s (1968) conception of semiotic symbol systems to describe how the components of a graph were attached to meaning. Goodman proposed a three-level hierarchy in which individual “marks” belonged to “character classes” which in turn belonged to “compliance classes”. The essence of Goodman’s system was that a set of lines in a graph (its

marks) should map into one and only one character class, which in turn should map into only one compliance class, and that the character class to which a mark belonged should be clear.

Applying Goodman's system, Kosslyn was able to determine that the specifiers and labels (see p. 6) in graphs must be distinguishable and unambiguously denote what they are intended to symbolize. Goodman's notion is similar to the concept of semantic congruity (Carswell, 1992) which concerns the relationship between a graph's specifier and the variable it represents. On this basis both Kosslyn's (1989) and Carswell's (1992) theories would predict that pictographs utilizing multiple coffee mugs to represent coffee consumption, for example, should be easier to recall and comprehend than a simple bar graph. Wickens & Carswell (1995) subsequently developed what they called the proximity compatibility principle which is concerned with multiple graphs (or displays) that might be found in aircraft cockpits or nuclear power plants. These kinds of graphs are beyond the scope of this thesis and will not be discussed further (see Wickens & Carswell, 1995; Carswell & Wickens, 1996 for reviews).

One of Kosslyn's contributions to theories of graph comprehension was the identification of syntactic "acceptability principles" to diagnose why some graphs facilitated fact extraction for communication better than others. These acceptability principles were based on perceptual notions of discriminability, distortion, and Gestalt laws of proximity, good continuation, similarity, and form (Kohler, 1947). These principles allowed Kosslyn to identify acceptability "violations" in real world graphs. For example, Kosslyn found graphs published by government agencies in which the labels and specifiers were not clearly matched, or where the specifiers of a stacked bar graph were not labeled clearly (Kosslyn, 1989). Since stacked bar graphs add multiple dependent variables together to "stack" the bars it is often difficult to determine if the total (as opposed to the individual components) is most meaningful.

Kosslyn's (1989) work might be interpreted as more of an empirically-based prescription for graph creation and less as a graph comprehension theory. Nonetheless, it set the stage for more cognitively-based theories to follow. Kosslyn's focus was primarily on the perceptual issues associated with graphs. Even his application of Goodman's scheme to the semantic level of graphs was confined to violations of what he termed "between-level mappings" in which the specifiers of a graph and their meaning were mismatched. At one point Kosslyn acknowledges that "...many of the 'violations' we identified would in fact be irrelevant, but this cannot be known without knowing the purpose of the display." (Kosslyn, 1989, p. 214). Clearly, context plays a critical role in graph comprehension.

Complementary to Kosslyn (1989), Pinker (1990) emphasized the cognitive processes associated with graph comprehension. Starting with initial perceptual processing, Pinker distinguished between two stages. The very earliest post-retinal processing he called the visual array which is equivalent to what Marr and Nishihara (1978) called the 2½-D sketch. The visual array captures the initial sensory, pictorial representation of visual input. Pinker hypothesized that the processing of the visual array is followed by a visual description which consists of propositional descriptions of the graph. This is the level at which the perceptual acceptability and Gestalt principles posited by Kosslyn (1989) operate. Pinker also saw this level as the point at which the fundamental visual structure of the graph is represented and where highly practiced automatic processes operate. For example, Shiffrin and Schneider (1977) and Schneider and Shiffrin (1977) distinguished between automatic and controlled processing where the former is a result of extended practice and can be executed without attentional control. The processing of graphs in Pinker's visual description would be automatic and therefore independent of attentional limitations.

After a visual description is created, Pinker recruits the notion of *schemata* to help explain how LTM might influence graph comprehension. This is a key component of Pinker's theory and represents a significant advance in graph comprehension theory. Schemata are LTM representations that embody domain knowledge and which entail the attributes of any given object as well as the attributes (or slots) that will vary from one set of object classes to another (e.g., Norman, 1976; Norman & Rumelhart, 1981). In Pinker's model, schemata mediate between the output from the visual description and LTM processes. Graph schemata help to assemble the bottom-up *conceptual messages* arising from a graph as well as the top-down *conceptual questions* that the graph user might hope to answer. Conceptual messages and conceptual questions are pairs or *n*-tuples of variables in an equation that code the relationships in the graph. In the simplest line graphs a conceptual message might be the equation that codes an independent variable such as "April" and a dependent variable such as "gas price high". The corresponding conceptual question for the independent variable "April" might be the dependent variable "What is the gas price?" Conceptual messages and conceptual questions are typically only those portions of an instantiated graph schema that memory or processing limitations will allow. In theory, the instantiated graph schema would contain every possible piece of information that a graph user could ever extract from a given graph with unlimited time and unlimited processing capacity.

Pinker's schemata operate by subjecting the intermediate products of the visual description of a graph to a parallel matching process. If a sufficient number of parameters fulfill a match criterion, the graph is recognized as a line graph, a histogram, or a pie chart. For example, once sufficient criteria are met for an object to be recognized as a line graph, the line graph schema is instantiated or populated by the specific graph and a set of conceptual messages

are created. Although Pinker does not specify how the graph recognition criteria are determined, the process seems to be flexible enough that a totally novel graph type can be identified and subsequently understood. The process of comprehension is iterative. If the desired answer is not initially found in the conceptual message, the graph user initiates a conceptual question which in turn leads to an interrogation of the instantiated graph schema. If it reveals no answer, the process works back through the visual description and visual array and ultimately back to the graph itself to find the answer. This process can be repeated as often as time or motivation allows.

Introducing schemata to his information processing theory of graph comprehension allowed Pinker to accommodate more aspects of people's interactions with graphs. For example, if a conceptual question remained unanswered by the current graph schema, a top-down process would be initiated to interrogate intermediate products or even the original graph itself. Thus, Pinker was able to reconcile his theory with the tendency of people to use both an internal and an external representation of the artifact (the graph) being considered (Zhang, 1997; Zhang & Norman, 1994). Pinker also proposed that, in addition to the perceptual issues identified by Kosslyn (1989) and others (e.g., Hoadley, 1990; Simkin & Hastie, 1987), graph difficulty may also be mediated by the linguistic propositions available to the user. Thus, the many words available to English speakers to describe lines and curves such as 'flat', 'steep', and 'undulating', should help to make the trends in line graphs easier to comprehend than trends in histograms where the vocabulary available to describe the height of bars is more impoverished. For example the vocabulary available to describe the shapes of the tops of bars in a histogram seem to be confined to words such as 'staircase', 'descending', and 'rectangular' (Pinker, 1990). Pinker also extended Kosslyn's research in terms of the type of graph behaviour he addressed. In

addition to fact extraction for communication, Pinker's iterative model also permits integration-type questions in order to construct trends. A summary representation of Pinker's (1990) model of graph comprehension is provided in Figure 1.

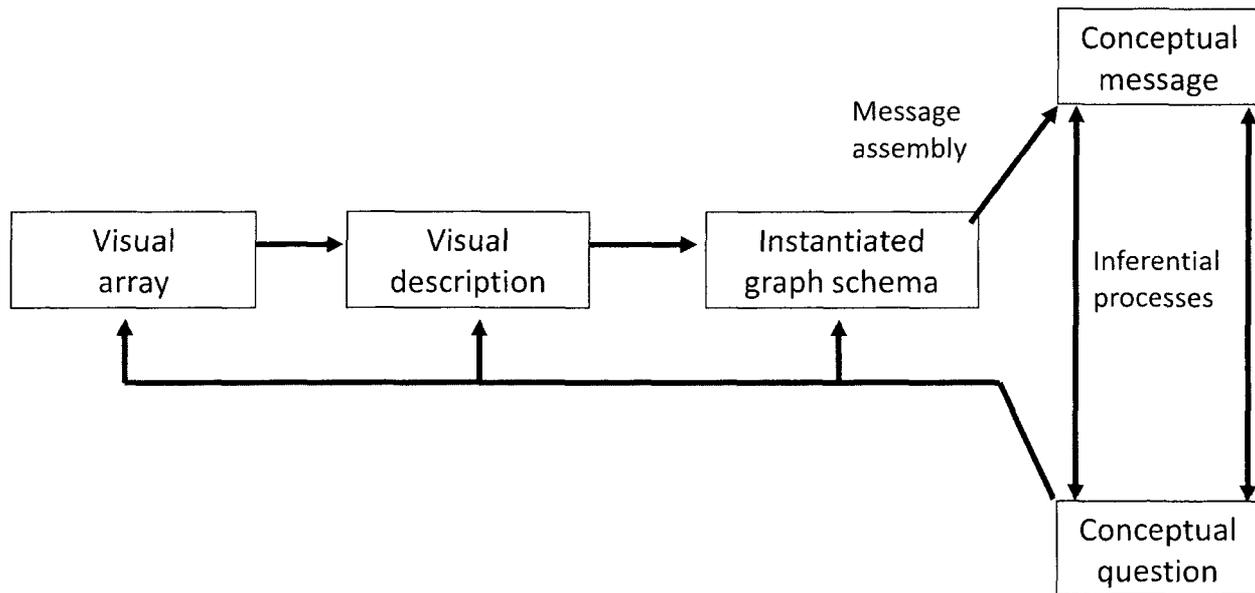


Figure 1. Pinker's (1990) graph comprehension model.

Pinker provided empirical support for his theory in the form of two sets of unpublished studies. In one (Pinker, 1983, cited in Pinker, 1990), graph familiarity was controlled through a novel form of graph in which rainfall over a 12-month period was coded by the length of line segments, and temperature was coded by the angle of those segments. Pinker's participants identified single values of line length faster than single angles. However, since a consistent pattern of temperature yielded a simple convex shape over 12 months, Pinker's participants could recognize a consistent set of temperatures but not a consistent set of rainfall denoted by the length of lines. Pinker interpreted this interaction as support for the idea that the convex line over a 12-month period was encoded as part of a graph schema thus permitting people to encode the trend directly. Unfortunately, since these graphs must have been quite novel to Pinker's

participants, it is difficult to determine if these were complex or simple graphical representations. Furthermore, as Pinker used novel graph forms, it is impossible to determine if his participants even treated them as graphs. Further discussion of that study is beyond the scope of this thesis.

Support for Pinker's theory was also provided by Simcox' unpublished dissertation research (Simcox, 1983, cited in Pinker, 1990) which measured response time in a card-sort task. In that case, sorting on the basis of the height of bars in a histogram was faster than sorting by the height of one end of a line graph. Likewise, sorting by the slope of a line graph was faster than sorting by the slope of two bars in a histogram.

Importantly, Pinker's theory allows for two loci of expertise: the visual description and the graph schema. If a person frequently processes a particular visual pattern, that processing becomes automatic (Shiffrin & Schneider, 1977) and a more elaborate and complete visual description results. For example, in the graphical representation of the Yerkes-Dodson law, the processing of the ubiquitous inverted U-shaped curve will be fast and completed with minimal attentional resources. Thus an individual who has had extensive experience with a particular shape of line graph will be able to encode a new instantiation with little effort; response time of a domain expert to a familiar graph should be shorter than to an unfamiliar graph, and also shorter than for a domain novice. This was tested in Experiment 2 in this thesis.

Pinker also hypothesized that a practiced graph reader will have a richer graph schema and will therefore associate a larger number of conceptual messages with a given graph than an inexperienced graph reader. As a consequence, an expert should be able to apply more descriptors to a given graph than a novice. No empirical verification of this point was found in the literature. However, it was explored here by comparing the number of conceptual messages generated by experts and novices in Experiment 2.

Independent empirical support for Pinker's model was provided by Lohse (1993) who developed a simulation of graph comprehension and evaluated it against empirical data. UCIE (Understanding Cognitive Information Engineering) is a GOMS-like (Goals, Objects, Methods, Selection rules) model (Card, Moran, & Newell, 1983) insofar as both UCIE and GOMS take estimates of the component times for various cognitive tasks and compares total task time between the model and empirical results. UCIE extends the original GOMS paradigm and models eye movements, cognitive operations (such as reading and comparing items in working memory) as well as the keystrokes of people using graphs. Like Kosslyn (1989), Lohse (1993) was primarily concerned with the perceptual aspects of graph comprehension and ignored such issues as expertise, context, or bias. Lohse compared the response times of his participants answering several hundred questions about graphs. Lohse utilized line graphs representing the market price of 6 minerals over a 12-year period and asked fact extraction questions such as "What was the price of zinc in 1981?" as well as integration questions such as "Did zinc cost more than copper in 1981?" or "From 1970 to 1975 did copper prices decrease?"

In comparing modelled with actual results, UCIE explained approximately 10% of the variance in individual participant response times. Lohse explained this low value by concluding that individual differences may have overwhelmed some of the perceptual constants in the model. Clearly, graph comprehension is more than visual perception and cognitive processes working in splendid isolation. Consistent with Lohse, graph comprehension theories need to accommodate the systematic differences that may appear between experts and novices. This is one of the purposes of this thesis.

Top-Down Influences On Graph Comprehension

Freedman and Shah (2002) explicitly acknowledged the importance of prior knowledge and biases operating on “real world” graph comprehension. They leveraged Kintsch’s (1988) construction-integration (CI) model of discourse comprehension wherein comprehension proceeds via iterative construction and integration phases. Construction phases create a network of possible meanings followed by integration phases in which specific propositions are created. Although Freedman and Shah’s CI model is quite similar to Pinker’s (1990) theory in the sense that both conceive of an iterative process of graph comprehension, they differ in that the CI model acknowledges the ambiguities of text. That is precisely why a variety of possible meanings is created by the construction phase. In contrast, Pinker assumes that *one* meaning is created deterministically by the instantiation of the graph schema and that meaning is elaborated by subsequent iterations.

Freedman and Shah (2002) built a CI model in which both text and graph comprehension consist of three components: visual feature units automatically process the visual characteristics of a graph; domain knowledge units integrate prior knowledge; and proposition units capture potential interpretations of a graph. In Freedman and Shah’s model prior knowledge operates on both construction and integration phases where that knowledge might include domain-specific information, graph-specific knowledge or expertise, and explanatory skills (Freedman & Shah, 2002). Empirical support for Freedman and Shah’s model is provided by Carpenter and Shah (1998) whose eye fixation data confirmed the iterative nature of the graph comprehension process. Similarly, Tversky and Schiano (1989, Experiment 5) demonstrated differential recall for a diagonal line positioned in an “L” framework, which was described as either a graph or a map. The graph context led to recall that was biased towards the 45 degree angle, whereas the

map instruction led to a more veridical judgment of the angle. This difference therefore demonstrated the influence of conceptual information on memory for graphs. It is the explicit acknowledgement of context and bias that differentiates Freedman and Shah from the theories of Kosslyn (1989) and Pinker (1990).

However, in a related study, Shah and Carpenter (1995, Experiment 3) failed to find any effect of expertise on graph interpretation. Shah and Carpenter recruited undergraduate and graduate psychology students and asked them to generate explanations to account for the data portrayed in graphs. This was clearly an example of exploratory, inferential graph behaviour. Participants were also asked to make a same/different judgment about underlying data in the graphs when two of the three variables depicted in the graph were transposed. Despite the graduate students' training in data analysis, their explanations of the data were indistinguishable from those of the undergraduates. Furthermore, the same/different judgments of both experts and novices reached only chance levels (50%). One possible explanation for the lack of effect may be that their graduate and undergraduate students were too similar in their graph comprehension skill. To address this issue, Experiment 2 in the current thesis recruited experts comprised of a combination of faculty and advanced (PhD candidate) graduate students. These experts were compared with first year, first semester undergraduate students. Another possibility is that Shah and Carpenter's task was more like exploratory data analysis whereas the emphasis in most graduate courses in experimental design in psychology is on confirmatory data analysis (e.g., Neter, Kutner, Nachtsheim, & Wasserman, 1996; Tukey, 1977). If true, then both the graduate and undergraduate participants in Shah and Carpenter's (1995) study would have been equally non-expert. Finally, with performance near chance levels, it is possible that Shah and Carpenter's failure to show expert superiority might be due to a floor effect. This was avoided in

the current thesis by employing both simple and complex graph stimuli to determine any expertise-based differences.

Very few researchers have acknowledged a role for spatial representations of graphs. This is puzzling since, by their very nature, graphs are a form of spatial representations. Most graph theories (Freedman & Shah, 2002; Kosslyn, 1989; Pinker, 1990) emphasize the propositional nature of graph comprehension. However, Tabachneck-Schijf, Leonardo, and Simon (1997) developed a computational model of graph communication, CaMeRa (Computation with Multiple Representations), in which a pictorial representation is explicitly acknowledged. In CaMeRa multiple representations interact. To explain how an expert economist communicates the law of supply and demand using graphs and text on a blackboard, CaMeRa requires both internal (cognitive) and external (blackboard) representations where the internal image is occasionally updated by the external version (e.g., Zhang & Norman, 1994). CaMeRa also has both verbal (propositional) and visual (see Pylyshyn, 2002) processes working in concert. Tabachneck-Schijf et al. further distinguish between modality-specific STM and LTM (verbal STM/LTM and visual STM/LTM), acknowledging that the two systems are computationally superior for different types of sub-tasks (Larkin & Simon, 1987). Tabachneck-Schijf et al. considered the differentiation between experts and novices as due to their differing abilities to integrate the separate verbal and visual representations: “What distinguishes the experts from the novices is not the sharpness of their vision...” (Tabachneck-Schijf et al., 1997, p. 336).

Although a review of the debate between propositional and visual representations in cognition is beyond the scope of this thesis (see Pylyshyn, 2002 for a review), Tabachneck-Schijf et al. and Trafton and his colleagues advocate a focus which is qualitatively different from the

existing graph comprehension models. Trafton and Trickett (2001) have directed their attention to real-world graph usage in which professionals in their respective domains conduct exploratory data analysis using graphs that are actually being utilized to extend the knowledge base of their users. For example, Trafton, Kirschenbaum, Tsui, Miyamoto, Ballas, and Raymond (2000) studied military weather forecasters using charts and graphs to create pilot briefs, that is, forecasts to guide pilots in the creation of their flight plans. Thus, the task demanded of their participants was clearly exploratory and inferential.

This approach of recruiting domain experts is similar to that of Roth and Bowen whose participants were working scientists in their respective ecology, physics, and forest engineering disciplines. Over a number of studies, Roth and his associates (Roth, 2004; Roth & Bowen, 2001, 2003; Roth, Bowen, & McGinn, 1999) have developed a semiotic interpretation of graphs by experts that encompasses the graph content, what the graph signifies to its users, the context, and the domain rules under which the expert operates. Roth and Bowen (2003) also observed that their scientists sometimes had difficulty with graphs that were representative of introductory textbook descriptions in their respective domains. In fact, they provided a correct interpretation of textbook graphs at an overall aggregate rate of only 50%! Roth and Bowen interpreted this as reflecting a distinction between general and specific graph skills. In fact, their participants appeared to apply unique interpretations to graphs of their own work as opposed to graphs on similar topics generated by others. An alternative explanation is that the graphs of their own research were exploratory and inferential. In contrast, the textbook graphs were communicative and intended for fact extraction. The most parsimonious explanation may be the overall complexity of the graphs utilized. Clearly a 50% correct interpretation rate introduces the possibility that the textbook graphs were considerably more complex than the experts' own

graphs. To overcome this potential confound the current thesis used empirically determined simple and complex graphs.

Graph Comprehension Theories Reconciled

Overall, it may be that the theories of Kosslyn (1989), Pinker (1990), Freedman and Shah (2002), Trafton et al. (2000; 2001), and Roth (2004) may represent more of a difference in emphasis than in kind. Lewandowsky and Behrens (1999) noted the profound effect of task type and task demands on graph performance. Because the Kosslyn, Pinker, and Lohse perspectives were primarily built on empirical data involving fact extraction and integration graph tasks, it might be expected that their theories reflect an emphasis on the processes necessary for fact extraction and integration. Similarly, the emphasis of Freedman and Shah, Trafton et al., and Roth and colleagues has been on graph exploration and the integration and inferential processes required. As a result, their theories have tended to emphasize those processes that are necessary for graph exploration.

An alternative way of looking at some of the dominant theories of graph comprehension is to first consider the graph behaviour they are modelled after. For fact extraction a process like that described by Pinker (1990) would involve bottom-up processing through the visual array then the creation of a visual description and eventually an instantiated graph schema from which conceptual messages are assembled. This relatively simple process would accrue in fact-extraction tasks where the graph maker's intent is purely communicative. However if, for the same graph, analytical inferential graph behaviour is expected (in the context of the task or by virtue of the domain knowledge brought to bear by the reader) then top-down conceptual question(s) would drive the entire process. This might also occur if the initial default, bottom-up

process failed to return an acceptable result. Then the graph user might consciously select an alternate process if their level of skill or training or expertise means that they have such alternatives available. Investigations of expert and novice participants' verbal protocols should reveal their graph-exploration behaviour.

According to Freedman and Shah's (2002) CI model, many possible meanings are constructed first, followed by an integration phase which reduces these to fewer, more coherent interpretations. This would be an adaptive characteristic in the context of exploratory inferential task requirements that would necessitate bottom-up as well as top-down processes. Pinker (1990) does not address these explicitly and it appears that there is a disconnect between the conceptual questions and conceptual messages in his model. That disconnect was addressed in this thesis.

It is proposed that differences between novices and experts in graph exploration might be explained in terms of a strategy selection process. The issue of expertise is examined next followed by a discussion of strategy use and selection by experts.

Expertise

As noted earlier, theories of graph comprehension have recently begun to encompass how experts behave with graphs. The issue of expertise will be examined more generally here, with a focus on how expertise pertains to graph exploration. Finally attention will be directed to the strategies that experts and novices employ when exploring graphs.

Apparently, the first known study of experts was conducted by Hughes in 1917 (cited in Shanteau, 1992). Hughes compared expert agricultural judges' evaluations of seed corn with the subsequent yields. He found that expert judgments did not correlate with seed corn yields. That result notwithstanding, many subsequent studies have established reliable domain-specific

performance differences between experts and novices in many fields including chess (Chase & Simon, 1973), physics (Chi, Feltovich, & Glaser, 1981), sports and dance (Allard & Starkes, 1991) and odour recognition by wine experts (Parr, Heatherbell, & White, 2002). For other examples see Ericsson (2005), Ericsson and Smith (1991), and Rikers and Paas (2005).

Ericsson and Smith (1991) define expertise as outstanding, stable performance in a particular domain. However, as expertise falls along a continuum of competence in which there are often no objective criteria for ascendance from one level to another (Salthouse, 1991), for the purposes of this thesis the following definition was applied:

1. Create graphs at least once per month.
2. On a 100-point visual analogue scale concerning ease of working with graphs, mark the scale at 50 or lower.

As an example, the expert participants recruited in Experiment 2 in this thesis were senior graduate students (PhD candidates) and tenured PhD research and teaching faculty at Carleton University. (Note that the heterogeneity of the experts was not ideal but was a function of the difficulty of recruiting a sufficient number of tenured faculty. However, there were no obvious differences in the dependent measures obtained from these two groups of experts in Experiment 2.) Although international rating systems for games such as bridge permit reliable assessments of expertise in the sense that a player with a higher rating will consistently defeat those with a lower rating (N. Charness, personal communication, September 26, 1978), in most domains researchers must be content with samples from different points on the expertise continuum. Alternatively, intensive training in longitudinal studies has enabled an assessment of some characteristics of the acquisition of expertise (Anzai, 1991; Charness & Campbell, 1988; Kleigl, Smith, Heckhausen, & Baltes, 1987).

Among the most enduring results of research into expert/novice differences are that expertise is domain-specific (Arbernethy, Baker, & Cote, 2005; Sims & Mayer, 2002); experts are faster than novices (Gobet & Simon, 2000); experts' representations of their domain is richer than that of novices (Glaser & Chi, 1988), and experts use forward search from the initial state to the goal state while novices search backwards from the goal state to the initial state of the problem space (Gobet, 1998).

Much of the theoretical development in expertise has been driven by research involving the game of chess. Chess has been used as the *Drosophila melanogaster* of expertise research in part because it has a widely accepted objective measure of skill (Waters, Gobet, & Leyden, 2002). The so-called Elo rating has a mean of 1500 and a standard deviation of 200 and grandmasters are defined as those with an Elo greater than 2500 (Gobet, 1998). While desirable, no such rating system exists for expertise in many other domains. Chess as a research domain also seems to strike a balance between the simplicity of its rules, a non-trivial level of difficulty, and the problem-space of possible games (Gobet, 2005). As a result, a large empirical corpus of research has contributed to theories of chess expertise.

The dominant model of chess expertise is the perceptual chunking theory of Chase and Simon (1973) and the schema variant espoused by Gobet and Simon (1996a, 1996b, Gobet, 1998, 2005; but see also Ericsson & Staszewski, 1989; Vicente & Wang, 1998). Both the Chase and Simon and Gobet and Simon theories include a discrimination network that indexes items in LTM. The discrimination net is conceived as a set of binary tests that when executed, recognizes chunks that are stored in LTM (e.g., Gobet et al., 2001).

Chase and Simon (1973) developed their chunking theory to accommodate two phenomena in chess expertise. The fact that chess grand masters are able to near-perfectly recall

briefly presented chess boards they ascribe to superior perceptual abilities in experts compared to novices. Chase and Simon postulated that experts encode the board positions as 7 ± 2 chunks (Miller, 1956) held in STM, and since the size of their chunks is larger than those of novices, they are able to recall a larger proportion of board positions. To explain how experts find good moves in the large problem space of chess, Chase and Simon hypothesized that the chunks may serve as triggers to production systems—sets of executable condition/action pairs (Newell & Simon, 1972). When planning their next move, experts may look ahead by executing a series of hypothetical moves where the size of the experts' perceptual chunks allow for more efficient execution than novices.

A chess expert exhibits superior recall of a chess board compared to that of a novice because experts have stored more chunks (approximately 50,000 according to Gobet & Simon, 1996b) that enable them to recognize more of the patterns on the chess board. However, when confronted with a chess board comprised of randomly-placed pieces, experts' performance decreases significantly although they are still somewhat better than less expert players.

Despite the general success of the perceptual chunking model in explaining chess expertise, a number of researchers have questioned the reliance on STM for chess recall. For example, Charness (1976) inserted an interference task between initial board presentation and subsequent recall. He observed that the interfering task had a negligible effect on recall accuracy even though it did increase the latency with which the first pieces were recalled. Charness interpreted the latency in his results as indicating that his experts were retrieving the board positions from LTM thus contradicting the original Chase and Simon (1973) STM chunking hypothesis. These findings caused Gobet and Simon to develop a hybrid model which places a greater emphasis on post-perceptual or cognitive processes.

Gobet and Simon (1996a, 1996b) hypothesized that, with the development of chess skill, multiple schemata are created. The schemata have slots in which individual game pieces or chunks may be placed. In the recall of multiple chessboards recognition is accomplished through the discrimination net and recall is executed by placing the recognized schemata in STM, one for each board recalled. The empty slots in Gobet and Simon's schemata have the attractive quality of allowing "near matches" thus permitting the recall of board positions with a few pieces that do not exactly match. This near match attribute of schemata also helps to explain the modest performance superiority of grand masters over novices even with random board positions because even such boards may have chunks that can be identified with one or more schemata thus facilitating recall.

This departure into a description of theories of chess expertise was necessitated by the absence of such theories in graph exploration. However, it is reasonable to believe that Gobet and Simon's (1996a, 1996b) schema theory might apply to graph exploration. In particular, it is proposed that Gobet and Simon's schemata are the key mechanism by which someone explores a graph. A partially instantiated schema is the indexing mechanism through which a particular graph is compared to other graphs in LTM as well as analogous situations and circumstances also stored in LTM (e.g., Gick & Holyoak, 1980, 1983). This explanation would predict that experts, by virtue of their larger corpus of LTM schemata would take longer to explore graphs than novices.

As another example, Trafton and Trickett (2001) recorded the verbal protocols of two PhD astronomers engaged in exploratory analysis of radio telescope data. Trafton and Trickett found that approximately half their participants' utterances were related to the direct fact extraction of numbers from their graphical displays. However, an almost equal proportion of

their utterances were related to “spatial transformations” of the data, creating an alternative mental image, modifying the existing image, or comparing the current image to a previously created mental image. The spatial transformations described by Trafton et al.’s participants suggest that perhaps one concomitant of expertise is having alternative meanings available. This capability would be facilitated by another schema attribute hypothesized by Gobet and Simon (1996a) which is that verbal labels may be attached to schemata. For example, “Queen’s Gambit with a minority attack” or “Grünfeld defense” may be attached to a schema. Thus, the alternative meanings may also have names. The possibility of the existence of such schema-naming in graph comprehension will be examined through the verbal protocols of participants in Experiment 2.

Although creative and insightful processes may be operating in both chess and graph exploration, discussion of these is beyond the scope of this thesis. However, see Chronicle et al. (2004), Isaak and Just (1995) and Mayer (1995) for reviews. One key shortcoming in applying Gobet and Simon’s (1996a, 1996b) theory to graph exploration is that there seems to be no accommodation for the strategies that experts and novices might adopt. This issue is addressed next.

Expert Problem-Solving Strategies

The kinds of problems addressed in this thesis concern the exploration of line graphs varying in complexity and domain familiarity. For the present purposes, graph exploration strategies were defined as the ability of experts to consistently produce an ordered list of questions that may be asked of a line graph. For example, one might expect experts to include

the graph title, questions about the ordinate and abscissa, and so forth whereas novices would not be expected to generate such ordered lists.

Although Gobet and Simon's (1996a, 1996b) schema theory appears to provide a good account of expertise in chess, it is difficult to know if it would be applicable to other unrelated domains (but see Gobet, 2005 for applications in education). Furthermore, as noted earlier, the role of strategies in Gobet & Simon's schema theory is unclear. Although strategies were clearly manifest in participants' comments about how they completed the recall task, Gobet and Simon (1996b) found no systematic effect of expertise on strategy use.

Gobet, Richman, Staszewski, and Simon (1997) addressed the issue of strategies, albeit independently of expertise. They modeled the concept attainment task of Medin and Smith (1981) in which participants learned to categorize a set of schematic faces which varied in terms of attributes like eyes high or low, nose long or short, mouth high or low, etc. Gobet et al. equated Medin and Smith's original instructions—guess and then attend to feedback; learn a rule plus exceptions; and learn a “prototype”—as strategies which resulted in the explicit and intentional formation of nodes in a discrimination net. For example, the rule plus exceptions strategy was implemented by Gobet et al. as two discrimination nets. The first net categorizes faces according to an arbitrary test of nose length. Because nose length did not perfectly characterize stimuli in this task, a second discrimination net is created to categorize the exceptions to the nose length rule. By assuming that a new stimulus face is noticed in 10 ms, entering a discrimination net is completed in 100 ms, and tests for each attribute (nose length, etc) requires 250 ms, Gobet et al. were able to simulate both total response time and errors in this task. Correlations between the Medin and Smith human data and the Gobet et al. simulation were high with Pearson's correlation coefficients for each of the feedback, rules plus exceptions,

and prototype strategies reaching values of .93, .93, and .77 respectively. Thus, it would seem that the notion of strategy as embodied in the instructions of a concept attainment task can be modeled as the deliberate construction of a discrimination net. Gobet et al.'s discrimination nets are equivalent to alternate sets of conceptual questions. For the present purposes, strategies are defined as a set of conceptual questions selected from amongst alternate sets which may differ in terms of the individual questions imbedded within each set. However, apart from strategy execution, the Gobet et al. model is silent on how strategies are selected, how they develop, and the interaction of strategies with expertise. Lemaire and Siegler (1995) explored the functions of a strategy selection process.

Lemaire and Siegler (1995) developed the Adaptive Strategy Choice Model (ASCM) to describe the operation of strategies in select domains. The ASCM distinguishes four components of strategy use:

1. What strategies are utilized?
2. When are particular strategies selected for use?
3. How efficient are strategies executed?
4. Meta-strategy—on what basis are strategies selected?

Lemaire and Siegler applied ASCM to the learning of multiplication amongst Grade 2 children and found that, over a six-month period, there seemed to be no changes in the repertoire of strategies available to students. However, the children improved in the relative frequency with which better strategies were selected, how the strategies were executed, and in their meta-strategy selection. Similar results have been reported in adults executing complex multiplication problems (Tronsky, 2005) as well as military leadership strategies (Schunn, McGregor, & Saner, 2005). Thus, as a model for the description of strategy selection and change, ASCM seems to be

capable of generalizing across domains. One of the goals of this thesis is to determine the extent to which the model may apply to graph exploration in familiar and unfamiliar domain tasks. If strategy selection is more efficient in experts then one testable prediction is that experts should require less time than novices in graph exploration and that familiar domains should accord an even greater advantage.

It is reasonable to argue that people may benefit from explicit instructions in strategy use. For example, Feldon, Timmerman, Stowe, and Showman (2010) developed a series of instructional videos based on a cognitive task analysis (Schraagen, 2009) of expert biologists and which focused on the selection of appropriate strategies for biological research. Feldon et al.'s experimental instructional videos were similar to Pinker's (1990) conceptual questions and conceptual messages/answers described earlier (p. 15). As an example:

Now you need to graph your data. To do this you should recall from your design how many dependent variables you have. If you have only one and you're (sic) measuring has changed in relation to the independent variable, plot the points on the graph where the "x" axis represents the value of the independent variable and the "y" axis represents the value of the dependent variable... (Feldon et al. 2010, p. 1185)

Feldon et al. also created control videos that represented traditional instructional methods dealing with principles and illustrative examples:

...and out of the pattern that appears of the plotting of the data you begin to see a relationship between unknown factors and known factors. This plotting of the data requires that you think about how to deal with individual points... (Feldon et al. 2010, p. 1185)

Feldon et al. compared the relative efficacy of the two instructional methods on the quality of the discussion section of biology laboratory reports produced after students viewed the experimental or control video instructions and subsequent attrition rates in an undergraduate biology course. Feldon et al. found that students who received the conceptual question/answer strategy instruction performed significantly better than controls on the quality of the conclusions they drew from data; the number of alternative explanations proposed; identifying limitations in the experimental design; and identifying the significance of the results. Controls were also six times more likely to withdraw from the course than those who received the question/answer strategy instruction.

Although Feldon et al.'s (2010) research was conducted in the domain of undergraduate biology, there are clear overlaps between the cognitive skills their tasks required and the current graph exploration paradigm. These were explored here.

Although strategy use in graph exploration has not been previously investigated, there are some domains which are similar. One of these is the design and interpretation of psychological experiments. Schraagen (1993) asked undergraduate psychology students (beginners); graduate students in psychology (intermediates); psychology faculty with at least 10 years of experience designing non-sensory psychology experiments (design experts); and psychology faculty with at least 10 years of experience in sensory psychology (domain experts) to design an experiment to evaluate the taste of a soft drink and compare it to alternate brands. Schraagen was interested in the strategies his different groups employed.

In his analysis of verbal protocols Schraagen (1993) found evidence for two strategies which controlled the overall problem solving task: progressive deepening and mental simulation.

Progressive deepening was characterized by a series of iterations in which experimental variables were kept 'open' but subsequently instantiated. Mental simulation was a strategy in which participants played out the procedure in their mind in order to determine if variables needed controlling or if methods needed clarification. Schraagen found that the progressive deepening strategy was used mostly by intermediates and design experts while beginners and domain experts used mental simulation less than the other two groups.

Interpreting Schraagen's results in terms of both Gobet and Simon's (1996a, 1996b) schema model and Lemaire and Siegler's (1995) ASCM several notable conclusions can be drawn. For one, Schraagen's progressive deepening strategy seems to echo Gobet and Simon's iterative schemata instantiation. By "keeping variables open" it would appear that Schraagen's participants adopted a process in which they developed a series of discrimination nets with empty slots. However, as they worked through the experimental design problem those empty slots become instantiated just as Gobet and Simon's schema theory would predict.

From the perspective of Lemaire and Siegler's (1995) ASCM, it appears that there were qualitative differences in the strategies selected by participants across the range of expertise sampled by Schraagen. Curiously though, instead of a linear change in strategy selection, Schraagen's results seemed to imply an inverted "U" shaped curve where beginners and domain experts (the two ends of the expertise spectrum sampled by Schraagen) utilized the mental simulation strategy less than the other groups.

Schraagen's (1993) conclusion was that one manifestation of expertise is the use of general strategies (see also Adelson & Soloway, 1985) as opposed to the view of Anderson and Lebiere (1998) and others who hold that expertise is primarily about increasing domain-specific knowledge. However, one of the missing pieces in Schraagen's task was feedback in the form of

the results of participants' experimental designs. This was rectified in the work of Schunn and Anderson.

Evaluating the role of domain knowledge in the design of experiments, Schunn and Anderson (1998, 1999) adopted an ACT-R (Anderson & Lebiere, 1998) model of expertise. ACT-R models the human as a perfectly rational being with declarative and procedural knowledge. ACT-R conceives of developing skill as the accretion of chunks in declarative knowledge. With increasing practice this declarative knowledge becomes automatic (Shiffrin & Schneider, 1977) and eventually becomes procedural in nature. Skill is the process of "compiling" many production rules and where those production rules are selected according to expected gain (Anderson & Labiere, 1998). In ACT-R the concept of "strategy" is one of selecting the highest expected gain which is equal to the product of the probability of achieving a production's goal and the value of that goal minus the cost.

Against the ACT-R backdrop, Schunn and Anderson (1998, 1999) asked their participants to design an experiment to compare theories. Participants were informed about two competing accounts of the distributed practice effect in memory and then they had to design experiments to evaluate the hypotheses. To do so they selected values for six independent variables and then made predictions about experimental outcomes in a data table. They were subsequently provided with feedback that represented the simulated empirical outcome of the study they had designed in conjunction with the values they had predicted. After several iterations participants were asked whether their results supported one or the other theory or neither. Schunn and Anderson's participants included domain experts who were well-published cognitive psychologists who had received their PhDs nearly two decades earlier. Schunn and Anderson also utilized task experts who were matched in terms of years since PhD and

publication record but whose specialty was social- or developmental psychology. (These individuals are equivalent to Schraagen's design experts.) Their novices were undergraduates distinguished on the basis of their Scholastic Aptitude Test (SAT) scores in Math. The logic behind their selection of two different types of PhDs was that domain-specific knowledge would be brought to the experimental task by the cognitive psychologists while experimental design skills but relatively little domain-specific knowledge would be brought to the task by the developmental- and social psychologists.

Consistent with Schunn and Anderson's (1998, 1999) expectations, the domain experts exhibited the best experimental design performance, followed by task experts, then by high math SAT undergraduates and finally by the low math SAT undergraduates. The verbal protocols revealed that domain- and task experts mentioned the competing theories whereas the undergraduates did not. Domain- and task experts both kept irrelevant variables constant across the experiments they designed and avoided floor and ceiling effects, whereas undergraduates did not. Both domain- and task experts demonstrated that they had noticed and understood main effects and interactions in the results of the experiments they designed, but the undergraduates did not. Finally, the experts were not fooled by small, insignificant interactions while the undergraduates were misled by these.

Overall, Schunn and Anderson's (1998, 1999) empirical results strongly supported the view that domain knowledge had a significant effect in their experimental design and results interpretation task. Unfortunately however, Schunn and Anderson's participants interacted with their results in tabular form and not as graphs. The current experiments partially replicated the results interpretation portion of Schunn and Anderson to determine if domain knowledge exerts a similar effect on graphs.

A Model Of Strategy And Expertise

Although Gobet and Simon's (1996a, 1996b) schema theory was developed in the context of memory for chess pieces, it is possible to extend it to graph exploration across domains. For business experts, chunks of business scenarios or cases might be developed that represent specific situations in product planning, finance, or marketing. On viewing a graph, a business expert might activate a discrimination net or set of conceptual questions which acts as an index to that set of case histories. If an exact match is not found then a schema with similar characteristics could be instantiated and used to guide interpretation. Retrieval structures would correspond to the deliberate learning of case histories in a fashion that has become popularized through the Harvard Business School (Christensen, 1987; Ronstadt, 1980).

Note that in the case of identical graph specifiers, but where the labels and framework convey a different domain, a different result should ensue. The set of conceptual questions or discrimination net that a business expert establishes when exploring a graph about psychology should be impoverished. One would predict that the cases or analogies that would be generated by a business context would be relatively primitive or missing altogether. The strategies and the performance of individuals confronted with graphs from an unfamiliar domain should reflect this as well.

If Gobet and Simon's schemata are an effective vehicle for strategy execution, the manner in which those strategies develop and are selected is somewhat indeterminate. The results of Schraagen and Schunn and Anderson suggest that there may be complex processes governing strategy development and selection. Perhaps the learning of multiplication tables by children (Lemaire & Siegler, 1995), learning longer multiplication problems by adults (Tronsky, 2005), or selecting appropriate military leadership strategies (Schunn, McGregor, & Saner, 2005)

are more deterministic than designing a good psychological experiment or exploring a business graph.

A primary goal of this thesis was to explore the strategies adopted by domain experts and novices in their exploration of familiar and unfamiliar domain graphs. The initial assumption is that Gobet and Simon's discrimination nets (what are referred to here as conceptual questions) are the vehicle or embodiment of strategies. Then, by sampling from business and psychology experts and novices it should be possible to identify their graph exploration strategies. To ensure that neither floor nor ceiling effects occlude the manifestation of expertise, the simplicity and complexity of graph stimuli were manipulated across experiments. Finally, the effect of domain was investigated by providing graphs that were either familiar or unfamiliar with an individual's domain of expertise. Verbal protocols were collected while participants were exploring a range of different graph types similar to the methods utilized by Preece and Janvier (1992, 1993).

Summary of Preliminary Studies

Three Preliminary Studies systematically converged on graph stimuli, methods, participant recruitment, and procedures that informed the formal experiments. Preliminary Study 1 (Appendix A) asked participants drawn from business- and psychology student populations to rate a corpus of 256 line graph stimuli developed by Oakley, Lindgaard, and Ferres (in preparation). These stimuli were prepared with white specifiers on a black background and had between one and seven variables, two to seven data points per variable, and zero to seven "crossovers" (interactions) plotted. An example graph stimulus is provided in Figure 2.

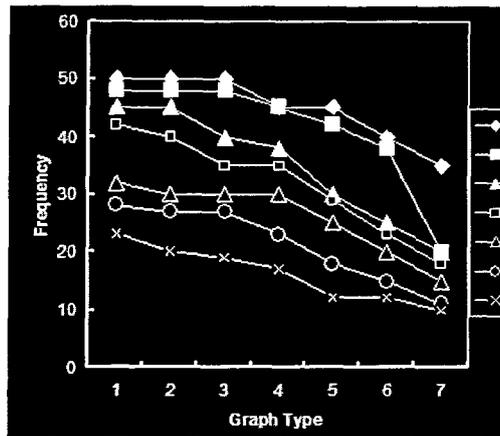


Figure 2. Preliminary Study 1, example graph stimulus

All graph stimuli with more than one variable also included a legend; the abscissa was labeled “Graph Type” and the ordinate “Frequency”, numbered zero to sixty. Participants were asked to rate each graph as “simple”, “moderate”, or “complex”. Although the results indicated that they were able to complete the task, some issues were identified. Participants appeared to base their complexity judgments on the graph stimuli attributes of (1) slope, (2) whether the z-variables were parallel, (3) whether the slope changed direction, (4) the number of common values, and (5) the number of crossovers. Unfortunately, judgments could have been made utilizing some one or all of the dimensions along which the graphs varied, or they may have used some other criterion altogether. In fact, a majority of participants commented that their criterion included how difficult it might be to explain the results to someone else. Since one purpose of the study was to develop a set of graph stimuli for further experiments, it was necessary to control the graph attributes more assiduously and to make more explicit the criterion on which the complexity judgments were made. This was attempted in Preliminary Study 2.

Preliminary Study 2 (Appendix B) reduced the number of graph stimuli from 256 (Preliminary Study 1) to 168 by excluding the graph stimuli that had more than three data points per z-variable. Thus, the graph stimuli included only two and three data points per z-variable. In response to participants' surprise, the graph stimuli were altered to a more common style in which black specifiers were plotted against a white background and the legend was removed to ensure that participants based their judgments on the specifiers only. The graph stimuli used one common title and labels, 'Recall By Internet Usage By Spatial Ability' with the abscissa labeled 'Internet Usage' and the ordinate 'Recall'. For stimuli with two data points, the abscissa was labeled "4 hours" and "8 hours" while three data point graphs had labels of "4 hours", "6 hours", and "8 hours". The ordinate had values ranging from 0 to 10. Examples of the stimuli may be found in Appendix B, Table B10, while an example is reproduced in Figure 3.

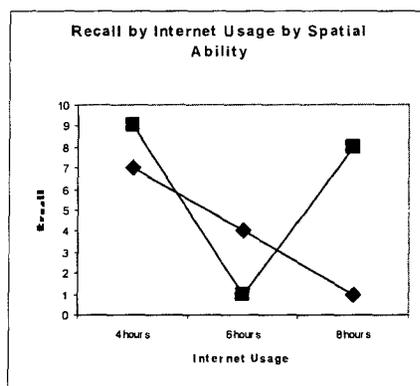


Figure 3. Preliminary Study 2, example graph stimulus

Participants rated the visual complexity of 168 graphs (plus 5 practice graphs) in a block first to avoid any potential contamination of their semantic ratings. After a participant-determined rest period, they then rated the semantic complexity of the same 168 graphs (plus 5 practice graphs). Participants were instructed to base their semantic complexity judgments in terms of how difficult it would be to explain the graph to someone else. Participants responded

to both visual and semantic trials using a seven-point rating scale. Finally, the Visual Patterns Test (VPT, Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999) was administered to determine the degree to which visuo-spatial memory might be associated with complexity ratings of the graph stimuli. A sample of 48 graphs representing the 24 most and 24 least visually and semantically complex was selected. However, while participants were asked to respond using visual or semantic criteria, the graph stimuli were displayed for as long as they desired, because the trials were self-paced. Because of this, the study could not preclude the possibility of semantic criteria encroaching on visual complexity judgments. While Preliminary Study 2 did serve to reduce the number of graph stimuli, it did not allow a clear and unambiguous determination of the criterion on which the graph complexity judgments were based. Preliminary Study 3 was therefore designed to resolve this by reducing the number of attributes differentiating the graphs, increasing the span of the rating scale, and placing strict limits on graph stimulus exposure duration.

Preliminary Study 3 (Appendix C) used 48 unique graphs: the 24 least visually and semantically complex and the 24 most visually and semantically complex graph stimuli selected from Preliminary Study 2. The graph stimulus characteristics are summarized in Table 2. The line graphs again varied in their slopes (positive, negative, zero), parallel versus non-parallel lines, whether they did or did not change slope, number of data points, number of points in common, and number of crossovers. However they were limited to two z -variables which was a departure from the previous study in which they varied from one to seven variables. In order to maximize potential statistical differences, judgments were made on a 9-point scale instead of the 7-point scale used in Preliminary Study 2.

Table 2

Preliminary Study 3, Mean attributes of the simple/complex, visual/semantic graphs selected from Preliminary Study 2

Graph Attributes	Visually	Visually	Semantically	Semantically
	Simple	Complex	Simple	Complex
Data Points	4.9	5.3	4.8	5.2
Crossovers	0.1	1.2	0.1	1.3
Common points	0.2	0.7	0.2	0.6
Slope changes	0.6	1.0	0.4	1.0
Parallel segments	0.2	0.0	0.3	0.0
Zero slope segments	1.0	0.5	1.0	0.4

The other key difference in Preliminary Study 3 was that visual criterion graphs were displayed for only 150 ms with a visual mask displayed at a stimulus onset asynchrony (SOA) of 150 ms because pilot testing indicated that this SOA made it much more difficult to discern the details of the graph apart from its physical characteristics. Semantic criterion graphs were also displayed for 150 ms but with a mask SOA of 3000 ms again because pilot testing indicated that people felt they could “get much more” information from the longer SOA display condition. From the starting set of 48 graphs the results enabled the identification of a subset of 25 line graphs that participants could differentiate on both visual and semantic criteria. The ratings enabled the selection of a set of line graphs shown empirically to vary along visual and semantic complexity

dimensions. In addition to providing a set of stimuli for the formal Experiment 1, the Preliminary Studies thus served to progressively refine the methods and procedure to be utilized.

Experiment 1: Judging Visual versus Semantic Graph Complexity

Preliminary Study 3 had identified a set of 25 graph stimuli to be used in the formal experiments. The purpose of Experiment 1 was to overcome some of the possible reasons for Carpenter and Shah's (1995) inability to find any effects of expertise. As argued earlier (p. 19) there could be several reasons why Carpenter and Shah found no effect of graph expertise in their study. One possibility is that their sample of undergraduate and graduate students was too similar for an expertise effect to occur. To test that, participants were selected here in an attempt to maximize probable differences in their previous exposure to graph exploration tasks. Therefore, Hypothesis 1 predicted that ratings of complexity obtained from senior (PhD candidate) graduate students in business and psychology would differ from those obtained from first-year, first-semester undergraduate students. Another possibility is that Carpenter and Shah's graph stimuli may have been too complex, as evidenced by participants' performance which only approached chance levels. In an effort to overcome that issue, the graph stimuli included here comprised graphs judged in the preliminary experiments to be least and most complex. Furthermore, Shah and Shellhammer (1999, cited by Freedman & Shah, 2002) reported that unskilled graph viewers tend to focus on surface or visual characteristics of graphs. One would therefore expect that the requirement to rate the visual aspects of a graph should seem easier, and therefore result in lower complexity ratings, than the requirement to rate the difficulty

of explaining a graph to another person. Accordingly, Hypothesis 2 predicted lower ratings for the visual graph reading task than for the semantic task.

Method

Participants

A total of 40 undergraduate and graduate students who had not participated in any of the three preliminary studies were recruited using posters and word-of-mouth. An equal number of participants were first-semester undergraduates majoring in psychology, first-semester undergraduates in business, senior students (PhD candidates) in psychology, and senior students (PhD candidates) in business. Undergraduate psychology students were granted 1% course credit for their participation. All other participants were given a \$10 coffee shop gift certificate. All had normal or corrected-to-normal vision. Mean age of experts was 28.7 yrs while the mean age of novices was 22.6. Experts included 13 females while the novices included 9 females. In response to the Graph Experience Questionnaire, (Appendix D7) ten novices reported that they had to create graphs and fifteen experts reported that they had to do so. All participants reported that the graphs with which they were most familiar were line graphs. Participants were tested individually in sessions lasting approximately 60 min.

Apparatus and Materials

The graph stimuli were the 25 line graphs selected in Preliminary Study 3, plotted on x - and y -coordinates and limited to two lines (z -variables). These varied in six ways: (1) by slope

(positive, negative, or zero); (2) by parallel or non-parallel z -variables; (3) by whether the slope changed direction, (4) by the number of data points; (5) the number of points in common, and (6) the number of crossovers. Although graph legends, titles, abscissa and ordinate labels were removed, a paper graph with legend, labels, and title was provided beside participants who were instructed to assume that the graph stimuli would otherwise have these labels associated with them. The title on the paper graph was 'Gas Revenue By Year By Market', and the two z -variables were identified in a legend as Brazil and China. The years displayed on the abscissa were 2005, 2006, and 2007. Seven of the graph stimuli from Preliminary Study 2 that were not included in the experimental stimuli here were reserved for practice trials. A visual mask was presented immediately after every graph stimulus. The mask was identical to that used in Preliminary Study 3 and was created by graphing a 15 x 15 matrix of random numbers (see Figure 4 on p. 46).

Stimulus presentation and timing was controlled using DirectRT™ on a Dell laptop computer with a 1400 x 1050 pixel resolution LCD display. Participants entered their responses using the digit keys 1-*very simple* through 9-*very complex* on a 3 x 3 numeric keypad placed under the participants' preferred hand.

Design and Procedure

The design and procedure were identical to Preliminary Study 3; a 2 (Expertise: novice, expert) x 2 (Domain: business, psychology) x (2) (Criterion: visual, semantic) repeated measures design was adopted. After obtaining Informed Consent and providing detailed instructions, 7 visual criterion practice trials were presented. These were followed by 25 visual criterion

experimental trials, and then by a participant-determined rest period. Next, 7 semantic criterion practice trials were given, followed by 25 semantic criterion experimental trials. Participants were then thanked, paid (if applicable) and debriefed.

Practice and experimental trials were conducted as per Figure 4. On each trial a 'ready' screen consisting of a 9-point scale was positioned in the center of the screen for 1000 ms. On visual criterion trials participants were asked to rate the visual complexity from 1 = least complex to 9 = most complex and the rating scale was replaced by a graph displayed for 150 ms, followed by a visual mask at a Stimulus Onset Asynchrony (SOA) of 150ms. On semantic criterion trials participants were asked to rate the semantic complexity on the basis of how difficult it might be to explain the results to someone else. Again, using 1 = least complex to 9 = most complex. On semantic criterion trials the mask SOA was 3000 ms. (Stimulus timing was determined in the Preliminary Study 3 and adjusted to obtain performance that was free from floor- and ceiling effects). However, if participants made a valid response while the graph was still displayed, it was immediately replaced with the visual mask. Each participant received the graph stimuli in a unique random order, but all participants rated each graph twice; first on the visual criterion and then on the semantic criterion. On each experimental or practice trial the graph was replaced by the ready screen and the next trial would begin immediately after responding with the digit rating. Participants were encouraged to rest their middle finger of their preferred hand on the "5" key of the keypad between trials in an attempt to reduce bias towards the high or low end of the rating spectrum and there was no evidence of such bias occurring (Marteniuk, Ivens, & Brown, 1996; Seibel, 1972).

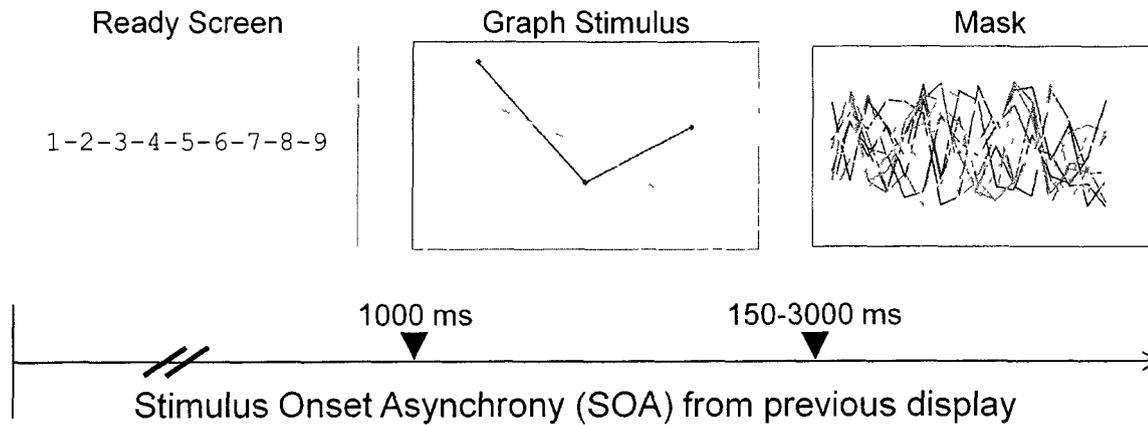


Figure 4. Experiment 1 trial procedure. On Visual Criterion trials the mask was displayed at 150 ms SOA. On Semantic Criterion trials the mask was displayed at 3000 ms SOA.

Results and Discussion

An alpha level of .05 was used for all statistical tests and only significant results are reported throughout this thesis. When 95% confidence intervals are displayed in the figures throughout the thesis, these were calculated according to the procedure of Jarmasz and Hollands (2009).

Complexity Ratings and Response Times

The mean visual- and semantic criteria ratings of the 25 graph stimuli, shown in Figure 5 below, were entered into an overall 2 (Expertise: novice, expert) x 2 (Domain: business, psychology) x 2 (Criterion: visual, semantic) ANOVA. Simple main effect analyses performed on the significant Expertise x Criterion interaction, $F(1, 36) = 4.55, p = .040, \eta_p^2 = .14$ indicated that novices rated graphs as more complex using the semantic criterion ($M = 3.64$) than the same graphs using the visual criterion ($M = 2.61$), as determined through a paired-samples t test, $t(49)$

= 7.12, $p = .0001$. Although the comparison for the expert senior PhD students was also highly significant, (semantic: $M = 2.81$; visual: $M = 2.46$) the difference was not as large as for the novices, $t(49) = 3.78$, $p = .0001$. This confirms Hypothesis 1 predicting that graph complexity ratings obtained from first-semester undergraduates would differ from those obtained from senior graduate students.

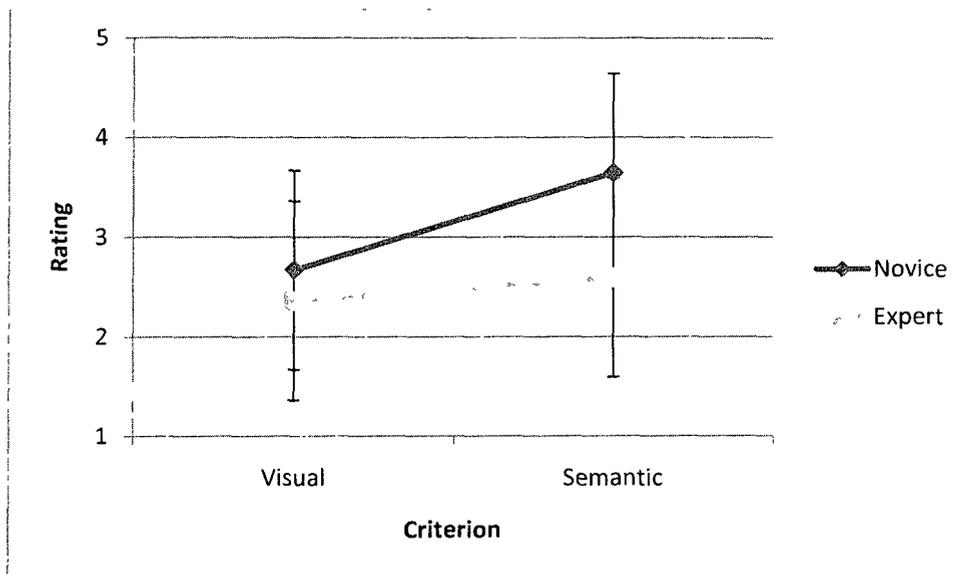


Figure 5. Experiment 1, Expertise x Criterion for mean complexity ratings. Error bars indicate 95% confidence interval (Jarmasz & Hollands, 2009) in this and all subsequent graphs.

The ANOVA also revealed a significant main effect of Criterion, $F(1, 36) = 19.38$, $p = .0001$, $\eta_p^2 = .39$ indicating that the semantic criterion ratings ($M = 3.23$) were higher (i.e. perceived to be more complex) than the visual criterion ratings ($M = 2.53$) thus confirming Hypothesis 2.

Response times were also submitted to a 2 (Expertise: novice, expert) x 2 (Domain: business, psychology) x (2) (Criterion: visual, semantic) ANOVA. The response times are summarized in Figure 6.

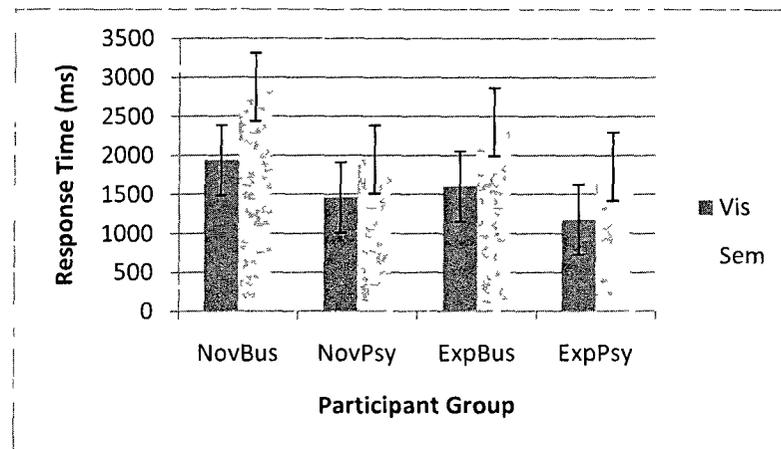


Figure 6. Experiment 1 response times.

The ANOVA yielded a main effect for domain: psychology participants ($M = 1589$ ms) responded faster than business participants ($M = 2212$ ms). The main effect for criterion was also significant, $F(1,36) = 8.39$, $p = .006$, $\eta_p^2 = .44$ revealing that participants rated the visual criterion ($M = 1545$ ms) faster than the semantic criterion ($M = 2255$ ms). No other response time main effects or interactions approached significance.

Summary

The ratings data demonstrated clearly that first-semester undergraduates rated the graphs as semantically more complex than senior graduate students but no differences were observed for the same graphs rated for visual complexity. The lower mean score for the visual criterion than for semantic criterion indicated that participants perceived the task of assessing the visual aspects of the graph stimuli as less difficult than the task of assessing the semantic aspects, confirming Hypothesis 2. Finally, the response time results supported the ratings data insofar as the higher

complexity ratings were associated with longer response times. Overall, a significant Pearson product-moment correlation between ratings and total response time ($r = .371, p < .0001$) indicated that participants did not “trade-off” ratings with the speed of their response.

Graph Stimulus Selection

Experiment 1 demonstrated differences between novices and experts’ semantic ratings. It also revealed a difference between visual and semantic rating criteria, enabling the selection of a smaller sample of stimuli for the subsequent experiments in this thesis. However, one important issue remained. Recall that Carpenter and Shah’s (1995, Experiment 3) participants’ performance approached chance levels, introducing the possibility that their failure to find a graph expertise effect may have been due to overly complex graph stimuli. The opposite problem might be encountered here, where the graphs could have been perceived to be too simple: the median visual complexity rating of all graph stimuli was 2.98 and the median semantic complexity rating was 3.93 on a 9-point scale. Such low complexity ratings might prevent Experiment 2 experts from demonstrating superior performance compared with novices. To address this issue, graph stimuli falling above the median on both visual and semantic criteria were selected such that one set of 12 graphs had a new median visual complexity rating of 3.42 and a semantic complexity rating of 4.62 (see Appendix D12). This set of graph stimuli was deemed sufficiently sensitive to test the manipulations introduced in Experiment 2.

Experiment 2: Graph Exploration By Domain Experts and Novices

The range of graph stimuli selected in Experiment 1 were used in Experiment 2 to identify the graph exploration and interpretation strategies of participants varying in expertise, and to learn more about how domain and expertise may interact. Research in well-defined domains such as arithmetic (Lemaire & Siegler, 1995; Tronsky, 2005) has found that, as skill increases, the strategies available remain constant but the *choice* and *execution* of those strategies improves. Similar research has not been conducted in graph exploration and interpretation. Lemaire and Siegler's results would suggest that experts and novices should exhibit a similar pattern of utterances in their exploration of graphs. However, if the experts' choice and execution of strategies is faster than novices then their response times will be shorter. Hypothesis 1 predicts that experts will take less time than novices to complete their graph explanations.

As discussed earlier (p. 19), Shah and Carpenter (1995, Experiment 3) compared psychology graduate- and undergraduate students using graphs from common-knowledge domains rather than domain-specific graphs as used here. They found no effect of expertise. Using business- and psychology-specific graphs, the present experiment explored the potential effects of graph-exploration expertise. Shah and Carpenter also asked participants to generate an explanation of graphs following the general form "variable *a* causes variable *b*". Again, they found no differences between experts and novices. Both the graph and the participants' domain were manipulated in the present experiment, leading to Hypothesis 2 predicting that experts would generate more explanations than novices.

As discussed earlier (p. 4), expertise tends to be domain-specific. However, the role of domain-specificity as a function of expertise has not been investigated in graph exploration

studies before and this formed another focus of the current experiment. Hypothesis 3 predicted that experts would provide more complete explanations of graphs in familiar than in unfamiliar domains.

Carpenter and Shah (1998) found that the proportion of nominal, ordinal, and metric descriptions of graphs varied across different graph types. Nominal utterances were defined as the names of z -variables without any ordinal or metric information about the z - y relation; ordinal utterances typically mentioned the explicit relationships between z -variables, and metric utterances included descriptions of the interval- or ratio relationship between z -variables. Carpenter and Shah reported that the proportion of ordinal and metric utterances varied with the perceptual salience of the z -variable as manipulated through (for example) the number of parallel x - y lines. Although Carpenter and Shah did not examine the interaction of description type and expertise, it is possible to equate their nominal, ordinal, and metric descriptions with different types of the conceptual messages proposed by Pinker (1990). In that case, extending Pinker's model to include expertise effects would lead to Hypothesis 4 predicting that experts would generate more conceptual messages (nominal, ordinal, and metric combined) than novices.

Finally, in order to better understand how expertise might exert its effects on graph exploration and to better control for potential floor- and ceiling effects, both simple and complex graphs were employed as stimuli in this experiment. Even if superior expert performance cannot be found with simple graphs, more complex graphs might allow experts to demonstrate more complete graph explanations than novices, as predicted by Hypotheses 5 and 6.

Thus, six hypotheses were tested:

1. Expert graph explanation response time will be faster than novice graph explanation response time.

2. Experts will generate more explanations than novices.
3. Experts will provide more complete explanations for graphs in a familiar domain than for graphs in an unfamiliar domain.
4. Experts will generate more conceptual messages than novices.
5. Experts will provide more complete graph explanations than novices.
6. Experts will provide more complete graph explanations for complex than for simple graphs.

Method

Participants

Twenty-six new participants were recruited in the same manner as before. Of the ten (seven female) undergraduate novices, six were majoring in psychology and four in business. The expert sample comprised eight of each psychology (seven female) and business (six female) faculty or senior (PhD candidate) graduate students. In response to the Graph Experience Questionnaire (Appendix D7) five novices, six business experts, and seven psychology experts reported that they had to create graphs. All reported that line graphs were the graphs most familiar to them. The novice undergraduate students were granted 1.0% course credit, and experts were given a \$10 coffee shop gift certificate for their participation. All had normal or corrected-to-normal vision. Participants were tested individually in sessions lasting 75 minutes on average.

Apparatus and Materials

The twelve graphs arising from Experiment 1 were used, six simple and six complex graphs. Each graph was assigned two sets of titles, labels and axes with one drawn from psychology and one from business. The business descriptors were selected from an undergraduate textbook on international business (Griffin & Pustay, 2007), and the psychology descriptors were drawn from an undergraduate textbook on psychology (Weiten & McCann, 2007). The 12 business graphs were the mirror images of the 12 psychology graphs as shown in the typical examples in Figure 7 below. The complete set of stimuli is presented in Appendix E2.

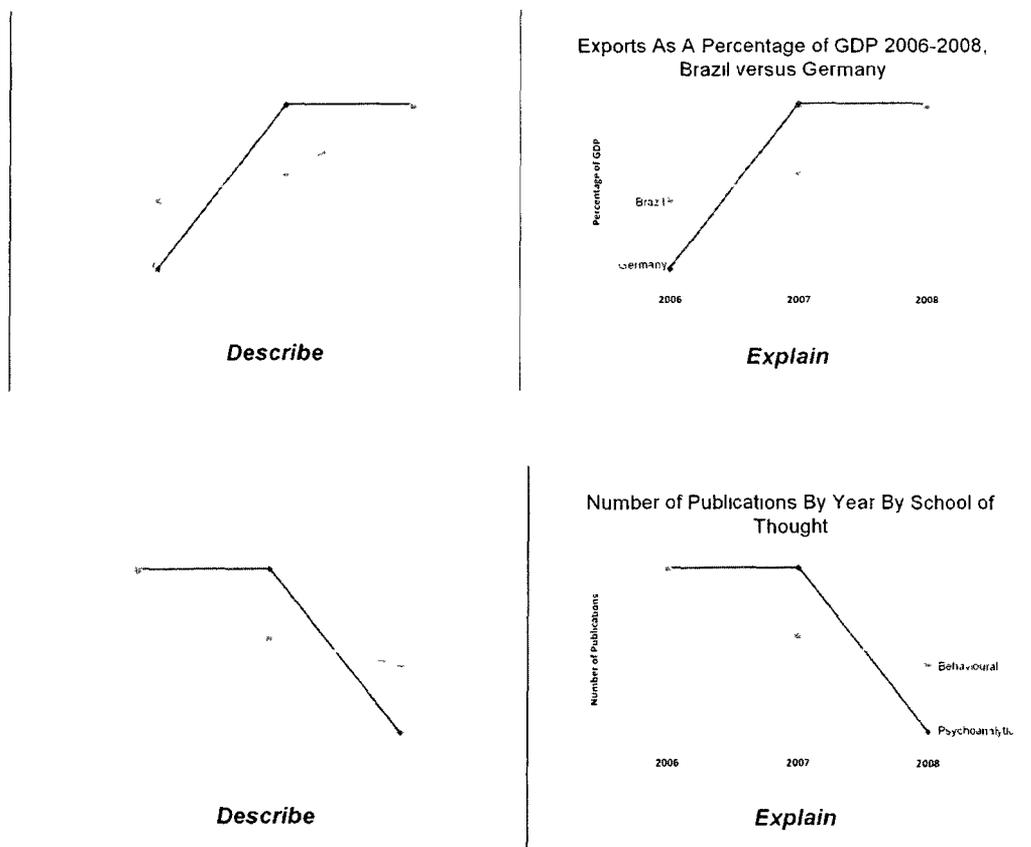


Figure 7. Example graph stimuli: business domain (top) and psychology domain (bottom).

The stimuli, presented one at a time, were randomized; stimulus presentation time as well as response times were controlled by DirectRT™ on a Dell Latitude D610 laptop computer with 1280 X 800 pixel screen resolution. Participant verbalizations were recorded on a Panasonic RR-US500 digital voice recorder.

Design and Procedure

A 3 (Expertise: business expert, psychology expert, novice) x (2) (Graph difficulty: simple, complex) x (2) (Domain: business, psychology) repeated measures design was adopted.

After Preliminary instructions and Informed Consent, the Graph Experience Questionnaire and the Visual Patterns Test (VPT, see Appendix D8) were administered in counterbalanced order, followed by detailed experimental instructions. Four practice trials were followed by 24 experimental trials, each initiated by pressing the spacebar. On each trial a randomly selected graph without labels or titles was displayed with the word “describe” played over the computer speakers as well as appearing at the bottom of the display. When done, participants were instructed to press the spacebar wherein the graph was re-displayed with the corresponding business or psychology labels and titles accompanied by the instruction to “explain” played over the computer speakers and on the screen as before. Total response time was the total time required for each explain-graph trial and was composed of two periods of interest: Silent time (ST), taken from the digital recordings, was defined as the silent period between the start of each trial and the initiation of the verbal explanation. Explain time (ET) was the duration in which participants were explaining each graph and was delineated by the start of a participant’s utterance and their spacebar press which ended the current trial and initiated the

next. Practice-trial data were not analyzed. At the end of the experiment participants were debriefed, thanked, and paid (if applicable).

Data Analysis

Verbal protocols were transcribed ad verbatim, coded, and analyzed with NVIVO™ Version 8.0. Frequency and order of occurrence of utterance-types were calculated, as was the presence/absence and completeness of explanations and the frequency of conceptual messages (sum of nominal, ordinal, and metric utterances combined). Interrater reliability was assessed by having an independent rater code a randomly selected 15% of the verbal protocols. Silent time and talk time were also analysed.

Results

Due to DirectRT™ software errors, two business- and two psychology graphs were removed, leaving 10 graphs representing each domain. The results are described in subsections corresponding with the visual descriptions, semantic explanations, serial order, and the various hypotheses. (Please note that the order of exposition is logical but does not adhere to the order of the Hypotheses in this experiment.) A penultimate section deals with time measures, and the final section briefly examines the VPT results. All descriptive statistics and ANOVA summary tables are provided in Appendices E3 through E17.

Visual Descriptions of Graphs

The number of words used to visually describe graphs showed that novices said less than experts, as shown in Figure 8. This was confirmed by a 3 (Expertise: novice, business expert, psychology expert) x 2 (Difficulty: simple, complex) x 2 (Domain: business, psychology) repeated-measures ANOVA on word count. The main effect of expertise was significant, $F(2, 23) = 3.82, p = .037, \eta_p^2 = .25$. Post hoc Tukey tests showed that novices used fewer words ($M = 37.79$) than psychology experts, ($M = 86.88$) $p = .031$, but not fewer than business experts, ($M = 52.71$) $p = .688$. This provided partial support for Hypothesis 2. No other effects approached significance.

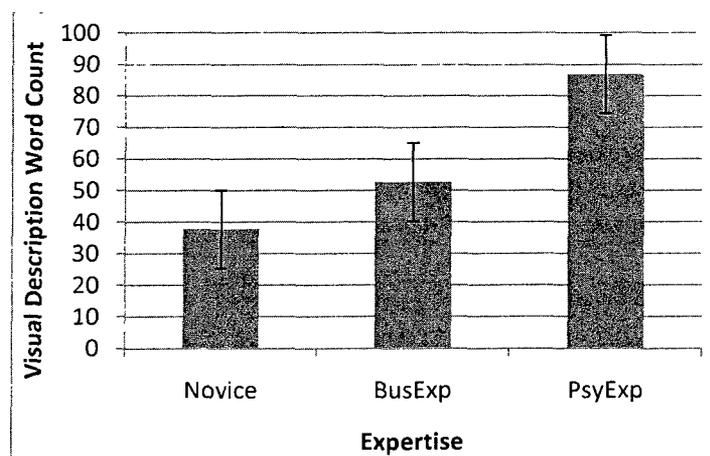


Figure 8. Word count for visual description task for novices, business, and psychology experts.

Semantic Explanations of Graphs

Initial coding of the graph explanation protocols resulted in 23 unique codes which were subsequently distilled into nine themes, all of which could be mentioned in each of the graphs. Typical examples of codes not included in the final analysis were interjections (“uh, um, so”); relative and absolute utterances (“lighter coloured line” versus “grey line”) and were

subsequently combined. Note that there was no evidence of the verbal labels or naming mentioned by Gobet and Simon (1996a) and discussed previously (p. 28). The final nine themes are shown for each expertise group in Table 3. A randomly selected sample of 15% of the graph explanations were coded by an independent rater and the percentage agreement was 90.0%. The values are given as proportions of the total number of trials per expertise group to enable direct comparison of the different expertise-group sizes. Because the themes are not mutually exclusive, they do not sum to 1.0.

Clearly, novices and experts differed in the frequency with which they mentioned most themes. Novices used the Because theme infrequently, and they were far less likely to propose potential inferences than the experts. Discussion of Between z-variables (here China and Japan) differed as well as shown in the typical example below in which one psychology expert explained:

...1996 to 2005 again both of their, their proportion of the populations of young people increased for both of them however increased far more for China and they overtook Japan whereas apparently the proportion of young people in China is higher than Japan.

In contrast, novices tended to explain z-variables one at a time, as one of the novices said of the same graph:

Okay, so this is in terms of population over years in China there were a lot of people over 30 in the year 1976 to 1985 and then it steadily decreased in 1986 1995 and then something happened and it rose steadily this in 1996 to 2005.

Table 3

Examples and Mean Proportion of Themes Uttered by Novices, Business Experts, and Psychology Experts

Theme	Examples	Novice	BusExp	PsyExp
BECAUSE: Attributions of causality	“don’t know whether they’ve had a change in government or if officials have just gotten a lot more corrupt but ... for whatever reason corruption was really low in 2007...”	.19	.41	.54
BETWEEN Z- VARIABLE: Comparisons between z-variables	“In 2008 the big 10 and the emerging economies have an equal amount of annual average growth in their Gross Domestic product”	.61	.83	.97
DIRECTION: Within a single z- variable	“Azerbaijan is expected to remain stable ...over um 2010 to 2011, but then is predicted to decrease their instability of oil pipelines in 2012.”	.52	.58	.46
QUANTITATIVE: Interval or ratio relationship	“the difference increases dramatically in 2007. It is maybe 5 times or 4 times greater in 2007 where Pakistan is doing better.”	.03	.25	.39
TITLE: Repeat the title of the graph	“hypnotic susceptibility by field dependence by gender”	.32	.91	.92
TREND: Overall direction	“over a 3-year span, both groups seem to be decreasing the number of publications they’re putting out”	.11	.18	.35
X-AXIS: References to the abscissa	“x-axis shows Day 1, Day 2, Day 3”	.25	.45	.36
Y-AXIS: References to the ordinate	“The y-axis shows GDP—adjusted GDP—in billions of US dollars.”	.23	.40	.28
Z-Variable: Number or name of z- variables	“The two lines represent—repre—ah, ah, respectively, the scores for males and for females...”	.18	.20	.25

The Quantitative theme, also mentioned more frequently by experts than by novices, included utterances such as “However by 2007 the number of publications for behavioural school decreases; practically, I would say, by 45 percent” as well as attempts to characterize the function such as “India, again, more of a linear pattern.” Likewise, attempts to relate statistical functions to the graphs were included in this category: “Don’t really know what the scale is, so it might not be different after all, statistically.”

As mentioned above the themes assigned to each participant’s semantic explanation of each graph were not mutually exclusive. In coding the verbal protocols, each participant statement was assigned to all the categories or themes to which it applied. For example, if a participant commented on the increase in gas prices (z -variable) over the course of several years (x - axis) the statement was coded as both a z -variable and x -axis. Nonetheless, the presence of more themes in a given participants’ graph explanation can be conceived as representing a more complete explanation of the graph than an explanation with fewer themes present. A completeness score was calculated by determining the proportion of all nine themes present in each participant’s explanation of each graph. The mean completeness scores are shown in Figure 9 for each expertise group and for business and psychology domains. The Figure suggests that the two expert groups’ explanations were more complete than those of novices and that this was more pronounced for psychology than for business graphs. This was confirmed by a repeated measures 3 (Expertise: novice, business expert, psychology expert) x (2) (Difficulty: simple, complex) x (2) (Domain: business, psychology) ANOVA on completeness scores. The main effect of expertise was significant, $F(2, 23) = 8.02, p = .002, \eta_p^2 = .41$ and independent Tukey post hoc comparisons confirmed that business experts ($M = .47$) provided more complete

explanations than novices ($M = .27$), $p = .014$, and the same was also true for the psychology experts ($M = .51$), $p = .004$. This confirms Hypothesis 5. There was also a main effect for domain, $F(1, 23) = 9.90$, $p = .005$, $\eta_p^2 = .30$ confirming only that stimulus differences existed since completeness scores were higher for psychology than for business graphs. There were no other effects. Because neither the main effect of graph difficulty nor its interactions approached significance, Hypothesis 6 predicting that experts would generate more complete explanations of complex than of simple graphs was refuted.

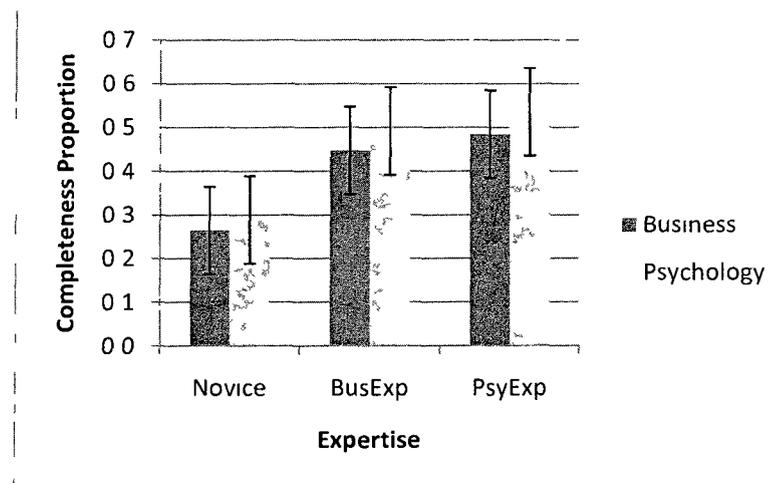


Figure 9. Mean Completeness Scores for Expertise and Graph Domain

Serial Order in Explanations of Graphs

The higher completeness scores for experts than for novices indicate that the *content* of their graph explanations differed. In addition, the *serial order* in which explanations were provided differed also between the groups. For example, the two experts achieving the highest mean completeness score (business .79, psychology .66) explained every graph (irrespective of domain) in a consistent serial order, namely:

- Step 1 Title
- Step 2 Y-axis
- Step 3 X-axis
- Step 4 Z-variable
- Step 5 Between z-variables
- Step 6 Because

The example below is a business expert's graph explanation. The numbers in parentheses indicate the themes listed above:

[1] The chart shows average annual GDP growth during the period from 2006 to 2008 for ah four countries: Brazil, Russia, India, and China. The emerging powers versus the Big10. [2] The Y graph—or the Y-axis on the graph—shows average annual growth in GDP as a percentage. But again, the specific index values along the Y-axis are missing. I don't know what the absolute values are, although you can see relative positioning of the, ah, scores for the two groups of countries. [3] The X-axis shows three time periods during which the data—er, ah, (that) the data represent...or the...yeah...that the data were taken for. [4] Ah, in 2006, the average annual growth in GDP for the Big10 was higher than for ah Brazil, Russia, India, and China. [5] Ah, in 2007, the relative positions were reversed. And the average annual GDP growth rate for Brazil, Russia, India, and China was higher than the average annual GDP growth rate for the Big10. And then in 2008, um, the average annual GDP growth rate for the Big10 had rebounded to a point. And the average annual growth rate—ah, GDP growth rate—for Brazil, Russia, and India, and China had continued to decline to a point where the ah, ah, the data for the two groups of

country in 2008 converged at what appears to be the same average annual growth rate for both groups of countries. That's it.

No such systematic order was observed in the novices' explanations, as shown in a typical example for the same graph:

The average annual growth in GDP percentage with Big10 decreases since, start decreasing start 2006 and, and uh, increasing after 2007. The average annual growth in GDP percentage with BRIC starts decreasing since 2007 and decreases more, decreasing faster after 2007.

Table 4 shows the frequency of initial (Step 1) themes for each expertise group. Thus, business- and psychology experts mentioned the Title first much more frequently than novices, who were equally likely to start by explaining the Between z -variable relationships or, and somewhat less likely, by discussing the direction of a single z -variable. The frequency of step 2, 3, 4, 5, and 6 utterances are presented in Appendix E18.

Taken together, experts and novices' explanations differed in terms of completeness, offering inferences, quantitative explanations, between z -variable relationships, and in the order in which these utterances were typically made. Thus, it is safe to conclude that the strategies experts adopted were more complete and that the order in which the component steps of those strategies differed systematically from novices' strategies.

Table 4

Mean Frequency of Step 1 Themes by Novices, Business, and Psychology Experts

Step 1 Themes	Novice	BusExp	PsyExp
Title	.34	.79	.95
Between z-Variable	.33	.00	.02
z-Variable	.10	.18	.02
x-Axis	.02	.00	.01
y-Axis	.04	.01	.01
Because	.00	.01	.00
Direction	.15	.00	.00
Quantitative	.00	.00	.00
Trend	.01	.01	.00

Proportion of Explanations

Although all participants were asked to “explain the graph as if you were the author and you were explaining the results to another person”, utterances of the general form “variable *a* causes variable *b*” were observed in fewer novice- than expert trials. For example, a psychology expert inferred that “insomniacs apparently have more work problems than controls do, uh probably because they just don’t get enough sleep so they have trouble at work.” Figure 10 illustrates that experts voiced more explanations than novices. A repeated measures 3 (Expertise: novice, business expert, psychology expert) x (2) (Difficulty: simple, complex) x (2) (Domain: business, psychology) ANOVA on explanations revealed a significant main effect of

expertise, $F(2, 23) = 4.73$, $p = .019$, $\eta_p^2 = .29$. Independent post hoc Tukey tests confirmed that psychology experts ($M = .57$) attempted more explanations than novices ($M = .18$), $p = .015$; the difference between business- and psychology experts was not significant ($p = .440$), and nor was the difference between business experts and novices ($p = .214$). Hypothesis 2 stating that experts would provide more explanations than novices was thus supported. No other effects approached significance.

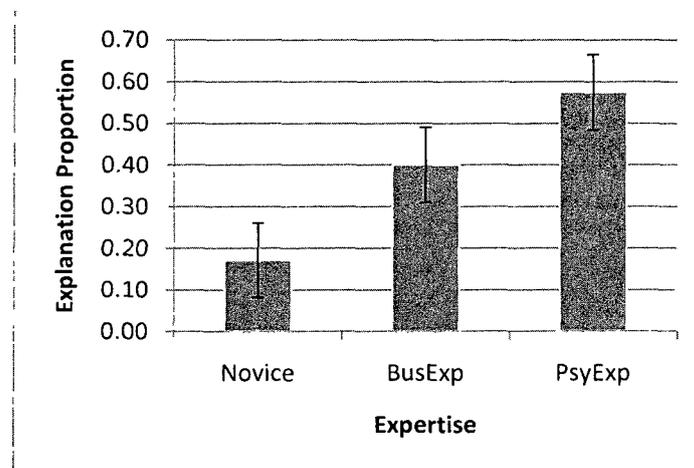


Figure 10. Proportion of explanations provided by novices, business, and psychology experts.

Familiar and Unfamiliar Domain Explanations

Hypothesis 3 predicted that experts would generate more complete explanations of graphs in familiar than unfamiliar domains. Excluding novices, a repeated measures 2 (Expertise, business, psychology) x 2 (Domain: familiar, unfamiliar) x 2 (Difficulty: simple, complex) ANOVA resulted in only one significant effect, the Expertise x Domain interaction, $F(1, 14) = 6.56$, $p = .023$, $\eta_p^2 = .56$ and there were no other significant main effects or interactions. Post hoc t -tests for independent samples confirmed that the interaction was due to significantly higher completeness scores for psychology experts on familiar domain graphs ($M =$

.53) compared to the unfamiliar domain ($M = .48$), $t(7) = 3.30$, $p = .013$, and business experts exhibited the opposite effect of significantly *lower* completeness scores on familiar domain graphs ($M = .45$) compared to the unfamiliar domain ($M = .49$), $t(7) = -2.71$, $p = .030$. Figure 11 illustrates the mean completeness scores for business- and psychology experts (excluding novices) for graph domain. Thus Hypothesis 3 was partially confirmed by psychology experts but refuted by business experts.

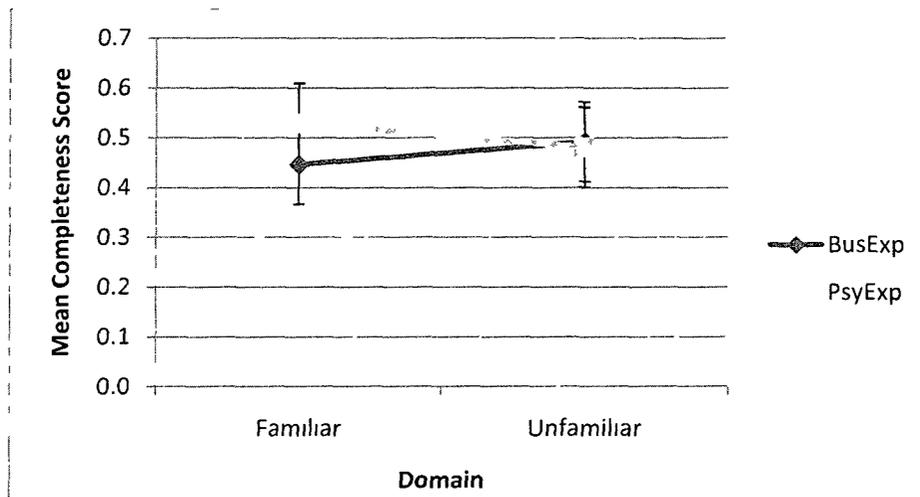


Figure 11. Business- and psychology expert mean completeness scores as a function of familiar versus unfamiliar graph domain.

Nominal, Ordinal, and Metric Conceptual Messages

In addition to the completeness score discussed previously (p. 59) another derived measure is the mean proportion of conceptual messages which is the sum of z-variable, between z-variable, and quantitative proportions. It will be recalled that conceptual messages were equated with discrimination nets discussed earlier (p. 30). The individual proportions are equivalent to Carpenter and Shah's (1995) nominal, ordinal, and metric measures respectively.

(Similar to the completeness score, the number of conceptual messages may not sum to 1.0.) A 3 (Expertise: novice, business expert, psychology expert) x (2) (Difficulty: simple, complex) ANOVA of the conceptual messages resulted in a significant expertise main effect, $F(2, 23) = 7.80, p = .003, \eta_p^2 = .40$ which is illustrated in Figure 12.

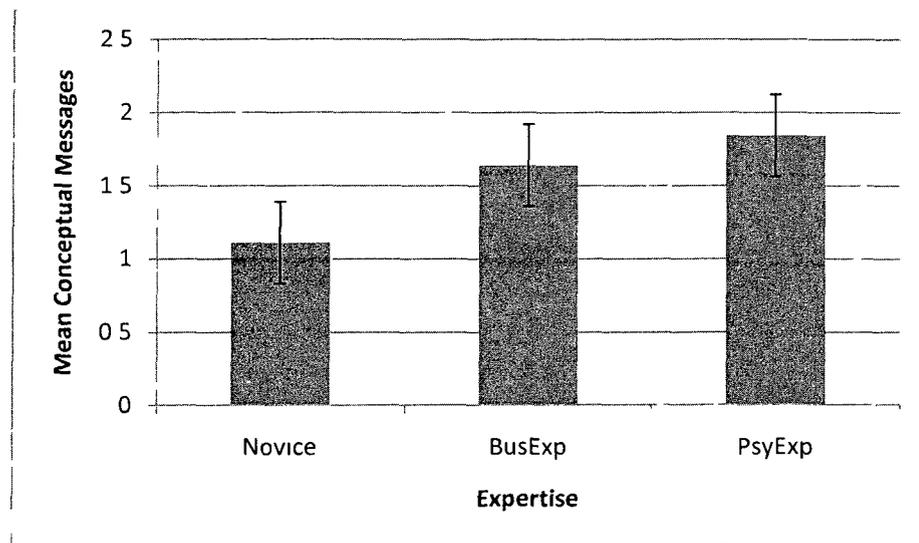


Figure 12. Mean conceptual messages generated by novices, business experts, and psychology experts.

Hypothesis 4 stated that experts would generate more conceptual messages than novices. Planned comparisons indicated that business experts ($M = 1.64$) generated more conceptual messages than novices ($M = 1.11$), $p = .012$, psychology experts ($M = 1.55$) generated more than novices, $p = .013$ but business and psychology experts did not differ from each other, $p = .59$. Thus Hypothesis 4 was confirmed. No other comparisons approached significance.

Time Measures

Recall that the Total Response Time (TRT) for the semantic explanation task was composed of a Silent Time (ST) plus an Explain Time (ET). ST was the silent period before

participants began their graph explanations and ET was the actual time during which participants voiced their semantic explanations. A repeated measures 3 (Expertise: novice, business expert, psychology expert) x (2) (Difficulty: simple, complex) x (2) (Domain: business, psychology) ANOVA for TRT revealed no main effect for Expertise ($p = .817$) nor any significant interactions with Expertise, suggesting that the efficiency with which experts completed the graph explanation task was no better than that of novices. The main effect for Difficulty was, however, significant, $F(1, 23) = 11.13, p = .003, \eta_p^2 = .33$ confirming that total response time was longer for complex ($M = 75$ s) than for simple graphs ($M = 70$ s). The main effect of Domain was also significant, $F(1, 23) = 5.83, p = .024, \eta_p^2 = .20$ confirming that participants took more time to explain psychology domain graphs ($M = 76$ s) than business graphs ($M = 70$ s). There were no other significant effects.

A repeated measures 3 (Expertise: novice, business expert, psychology expert) x (2) (Difficulty: simple, complex) x (2) (Domain: business, psychology) ANOVA for ET revealed no main effect for Expertise ($p = .478$) nor any significant interactions with Expertise. The only significant effect was Difficulty, $F(1, 23) = 9.29, p = .006, \eta_p^2 = .29$ reflecting that participants required less time to explain simple graphs ($M = 64.87$ s) than complex graphs ($M = 69.85$ s). One other measure suggests that experts were more efficient in their graph explanations than novices. A 'words per second' measure was calculated for ET by a repeated measures 3 (Expertise: novice, business expert, psychology expert) x (2) (Difficulty: simple, complex) x (2) (Domain: business, psychology) ANOVA for words per ET seconds. It revealed a significant main effect of Expertise, $F(2, 23) = 8.64, p = .002, \eta_p^2 = .43$. An independent post hoc Tukey test confirmed that the number of words/ET second was significantly lower for novices ($M = 1.44$) than for business experts ($M = 1.83$), $p = .002$ and for psychology experts ($M = 2.03$), $p = .000$.

Finally, a repeated-measures 3 (Expertise: novice, business expert, psychology expert) x (2) (Difficulty: simple, complex) x (2) (Domain: business, psychology) ANOVA on ST revealed a significant main effect of expertise, $F(2, 23) = 7.71, p = .003, \eta_p^2 = .41$ shown in Figure 13 below. Independent Tukey post hoc tests confirmed that novices had longer silent periods before beginning their semantic explanations ($M = 8.95$ s) than business ($M = 2.17$ s), $p = .003$, or psychology experts, ($M = 3.37$ s), $p = .016$. If it is assumed that ST represents the time required to select and/or initiate a strategy then Hypothesis 1 is confirmed: Experts require less time to select their graph explanation strategies than novices. This notion is partially supported by the finding that even though novices paused a significantly longer time than experts before beginning their graph explanations, their total response time was not different from either of the expert groups. This suggests that novices were uncertain about what to say or perhaps where to start their graph explanations, which is also reflected in their words per ET seconds measure.

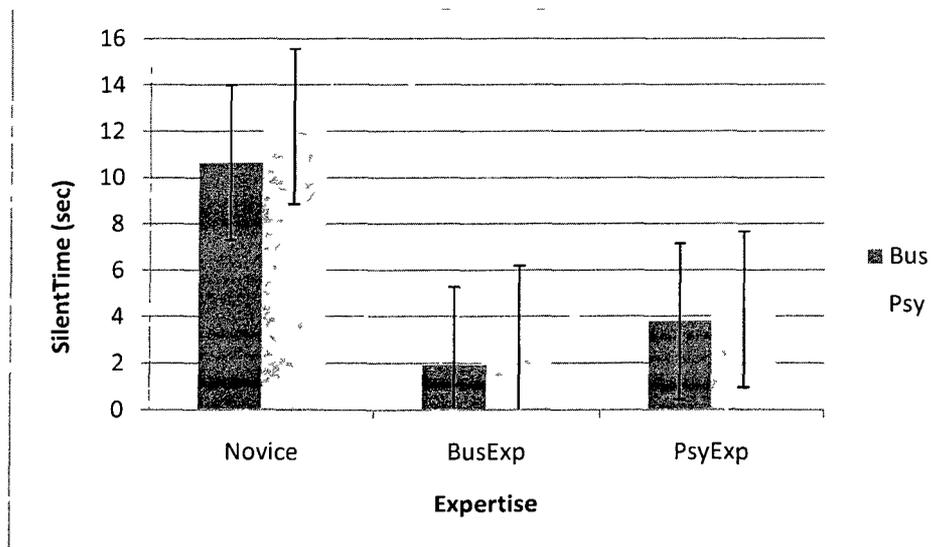


Figure 13. Mean silent time by Expertise (novices, business, and psychology experts) x Domain (business, psychology).

Visual Patterns Test

The VPT was administered to determine if short-term visual memory might be associated with semantic explanations of graphs. However, VPT scores failed to correlate significantly with completion scores and will therefore not be discussed further. The VPT results are shown in Appendix E3.

Results Summary

The results of Experiment 2 are summarized in Table 5 and the relationship of the six hypotheses to those results are summarized in Table 6.

Table 5
Experiment 2 Summary of Results

Measures	Results
Experts' graph exploration strategies	<ul style="list-style-type: none"> • More complete explanations by experts than novices • More titles by experts than novices • More consistent order for experts than novices • More between z- comparisons for experts than novices • More quantitative themes for experts than novices • More causal inferences for experts than novices
Time Measures	<ul style="list-style-type: none"> • Total response time for complex graphs longer than simple graphs • Total response time for psychology graphs greater than business graphs • Explain time greater for difficult graphs than simple • Silent time for novices greater than experts • More words per second for experts than novices
Visual descriptions	<ul style="list-style-type: none"> • Visual descriptions by experts longer than novices

Table 6

Experiment 2 Hypotheses, Results, and Conclusions

Hypotheses	Results	Conclusions
H1. Experts' strategy choice and execution superior to novices	<ul style="list-style-type: none"> • Experts used more words to visually describe graphs • Experts made more Between z-variable comparisons • Experts made more quantitative explanations • Experts made more complete explanations • Experts think time faster than novices 	H1 supported as measured by silent time and words per second, expert strategy selection is more efficient. Experts use more themes than novices.
H2. Experts will generate more explanations than novices	<ul style="list-style-type: none"> • More explanations by BusExp and PsyExp than novices • Similar number of explanations by BusExp and PsyExp 	H2 supported
H3. Experts will provide more complete graph explanations for graphs in their own domain for than for graphs in another domain.	<ul style="list-style-type: none"> • PsyExp same domain explanations more complete than business domain explanations • BusExp same domain explanations similar completeness scores as psychology domain 	H3 supported by PsyExp explaining same domain graphs but refuted by BusExp
H4. Experts will generate more conceptual messages (nominal, ordinal, and metric) than novices.	<ul style="list-style-type: none"> • BusExp and PsyExp generated more conceptual messages than novices 	H4 supported
H5. Experts will generate more complete explanations than novices	<ul style="list-style-type: none"> • Higher completeness scores for experts than novices 	H5 supported
H6. Experts generate more complete explanations for complex than simple graphs	<ul style="list-style-type: none"> • No difference between simple and complex graphs. 	H6 refuted

Discussion

In contrast to Shah and Carpenter (1995), who found no differences between experts and novices on any of their dependent variables, in Experiment 2 experts used more words than novices to visually describe graphs: they made more between z -variable comparisons; they generated more quantitative explanations; and their explanations were more complete than those of novices. In contradiction to Lemaire and Siegler's (1995) results in children's arithmetic, the current experiment showed that experts differed from novices in the number and order of themes mentioned. In addition, the dependent measures of silent time and words per second suggested that experts were more efficient as well. The levels of participants' experience in Shah and Carpenter's research may have been insufficient to reveal any differences in the level of expertise. In addition to recruiting participants here who were more likely to differ in graph-interpretation expertise, the graphs were also selected specifically to avoid floor effects or ceiling effects. The absence of any graph complexity effects refutes Hypothesis 6 and contradicts the results of Schunn and Anderson (1998, 1999). One explanation for the different findings is that Schunn and Anderson's participants interacted with tabular data as opposed to the line graphs employed here. Also, the wider range of expertise among the participants employed here may account for the difference between the current results and Shah and Carpenter's research.

Hypothesis 2 predicting that experts would provide more causal inferences ('Because' themes) than novices was clearly supported, again in contradiction with Shah and Carpenter's results (1995, Experiment 3) but consistent with expectations arising from Feldon et al.'s (2010) research in the domain of biology. One possible explanation for the differing results is that novices in the current experiment may have found the graphs somewhat more difficult to explain than those employed by Shah and Carpenter, therefore providing a lower proportion of

explanations ($M = 0.17$) compared to Shah and Carpenter's novice and expert participants ($M = 66\%$). (Shah and Carpenter did not report the proportion of explanations provided by their experts and novices separately.)

Hypothesis 3 predicted that the business and psychology experts' completeness scores would differ by familiarity of domain. Although psychology experts did score higher on familiar domain graphs, the opposite was true for business experts. It is possible that the business graph stimuli were insufficiently domain-specific to yield any domain-specific differences. Therefore Hypothesis 3 was only partially supported.

Hypothesis 4, which predicted that experts would generate more conceptual messages than novices, was supported. However, Carpenter and Shah (1998) found more ordinal and metric explanations when multiple x - y lines were presented. They argued that three lines made the ordinal and metric z -variables more perceptually salient than one or two lines. In the current experiment, all graphs had two lines. Thus, if Carpenter and Shah's argument is correct, there should be no difference in the perceptual salience of these graphs. Therefore, it could be argued that the higher mean proportion of metric utterances ($M = .25$) observed with complex graphs compared to simple graphs ($M = .19$), was due to differences in processing at a conceptual rather than merely at a perceptual level. This would support Pinker's (1990) post-perceptual notion of conceptual messages created in the interpretation of graphs.

Experiment 2 confirmed that experts' and novices' graph explanation strategies differed. The observation that novices took longer than experts to begin their explanation invites three possible explanations. First, it is likely that their memory for graph explanation strategies is limited, rendering retrieval more time-consuming than for experts. Second, one may assume that novices' memory for such strategies, presumably stored in graph schemata, would be less well

organized than those of the experts. This was evidenced in the novices' lack of systematic approach to explaining graphs, which also differed from that of the experts. Third, it is possible that the novices tested here had inappropriate meta-strategies (Lemaire & Siegler, 1995) for graph exploration. In that case, they would have been unable to identify the graph features that would indicate an optimal strategy. The present data are insufficient to clearly distinguish among the three possibilities. Nonetheless, Shah and Carpenter's (1995, Experiment 3) results led them to suggest that novices might be trained to perform as well as experts. The effect of simple instruction on graph exploration strategies was explored in Experiment 3.

Experiment 3: Graph Exploration Instruction For Novices

In the domain of simple arithmetic, Lemaire and Siegler (1995) found that, with increasing expertise, children improved the frequency with which better strategies were selected, in how the strategies were executed, and in their meta-strategic strategy selection. Similar results have been reported by Tronsky (2005) using adult students solving multiplication problems and by Schunn, McGregor, and Saner (2005) who studied adults solving problems in the domain of military leadership. In addition, Feldon et al. (2010) found that certain types of instructions facilitated the acquisition of data exploration skills that bear considerable similarity to the graph exploration skills that are the subject of this thesis. Extrapolating from that research, it is reasonable to assert that appropriate instructions might enable novices to perform like experts in this experiment. The following six Hypotheses, all of which are based on a comparison of before- and after instruction performance, were tested:

1. The length of visual descriptions can be increased by instructions

2. Completeness scores can be increased by instructions
3. Because explanations can be increased by instructions
4. Between z-variable comparisons can be increased by instructions
5. Graph titles can be increased by instructions
6. The number of conceptual messages can be increased by instructions.

Method

Participants

A total of 21 new undergraduate novice participants (11 female, 10 male) majoring in psychology (n=10, 7 female, 3 male) or in business (n=3, 0 female, 3 male) or science (n=8, 3 female, 5 male) were recruited and rewarded for their participation in the same manner as before. All participants said they created graphs and that they were most familiar with line graphs. Participants' mean age was 21.5 years; all were tested individually in sessions taking 45 minutes each.

Apparatus and Materials

Because the experiment included an instructional session, only a subset of eight semi-randomly selected graphs was included from Experiment 2 to avoid fatigue effects (see Appendix F2). The same practice stimuli were used as before. The instructional materials were presented in eleven screens (Appendix F3) accompanied by example graphs informing participants to:

- A. Describe as much of the visual characteristics as you can. It will help you later when you explain the graphs.

- B. Read the graph title out loud and then match parts of the title to the vertical axis, horizontal axis, and each of the two lines in the graph.
- C. Look at the area between the lines. Are the lines parallel? Do they cross? Does one go up and the other down? Do they share any points in common?
- D. Consider cause and effect. What causes the lines to go up, go down, or to stay the same? Is that cause operating equally on one line versus the other?

All other materials were identical to Experiment 2.

Design and Procedure

After preliminary instructions and Informed Consent, the Graph Experience Questionnaire was administered. Four practice trials were then followed by eight experimental graph trials, presented in random order as before. All participants then worked through the instructional session of 11 self-paced screens. This was followed immediately by eight different randomly presented graph stimuli that included different titles and axis labels. No rest period was offered in this experiment. At the end of the experiment participants were debriefed, thanked, and paid (if applicable).

The practice trials were again discarded; verbal protocols were transcribed, and the same nine themes as before were used in the data interpretation. Like Experiment 2, 15% of the verbal protocols were coded by an independent rater and interrater agreement was 76.0%. A within-subjects design was adopted. Although the graph stimuli included both business and psychology domains and simple and complex graphs these attributes were ignored in the analyses performed. Instead, planned comparisons focused on the six hypotheses (p. 74).

Results and Discussion

Visual Descriptions of Graphs

The mean number of words describing the visual aspects of the graphs before- and after instruction showed that novices said more after ($M = 47.10$) than before instruction ($M = 33.15$). A t -test for dependent samples confirmed this difference, $t(20) = 3.18$, $p = .005$ thus supporting Hypothesis 1.

Semantic Explanations of Graphs

The proportion of the nine themes observed in the before and after instruction phases are shown in Table 7. Significant before and after instruction differences were determined with dependent t tests. Comparable means from Experiment 2 are also included as a check that the two groups of novices were drawn from similar populations, and to determine how similar the after instruction novices were to Experiment 2 business and psychology experts combined. As before, a completeness score was calculated. The mean completeness score clearly increased as a consequence of instruction ($p = .001$) thereby supporting Hypothesis 2. In addition, the after instruction completeness score was identical to that of Experiment 2 experts.

Table 7

Experiment 3 before- and after-instruction means. Experiment 2 novice and expert means are included for comparison purposes.

	Exp 3	Exp 3	Exp 2	Exp 2
Theme	Before-	After-	Novices	Experts
	Instruction	Instruction		
Completeness	.36	.49**	.27	.49
Because	.01	.23**	.19	.47
Between z-variable	.60	.76	.61	.90
Direction	.52	.43	.52	.53
Quantitative	.05	.08	.03	.32
Title	.64	.73	.32	.92
Trend	.33	.07*	.11	.27
x-axis	.16	.75***	.25	.40
y-axis	.17	.77***	.23	.34
z-variable	.77	.56	.18	.23

* $p < .01$

** $p < .001$

*** $p < .0001$

Novice participants also significantly increased their proportion of 'Because' themes after instruction, consistent with the predictions of Hypothesis 3. However, their final proportion of Because themes was still only half that of Experiment 2 experts. One possible explanation of that difference is the far more labile nature of novice strategies. A few brief minutes of

instruction is capable of substantially changing the way that novices explain graphs in part because of their strategies were less well-developed than those of the experts. As a consequence, some random variation in the proportion of explanation themes may be expected.

Hypothesis 4 predicted that participants would increase their proportion of 'Between z -variable' themes after instruction. Although 'Between z -variable' themes increased, this increase did not achieve significance thus refuting Hypothesis 4. Following Shah and Carpenter's (1995) line of argument, this may be due to fundamental processing limitations associated with z -variables. If this is true, the simple instructions provided here would be insufficient to substantially increase the proportion of 'Between z -variable' themes. Similarly, participants' reiteration of the Title of a graph increased slightly after instruction but this increase was not statistically significant. Therefore, Hypothesis 5 was refuted. One likely explanation for this result is that instructions stressed the relationship of the graph title to the axes. Rather than simply repeat the title, these novices attended to the x - and y -axes instead. In fact, a substantial increase in x -axis and y -axis themes was observed, consistent with the notion of the development of description level knowledge (Pinker, 1990; van Lehn & van de Sande, 2009) mentioned earlier (p. 13). Clearly, the instructions encouraged participants to attend much more to these aspects of the graph than before training. The x - and y -axis themes were well in excess of the similar values of Experiment 2 experts for reasons similar to those mentioned earlier in the context of Because themes.

The Trend themes decreased significantly after instruction. Trend themes included utterances such as "Females grow gradually more...more...le—sorry, females grow, like, gradually less alert throughout the day." Since trend was not explicitly mentioned in the instruction, one explanation for the decrease is that participants followed the instructions and

prioritized other aspects of the graphs (e.g., x -, and y -variables) above noting the trend. No other before/after instruction comparisons approached significance.

Serial Order of Themes

A comparison of the most common serial order of themes before and after instruction is provided in Table 8. The values in parentheses represent the proportion of time the theme appeared in that serial position. The graph title was thus the most common theme in Step 1 but it appeared in only 0.14 of all trials in the after instruction phase. Likewise, y -axis was the most common Step 2 theme but it only occurred in this position in 0.07 of all after instruction trials. Apart from Step 1, there appears to be little consistency in the order of themes either before or after instruction. Nor was there any apparent relationship between the most common Step 1 themes and the order of themes prescribed in the instructions (see p. 75-76).

The proportion of trials in which a given theme was mentioned in Step 1 is shown in Table 9. Although the graph title was mentioned in Step 1 more often after instruction than before, this difference was not statistically significant. However, participants mentioned the y -axis in Step 1 more frequently after instruction than before. A plausible explanation for these differences is merely the overall frequency of these themes as reflected in Table 7. There is no evidence of any systematic strategy of participants to mention a particular theme in a particular order apart from the graph title.

Table 8

Most common themes before and after instruction by Step. Each theme listed was the most common observed for each step. Values in parentheses are proportion of instances each theme was observed in each step.

Step	Before-Instruction	After- Instruction
1	Title (.12)	Title (.14)
2	z-Variable (.06)	y-Axis (.07)
3	Between (.04)	x-Axis (.08)
4	Direction (.01)	z-Variable (.07)
5	Trend (.001)	Between (.05)
6	Quantitative (.001)	Because (.01)

Table 9

Mean proportion of initial (Step 1) graph explanation themes before and after instruction

Initial (Step 1) Themes	Before- Instruction	After- Instruction
Because	.00	.01
Between z-	.07	.04
Direction	.07	.05
Quantitative	.01	.00
Title	.59	.68
Trend	.10	.00***
x-Axis	.03	.03
y-axis	.02	.19***
z-variable	.10	.00***

*** $p = .0001$

Nominal, Ordinal, and Metric Conceptual Messages

If the instruction had been completely successful in transforming the novice participants into experts, the number of conceptual messages (combined nominal, ordinal, and metric graph-explanation utterances) would have increased. However a dependent t test on Conceptual Messages (before, after instruction) revealed no significant effect of instruction, $t(83) = 0.27, p = .611$. Thus, Hypothesis 6 was refuted. No other effects approached significance.

Time Measures

Because the instruction encouraged participants to say more about each graph, it is not surprising that total response time was shorter before than after instruction ($M = 40.4$ s, $SD = 18.1$; $M = 66.3$ s, $SD = 27.3$ respectively). By contrast, instructions led to a significant *decrease* in silent time, from an average of 6.07 s ($SD = 4.07$) to an average of 4.13 s ($SD = 3.32$), $t(83) = 4.64, p = .000$. Thus, even though participants said more after instruction, they required *less* time to begin their explanations. A plausible explanation of the silent time results is that the instructions gave participants some of the strategies they required to explain the graphs. As a consequence, they required less time to retrieve those strategies from LTM.

Summary

Taken together, the above results demonstrate that even a small amount of instruction based on experts' strategies as uncovered in Experiment 2 could help novices perform more like experts in some aspects of graph exploration. These aspects included completeness scores

(Table 7, p. 77), the length of their visual descriptions (p. 76), the proportion of x -axis and y -axis themes mentioned in their graph explanations (Table 7, p. 77) and in the silent time they took in preparation for their graph explanations (p. 81). However, the instructions did not help novices achieve greater consistency in the serial order of themes as exhibited by the experts in Experiment 2 (see p. 60), perhaps because they were focusing so strongly on the aspects in which they did improve. It is likely that for similar reasons, the instructions also failed to contribute to an increase in the number of conceptual messages to the levels exhibited by experts in Experiment 2. In conclusion, therefore, the brief instructions were sufficient to initiate the construction of some expert strategies in terms of their graph-explanation themes, but those strategies did not include procedures executed in a particular serial order that may be an essential component of graph exploration expertise in particular, and perhaps domain expertise more generally. This is an important aspect of instruction in the use of graphs and interpretation of data that was overlooked in Feldon et al's (2010) instructional regimen for science, technology, engineering, and math.

General Discussion

This series of experiments was designed to investigate the strategies used by novices and experts in the exploration of line graphs. In particular, its goals were threefold:

1. Identify performance and strategy differences (if any) between novices and experts in their exploration of simple and more complex graphs drawn from familiar and unfamiliar domains.

2. Determine if brief instructions about graph exploration strategies might make novices perform more like experts when they are analyzing graphs.
3. Contribute to theories of graph comprehension.

Experiment 1 found that experts (business- and psychology PhD graduate students) rated the visual and semantic complexity of graphs differently from novices (business- and psychology first-semester undergraduate students). The experts rated the graphs as less complex than novices, and the difference between ratings of visual and semantic complexity was smaller for the experts than for the novices. Experiment 1 also identified graph stimuli of moderate complexity to avoid floor- and ceiling effects in the subsequent thesis experiments.

Experiment 2 then investigated the differences between novices (business- and psychology first-semester undergraduate students) and experts (business and psychology faculty and PhD graduate students) when they were asked to visually describe and semantically explain a subset of line graphs selected from Experiment 1. Experts said significantly more in their visual descriptions than novices. They also began their semantic explanations with less hesitation, and their semantic explanations were more complete, than those provided by the novices.

Examination of the verbal protocols determined that experts also differed from novices in the types of utterances in their graph explanations. Experts used more graph titles; more causal inferences; made more comparisons between z -variables; and made more quantitative references in their explanations. Experts also exhibited consistency in the sequential order of themes in their graph explanations; the majority began their explanations by reiterating the graph title, then systematically addressed the ordinate, abscissa, z -variables, and finally made causal inferences. Novices exhibited no such consistency. Experts also provided more conceptual messages in their

graph explanations than novices, although overall, the graph domain manipulation did not systematically affect the behaviour of experts or novices.

The results were interpreted as reflecting strategy differences between novices and experts and can be understood in reference to Figure 14. When confronted with a simple line graph, both experts and novices first appeared to build a visual description sufficient to match the graph to a line graph schema (Pinker, 1990). That partially instantiated schema allows experts to initiate a top-down strategy selection mechanism to ask a series of specific conceptual

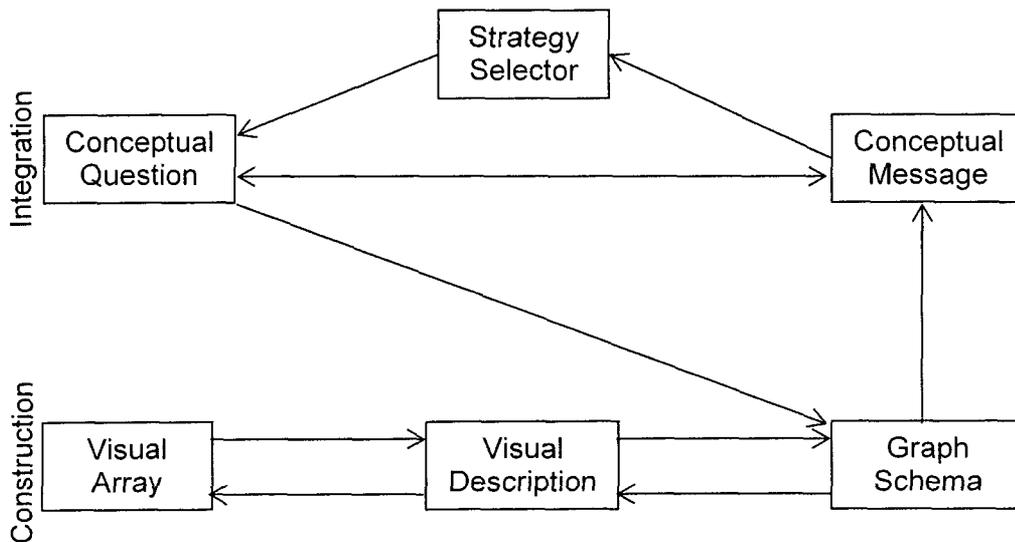


Figure 14. Hybrid graph comprehension model.

questions: 1. What is the graph title? 2. What is the y-axis? 3. What is the x-axis? 4. What are the z-variables? 5. What are the relationships between the z-variables? In response to each conceptual question an attempt is made to instantiate the relevant portions of the graph schema as an integrative process (Freedman & Shah, 2002). They read the graph title and match the relevant parts of the title to the parts of the graph. Next they attend to different parts of the graph

in a systematic manner: first the y -axis, then the x -axis, followed by the z -variables, and finally, the between z -variable comparisons. Although experts were relatively reluctant to attempt causal inferences of the data shown in the graphs, when they did, they typically provided these at the end of their semantic explanations. In contrast, novices' strategies were incomplete and ill-formed. They failed to utilize the visual description task to encode the visual characteristics of the graphs and when asked to semantically explain graphs, they paused much longer than the experts. Despite the pause, novices were often uncertain about the best graph exploration strategy, as evidenced by the inconsistency and in the order of their utterances. Although novices did sometimes start their semantic explanations by reiterating the graph title, they did so less often than experts and less consistently. Next novices might ask the conceptual question: What are the z -variables? Followed by one or another conceptual question where the conceptual questions may or may not appear from trial to trial and where their order may differ from one trial to the next.

To determine if the novice-expert differences uncovered in Experiment 2 might be attributable to a deliberate selection amongst available strategies, Experiment 3 gave a new sample of novices information about the experts' strategies uncovered in Experiment 2. After an average of 3.5 minutes of instruction, novices' visual descriptions lengthened, the completeness of their semantic explanations of graphs increased, and the proportion of between z -variable comparisons increased consistent with the findings of Feldon et al. (2010). Silent time also decreased as a result of the instruction, becoming similar to expert silent time observed in Experiment 2. The conclusion drawn was that the reduced silent time was a reflection of greater certainty about a strategy to be utilized for graph exploration.

Future Research

These experiments yield a picture of graph expertise that is most parsimoniously explained by Pinker's (1990) graph schema model. Pinker's model of graph comprehension starts with a 'bottom-up' process to build a visual description. That is exactly what the experts in Experiment 2 seem to do in their longer (than novices) visual descriptions. Then, in their semantic explanations we see evidence for a strategy that includes a consistent and relatively complete set of conceptual questions. In contrast, novices began with an impoverished visual description and then in their incomplete and inconsistent semantic explanations we observe an incomplete and inconsistent set of conceptual questions. The poorer visual descriptions of novices seems to be at odds with Freedman and Shah's (2002) and Preece and Janvier's (1992; 1993) findings that novices tend to focus on surface physical, or graphical content. It may be that the graph stimuli in the current experiments were simple enough that novices thought they could afford to ignore the visual content and thus foreshortened their visual descriptions. At the beginning of each semantic explanation trial, novices had longer silent times than experts. One possible explanation of these longer silent times is that novices needed to re-create a visual description of the graph, something which the experts did as part of the immediately preceding visual description. A test of this assertion would be to ask novices to reproduce the visual characteristics of a graph immediately after their silent time. Pinker's model would predict that if novices failed to build a visual description they would be forced to do so as part of the semantic explanation task. If novices' silent time is being used to create a visual description of the graph, then if asked to draw the graph immediately after their silent time (controlling for overall exposure time) then their graph drawings should be similar to those of experts.

It was interesting to note that novices' performance could become more expert-like in Experiment 3 after a minimum of instruction, compared with Experiment 2. Apparently, expert strategies can be taught quite readily in the area of graph exploration, thereby improving the performance (completeness) of novices considerably. One potential direction for future research would be to follow the development of novice strategies over an extended period of time to determine if strategy development is synchronized with the trajectory of expertise or if one precedes the other.

These experiments also point to a temporal dimension of expertise. Experts were quite consistent in the order in which the elements of semantic explanations were given. Presumably some kind of order information is stored with elements in a graph schema but it might be useful to explore this issue further with sequential analysis techniques (Gottman & Roy, 1990).

In Experiment 2 it was predicted that experts would provide more complete graph explanations better in their familiar domain than in their unfamiliar domain. However, the graph domain manipulation had no effect. It may be that the business and psychology introductory textbook graphs were an inappropriate domain in the context of graph exploration. Alternately, it is possible that graph expertise is domain independent and the current results together with those of Roth and Bowen (2003) and Trafton et al. (2002) are an expression of general expertise. Roth and Bowen (2003) found a key distinction between graphs of one's own work compared to those prepared by others. By selecting graph stimuli that are more personally relevant to individual participants it might be possible to distinguish between the domain independence versus personal domain alternatives.

Contributions of the Research

The present experiments contributed to an understanding of graph exploration in four ways. First, the experiments in this thesis are among the first demonstrations of an expertise “effect” in the domain of graph explanation. Although others have studied graph expertise (e.g., Roth, 2004; Roth & Bowen, 2003) they had not directly contrasted novice and expert performance. Previous attempts to distinguish novice and expert graph comprehension (Shah & Carpenter, 1995) found no differences between the two types of participants.

Second, this thesis showed that experts adopt a graph exploration strategy in which steps are executed in a particular order. It is proposed that these steps represent an ordered list of conceptual questions that is the embodiment of a graph exploration strategy. The addition of this top-down process adds clarity to Pinker’s (1990) model of graph comprehension by introducing a mechanism for the operation of expertise. In contrast, the strategies of novices were inconsistent. As a consequence, their graph exploration took longer to initiate, and were less complete than experts.

A third contribution of this thesis was the generation of a “completeness” score as a measure of expertise. In more well-defined domains such as chess, independent measures such as the Elo rating help to distinguish those with greater or lesser expertise (Gobet, 1998). However many real-world domains do not have that luxury. An intriguing possibility would be to attempt to generate a completeness score in other domains such as avionics (Lajoie, 2009), empirical biology (Feldon et al., 2010), or military leadership (Schunn, McGregor, & Saner, 2005) where metrics of expertise may be in short supply but sorely needed.

Finally, the observation that the brief instructions provided in Experiment 3 facilitated the graph explanation behaviour of a group of undergraduate novices is important. Those instructions enriched their visual descriptions of graphs (Pinker, 1990) as well as promoting rapid development of an effective strategy thereby increasing the number of conceptual questions. No claim can be made that these instructions turned the novices into fully-fledged graph experts. However, like Feldon et al., (2010) this manipulation may have facilitated an expertise-development process. This could potentially be an important adjunct to efforts to improve the data literacy of children (Siegler & Ramani, 2008), and older students (Feldon et al., 2010).

Conclusions

The goals of the thesis were:

1. To identify performance and strategy differences (if any) between novices and experts in their exploration of simple and more complex graphs drawn from familiar and unfamiliar domains.
2. Determine if brief instructions about graph exploration strategies might make novices perform more like experts when they are analyzing graphs.

An additional theoretical goal of the thesis was to contribute to theories of graph comprehension.

The experiments addressed these goals in the following ways: First, an expertise effect was identified in terms of the completeness of the explanations given by experts, the length of the visual descriptions offered by experts, and the shorter silent time of experts compared to novices. Second, the experts adopted a strategy in which the visual characteristics of graphs were

described, followed by the title, the x -, y -, and z -variables, the variations between z -variables, and finally causal inferences. Third, if novices have a graph exploration strategy, it is one which is inconsistent and impoverished compared to experts. Finally, at least some of the behaviours associated with expert data and graph exploration can be quickly and easily trained. These conclusions lend themselves to a view of expertise which must include a mechanism by which people select an exploration strategy comprised of an ordered and consistent set of executable steps. These are conceived as a set of conceptual questions that are initiated in a top-down fashion.

References

- Abernethy, B., Baker, J., & Cote, J. (2005). Transfer of pattern recall skills may contribute to the development of sport expertise. *Applied Cognitive Psychology, 19*, 705-718.
- Adelson, B., & Soloway, E. (1985). The role of domain experience in software design. *IEEE Transactions on Software Engineering, 11*, 1351-1360.
- Allard, F., & Starkes, J. L. (1991). Motor-skill experts in sports, dance, and other domains. In K. A. Ericsson, & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits*. (pp. 126-152). Cambridge, UK: Cambridge University Press.
- Anderson, J. R., & Lebiere, C. (1998). Knowledge representation. In J. R. Anderson, & C. Lebiere (Eds.), *The atomic components of thought*. (pp. 19-55). Mahwah, NJ: Lawrence Erlbaum Associates.
- Anzai, Y. (1991). Learning and use of representations for physics expertise. In K. A. Ericsson, & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits*. (pp. 64-92). Cambridge, UK: Cambridge University Press.
- Baddeley, A. D. (1976). *The psychology of memory*. New York, NY: Basic Books Inc.
- Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods, 2*(2), 131-160.
- Bertin, J. (1981). *Graphics and graphic information-processing*. Berlin: Walter de Gruyter.

Bertin, J. (1983). *Semiology of graphics*. Madison, WI: The University of Wisconsin Press.

Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(4), 640-645.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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, C. M. (1992). Reading graphs: Interactions of processing requirements and stimulus structure. In B. Burns (Ed.), *Percepts, concepts and categories: The representation and processing of information*. (pp. 605-645). Amsterdam: Elsevier Science Publishers.

Carswell, C. M., & Wickens, C. D. (1996). Mixing and matching lower-level codes for object displays: Evidence for two sources of proximity compatibility. *Human Factors*, 38, 1-22.

Chambers, J. M., Cleveland, W. S., Kleiner, B., & Tukey, P. A. (1983). *Graphical methods for data analysis*. Belmont, CA: Wadsworth International Group.

Charness, N. (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 641-653.

- Charness, N., & Campbell, J. I. D. (1988). Acquiring skill at mental calculation in adulthood. *Journal of Experimental Psychology: General*, *117*(2), 115-129.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55-81.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121-152.
- Christensen, C. R. (1987). *Teaching and the case method*. Boston, MA: Harvard Business School.
- Chronicle, E. P., MacGregor, J. N., & Ormerod, T. C. (2004). What makes an insight problem? the roles of heuristics, goal conception, and solution recoding in knowledge-lean problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 14-27.
- Cleveland, W. S. (1984). Graphs in scientific publications. *The American Statistician*, *38*(4), 261-269.
- Cleveland, W. S. (1987). Research in statistical graphics. *Journal of the American Statistical Association*, *82*(398), 419-423.
- Cumming, G., & Finch, S. (2005). Inference by eye : Confidence intervals and how to read pictures of data. *American Psychologist*, *60*(2), 170-180.
- Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span : a tool for unwelding visuo-spatial memory. *Neuropsychologia*, *37*, 1189-1199

du Toit, S. H. C., Steyn, A. G. W., & Stumpf, R. H. (1986). *Graphical exploratory data analysis*. New York, NY: Springer-Verlag.

Ericsson, K. A. (2005). Recent advances in expertise research: A commentary on the contributions to the special issue. *Applied Cognitive Psychology, 19*, 231-241.

Ericsson, K.A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49*, 725-747.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. (Revised ed.). Cambridge MA: MIT Press.

Ericsson, K. A., & Smith, J. (1991). Prospects and limits of the empirical study of expertise: An introduction. In K. A. Ericsson, & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 1-38). Cambridge, UK: Cambridge University Press.

Ericsson, K. A., & Staszewski, J. J. (1989). Skilled memory and expertise: Mechanisms of exceptional performance. In D. Klahr, & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon*. (pp. 235-267). Hillsdale, NJ: Lawrence Erlbaum Associates.

Feldon, D.F., Timmerman, B.C., Stowe, K.A., & Showman, R. (2010). Translating expertise into effective instruction: The impacts of cognitive task analysis (CTA) on lab report quality and student retention in the biological sciences. *Journal of Research in Science Teaching, 47*(10), 1165-1185.

Freedman, E. G., & Shah, P. (2002). Toward a model of knowledge-based graph comprehension.

In M. Hegarty, B. Meyer & N. H. Narayanan (Eds.), *Diagrams 2002, LNAI 2317* (pp. 18-30). Berlin: Springer-Verlag.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*, 306-355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1-38.

Glaser, R., & Chi, M. T. H. (1988). Overview. In Chi, M. T. H., R. Glaser & M. Farr (Eds.), *The nature of expertise*. (pp. xv-xxviii). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gobet, F. (1998). Expert memory: A comparison of four theories. *Cognition, 66*(2), 115-152.

Gobet, F. (2005). Chunking models of expertise: Implications for education. *Applied Cognitive Psychology, 19*, 183-204.

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C. -H., Jones, G., & Oliver, I., et al. (2001).

Chunking mechanisms in human learning. *Trends in Cognitive Sciences, 5*(6), 236-243.

Gobet, F., Richman, H., Staszewski, J., & Simon, H. A. (1997). Goals, representations, and strategies in a concept attainment task: The EPAM model. *The Psychology of Learning and Motivation, 37*, 265-290.

- Gobet, F., & Simon, H. A. (1996a). Recall of random and distorted chess positions: Implications for the theory of expertise. *Memory and Cognition*, 24(4), 493-503.
- Gobet, F., & Simon, H. A. (1996b). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology*, 31, 1-40.
- Gobet, F., & Simon, H. A. (2000). Five seconds or sixty? Presentation time in expert memory. *Cognitive Science*, 24(4), 651-682.
- Goodman, N. (1968). *Languages of art: An approach to a theory of symbols*. Indianapolis, IN: Bobbs-Merrill Company Inc.
- Gottman, J.M., & Roy, A.K. (1990). *Sequential analysis: A guide for the behavioral sciences*. New York, NY: Cambridge University Press,
- Griffin, R. W., & Pustay, M.W. (2007). *International Business: A Managerial Perspective* (5th ed.). Upper Saddle River, NJ: Pearson Education.
- Haig, B. D. (2005). An abductive theory of scientific method. *Psychological Methods*, 10(4), 371-388.
- Heer, J., Bostock, M., & Ogievetsky, V. (2010). A tour through the visualization zoo: A survey of powerful visualization techniques from the obvious to the obscure. *ACMQueue*, 8(5), 1-22.

- Hoadley, E. D. (1990). Investigating the effects of color. *Communications of the ACM*, 33(2), 120-139.
- Huff, D. (1954). *How to lie with statistics*. New York, NY: W. W. Norton & Company Inc.
- Isaak, M. I., & Just, M. A. (1995). Constraints on thinking in insight and invention. In R. J. Sternberg, & J. E. Davidson (Eds.), *The nature of insight* (pp. 281-325). Cambridge MA: MIT Press.
- Jarmasz, J., & Hollands, J.G. (2009). Confidence intervals in repeated-measures designs: The number of observations principle. *Canadian Journal of Experimental Psychology*, 63(3), 124-138.
- Kimball, R. (1996). *The data warehouse toolkit*. New York, NY: John Wiley & Sons, Inc.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163-182.
- Kliegl, R., Smith, J., Heckhausen, J., & Baltes, P. B. (1987). Mnemonic training for the acquisition of skilled digit memory. *Cognition and Instruction*, 4(4), 203-223.
- Kohler, W. (1947). *Gestalt psychology: An introduction to new concepts in modern psychology*. New York, NY: Liveright Publishing Corporation.
- Kosslyn, S. M. (1993). *Elements of graph design*. New York, NY: W. H. Freeman.

- Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, 3, 185-226.
- Lajoie, S. (2009). Developing professional expertise with a cognitive apprenticeship model: Examples from avionics and medicine. In K. A. Ericsson (Ed.) *Development of professional expertise: Toward measurement of expert performance and design of optimal learning environments*. New York, NY: Cambridge University Press, 61-83.
- Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11(1), 65-99.
- Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General*, 124(1), 83-97.
- Lewandowsky, S., & Behrens, J. T. (1999). Statistical graphs and maps. In F. T. Durso, R. S. Nickerson, R. W. Schvaneveldt, S. T. Dumais, D. S. Lindsay & M. T. H. Chi (Eds.), *Handbook of applied cognition* (pp. 513-549). New York, NY: John Wiley & Sons Ltd.
- Lohse, G. L. (1993). A cognitive model for understanding graphical perception. *Human-Computer Interaction*, 8, 353-388.
- Marr, D., & Nishihara, H. K. (1978). Representation 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.

Marteniuk, R.G., Ivens, C.J., & Brown, B.E. (1996). Are there task specific performance effects for differently configured numeric keypads? *Applied Ergonomics*, 27(5), 321-325.

Mayer, R. E. (1995). The search for insight: Grappling with gestalt psychology's unanswered questions. In R. J. Sternberg, & J. E. Davidson (Eds.), *The nature of insight* (pp. 3-62). Cambridge, MA: MIT Press.

Medin, D. L., & Smith, E. E. (1981). Strategies and classification learning. *Journal of Experimental Psychology: Human Learning and Memory*, 7 (4), 241-253.

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

Neil, T. (2009). 28 rich data visualization tools. In DevelopRIA
<http://www.developria.com/2009/12/28-rich-data-visualization-too.html> retrieved March 01, 2011.

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear statistical models*. (4th ed.). Boston, MA: WCB McGraw-Hill.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Norman, D. A. (1976). *Memory and attention*. (2nd ed.) New York, NY: John Wiley & Sons, Inc.

- Norman, D. A., & Rumelhart, D. E. (1981). The LNR approach to human information processing. *Cognition, 10*, 235-240.
- Parr, W. V., Heatherbell, D., & White, K. G. (2002). Demystifying wine expertise: Olfactory threshold, perceptual skill and semantic memory in expert and novice wine judges. *Chemical Senses, 27*, 747-755.
- Pinker, S. (1990). A theory of graph comprehension. In R. Freedle (Ed.), *Artificial intelligence and the future of testing*. (pp. 73-126). Hillsdale NJ: Lawrence Erlbaum.
- Postigo, Y., & Pozo, J. I. (2004). On the road to graphicacy: The learning of graphical representation systems. *Educational Psychology, 24*(5), 623-644.
- Preece, J., & Janvier, C. (1992). A study of the interpretation of trends in multiple curve graphs of ecological situations. *School Science and Mathematics, 92*(6 (October)), 299-306.
- Preece, J., & Janvier, C. (1993). Interpreting trends in multiple-curve graphs of ecological situations: The role of context. *International Journal of Science Education, 15*(2), 199-212.
- Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences, 25*, 157-238.
- Ratwani, R. M., & Trafton, J. G. (2004). Making graphical inferences: A hierarchical framework. *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society*, Chicago, IL. 1119-1124.

- Rikers, R. M. J. P., & Paas, F. (2005). Recent advances in expertise research. *Applied Cognitive Psychology, 19*, 145-149.
- Ronstadt, R. (1980). *The art of case analysis: A guide to the diagnosis of business situations*. (2nd ed.). Dover, MA: Lord Publishing.
- Roth, W. (2004). Emergence of graphing practices in scientific research. *Journal of Cognition and Culture, 4*(3-4), 595-627.
- Roth, W., & Bowen, G. M. (2001). Professionals read graphs: A semiotic analysis. *Journal for Research in Mathematics Education, 32*(2), 159-194.
- Roth, W. M., & Bowen, G. M. (2003). When are graphs worth ten thousand words? an expert-expert study. *Cognition and Instruction, 21*(4), 429-473.
- Roth, W., Bowen, G. M., & McGinn, M. K. (1999). Differences in graph-related practices between high school biology textbooks and scientific ecology journals. *Journal of Research in Science Teaching, 36*(9), 977-1019.
- Salthouse, T. A. (1991). Expertise as the circumvention of human processing limitations. In K. A. Ericsson, & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits*. (pp. 286-300). Cambridge, UK: Cambridge University Press.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. detection, search, and attention. *Psychological Review, 84*, 1-66.

Schraagen, J. M. (1993). How experts solve a novel problem in experimental design. *Cognitive Science*, 17(2), 285-309.

Schraagen, J.M. (2009). Designing training for professionals based on subject matter experts and cognitive task analysis. In K.A. Ericsson (Ed.) *Development of professional expertise: Toward measurement of expert performance and design of optimal learning environments*. New York, NY: Cambridge University Press, pp. 157-179.

Schunn, C. D., & Anderson, J. R. (1998). Scientific discovery. In J. R. Anderson, & C. Lebiere (Eds.), *The atomic components of thought*. (pp. 385-427). Mahwah, NJ: Lawrence Erlbaum Associates.

Schunn, C. D., & Anderson, J. R. (1999). The generality/specificity of expertise in scientific reasoning. *Cognitive Science*, 23(3), 337-370.

Schunn, C. D., McGregor, M. U., & Saner, L. D. (2005). Expertise in ill-defined problem-solving domains as effective strategy use. *Memory and Cognition*, 33(8), 1377-1387.

Seibel, R. (1972.) Data entry devices and procedures. In H.P. Van Cott & R.G. Kinkade (Eds.) *Human engineering guide to equipment design (revised ed.)* Washington, DC: American Institutes for Research, pp.311-344.

Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, 14(1), 47-69.

Shah, P., & Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs.

Journal of Experimental Psychology: General, 124(1), 43-61.

Shanteau, J. (1992). The psychology of experts: An alternative view. In G. Wright, & F. Bolger

(Eds.), *Expertise and decision support*. (pp. 11-23). New York, NY: Plenum Press.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information

processing: II. perceptual learning, automatic attending, and a general theory. *Psychological*

Review, 84, 127-190.

Siegler, R. S., & Ramani, G.B. (2008). Playing linear numerical board games promotes low-

income children's numerical development. *Developmental Science*, 11(5), 655-661.

Simkin, D., & Hastie, R. (1987). An information-processing analysis of graph perception.

Journal of the American Statistical Association, 82(398), 454-465.

Sims, V. K., & Mayer, R. E. (2002). Domain specificity of spatial expertise: The case of video

game players. *Applied Cognitive Psychology*, 16, 97-115.

Stutely, R. (1993). *The Economist guide to business numeracy*. Chichester, UK: John Wiley &

Sons, Inc.

Tabachneck-Schijf, Hermina J. M., Leonardo, A. M., & Simon, H. A. (1997). CaMeRa: A

computational model of multiple representations. *Cognitive Science*, 21(3), 305-350.

- Trafton, J. G., Kirschenbaum, S. S., Tsui, T. L., Miyamoto, R. T., Ballas, J. A., & Raymond, P. D. (2000). Turning pictures into numbers: Extracting and generating information from complex visualizations. *International Journal of Human-Computer Studies*, 53(5), 827-850.
- Trafton, J. G., Marshall, S., Mintz, F., & Trickett, S. B. (2002). Extracting explicit and implicit information from complex visualizations. In M. Hegarty, B. Meyer, N. Narayanan (Eds.) *Diagrams 2002, Lecture Notes in Computer Science*, 2317/2002, 206-220.
- Trafton, J. G., & Trickett, S. B. (2001). A new model of graph and visualization usage. In J. D. Moore, & K. Stenning (Eds.), *Proceedings of the twenty second annual conference of the cognitive science society* (pp. 1048-1053). Mahwah, NJ: Lawrence Erlbaum Associates.
- Tronsky, L. N. (2005). Strategy use, the development of automaticity, and working memory involvement in complex multiplication. *Memory and Cognition*, 33(5), 927-940.
- Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
- Tversky, B., & Schiano, D. J. (1989). Perceptual and conceptual factors in distortions in memory for graphs and maps. *Journal of Experimental Psychology: General*, 118(4), 387-398.
- VanLehn, K., & van de Sande, B. (2009). Acquiring conceptual expertise from modeling: The case of elementary physics. In K.A. Ericsson (Ed.) *Development of professional expertise:*

Toward measurement of expert performance and design of optimal learning environments.

New York, NY: Cambridge University Press, pp. 356-378.

Vicente, K. J., & Wang, J. H. (1998). An ecological theory of expertise effects in memory recall.

Psychological Review, 105(1), 33-57.

Waters, A. J., Gobet, F., & Leyden, G. (2002). Visuospatial abilities of chess players. *British*

Journal of Psychology, 93, 557-565.

Weiten, W., & McCann, D. (2007). *Psychology: Themes and Variations*. (1st Canadian ed.).

Toronto, ON: Nelson Education.

Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: Its

psychological foundation and relevance to display design. *Human Factors*, 37(3), 473-494.

Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science*,

21(2), 179-217.

Zhang, J., & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive*

Science, 18, 87-122.

Appendix A: Preliminary Study 1

This preliminary study was designed to test the methodology to be employed throughout this thesis and to sort a set of line graph stimuli into two categories representing simple and complex graphs as determined by a sample of business and psychology participants. Oakley, Lindgaard, and Ferres (in preparation) have attempted to establish an unambiguous set of perceptually simple and complex line graphs. The purpose of this study was to partially replicate Oakley et al and to isolate a set of graphs varying in perceptual complexity as judged by a sample of participants drawn from expert/novice populations. The goal was to assemble a set of 26 of each, unambiguously simple and complex graphs which will be utilized in subsequent experiments in this thesis. Because the psychology and business experts who will be recruited for the main Experiments in this thesis are in short supply, this preliminary study utilized graduate students in psychology and business as experts and compared their performance to that of undergraduate novices. The expert/novice distinction will be increased in the subsequent experiments in which faculty in psychology and business will be recruited as expert participants.

Method

Participants

A sample of 6 expert and 6 novice participants were recruited. Graduate students acting as experts were recruited from the Carleton University Sprott School of Business ($n = 3$: Business Experts or BE) and from the Carleton University Department of Psychology ($n = 3$: Psychology Experts or PE). Novices were undergraduates majoring

either in business (Business Novices or BN) or psychology (Psychology Novices or PN) as shown in Table 1 below. Experts were paid for their participation in the form of a \$5 coffee shop gift certificate. Novices received 1% course credit in return for their participation. Participants were tested individually in sessions of approximately 60 minutes.

Table A1
Preliminary Study 1 Participants

Expertise	Business	Psychology
Novice	BN: business undergraduates $N = 3$	PN: psychology undergraduates $N = 3$
Expert	BE: business graduate students $N = 3$	PE: psychology graduate students $N = 3$

Apparatus & Materials

The line graph stimuli were drawn from the pool of 256 unique graphs generated and utilized by Oakley et al. (in preparation). All graphs were monochrome (white on black) and varied in terms of the number of variables (1, 2, 3, 5, or 7 lines), the number of data points per variable (2, 3, 5, or 7), and the number of “crossovers”, i.e. interactions, between variables (0, 1, 2, 3, 5, or 7). The size of each graph was kept constant across the stimulus set. The graphs were displayed on a Dell laptop computer with resolution of 1400 x 1050 pixels and all timing and display managed by DirectRT™. Participants entered their responses on a numeric keypad using the digit keys 4, 5, and 6. A sample of some of the current graph stimuli is provided in Figure 1.

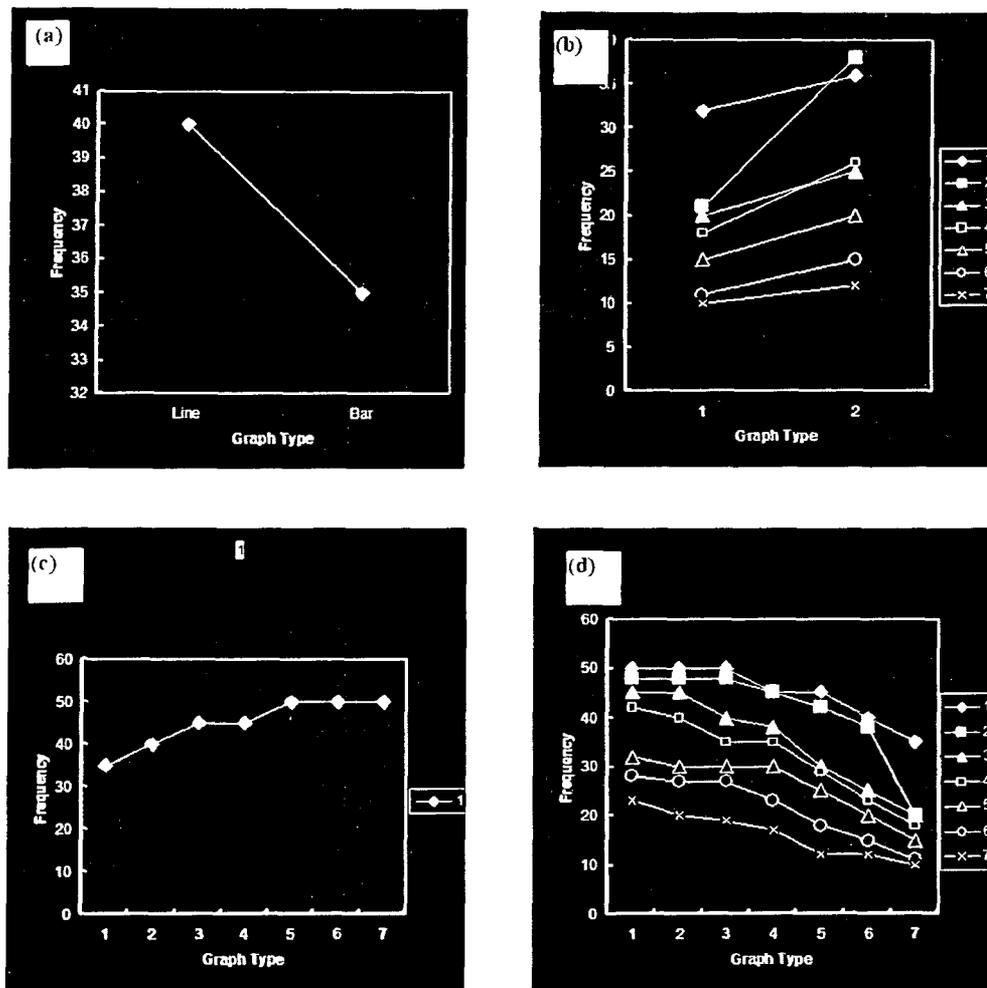


Figure A1. Preliminary Study 1 example stimuli. Figure 1(a) has 1 variable, 2 data points, and 0 crossovers. Figure 1(b) has 7 variables, 2 data points, and 2 crossovers. Figure 1(c) has 1 variable, 7 data points, and 0 crossovers. Figure 1(d) has 7 variables, 7 data points, and 0 crossovers.

Design and Procedure

A 2 x 2 between-subjects design for expertise (novice, expert) and domain (business, psychology) was adopted. After Preliminary instructions and Informed Consent, participants had nine practice trials to ensure that they understood instructions

and to acclimatize to the numeric keypad. They then judged each of the 256 unique graph stimuli in random order followed by a brief rest period, then followed by a new random order of the same graph stimuli for a grand total of 512 trials. Participants used their preferred hand with the second, third, and fourth fingers resting on the digits four, five, or six on the numeric keypad. Instructions asked participants to judge each graph stimulus as simple, moderate, or complex corresponding with the three numeric keys. After the keypress the display was blank for 1s and the next graph was displayed. Participants were instructed to work at their own pace.

Results

The 13 participants recruited for this Preliminary Study included three from the business domain and the remaining nine were drawn from the psychology domain. One psychology expert was discarded when it was determined that every single response they made was identical. No other systematic differences in complexity judgments between the two BE and PE groups were observed. No other systematic trends in judgments made by participants were observed. The purpose of Preliminary Study 1 was to select a sample of simple and complex graphs for the subsequent experiments in this thesis. Of the 256 graph stimuli tested, each graph stimulus was judged a total of 24 times. The graphs judged “simple” most consistently were so judged on 735 instances out of 768 or 95.7%. Figure 2 illustrates that there was substantial agreement among participants for these 32 graphs. However, there was somewhat less agreement on complex graphs. The graphs judged “complex” most consistently were so judged on 478 instances out of 720 or 66.4%. The degree of agreement of judgments for the complex graphs is depicted in Figure 3.

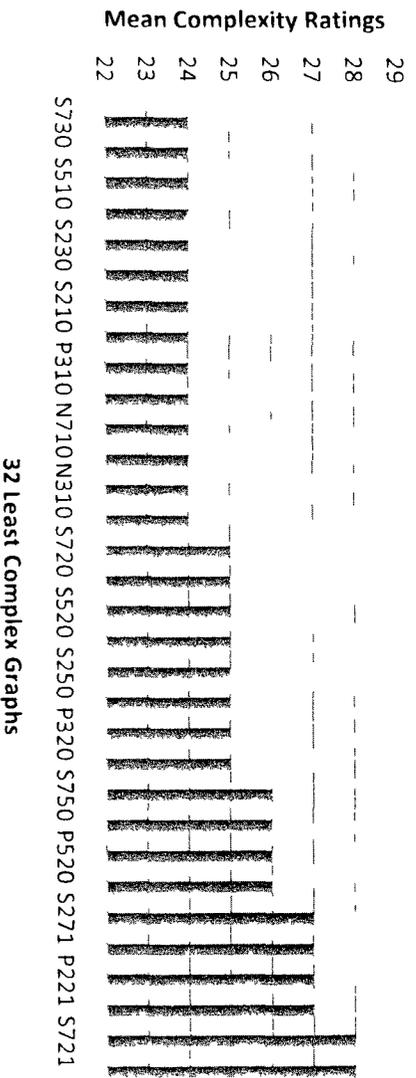


Figure A2 32 Least Complex Graphs

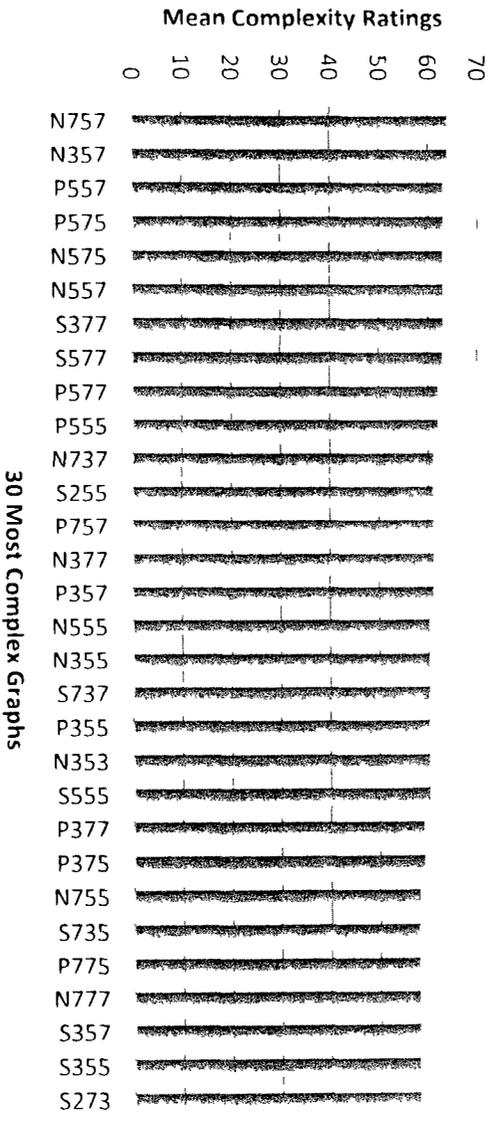


Figure A3 30 Most Complex Graphs

Examples of two of the stimuli judged simple and two judged complex are provided in Figures 4 and Figure 5 below.

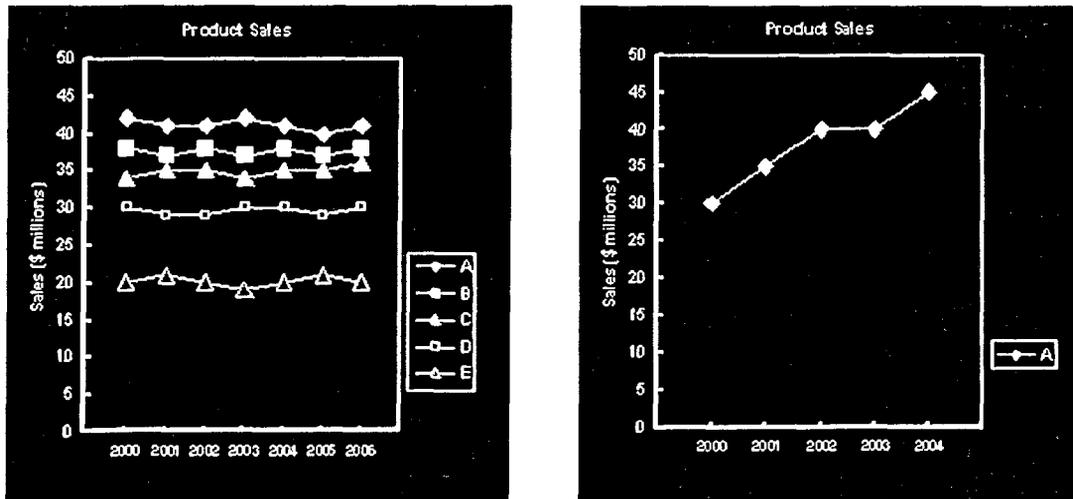


Figure A4. Simple graph stimuli examples.

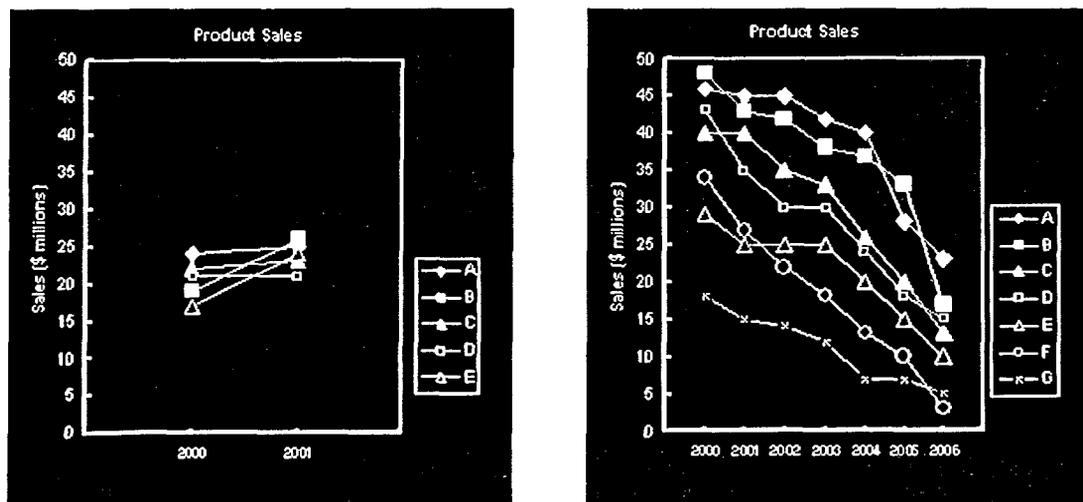


Figure A5. Complex graph stimuli examples.

Graph attributes

The graph stimuli utilized in this study varied in terms of slope (positive, negative, zero); the number of lines per graph (1, 2, 3, 5, or 7); the number of points plotted per line (2, 3, 5, or 7); and the number of crossovers (0, 1, 2, 3, 5, or 7). Note that the number of crossovers was defined as the total number of lines crossed. If a line crossed three others the graph was classified as a three crossover graph. Thus, in the

leftmost of the complex graphs in Figure 4 the number of crossovers is five and in the rightmost it is seven.

Since Preliminary Study 1 was conducted primarily to select stimuli for subsequent experiments in this thesis no statistical tests were conducted. There appeared to be no notable differences between simple and complex judgments in terms of slope except that a large proportion of the 32 graphs judged simple had zero slope. Of the 30 graphs judged complex there appeared to be an approximately equivalent number of positive, negative, and zero slope graphs as indicated in Table 2.

Table A2

Proportion of 32 simple & 30 complex graph stimuli with positive, negative, & zero slope

	Positive	Negative	Zero
Simple			
/32	25%	22%	53%
Complex			
/30	37%	33%	27%

Tables 3, 4, and 5 show respectively the proportion of 32 simple and 30 complex graphs with the number of data points depicted: the number of variables depicted: and the number of crossovers. Overall, the data suggest that participants seem to make their simple/complex judgments partly on the basis of the number of data points, partly on the number of lines, but mostly on the basis of the number of crossovers.

Table A3

Proportion of 32 simple & 30 complex graph stimuli by number of data points

	2 points	3 points	5 points	7 points
Simple				
/32	38%	19%	22%	19%
Complex				
/30	3%	37%	30%	30%

Table A4

Proportion of 32 simple & 30 complex graph stimuli by number of lines

Graphs	1 line	2 lines	3 lines	5 lines	7 lines
Simple /32	38%	34%	9%	6%	9%
Comple x /30	0%	0%	10%	57%	33%

Table A5

Proportion of 32 simple & 30 complex graph stimuli by number of crossovers

Graphs	0 crossovers	1 crossover	3 crossovers	5 crossovers	7 crossovers
Simple /32	91%	9%	0%	0%	0%
Complex /30	0%	0%	3%	47%	50%

Table A6.

Graph stimulus attributes in total stimulus corpus (/516)

	Negative Slope	Positive Slope	Zero Slope	Row Totals
Data points=2	36	36	36	108
Data points=3	42	42	42	126
Data points=5	46	46	46	138
Data points=7	48	48	48	144
Lines=1	8	6	8	22
Lines=2	28	30	28	86
Lines=3	42	42	42	126
Lines=5	46	46	46	138
Lines=7	48	48	48	144
Crossovers=0	40	38	40	118
Crossovers=1	32	34	32	98
Crossovers=2	30	30	30	90
Crossovers=3	28	28	28	84
Crossovers=5	24	24	24	72
Crossovers=7	18	18	18	54

In the original set of 256 graph stimuli there were an equal number of graphs with positive, negative, and zero slopes. However, there were differences in the total number of graph stimuli across data points, lines, and crossovers. Since participants encountered

each graph twice, the incidence of each graph attribute in the total $258 \times 2 = 516$ presentations is shown in Table A6.

Since there was lower overall agreement with the 30 complex graphs this analysis will focus on this subset. In Table A7 the attributes of each of the 30 complex graphs are summarized. The column titled Data points has a sum of 29 because one graph in this complex set had two data points only. For present purposes, this particular graph stimulus will be ignored.

Table A7
Complex graph attributes

Number of...	Data points	Crossovers	Lines
7	9	15	10
5	9	14	17
3	11	1	3
Total	29	30	30

Table A7 indicates that the number of crossovers and the number of lines seem to be strong criteria for participants' judgments of complexity. Specifically, 29 out of 30 complex graphs (97%) had five or seven crossovers and 27 out of 30 complex graphs (90%) had five or seven lines. The number of data points seems to be more weakly associated with complexity. As indicated earlier in Table 2, there was an approximately equal number of complex graphs with positive (11), negative (11), and zero (8) slope so this attribute did not appear to be associated with judgments of complexity.

To determine if participants reacted to this subset of graph stimuli in an idiosyncratic way, each participants' judgments were assigned 1 = simple; 2 = moderate; and 3 = complex. Then, since every participant judged each graph stimulus twice, the simple, moderate, and complex codes were summed for each of the 30 complex graphs.

By calculating the total number of simple, moderate, and complex judgments on a participant-by-participant basis (Table A8) it appears that Participants 9 and 12 utilized different complexity criteria from the rest of the participants. The only graphs that Participant 12 judged to be complex were Pgraph357, Sgraph232, and S357. Although Pgraph357 and S357 have seven crossovers each they only have three data points and five lines each. Stimulus Sgraph232 is a two data point, three line and two crossover graph that does not seem to consistently represent any particular attribute of graph complexity.

Participant 9 only judged stimuli Ngraph575; Pgraph737; and Sgraph273 as complex. These three graphs have five, seven, and two data points, seven, three, and seven lines; and five, seven, and three crossovers. Thus, there are no self-evident complexity criteria apparent from the judgments of Participants 9 and 12.

Table A8
Participant judgments of simple, moderate, & complex

Participant	Simple	Moderate	Complex
#01	298	139	79
#02	164	175	177
#03	256	235	25
#04	199	159	158
#05	179	228	109
#06	189	74	253
#07	128	181	207
#08	210	171	135
#09	459	54	3
#10	74	122	320
#11	177	252	87
#12	426	86	4

Instead, it appears that Participants 9 and 12 exhibited a response bias in which they favoured judgments of simple over complex.

Response times

The overall mean response time of all participants to all graph stimuli was 1837 ms. The mean response time for the subset of 32 simple graph stimuli was 1080 ms while the mean for the 30 complex graph stimuli was nearly twice that number at 2064 ms. An examination of response times averaged over the attributes of the most complex graphs (see Table 9) indicated no consistent pattern. Clearly, participants required more time to make a judgment about complex graph stimuli than simple.

Table A9.

Response times for complex graph attributes

Number of...	Data points	Lines	Crossovers
7	2029 ms	2057 ms	2089 ms
5	1876 ms	2047 ms	1943 ms
3	2123 ms	1596 ms	1613 ms
2	1707 ms		

Participant Comments

Immediately after the study was completed participants were asked about the subjective criteria they used to judge the graphs. The most common criterion in this study had to do with how easy it might be to explain the graphs either to someone else or in the context of the Results section of a paper. This criterion was cited by 10 of the 12 participants. Clearly this criterion is associated with expert participants who are relatively well-versed in the reading and preparation of scientific papers. Also, every

participant also included the criterion associated with the number of crossovers in the graphs.

Discussion

This Preliminary Study confirmed that complexity judgments of graph stimuli was a straightforward task and that the selection of stimuli for the subsequent experiments in this thesis should not be problematic. While the judgments of simple graphs were both fast and very consistent, the judgments of complex graphs were less clear. A closer examination of the subset of 30 graphs most consistently judged to be complex suggests that the criteria employed by participants were: (1) slope (positive, negative, or zero); (2) whether the z -variables are parallel or non-parallel; (3) whether the slope changed direction; (4) the number of data points; (5) the number of common values; and (6) the number of crossovers. Bertin's (1981, 1983) and Ratwani and Trafton's (2004) research (see Table 1, p. X) suggests that participants may have been responding, at least sometimes, on the basis of the purely visual criteria associated with the ease with which numerical values can be extracted from line graphs. A better understanding of criteria that participants apply when they judge complex graph stimuli should obtain in Preliminary Study 2 by focusing on the above six criteria and by distinguishing between visual and semantic criteria.

Table A10

Graph attributes and complexity judgments.

Graph #	Slope	Data-points	Lines	Cross-overs	Simple	Moderate	Complex
Pgraph557	+ve	5	5	7	4	1	19
Pgraph575	+ve	5	7	5	3	3	18
Ngraph575	-ve	5	7	5	3	3	18
Ngraph557	-ve	5	5	7	3	3	18
Ngraph757	-ve	7	5	7	2	4	18
Ngraph357	-ve	3	5	7	2	4	18
Ngraph737	-ve	7	3	7	4	3	17
Pgraph577	+ve	5	7	7	3	4	17
Pgraph555	+ve	5	5	5	3	4	17
Sgraph377	zero	3	7	7	2	5	17
Ngraph755	-ve	7	5	5	6	2	16
Ngraph555	-ve	5	5	5	4	4	16
Ngraph355	-ve	3	5	5	4	4	16
Sgraph255	zero	2	5	5	3	5	16
Pgraph757	+ve	7	5	7	3	5	16
Ngraph377	-ve	3	7	7	3	5	16
Sgraph577	zero	5	7	7	1	7	16
Sgraph735	zero	7	3	5	5	4	15
Pgraph775	+ve	7	7	5	5	4	15
Ngraph777	-ve	7	7	7	5	4	15
Sgraph737	zero	7	3	7	3	6	15
Pgraph355	+ve	3	5	5	3	6	15
Ngraph353	-ve	3	5	3	3	6	15
Pgraph357	+ve	3	5	7	2	7	15
Pgraph755	+ve	7	5	5	5	5	14
Sgraph357	zero	3	5	7	4	6	14
Sgraph355	zero	3	5	5	4	6	14
Pgraph377	+ve	3	7	7	3	7	14
Pgraph375	+ve	3	7	5	3	7	14
Sgraph555	zero	5	5	5	2	8	14

Table A11

Participant judgments of complex graphs.

Graph #	P't #1	P't #2	P't #3	P't #4	P't #5	P't #6	P't #7	P't #8	P't #9	P't #10	P't #11	P't #12
Pgraph557	6	6	5	6	6	6	6	6	2	6	6	2
Ngraph557	6	6	4	6	6	6	6	6	3	6	5	3
Ngraph575	6	6	5	6	5	6	6	5	4	6	6	2
Pgraph575	6	6	5	6	6	6	6	6	3	6	5	2
Ngraph357	6	6	5	6	6	6	5	6	2	6	6	4
Ngraph757	5	6	5	6	6	6	6	6	3	6	6	3
Ngraph737	6	6	2	6	6	6	5	6	4	6	6	2
Pgraph555	6	6	3	6	6	6	6	5	4	6	6	2
Pgraph577	6	6	5	6	4	6	6	6	2	6	6	3
Sgraph377	6	6	5	6	6	6	6	6	3	6	4	3
Ngraph755	6	5	2	6	5	6	6	6	2	6	6	2
Ngraph355	5	5	3	6	6	6	6	6	3	6	6	2
Ngraph555	6	6	3	6	6	6	6	5	3	5	6	2
Ngraph377	5	6	4	6	6	6	6	6	2	5	6	3
Pgraph757	6	6	3	6	6	6	6	6	3	5	5	3
Sgraph255	6	6	2	6	6	6	6	6	4	5	4	4
Sgraph577	5	6	4	5	6	6	6	6	3	6	6	4
Ngraph777	6	6	3	5	5	4	6	6	3	6	6	2
Pgraph775	5	6	6	5	3	5	6	6	2	6	6	2
Sgraph735	4	6	2	6	6	6	4	6	4	6	6	2
Ngraph353	4	6	4	6	6	6	6	6	2	6	5	3
Pgraph355	6	6	4	6	6	6	5	5	2	6	5	3
Sgraph737	5	6	3	6	6	6	6	4	4	6	6	2
Pgraph357	4	6	5	6	6	6	5	5	2	6	5	5
Pgraph755	5	6	5	5	6	6	6	5	2	4	5	2
Sgraph355	5	5	5	6	6	6	6	5	2	6	3	3
Sgraph357	3	6	4	6	4	6	6	5	2	5	6	5
Pgraph375	5	5	4	6	5	6	6	6	2	6	5	3
Pgraph377	6	5	4	5	5	6	6	5	2	6	6	3
Sgraph555	6	6	4	6	5	6	5	4	4	6	6	2

Appendix B: Preliminary Study 2

The purpose of Preliminary Study 2 was to identify a subset of graph stimuli that could be utilized in subsequent experiments utilizing the criteria identified in Preliminary Study 1 and to distinguish between the visual and semantic characteristics of those graphs. Preliminary Study 1 asked participants to judge if graphs were simple, moderate, or complex without specifying a particular *kind* of complexity. Ten out of twelve participants commented that they had based at least some of their judgments on the difficulty with which the data might be explained to others. It is unclear if participants utilized that criterion exclusively or even consistently. Bertin's (1981, 1983) and Ratwani and Trafton's (2004) research suggests that participants may have been responding, at least sometimes, on the basis of the purely visual criteria associated with the ease with which numerical values can be extracted from line graphs. Preliminary Study 2 explicitly manipulated the judgment criteria. In one block of trials participants were asked to judge on the basis of visual complexity only. In another block they were asked to judge on the basis of the imagined difficulty they would have in explaining the meaning of the data in the graph to others. This semantic complexity is expected to be closer to the cognitive processes associated with the inferences made in graph exploration. Carpenter and Shah (1998) established the importance of different slopes, number of crossovers, and the number of z variable lines on viewing times, verbal descriptions, and gaze durations of line graphs as discussed earlier. (p. XX). While these visual characteristics of graphs lead to increased viewing times, more complex verbal descriptions, and increased gaze duration, another level of cognitive processing (Craik &

Lockhart, 1972) is typically relevant to graph exploration. That is the level at which understanding of the causal mechanisms is sufficient to explain the effect to others. To distinguish this kind of judgment from Carpenter and Shah's visual judgments, this will be referred to as the *semantic* level. In the present study, participants will be asked to judge graph stimuli on both the visual criteria, as in Carpenter and Shah's research, and the semantic criteria associated with the difficulty of explaining the data to others. Preliminary Study 1 showed that participants agreed less than anticipated on judgments of complex graphs compared to simple graphs, suggesting that they might be applying different criteria to judging these two kinds of graphs. In particular, graph slope, crossovers, common points, and slope changes seemed to contribute to judgments of graph complexity. These visual graph characteristics also have a semantic aspect. For example, a line graph with three points that resembles a "V" starting with a negative slope, then an inflection point followed by a positive slope, has one semantic interpretation that is "quadratic function" while its more visual characteristics might be described as a "V". The present study was designed specifically to separate the graph attributes of slope, crossovers, common points, and slope changes by manipulating these variables explicitly.

One possible explanation for expert/novice differences in the exploration of graphs may be individual differences in visual processing capacity. Indeed, Tabachneck-Schijf, Leonardo, and Simon's (1997) CaMeRa model noted the importance of pictorial representations in graph communication. This study will measure the visual memory of participants to determine if any systematic differences characterize participant performance as measured by the Visual Patterns Test (Della Sala, Gray, Baddeley, &

Wilson, 1997). The Visual Patterns Test has been shown to measure short-term visual memory exclusive of its spatio-sequential component (Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999).

Finally, two stimulus-specific issues were explored. First, black specifiers against a white background were utilized in the stimulus graphs. Second, the graph legends were removed to ensure that participant attention is directed at the graph specifiers.

Method

Design

This study differed from Preliminary Study 1 in three ways. First, it required participants to judge the stimulus graphs in both visual and semantic terms: second, a 7-point rating scale was utilized: third, the Visual Patterns Test assessed the visual-spatio memory of participants in the experiment.

Participants

A sample of 10 graduate and undergraduate students who did not participate in Preliminary Study 1 were recruited using the local Department of Psychology system for participation in psychological experiments and through posters. They were paid \$5.00 or given a 1% course credit for participation. English was the first language of all participants and all had normal or corrected-to-normal vision.

Apparatus & Materials

The graph stimuli were derived from the criteria arising from Preliminary Study 1 and varied in visual and semantic complexity. All were line graphs plotted on x - and y -coordinates, and all were limited to two lines (z -variables). In addition, they varied in six ways: (1) by slope (positive, negative, or zero); (2) by whether the z -variables are parallel

or non-parallel: (3) by whether the slope changed direction, (4) by the number of data points: (5) the number of common values, and (6) the number of crossovers. Slope changes were limited to changes amongst positive, negative, and zero slopes: obviously lines with two data points could not have slope changes. The graph stimuli of Preliminary Study 2 are summarized in Table 9.

Table B1:

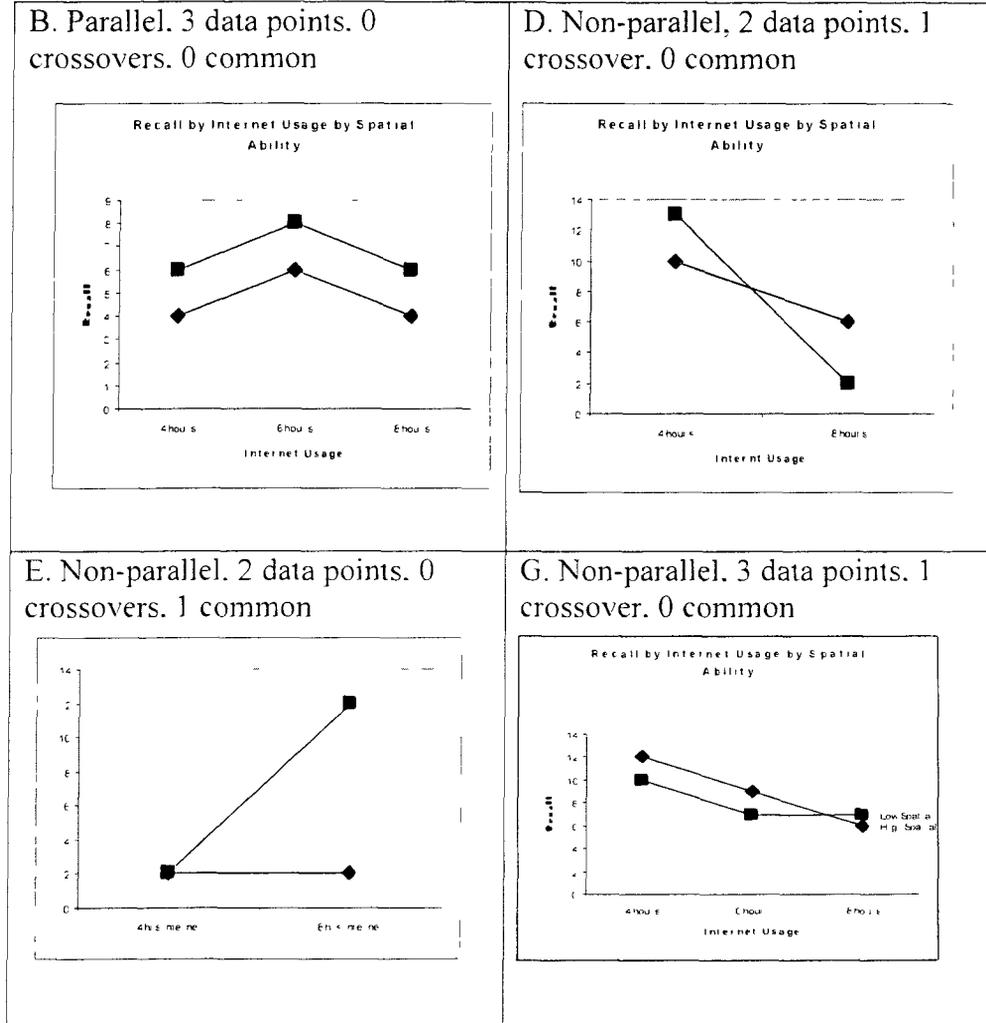
Preliminary Study 2 graph stimuli manipulations :

GRAPH STIMULI VARIABLES	LEVELS
Slope	positive, negative, zero
Parallel	both z-variables parallel; both z-variables non-parallel
Slope changes	no slope changes, positive to negative, positive to zero, negative to positive, negative to zero, zero to positive, zero to negative
Number of data points	2: 3
Number of Common Values	none, 1 left, 1 center, 1 right
Number of crossovers	none: 1 left of center: 1 at center: 1 right of center: 2 crossovers

These manipulations resulted in 173 unique graphs. Examples of these are provided in Table B2 below.

Table B2.

Preliminary Study 2 graph stimuli examples.



All graph stimuli were identical in size and all z variables were labeled on the right hand side of the graph. All graphs had the title *Recall By Internet Usage By Spatial Ability*. The ordinate was labeled *Recall* and values ranged from 0 to 14. The abscissa was labeled *Internet Usage* and included the values 4 hours, 6 hours, or 8 hours. Five of the 173 graph stimuli were reserved for practice trials. Each of the remaining 168 graph stimuli were presented in random order twice. In the first block of trials

participants were instructed to apply a visual criterion to judgments of complexity and in the second block of trials to apply a semantic criterion to judgments of complexity. All participants judged all the graphs in the same order. Stimulus presentation and timing was controlled using DirectRT™ on a Dell laptop computer with a 1400 X 1050 pixel resolution LCD display. In addition to the visual and semantic complexity judgments, participants were also required to increase their range of judgments from 3-points to 7. Participants entered their responses using the digit keys 1 (very simple) through 7 (very complex) on the top row of the keyboard. No other labels were attached to the keys. A cardboard overlay prevented participants from pressing any keys other than the valid digits and the spacebar to initiate the experiment.

Procedure

Upon entering the lab participants were given general instructions then asked to sign an Informed Consent form. They then filled out the Experience With Graphs questionnaire (Oakley & Tsuji, 2007) concerned with their prior experience with graphs and quantitative methods. This questionnaire is also found in Appendix A.

Participants then read specific visual instructions on the computer screen, which read as follows:

Please rate the graph's visual complexity.

What does that mean? Some graphs just look "simple". Other graphs just look "complex".

Rate the graphs on a 7-point scale where

1=very simple and 7=very complex visually.

This was followed by five visual practice trials followed by the 168 visual experimental trials, presented in random order. Following a participant-determined rest period they read the semantic instructions:

Please rate how complex the graph's meaning is to explain.

What does that mean? Imagine that you are a researcher who has collected the data in these graphs and you had to interpret the results to, say recommend internet usage in public schools.

Rate the graphs on a 7-point scale where

1=very simple and 7=very complex to explain.

These were followed by five semantic practice trials and another 168 experimental trials. On each trial the procedure was as follows: A ready screen with a "+" fixation point was the default display. After .5 s the default screen was replaced by a stimulus graph accompanied by one of two rating scale instructions: "Please rate the visual complexity (1-7)." On semantic rating trials the procedure was identical but the message at the bottom of the screen read "Please rate the 'explain-it' complexity (1-7)." All participants completed the visual criterion-trials first to prevent the semantic criterion from contaminating the purely visual criterion block of trials. On each experimental or practice trial the graph was replaced by the 'ready' screen and the next trial would begin immediately after responding with the digit rating.

At the end of the experimental trials participants completed the Visual Patterns Test. Participants viewed each checkerboard stimulus for 3 seconds and then attempted to duplicate it by marking the appropriate squares on a blank checkerboard (see Appendix A for examples). Finally, Participants were invited to ask questions, comment on the research, debriefed (Appendix A) and thanked for their participation.

Results and Discussion

Ratings

The mean ratings of each of the 168 graph stimuli were calculated for the visual ($M = 2.84$) and semantic ($M = 2.94$) criteria. The ratings were entered into a (2 x 2) ANOVA for criterion (visual, semantic) and rating (visual, semantic) where criterion approached significance, $F(1, 167) = 3.032, p = .08$. Since the purpose of Preliminary Study 2 was to identify a subset of graph stimuli that could be utilized in subsequent experiments, the 24 graphs rated most and the 24 rated least complex were selected for further exploration. The ANOVA yielded a significant criterion x rating interaction, $F(1, 956) = 49.16, p = .001$ which can be accounted for by the more extreme ratings for the semantic than for the visual criterion at both the high- and the low end of the scale. The criterion main effect was non-significant ($F(1, 167) = 3.03, p > .05$). However, as expected, the main effect for complexity was significant, $F(1, 956) = 855.4, p = .001$, with complex stimuli resulting in higher ratings than simple ones. This is shown in Figure 1 below.

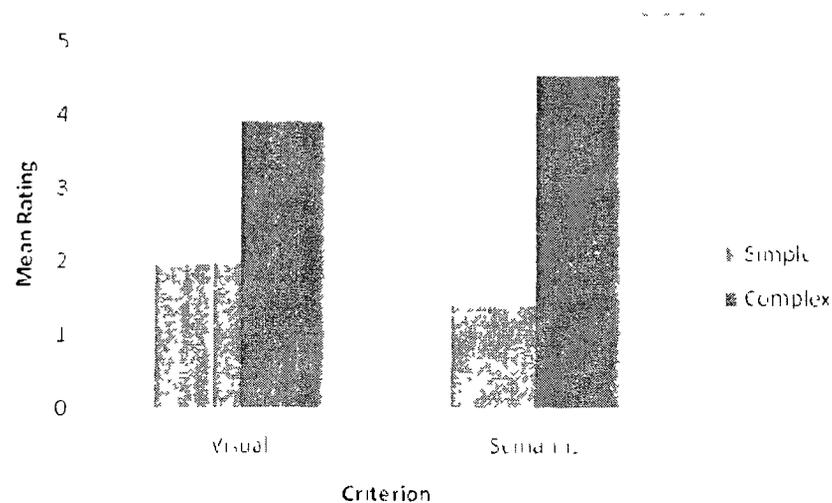


Figure B1. Visual and semantic criteria ratings of 24 simple and 24 complex graphs.

Response Times

A (2 x 2) ANOVA carried out for the same variables as above was also conducted for the mean response times associated with the 24 simple and 24 complex graphs. No interactions or main effects were significant. There was thus no evidence suggesting that participants gave considerably more thought to semantic than to visual judgments. Since the visual condition was designed to evoke an immediate response with a minimum of cognitive processing, the absence of a significant difference between the two judgmental criteria suggests that participants actually may have processed the stimuli equally deeply in both conditions. In an effort to prevent semantic processing of the stimuli, Preliminary Study 3 shortened the stimulus exposure time to a level at which no cognitive processing should be possible.

Graph attributes

In addition to the ratings and response time differences between simple and complex graphs observed in this preliminary study, several attributes of the graph stimuli

distinguished the two sets of stimuli. Recall that the graph stimuli were selected in terms of the number of crossovers, common points, slope changes, the number of data points, the number of parallel z-variables, and the number of segments with zero slope. The mean number of these attributes per graph is captured in Table 1 below. Overall, these attributes distinguished between the simple and complex graphs but failed to distinguish between visual and semantic graph criterion; no graph stimuli distinguished visual compared to semantic complexity in this Preliminary Study.

Table B3: Mean attributes of the simple/complex, visual/semantic graphs selected for further study

Graph Attributes	Visually	Visually	Semantically	Semantically
	Simple	Complex	Simple	Complex
Points	4.9	5.3	4.8	5.2
Crossovers	0.1	1.2	0.1	1.3
Common points	0.2	0.7	0.2	0.6
Slope changes	0.6	1.0	0.4	1.0
Parallel segments	0.2	0.0	0.3	0.0
Zero slope segments	1.0	0.5	1.0	0.4

Visual Patterns Test

The Visual Patterns Test scores can range from 1 to 16. In this study the observed range was 4 to 10 and the overall mean was 6.8. In order to explore the role that visual memory (as measured by the VPT) might play in participant ratings, the data from the 5 participants who scored less than the mean and the 5 who scored above the mean were

grouped together and the data entered into a $2 \times (2)$ mixed ANOVA for VPT (high, low), and criterion (visual, semantic). Only the main effect of VPT was significant, $F(1, 3)=93.82, p = .001$. As Figure 2 indicates, participants with higher VPT scores rated the graph stimuli higher than those with lower VPT on both criteria.

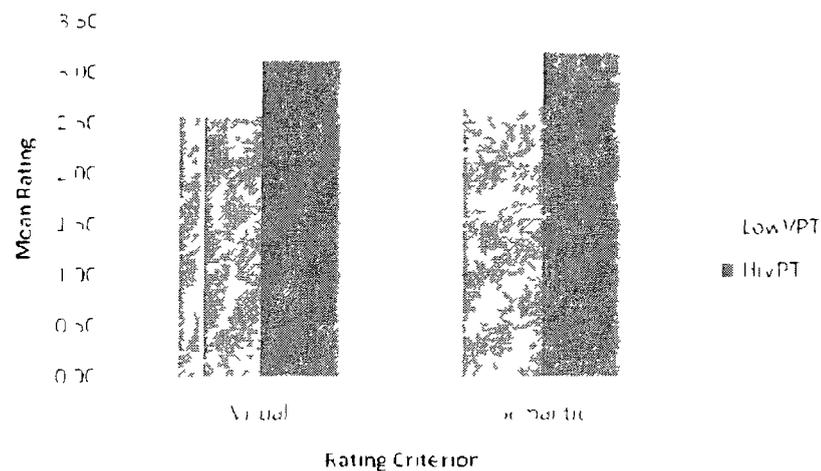


Figure B2 Graph rating as a function of VPT scores

A similar $2 \times (2)$ ANOVA conducted on the response time data revealed a significant VPT \times Criterion interaction ($F(1, 3)=68.64, p = .01$) and VPT main effect ($F(1, 3)=20.75, p = .01$). As demonstrated in Figure 3 below, the High VPT participants took more time to make their complexity judgments than Low VPT participants and this was particularly pronounced when the judgment was using a Visual criterion.

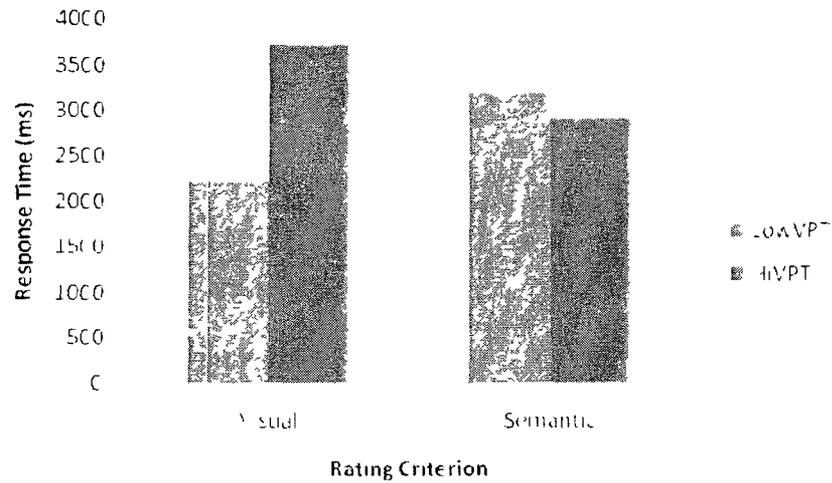


Figure B3. Response time as a function of ratings criterion.

Overall, Preliminary Study 2 enabled the selection of a set of simple and complex graph stimuli. As indicated by the results of Preliminary Study 1, simple and complex graphs were differentiated on the basis of the number of crossovers, the number of parallel line segments, slope changes, and the number of common points. Preliminary Study 2 was less successful in differentiating visual from semantic criteria for complexity. In an attempt to distinguish visual from semantic graph criteria, Preliminary Study 3 will force a more strictly visual processing regimen by reducing the exposure time of the graph stimuli.

Appendix C: Preliminary Study 3

Preliminary Study 2 asked participants to judge if graphs were simple or complex while adopting either a visual or semantic criterion for their ratings. That study demonstrated that participants could meaningfully apply those criteria and from an initial corpus of 168 graph stimuli a subset of 24 simple and 24 complex graphs based on visual criteria was identified and a similar subset of 24 simple and 24 complex graphs based on semantic criteria was also identified. However, since these graphs differed on the basis of many attributes such as the number of crossovers or number of slope changes, it remains possible that participants failed to distinguish between visual and semantic criteria and were simply responding to the graph attributes. Indeed, although the differences in graph *attributes* were substantial between simple and complex, the criteria by which these differences were rated along the visual and semantic dimensions were unclear. Furthermore, participant response times to visual criterion ratings were 2613 ms on average compared to an average of 3600 ms for semantic-criterion ratings. The ample viewing time of visual criterion graphs suggests that there is at least the possibility some of the information may be processed semantically. The present study was therefore designed specifically to resolve this issue by shortening the viewing time of visual criterion graphs to 150 ms followed by a visual mask. The viewing time and visual mask were selected after pilot testing indicated that participants had difficulty recalling the visual characteristics of graphs so displayed.

Preliminary Study 3 also asked participants to rate graphs on a 9-point scale as opposed to the 7-point scale utilized previously.

Method

Design

Preliminary Study 3 was entirely within-subjects, similar to that of Preliminary Study 2. Visual and Semantic ratings were the two within-subject levels.

Participants

A sample of 13 undergraduate students who did not participate in either of the two previous studies was recruited using the Department of Psychology system for participation in psychological experiments and were given a 1% course credit for participation. All had normal or corrected-to-normal vision. Data from two participants were excluded when it was clear that they had not understood the instructions and gave ratings of '1' to almost all stimuli. Data from the first participant was discarded due to a technical error in the setting of display parameters. Therefore, the analysis includes data from 10 participants. Mean age of participants was 24.9 yrs and the sample included an equal number of males and females.

Apparatus & Materials

The graph stimuli were a subset of graphs from Preliminary Study 2 comprising the 24 graphs rated most complex and 24 rated most simple visually as well as a similar set of 24 semantically most complex and 24 semantically most simple. As before, all were black line graphs plotted on x - and y -coordinates against a white background, all were limited to two lines (z -variables) and were the same size. As before, they varied in six ways: (1) by slope (positive, negative, or zero): (2) by whether the z -variables are

parallel or non-parallel; (3) by whether the slope changed direction, (4) by the number of data points; (5) the number of points in common, and (6) the number of crossovers. As before, no graph legends were displayed but in addition all titles and abscissa and ordinate labels were also removed. Instead, a paper graph titled Gas Prices By Time of Year and accompanied by an appropriate legend and axis labels was provided and participants were instructed to assume that this title, labels, and legend were associated with the graph stimuli presented. Seven of the 168 graph stimuli from Preliminary Study 2 that were not included in the present study were reserved for practice trials.

Stimulus presentation and timing was controlled using DirectRT™ on a laptop computer with a 1400 X 1050 pixel resolution LCD display. Participants entered their responses using the digit keys 1 (very simple) through 9 (very complex) on the top row of the keyboard. No other labels were attached to the keys. A cardboard overlay prevented participants from pressing any keys other than the valid digits or the spacebar to initiate the experiment.

Procedure

The procedure was identical to Preliminary Study 2 with the following exceptions: Seven visual criterion practice trials were presented, followed by 48 experimental trials, followed by a participant-determined rest period. Then seven semantic criterion practice trials were given, followed by 48 experimental trials. On each trial a ready screen consisting of a 9-point scale was positioned in the center of the screen for 1000 ms. On visual criterion trials this was replaced by the visual criterion graph displayed for 150 ms and a mask consisting of 100 random lines was presented for

100 ms at a Stimulus Onset Asynchrony (SOA) of 150ms. On semantic criterion trials the mask SOA was 3000 ms. However, if participants made a valid response within that time period the SOA was immediately shortened and the mask was displayed. Each participant received the graph stimuli in a unique random order, but all participants rated the graphs on the visual criterion first followed by the semantic criterion as in Preliminary Study 2. On each experimental or practice trial the graph was replaced by the ready screen and the next trial would begin immediately after responding with the digit rating.

Results and Discussion

Ratings

The mean ratings of each of the 48 graph stimuli were calculated for the visual ($M = 2.87$) and semantic ($M = 3.49$) criteria. The ratings were entered into an overall paired-samples t -test that included all the experimental stimuli. It yielded a significant difference, $t(479) = 6.179$, $p = .0001$. A comparison of the 25 graphs that were common to both the visual ($M = 2.9$) and semantic ($M = 3.5$) graph sets yielded a similar result, namely $t(249) = 6.028$, $p = .0001$. Thus, the different viewing times and the different instructions for rating the two sets of identical graphs resulted in different judgments in the two conditions. These 25 common graph stimuli may be found in Table C2. The remaining analyses will focus on this subset of 25 graph stimuli.

Response Times

Although response times were collected, the 9-point rating scale made interpretation difficult. Most participants were observed resting the fingers of one hand

on the lower portion of the response range (keys 1-4) and so the fact that the semantic criterion graph ratings were skewed towards the upper range of responses (keys 5-9) meant that response times for semantic graphs would be longer than for visual criterion graphs. Thus, the response times were not analyzed in this Preliminary Study.

Graph attributes

Recall that the graph stimuli were selected in terms of the number of crossovers, common points, slope changes, the number of data points, the number of parallel z-variables, and the number of segments with zero slope. Dividing the 25 common graph stimuli into those below and those above the overall mean complexity ratings, a subset of 12 simple and 13 complex graphs results. The mean number of graph attributes for these simple and complex graphs is captured in Table C1 below. Complex graphs appeared to have more points, more crossovers, and more slope changes while Simple graphs more often had zero slope segments and segments that were parallel.

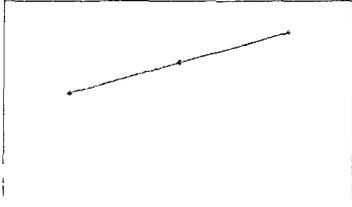
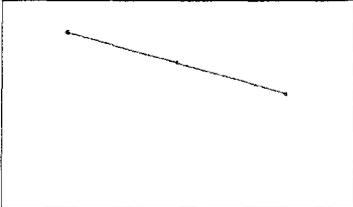
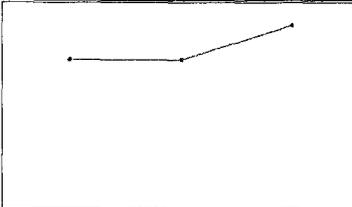
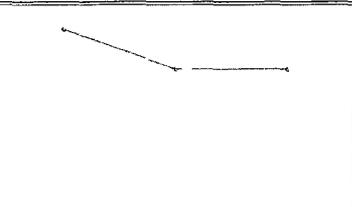
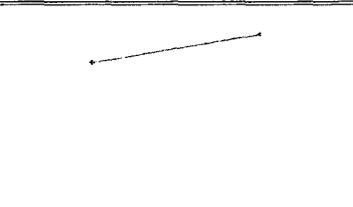
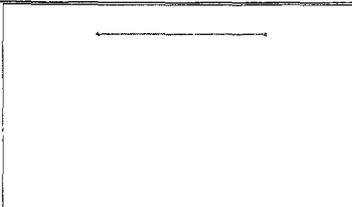
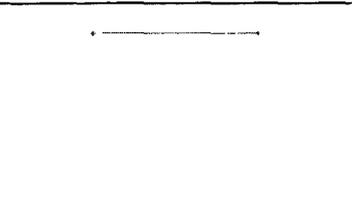
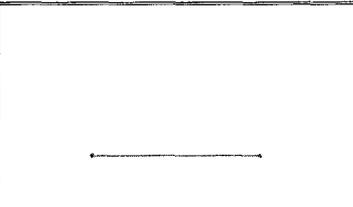
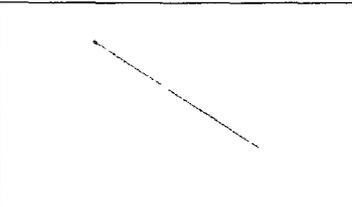
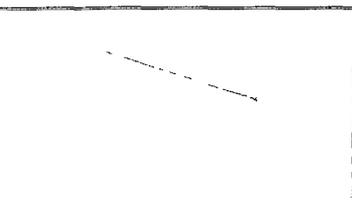
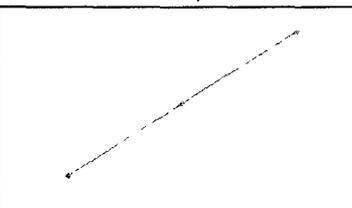
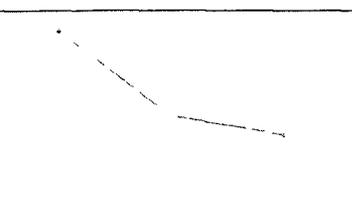
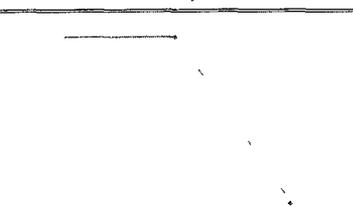
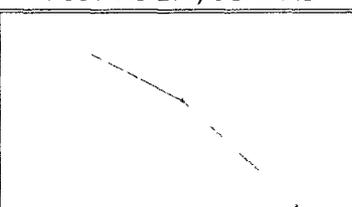
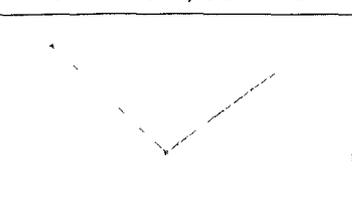
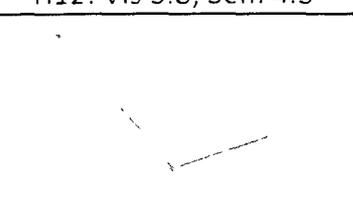
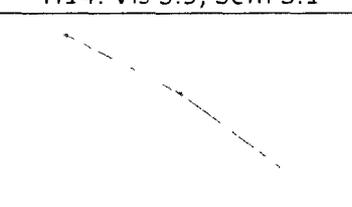
Table C1

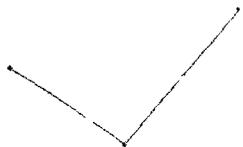
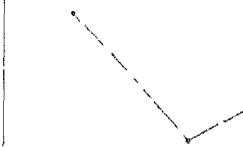
Mean attributes of the 25 common simple and complex graphs selected for further study

Graph attributes	Simple	Complex
Points	2.5	3.0
Crossovers	0.0	1.5
Common points	0.2	0.5
Slope changes	0.3	1.0
Parallel segments	0.7	0.0
Zero slope segments	1.0	0.6

Overall, Preliminary Study 3 provided a set of 25 graph stimuli that could be differentiated on the basis of visual and semantic criteria thus increasing the probability that a larger study with more participants (Experiment 1) might successfully explore this characteristic of graph complexity . Furthermore, the range of graph attributes in these stimuli allow participants to identify simple and more complex subsets. As a result, a graph stimulus set has been obtained that might be effectively utilized in the subsequent experiments of this thesis.

Table C2: Preliminary Study 3, 25 Common Graph Stimuli (with mean Visual and Semantic ratings).

 <p>B01: Vis 1.6; Sem 1.4</p>	 <p>B02: vis 2.1; Sem 1.5</p>	 <p>B06: Vis 2.4; Sem 2.2</p>
 <p>B09: Vis 2.1; Sem 2.0</p>	 <p>C02: Vis 1.6; Sem 1.8</p>	 <p>C03: Vis 1.5; Sem 2.2</p>
 <p>C04: Vis 1.6; Sem 1.6</p>	 <p>c05: Vis 1.8; Sem 1.9</p>	 <p>E02: Vis 2.2; Sem 2.1</p>
 <p>E07: Vis 1.9; Sem 1.8</p>	 <p>F04: Vis 1.7; Sem 2.1</p>	 <p>F05: Vis 2.4; Sem 2.5</p>
 <p>H04: Vis 3.0; Sem 4.3</p>	 <p>H12: Vis 3.8; Sem 4.5</p>	 <p>H14: Vis 3.9; Sem 5.1</p>
 <p>H22: Vis 4.4; Sem 5.2</p>	 <p>H25: Vis 4.0; Sem 4.0</p>	 <p>H26: Vis 3.8; Sem 5.5</p>

 <p>J01: Vis 3.2; Sem 5.3</p>	 <p>J02: Vis 3.8; Sem 4.9</p>	 <p>J03: Vis 4.3; Sem 4.3</p>
 <p>J04: Vis 3.8; Sem 4.7</p>	 <p>J05: Vis 3.6; Sem 5.6</p>	 <p>J06: Vis 3.0; Sem 4.7</p>
 <p>J10: Vis 4.2; Sem 5.3</p>		

Appendix D: Experiment 1

D1: Instructions

D2: stimuli

D3: Ratings descriptive statistics

D4: Ratings ANOVA

D5: Response time descriptive statistics

D6: Response time ANOVA

D7: Experience with graphs questionnaire

D8: Visual patterns test example

D9: Visual mask

APPENDIX D1: Experiment 1 Instructions

In this study we are investigating how people perceive pictures of graphs.

Your task is to decide on the difficulty level of each graph.

If you think that the graph is simple, press the “4” key on the left.

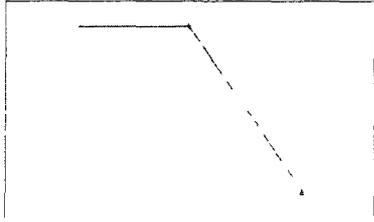
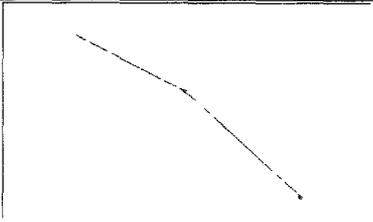
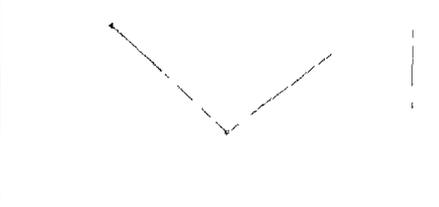
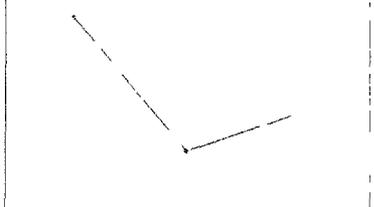
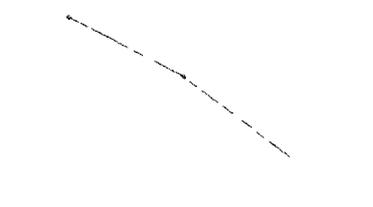
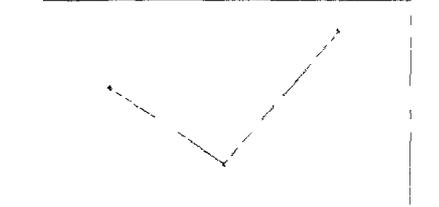
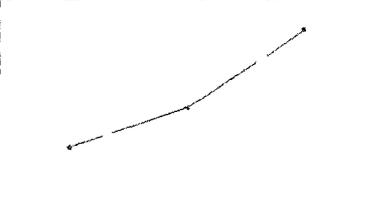
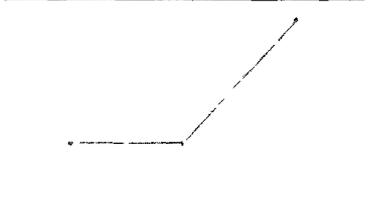
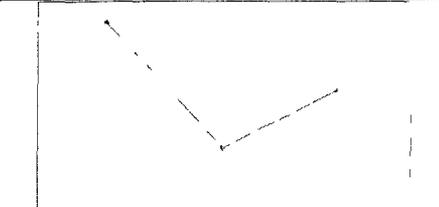
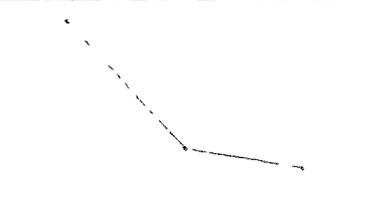
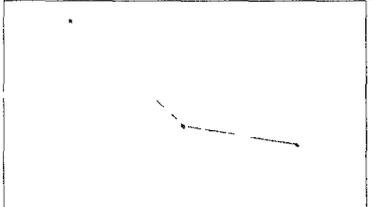
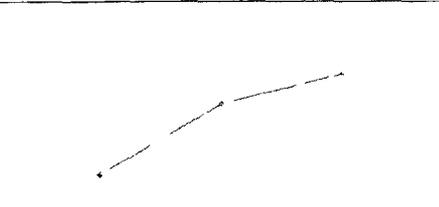
If you think that the graph is complex, press the “6” key on the right.

If you think that the graph is somewhere between simple and complex, press the “5” key in the center.

First you will do some practice trials just to make sure the instructions are clear. This will be followed by the experimental trials with a short break halfway through.

Any questions before you begin the practice trials?

Appendix D2: Experiment 1, 12 Common Graph Stimuli

	 <p>H12: Vis 3.8; Sem 4.5</p>	 <p>H14: Vis 3.9; Sem 5.1</p>
 <p>H22: Vis 4.4; Sem 5.2</p>	 <p>H25: Vis 4.0; Sem 4.0</p>	 <p>H26: Vis 3.8; Sem 5.5</p>
 <p>J01: Vis 3.2; Sem 5.3</p>	 <p>J02: Vis 3.8; Sem 4.9</p>	 <p>J03: Vis 4 3; Sem 4.3</p>
 <p>J04: Vis 3.8; Sem 4.7</p>	 <p>J05: Vis 3.6; Sem 5.6</p>	 <p>J06: Vis 3.0; Sem 4.7</p>
 <p>J10: Vis 4.2; Sem 5.3</p>		

Appendix D3: Experiment 1, Descriptive Statistics Ratings

Criterion	Expertise	Domain	Mean	SD	N
Visual	Novice	Business	2.705	1.063	10
		Psychology	2.519	0.921	10
		Total	2.612	0.973	20
	Expert	Business	2.752	1.351	10
		Psychology	2.161	0.616	10
		Total	2.457	1.066	20
	Total	Business	2.729	1.184	20
		Psychology	2.340	0.785	20
		Total	2.534	1.011	40
Semantic	Novice	Business	3.748	1.146	10
		Psychology	3.536	0.932	10
		Total	3.642	1.022	20
	Expert	Business	3.028	1.728	10
		Psychology	2.600	1.063	10
		Total	2.814	1.414	20
	Total	Business	3.388	1.474	20
		Psychology	3.068	1.085	20
		Total	3.228	1.288	40

Appendix D4: Experiment 1, Ratings ANOVA

Source	<i>df</i>	<i>F</i>	<i>p</i>
Criterion	1	19.375	.0001
Criterion x Expertise	1	4.551	.040
Criterion x Domain	1	0.047	.829
Criterion x Expertise x Domain	1	0.090	.766
Error	36		
Expertise	1	2.277	.140
Domain	1	1.181	.284
Expertise x Domain	1	0.227	.637
Error	36		

Appendix D5: Experiment 1, Descriptive Statistics Response Time

Criterion	Expertise	Domain	Mean	SD	N
Visual	Novice	Business	1937	1028.5	10
		Psychology	1460	350.9	10
		Total	1698	786.9	20
	Expert	Business	1605	557.7	10
		Psychology	1179	364.2	10
		Total	1392	508.0	20
	Total	Business	1771	823.1	20
		Psychology	1320	376.9	20
		Total	1545	672.0	40
Semantic	Novice	Business	2878	1212.5	10
		Psychology	1948	586.4	10
		Total	2413	1042.5	20
	Expert	Business	2428	940.8	10
		Psychology	1768	831.6	10
		Total	2098	928.2	20
	Total	Business	2653	1081.2	20
		Psychology	1858	706.4	20
		Total	2255	987.3	40

Appendix D6: Experiment 1, Response Time ANOVA

Source	<i>df</i>	<i>F</i>	<i>p</i>
Criterion	1	30.746	.0001
Criterion x Expertise	1	0.001	.974
Criterion x Domain	1	1.797	.189
Criterion x Expertise x Domain	1	0.184	.671
Error	36		
Expertise	1	2.088	.157
Domain	1	8.392	.006
Expertise x Domain	1	.138	.712
Error	36		

Appendix D7 : Graph Use Questionnaire
(Oakley & Tsuji, 2008)

1. Gender: Male Female
2. Age: 18-25 26-35 36-50 over 50
3. How many years of post-secondary education have you completed? (select one only)
 - < 1 year 1-4 years 5-8 years > 8 years
4. What is your primary area of study? (select one only)
 - psychology business science engineering
 - mathematics statistics humanities english
 - other (please specify): _____
5. Which kind of graph are you most familiar with? (select one only)
 - line graphs bar graphs pie charts maps
 - polar charts waterfall chart scatter diagram quadrant chart
 - other (please specify): _____
6. How frequently do you *encounter* graphs in a typical week?
 - once 2-3 times 4-5 times daily several times a day
7. Are you required to *create* graphs?
 - yes no

If yes,

 - a) How often? _____
 - b) what type of graph(s)? _____
8. In general, how easy or difficult do you find it to work with graphs? (draw a line through the scale below to indicate your general level of difficulty)

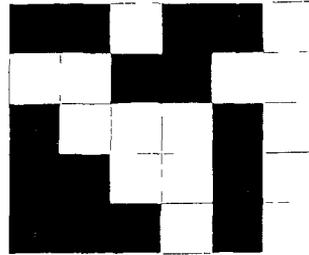
-----|-----

Easy **Difficult**

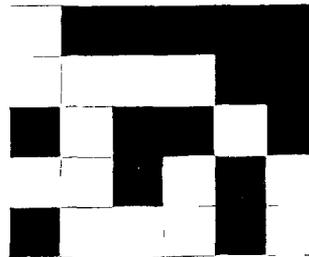
Thank you!

Appendix D8: Visual Patterns Test Examples

(Della Sala, Gray, Baddeley, & Wilson, 1997)

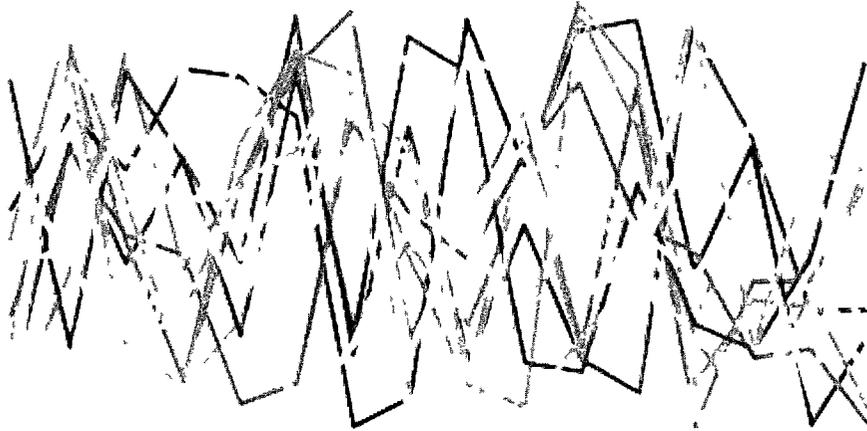


546/15



547/15

Appendix D9: Visual Mask



Appendix E: Experiment 2

E1: Experiment 2 Instructions

E 2: Graphs Stimuli for Experiment 2

E 3: Visual Patterns Test descriptive statistics

E4: Word count descriptive statistics

E 4: Word count ANOVA

E 5: Because descriptive statistics

E 6: Because ANOVA

E 7: Between descriptive statistics

E 8: Between ANOVA

E 9: Direction descriptive statistics

E 10: Direction ANOVA

E 11: Quantitative descriptive statistics

E 12: Quantitative ANOVA

E 13: Title descriptive statistics

E 14: Title ANOVA

E 15: Trend descriptive statistics

E 16: Trend ANOVA

E 17: x-axis descriptive statistics

E 18: x-axis ANOVA

E 19: y-axis descriptive statistics

E 20: y-axis ANOVA

E 21: z-variable descriptive statistics

E 22: z-variable ANOVA

E 23:Completeness descriptive statistics

E 24 Completeness ANOVA

E 25:Nominal, Ordinal, Metric descriptive statistics

E 26: Nominal, Ordinal, Metric ANOVA

E 27:Response Time descriptive statistics

E 28:Response Time ANOVA

Appendix E1: Detailed instructions for Experiment 2

You will now have four practice trials followed by 24 experimental trials. On each trial the following will occur: First, I will ask you to watch the plus sign currently in the middle of the computer screen. When you are ready, press the SPACEBAR and the plus sign will be replaced by a graph. Please examine this graph and when you are ready to describe it, press the SPACEBAR and immediately begin describing that graph and when you are done press SPACEBAR again. Now, I would like you to try to provide an explanation of what is happening in that graph. Talk aloud with your explanation and when you feel you are done, press SPACEBAR again. Then, the screen will go blank and you can rest until you feel ready for the next graph. The next graph will be displayed when you press SPACEBAR again and the same procedure will be followed. You will now have four practice trials followed by 24 experimental trials. You can work as fast or as slow as you like. Is that clear? Are you ready to begin?

Appendix E2: Experiment 2 Graph Stimuli

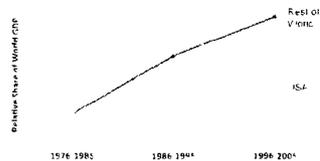
		<p>Exports As A Percentage of GDP 2006 2008 Brazil versus Germany</p>	
<p>Describe</p>	<p>Explain</p>		
		<p>World Exports As A Function of GDP 2006 2008 Goods versus Services</p>	
<p>Describe</p>	<p>Explain</p>		
		<p>Foreign Direct Investment As A Proportion of GDP 2006 2008 Brazil versus China</p>	
<p>Describe</p>	<p>Explain</p>		
		<p>Average Annual GDP Growth 2006 2008 Brazil/Russia/India/China versus Big10</p>	
<p>Describe</p>	<p>Explain</p>		



Describe



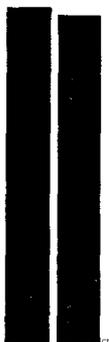
Relative Share of World GDP Over 3 Decades
USA versus Rest of World



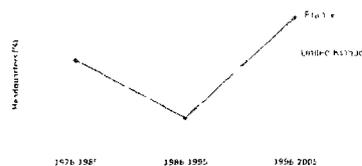
Explain



Describe



Headquarters of World's Largest Corporations
Over 3 Decades France versus United Kingdom



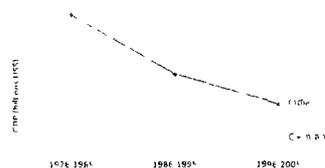
Explain



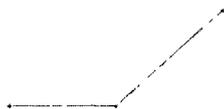
Describe



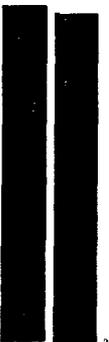
GDP (billions US\$) Adjusted for Purchasing Power
Over Three Decades India versus Germany



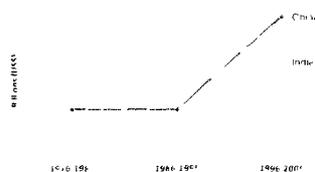
Explain



Describe



Foreign Direct Investment (billions US\$) Over 3
Decades China versus India



Explain



	<p><i>Describe</i></p>			<p>Proportion of Total Population < 30 Years Old Over 3 Decades: Japan versus China</p> <p><i>Explain</i></p>	
	<p><i>Describe</i></p>			<p>Software Piracy Rates: 2006-2008 Pakistan versus Nigeria</p> <p><i>Explain</i></p>	
	<p><i>Describe</i></p>			<p>Forecast Stability of Caspian Sea Oil Pipelines 2010-2012: Azerbaijan versus Chechnya</p> <p><i>Explain</i></p>	
	<p><i>Describe</i></p>			<p>Enkes As A Percentage of Annual Revenues 2006-2008: Armenia versus Kazakhstan</p> <p><i>Explain</i></p>	



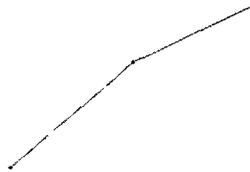
Describe



Number of Publications By Year By School of Thought



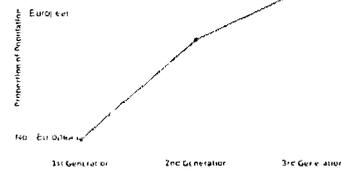
Explain



Describe



Canadian Population by Generational Status by Ethnic Origin



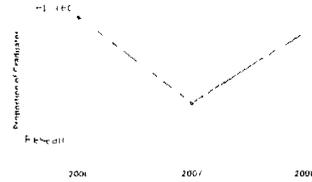
Explain



Describe



Psychology Graduates by Year by Employment Setting



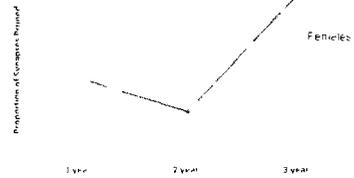
Explain



Describe



Pruning of Visual Cortex Synapses By Age By Gender



Explain

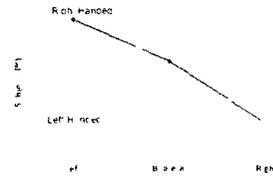




Describe



Proportion of Subjects By Speech Localization By Handedness



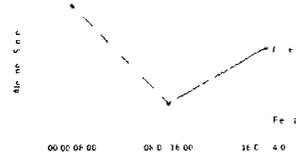
Explain



Describe



Alertness By Time of Day By Gender



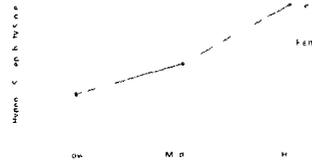
Explain



Describe



Hypnotic Susceptibility By Field Dependence By Gender



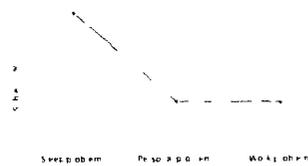
Explain



Describe

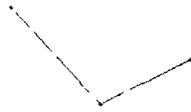


Subjects (%) Reporting Problems Insomniacs versus Controls



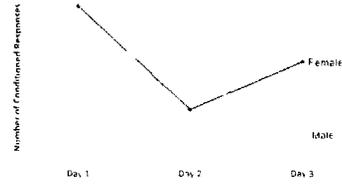
Explain





Describe

Spontaneous Recovery of Conditioned Responses
By Days After Extinction By Gender

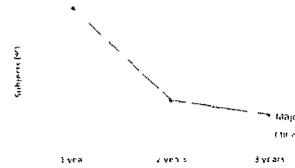


Explain



Describe

Proportion of Subject Flashbulb Memories By
Years Since Event Major versus Minor Distortions

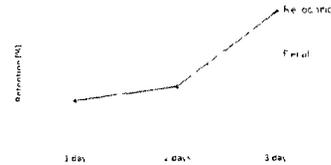


Explain



Describe

Retention By Retention Interval
Recognition versus Recall

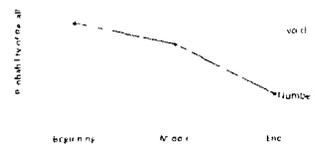


Explain



Describe

Probability of Recall By Serial Position
Words versus Numbers



Explain

Appendix E3: Experiment 2 Visual Patterns Test Descriptive Statistics

Expertise	Easy/Difficult	VPT
Novice	40	7
	50	3
	36	6
	30	5
	33	4
	74	5
	63	7
	50	7
	69	7
	41	6
BusExp	15	1
	44	8
	50	6
	5	6
	20	6
	30	5
	50	7
	4	8
PsyExp	17	12
	22	6
	33	8
	57	5
	55	10
	91	8
	9	7
	59	5

Appendix E4: Experiment 2 Descriptive statistics Describe Words

Graph Type	Expertise	Mean	<i>SD</i>	<i>N</i>
SimBus	Novice	37.44	19.58	10
	BusExp	50.71	44.77	8
	PsyExp	90.08	50.93	8
	Total	57.72	44.05	26
SimPsy	Novice	36.47	20.46	10
	BusExp	55.95	40.60	8
	PsyExp	88.75	50.43	8
	Total	58.55	42.58	26
ComBus	Novice	39.13	25.17	10
	BusExp	52.20	37.09	8
	PsyExp	83.58	50.80	8
	Total	58.83	41.19	26
ComPsy	Novice	38.12	22.08	10
	BusExp	51.99	39.48	8
	PsyExp	85.10	53.88	8
	Total	56.84	42.76	26

Appendix E5: Experiment 2, Describe Words ANOVA

Source	<i>df</i>	<i>F</i>	<i>p</i>
Complexity	1	0.626	.437
Complexity x Expertise	2	1.028	.374
Error	23		
Domain	1	0.128	.724
Domain x Expertise	2	0.475	.628
Error	23		
Complexity x Domain	1	0.052	.821
Complexity x Domain x Expertise	2	0.376	.691
Error	23		
Expertise	2	3.824	.037
Error	23		

Appendix E6: Experiment 2 Descriptive statistics Think Time

Graph Type	Expertise	Mean	<i>SD</i>	<i>N</i>
SimBus	Novice	10.57	8.89	10
	BusExp	2.09	0.69	8
	PsyExp	4.33	5.04	8
	Total	6.04	7.06	26
SimPsy	Novice	11.09	5.95	10
	BusExp	2.71	1.43	8
	PsyExp	4.13	3.92	8
	Total	6.37	5.69	26
ComBus	Novice	10.71	7.32	10
	BusExp	1.79	0.73	8
	PsyExp	3.28	2.59	8
	Total	5.68	6.18	26
ComPsy	Novice	13.33	8.24	10
	BusExp	3.00	1.76	8
	PsyExp	4.53	4.29	8
	Total	7.44	7.31	26

Appendix E7: Experiment 2, Think Time ANOVA

Source	<i>df</i>	<i>F</i>	<i>p</i>
Complexity	1	1.162	.202
Complexity x Expertise	2	3.23	.058
Error	23		
Domain	1	9.37	.006
Domain x Expertise	2	0.911	.416
Error	23		
Complexity x Domain	1	2.76	.110
Complexity x Domain x Expertise	2	0.287	.753
Error	23		
Expertise	2	8.063	.002
Error	23		

Appendix E8: Experiment 2 Descriptive statistics Explain Response Time

Graph Type	Expertise	Mean	<i>SD</i>	<i>N</i>
SimBus	Novice	63.6	38.8	10
	BusExp	67.4	32.3	8
	PsyExp	74.8	25.7	8
	Total	68.2	32.3	26
SimPsy	Novice	69.4	44.6	10
	BusExp	71.5	43.0	8
	PsyExp	77.4	30.6	8
	Total	72.5	38.8	26
ComBus	Novice	69.0	45.5	10
	BusExp	71.6	40.1	8
	PsyExp	77.0	26.1	8
	Total	72.3	37.4	26
ComPsy	Novice	69.6	36.4	10
	BusExp	82.7	39.9	8
	PsyExp	85.8	26.1	8
	Total	78.6	34.2	26

Appendix E9: Experiment 2, Explain Response Time ANOVA

Source	<i>df</i>	<i>F</i>	<i>p</i>
Complexity	1	11.13	.003
Complexity x Expertise	2	0.853	.439
Error	23		
Domain	1	5.825	.024
Domain x Expertise	2	.334	.720
Error	23		
Complexity x Domain	1	0.419	.524
Complexity x Domain x Expertise	2	0.970	.394
Error	23		
Expertise	2	0.204	.817
Error	23		

Appendix E10: Experiment 2 Descriptive statistics Completeness

Graph Type	Expertise	Mean	<i>SD</i>	<i>N</i>
SimBus	Novice	.262	.14	10
	BusExp	.431	.21	8
	PsyExp	.476	.12	8
	Total	.380	.18	26
SimPsy	Novice	.268	.11	10
	BusExp	.479	.13	8
	PsyExp	.541	.12	8
	Total	.417	.17	26
ComBus	Novice	.268	.13	10
	BusExp	.463	.18	8
	PsyExp	.486	.12	8
	Total	.395	.17	26
ComPsy	Novice	.286	.139	10
	BusExp	.505	.19	8
	PsyExp	.519	.11	8
	Total	.425	.18	26

Appendix E11: Experiment 2, Completeness ANOVA

Source	<i>df</i>	<i>F</i>	<i>p</i>
Complexity	1	0.744	.397
Complexity x Expertise	2	0.502	.612
Error	23		
Domain	1	9.897	.005
Domain x Expertise	2	1.187	.323
Error	23		
Complexity x Domain	1	0.157	.695
Complexity x Domain x Expertise	2	0.372	.694
Error	23		
Expertise	2	8.020	.002
Error	23		

Appendix E12: Experiment 2 Descriptive statistics Nominal, Ordinal, Metric

	Expertise	Mean	Std. Deviation	N
NomSimBus	Novice	.5250	.39878	10
	BusExp	.5000	.32733	8
	PsyExp	.5613	.37460	8
	Total	.5285	.35658	26
NomSimPsy	Novice	.5250	.38097	10
	BusExp	.6563	.29693	8
	PsyExp	.4988	.33215	8
	Total	.5573	.33538	26
NomComBus	Novice	.5820	.36150	10
	BusExp	.5612	.19802	8
	PsyExp	.4688	.41052	8
	Total	.5408	.32816	26
NomComPsy	Novice	.4660	.35930	10
	BusExp	.5625	.30807	8
	PsyExp	.4375	.32043	8
	Total	.4869	.32339	26
OrdSimBus	Novice	.5000	.31180	10
	BusExp	.8438	.35197	8
	PsyExp	.9788	.06010	8
	Total	.7531	.33948	26
OrdSimPsy	Novice	.5750	.42573	10
	BusExp	.8125	.25877	8
	PsyExp	.9150	.09087	8
	Total	.7527	.32942	26
OrdComBus	Novice	.5990	.39422	10
	BusExp	.7900	.29345	8
	PsyExp	.9687	.08839	8
	Total	.7715	.32666	26
OrdComPsy	Novice	.5500	.36065	10
	BusExp	.8325	.26816	8
	PsyExp	1.0000	.00000	8
	Total	.7754	.32321	26
MetSimBus	Novice	.0000	.00000	10
	BusExp	.1250	.26726	8

	PsyExp	.3963	.34446	8
	Total	.1604	.28586	26
MetSimPsy	Novice	.0250	.07906	10
	BusExp	.2813	.38816	8
	PsyExp	.2938	.27882	8
	Total	.1865	.28843	26
MetComBus	Novice	.0670	.11634	10
	BusExp	.3138	.34990	8
	PsyExp	.4063	.26517	8
	Total	.2473	.28520	26
MetComPsy	Novice	.0170	.05376	10
	BusExp	.2713	.31980	8
	PsyExp	.4375	.17678	8
	Total	.2246	.26621	26

Appendix E13: Experiment 2, Nominal, Ordinal, Metric ANOVA

Source	<i>df</i>	<i>F</i>	<i>p</i>
NomOrdMet	2	29.07	.0001
NomOrdMet x Expertise	4	1.863	.133
Error	46		
Complexity	1	1.13	.297
Complexity x Expertise	2	0.028	.973
Error	23		
Domain	1	0.007	.932
Domain x Expertise	2	1.538	.236
Error	23		
NomOrdMet x Complexity	2	3.267	.047
NomOrdMet x Complexity x Expertise	4	0.827	.515
Error	46		
NomOrdMet x Domain	2	0.025	.975
NomOrdMet x Domain x Expertise	4	0.310	.870
Error	46		
Complexity x Domain	1	0.533	.473
Complexity x Domain x Expertise	2	1.445	.256
Error	23		
NomOrdMet x Complexity x Domain	2	0.578	.565
NomOrdMet x Complexity x Domain x Expertise	4	0.727	.578
Error	46		
Expertise	2	7.854	.003
Error	23		

Appendix E14: Experiment 2 Descriptive statistics Domain

Graph Type	Expertise	Mean	<i>SD</i>	<i>N</i>
Simple Same	BusExp	.438	.207	8
	PsyExp	.540	.124	8
	Total	.489	.173	16
Complex Same	BusExp	.483	.177	8
	PsyExp	.521	.105	8
	Total	.502	.142	16
Simple Different	BusExp	.472	.130	8
	PsyExp	.476	.123	8
	Total	.474	.122	16
Complex Different	BusExp	.484	.202	8
	PsyExp	.484	.117	8
	Total	.484	.160	16

Appendix E15: Experiment 2, Domain ANOVA

Source	<i>df</i>	<i>F</i>	<i>p</i>
Same/Different	1	1.590	.228
Same/Different x Expertise	1	6.561	.023
Error	14		
Complexity	1	0.333	.573
Complexity x Expertise	1	0.744	.403
Error	14		
Same/Different x Complexity	1	0.014	.908
Same/Different x Complexity x Expertise	1	0.981	.339
Error	14		
Expertise	1	0.256	.621
Error	14		

Appendix E16: Experiment 2 Descriptive Statistics Explanations

Graph Type	Expertise	Mean	<i>SD</i>	<i>N</i>
SimBus	Novice	.150	.34	10
	BusExp	.375	.38	8
	PsyExp	.521	.37	8
	Total	.333	.38	26
SimPsy	Novice	.158	.24	10
	BusExp	.344	.27	8
	PsyExp	.667	.30	8
	Total	.372	.33	26
ComBus	Novice	.240	.28	10
	BusExp	.354	.39	8
	PsyExp	.542	.34	8
	Total	.368	.35	26
ComPsy	Novice	.170	.21	10
	BusExp	.542	.34	8
	PsyExp	.563	.37	8
	Total	.405	.35	26

Appendix E17: Experiment 2, Explanations ANOVA

Source	<i>df</i>	<i>F</i>	<i>p</i>
Complexity	1	0.817	.375
Complexity x Expertise	2	1.082	.356
Error	23		
Domain	1	1.339	.259
Domain x Expertise	2	1.052	.365
Error	23		
Complexity x Domain	1	0.004	.950
Complexity x Domain x Expertise	2	1.702	.204
Error	23		
Expertise	2	4.728	.019
Error	23		

Appendix E18: Step 2, 3, 4, 5 themes by novices, business-, and psychology experts.

Step2 Themes	Novice	BusExp	PsyExp
BetweenZ	21%	26%	23%
y-Axis	19%	36%	23%
z-Variable	10%	9%	20%
x-Axis	19%	19%	17%
Trend	19%	2%	10%
Because	7%	6%	3%
Title	3%	1%	2%
Direction	18%	1%	1%

Step3	Novice	BusExp	PsyExp
BetweenZ	19%	15%	32%
x-Axis	26%	38%	30%
Because	13%	11%	16%
y-Axis	16%	18%	10%
Trend	6%	7%	6%
z-Variable	9%	7%	6%
Title	2%	2%	1%
Direction	8%	1%	0%

Step4	Novice	BusExp	PsyExp
4.BetweenZ	29%	46%	30%
4.z-variables	12%	10%	26%
4.Because	14%	17%	22%
4.y-axis	0%	3%	13%
4.x-axis	6%	12%	6%
4.Trend	12%	9%	2%
4.Direction	27%	0%	0%
4.Quant	0%	0%	0%
4.Title	0%	2%	0%

Step5	Novice	BusExp	PsyExp
5.BetweenZ	38%	49%	55%
5.Because	19%	22%	30%
5.Trend	13%	4%	6%
5.x-axis	0%	6%	4%
5.Direction	28%	0%	3%
5.y-axis	3%	8%	1%
5.z-variables	0%	6%	1%
5.Quant	0%	0%	0%
5.Title	0%	4%	0%

Appendix F: Experiment 3

- F1: Experiment 3 Instructions
- F2: Graphs Stimuli for Experiment 3
- F3: Instructional Materials Experiment 3
- F4: Word count descriptive statistics
- F5: Word count ANOVA
- F6: Because descriptive statistics
- F7: Because ANOVA
- F8: Between descriptive statistics
- F9: Between ANOVA
- F10: Direction descriptive statistics
- F11: Direction ANOVA
- F12: Quantitative descriptive statistics
- F13: Quantitative ANOVA
- F14: Title descriptive statistics
- F15: Title ANOVA
- F16: Trend descriptive statistics
- F17: Trend ANOVA
- F18: x-axis descriptive statistics
- F19: x-axis ANOVA
- F20: y-axis descriptive statistics
- F21: y-axis ANOVA
- F22: z-variable descriptive statistics
- F23: z-variable ANOVA
- F24: Completeness descriptive statistics
- F25: Completeness ANOVA
- F26: Nominal. Ordinal. Metric descriptive statistics
- F27: Nominal. Ordinal. Metric ANOVA
- F28: Response Time descriptive statistics
- F29: Response Time ANOVA

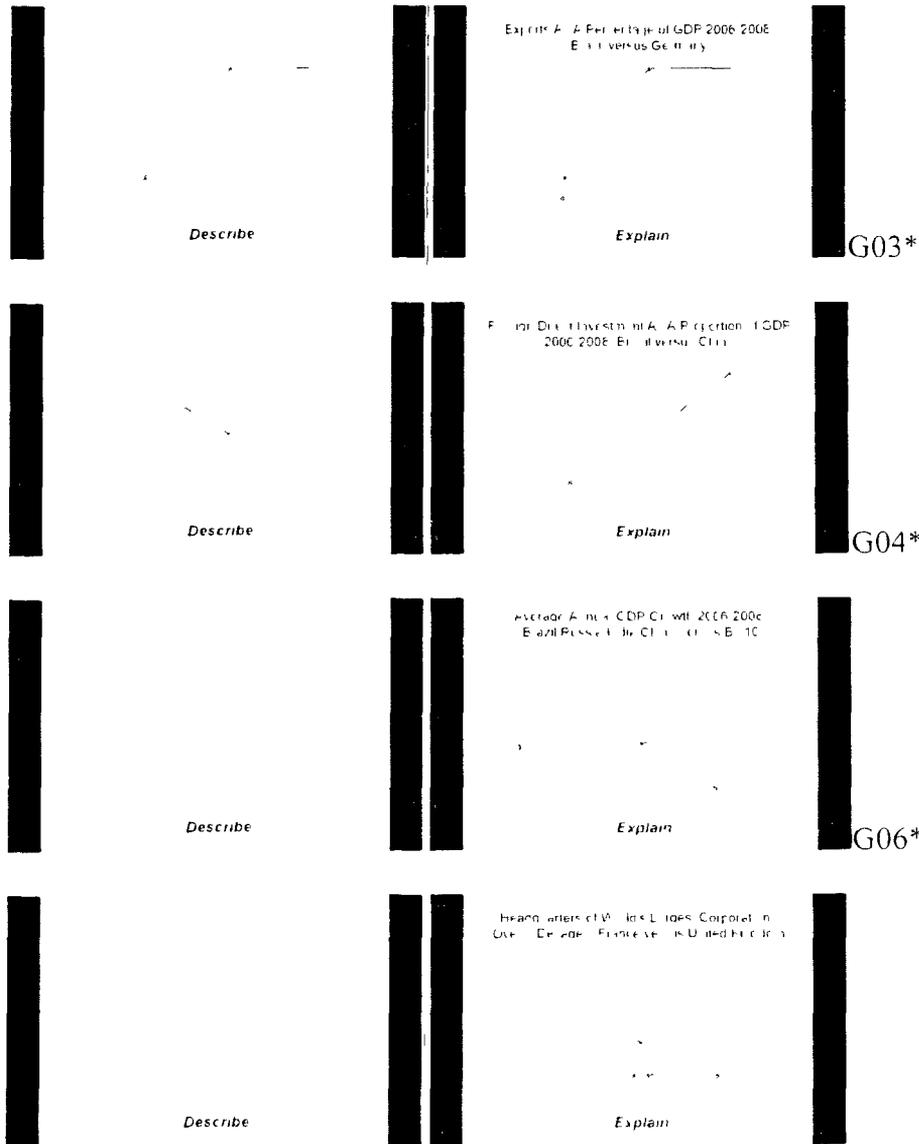
Appendix F1: Experiment 3 Instructions

In this study I will be showing you a number of line graphs. First, I will ask you to describe the visual characteristics of these graphs. These visual characteristics include the number of lines, how many times they cross over each other, what colour they are, whether they are straight or crooked, whether they go up or go down.

I will also ask you to explain those graphs. When you draw your conclusions, please take into account the title and the labels of the vertical and horizontal axes. Please pretend as if you were a business person or a scientist drawing conclusions about the graphs to another business person or to another scientist.

Appendix F2: Graph Stimuli for Experiment 3

The 16 graph stimuli used for Experiment 3 included 4 simple business (G01, G03, G04, G06); 4 complex business (G09, G10, G11, G12); 4 simple psychology (G13, G15, G16, G18); and 4 complex psychology graphs (G21, G22, G23, G24) G01*



G09*



Describe

Proportion of Total Fertilizer < 30 Years Old Over 2 Decades (1941 versus 2011)



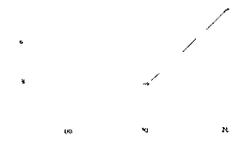
Explain

G10*



Describe

Software Piracy Rates 2006-2008 Pakistan versus Nigeria



Explain

G11*



Describe

Forecast Stability of Crispin's Senior Pipeline 2010-2012 Azerbaijan versus Czechia



Explain

G12*



Describe

Bride As A Percentage of Annual Revenue 2000-2008 Armenia versus Kazakhstan



Explain

G13*



Describe

Number of Full-time Employees Held Total Through



Explain

G15* [Redacted]

Describe

[Redacted]

Psychology Graduates by Year by Employment Setting

Explain

[Redacted]

G16* [Redacted]

Describe

[Redacted]

Function of Visual Cortex by Age by Gender

Explain

[Redacted]

G18* [Redacted]

Describe

[Redacted]

Air Fitness by Time of Day by Gender

Explain

[Redacted]

G21* [Redacted]

Describe

[Redacted]

Significance of Personality Characteristics by Age by Ability

Explain

[Redacted]

G22* [Redacted]

Describe

[Redacted]

Proportion of Subjective Feedback Metrics by Years Since Event Major versus Minor Distortions

Explain

[Redacted]

G23*



Describe

Retention By Recall Interval
Recall Interval versus R.C.I.



Explain

G24*



Describe

Probability of Recall By Serial Position
Word versus Numbers



Explain

Appendix F3: Experiment 3 Instructional Materials

Exp3Train01

Explaining Graphs

- The next slides contain 4 instructions I would like you to follow. Please read through them carefully.
- Press SPACE

Exp3Train02

Explaining Graphs

1. Describe: When asked to "describe", describe as much of the visual characteristics as you can. It will help you later when you "explain".
2. Title: Read the graph title out loud and match parts of the title to vertical axis, horizontal axis, each of the two lines in graph.
3. Between the lines: Are they parallel? Do they cross? Does one go up while the other goes down? Do they share any points in common?
4. Cause and effect: What causes the lines to go up, go down, or to stay the same? Is that cause operating equally on one line versus the other?

Press SPACE for a title more detail.

Exp3Train03

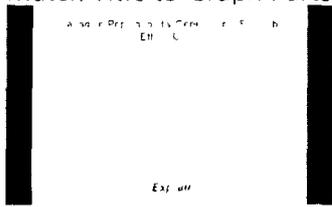
Describe Graphs

1. When asked to "describe", describe as much of the visual characteristics of the graphs as you can. It will help you later when you "explain".

Press SPACE

Exp3Train04

Match Title to Graph Parts

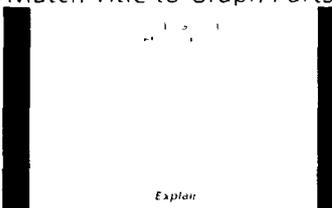


Explain

2 Title Read the graph title out loud and then match parts of the title to the vertical axis horizontal axis & the two lines in the graph
FIGURE 5.40

Exp3Train05

Match Title to Graph Parts

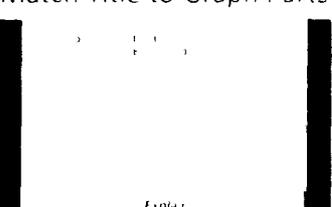


Explain

2 Read the graph title out loud and then match parts of the title to the vertical axis horizontal axis & the two lines in the graph
FIGURE 5.41

Exp3Train06

Match Title to Graph Parts



Explain

2 Read the graph title out loud and then match parts of the title to the vertical axis horizontal axis & the two lines in the graph
FIGURE 5.42

Exp3Train07

Match Title to Graph Parts

Expla

2 Read the graph title out loud and then match parts of the title to the vertical axis, horizontal axis & the two lines in the graph

Exp3Train08

Between The Lines

Expla

3 Identify the relationship between the lines. Are they parallel? Do they cross? Does one go up while the other goes down? Do they share any points in common?

Exp3Train09

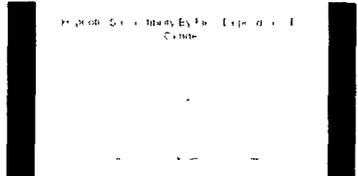
Between The Lines

Expla

3 Identify the relationship between the lines. Are they parallel? Do they cross? Does one go up while the other goes down? Do they share any points in common?

Exp3Train10

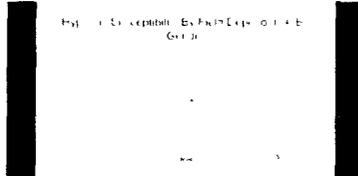
Cause & Effect

Explain Cause & Effect by Exploring Cause

Explain

4 What is causing the lines to go up, down, or stay the same? Is that cause operating equally on one line versus the other?
Fresh S-ACE

Exp3Train11

Cause & Effect

Explain Cause & Effect by Exploring Cause

Explain

4 What is causing the lines to go up, down, or stay the same? Is that cause operating equally on one line versus the other?
Fresh S-ACE

Exp3Train12

Explaining Graphs

- 1 Describe: When asked to "describe" describe as much of the visual characteristics as you can. It will help you later when you "explain."
- 2 Title: Read the graph title out loud and match parts of the title to vertical axis, horizontal axis, each of the two lines in graph.
- 3 Between the lines: Are they parallel? Do they cross? Does one go up while the other goes down? Do they share any points in common?
- 4 Cause and effect: What causes the lines to go up, go down, or to stay the same? Is that cause operating equally on one line versus the other?

Let you understand Fresh S-ACE and why it's important!

Appendix F4: Experiment 3 Descriptive statistics Word Count

Condition	<i>M</i>	<i>SD</i>	<i>N</i>
PreSimBus	36.571	32.5835	21
PreSimPsy	30.619	26.1237	21
PreComBus	32.881	28.4029	21
PreComPsy	32.548	24.2110	21
PosSimBus	46.333	26.8330	21
PosSimPsy	51.381	29.5080	21
PosComBus	45.238	23.2603	21
PosComPsy	45.429	27.5219	21

Appendix F5: Experiment 3, Word Count ANOVA

Source	<i>df</i>	<i>F</i>	<i>p</i>
Pre/Post	1	10.064	.005
Error	20		
Complexity	1	2.387	.138
Error	20		
Domain	1	0.075	.787
Error	20		
Pre/Post x Complexity	1	0.779	.388
Error	20		
Pre/Post x Domain	1	3.707	.069
Error	20		
Complexity x Domain	1	0.043	.839
Error	20		
Pre/Post x Complexity x Domain	1	6.803	.017
Error	20		

Appendix F6: Experiment 3 Descriptive statistics Because

Condition	<i>M</i>	<i>SD</i>	<i>N</i>
PreSimBus	.000	.000	21
PreSimPsy	.000	.000	21
PreComBus	.024	.109	21
PreComPsy	.024	.109	21
PosSimBus	.214	.373	21
PosSimPsy	.190	.370	21
PosComBus	.310	.580	21
PosComPsy	.214	.373	21

Appendix F7: Experiment 3, Because ANOVA

Source	<i>df</i>	<i>F</i>	<i>p</i>
Pre/Post	1	7.333	.014
Error	20		
Complexity	1	2.800	.110
Error	20		
Domain	1	0.075	.787
Error	20		
Pre/Post x Complexity	1	.323	.576
Error	20		
Pre/Post x Domain	1	.799	.382
Error	20		
Complexity x Domain	1	.380	.545
Error	20		
Pre/Post x Complexity x Domain	1	.380	.545
Error	20		

Appendix F8: Experiment 3 Descriptive statistics Between

Condition	<i>M</i>	<i>SD</i>	<i>N</i>
PreSimBus	.857	.808	21
PreSimPsy	.595	.605	21
PreComBus	.429	.576	21
PreComPsy	.524	.559	21
PosSimBus	.738	.407	21
PosSimPsy	.810	.432	21
PosComBus	.833	.508	21
PosComPsy	.643	.451	21

Appendix F9: Experiment 3, Between ANOVA

Source	<i>df</i>	<i>F</i>	<i>p</i>
Pre/Post	1	2.401	.137
Error	20		
Complexity	1	7.559	.012
Error	20		
Domain	1	1.611	.219
Error	20		
Pre/Post x Complexity	1	3.484	.077
Error	20		
Pre/Post x Domain	1	0.053	.820
Error	20		
Complexity x Domain	1	0.241	.629
Error	20		
Pre/Post x Complexity x Domain	1	4.884	.039
Error	20		

Appendix F10: Experiment 3 Descriptive statistics Direction

Condition	<i>M</i>	<i>SD</i>	<i>N</i>
PreSimBus	.571	.884	21
PreSimPsy	.476	.698	21
PreComBus	.571	.811	21
PreComPsy	.476	.661	21
PosSimBus	.476	.460	21
PosSimPsy	.429	.427	21
PosComBus	.429	.427	21
PosComPsy	.405	.407	21

Appendix F11: Experiment 3, Direction ANOVA

Source	<i>df</i>	<i>F</i>	<i>p</i>
Pre/Post	1	0.255	.619
Error	20		
Complexity	1	0.280	.602
Error	20		
Domain	1	1.352	.259
Error	20		
Pre/Post x Complexity	1	0.280	.602
Error	20		
Pre/Post x Domain	1	0.323	.576
Error	20		
Complexity x Domain	1	0.024	.877
Error	20		
Pre/Post x Complexity x Domain	1	0.020	.888
Error	20		

Appendix F12: Experiment 3 Descriptive statistics Quantitative

Condition	<i>M</i>	<i>SD</i>	<i>N</i>
PreSimBus	.048	.150	21
PreSimPsy	.071	.327	21
PreComBus	.048	.150	21
PreComPsy	.024	.109	21
PosSimBus	.071	.239	21
PosSimPsy	.119	.312	21
PosComBus	.071	.239	21
PosComPsy	.048	.218	21

Appendix F13: Experiment 3, Quantitative ANOVA

Source	<i>df</i>	<i>F</i>	<i>p</i>
Pre/Post	1	0.284	.600
Error	20		
Complexity	1	1.202	.286
Error	20		
Domain	1	0.074	.789
Error	20		
Pre/Post x Complexity	1	0.106	.748
Error	20		
Pre/Post x Domain	1	0.056	.815
Error	20		
Complexity x Domain	1	1.000	.329
Error	20		
Pre/Post x Complexity x Domain	1	0.033	.858
Error	20		

Appendix F14: Experiment 3 Descriptive statistics Title

Condition	<i>M</i>	<i>SD</i>	<i>N</i>
PreSimBus	.595	.436	21
PreSimPsy	.691	.445	21
PreComBus	.667	.456	21
PreComPsy	.667	.456	21
PosSimBus	.786	.373	21
PosSimPsy	.714	.463	21
PosComBus	.714	.405	21
PosComPsy	.714	.405	21

Appendix F15: Experiment 3, Title ANOVA

Source	<i>df</i>	<i>F</i>	<i>p</i>
Pre/Post	1	0.946	.342
Error	20		
Complexity	1	0.113	.741
Error	20		
Domain	1	0.113	.741
Error	20		
Pre/Post x Complexity	1	2.656	.119
Error	20		
Pre/Post x Domain	1	0.604	.446
Error	20		
Complexity x Domain	1	0.175	.680
Error	20		
Pre/Post x Complexity x Domain	1	0.521	.479
Error	20		

Appendix F16: Experiment 3 Descriptive statistics Trend

Condition	<i>M</i>	<i>SD</i>	<i>N</i>
PreSimBus	.167	.329	21
PreSimPsy	.429	.482	21
PreComBus	.333	.483	21
PreComPsy	.381	.416	21
PosSimBus	.048	.150	21
PosSimPsy	.071	.239	21
PosComBus	.095	.256	21
PosComPsy	.071	.239	21

Appendix F17: Experiment 3, Trend ANOVA

Source	<i>df</i>	<i>F</i>	<i>p</i>
Pre/Post	1	10.807	.004
Error	20		
Complexity	1	0.824	.375
Error	20		
Domain	1	5.972	.024
Error	20		
Pre/Post x Complexity	1	0.323	.576
Error	20		
Pre/Post x Domain	1	3.159	.091
Error	20		
Complexity x Domain	1	3.941	.061
Error	20		
Pre/Post x Complexity x Domain	1	1.429	.246
Error	20		

Appendix F18: Experiment 3 Descriptive statistics x-axis

Condition	<i>M</i>	<i>SD</i>	<i>N</i>
PreSimBus	.167	.329	21
PreSimPsy	.167	.329	21
PreComBus	.143	.359	21
PreComPsy	.167	.365	21
PosSimBus	.690	.432	21
PosSimPsy	.786	.405	21
PosComBus	.762	.436	21
PosComPsy	.762	.375	21

Appendix F19: Experiment 3, x-axis ANOVA

Source	<i>df</i>	<i>F</i>	<i>p</i>
Pre/Post	1	40.185	.0001
Error	20		
Complexity	1	0.087	.771
Error	20		
Domain	1	1.724	.204
Error	20		
Pre/Post x Complexity	1	0.811	.379
Error	20		
Pre/Post x Domain	1	0.588	.452
Error	20		
Complexity x Domain	1	0.380	.545
Error	20		
Pre/Post x Complexity x Domain	1	1.092	.309
Error	20		

Appendix F20: Experiment 3 Descriptive statistics y-axis

Condition	<i>M</i>	<i>SD</i>	<i>N</i>
PreSimBus	.167	.329	21
PreSimPsy	.167	.329	21
PreComBus	.167	.365	21
PreComPsy	.167	.365	21
PosSimBus	.738	.375	21
PosSimPsy	.786	.405	21
PosComBus	.762	.436	21
PosComPsy	.810	.370	21

Appendix F21: Experiment 3, y-axis ANOVA

Source	<i>df</i>	<i>F</i>	<i>p</i>
Pre/Post	1	40.848	.0001
Error	20		
Complexity	1	0.175	.680
Error	20		
Domain	1	1.649	.214
Error	20		
Pre/Post x Complexity	1	0.388	.540
Error	20		
Pre/Post x Domain	1	0.717	.407
Error	20		
Complexity x Domain	1	0.000	1.000
Error	20		
Pre/Post x Complexity x Domain	1	0.000	1.000
Error	20		

Appendix F22: Experiment 3 Descriptive statistics z-variable

Condition	<i>M</i>	<i>SD</i>	<i>N</i>
PreSimBus	.714	.663	21
PreSimPsy	.738	.785	21
PreComBus	.833	.796	21
PreComPsy	.810	.733	21
PosSimBus	.548	.472	21
PosSimPsy	.571	.455	21
PosComBus	.524	.487	21
PosComPsy	.595	.436	21

Appendix F23: Experiment 3, z-variable ANOVA

Source	<i>df</i>	<i>F</i>	<i>p</i>
Pre/Post	1	2.105	.162
Error	20		
Complexity	1	0.884	.358
Error	20		
Domain	1	0.154	.699
Error	20		
Pre/Post x Complexity	1	1.356	.258
Error	20		
Pre/Post x Domain	1	0.167	.687
Error	20		
Complexity x Domain	1	0.000	1.000
Error	20		
Pre/Post x Complexity x Domain	1	0.167	.687
Error	20		

Appendix F24: Experiment 3 Descriptive statistics Completeness

Condition	<i>M</i>	<i>SD</i>	<i>N</i>
PreSimBus	.364	.125	21
PreSimPsy	.362	.144	21
PreComBus	.356	.127	21
PreComPsy	.359	.115	21
PosSimBus	.479	.135	21
PosSimPsy	.498	.150	21
PosComBus	.501	.131	21
PosComPsy	.474	.156	21

Appendix F25: Experiment 3, Completeness ANOVA

Source	<i>df</i>	<i>F</i>	<i>p</i>
Pre/Post	1	15.207	.001
Error	20		
Complexity	1	0.094	.060
Error	20		
Domain	1	0.026	.053
Error	20		
Pre/Post x Complexity	1	0.081	.058
Error	20		
Pre/Post x Domain	1	0.032	.053
Error	20		
Complexity x Domain	1	1.202	.181
Error	20		
Pre/Post x Complexity x Domain	1	1.033	.162
Error	20		

Appendix F26: Experiment 3, Nominal, Ordinal, Metric Descriptive Statistics

Condition	<i>M</i>	<i>SD</i>	<i>N</i>
PreNomSP	.2381	.34889	21
PreNomSB	.2857	.44219	21
PreNomCP	.2381	.33049	21
PreNomCB	.2857	.40532	21
PreOrdSP	.2976	.30227	21
PreOrdSB	.4286	.40422	21
PreOrdCP	.2619	.27924	21
PreOrdCB	.2143	.28816	21
PreMetSP	.0357	.16366	21
PreMetSB	.0238	.07520	21
PreMetCP	.0119	.05455	21
PreMetCB	.0238	.07520	21
PosNomSP	.2143	.21339	21
PosNomSB	.2381	.23017	21
PosNomCP	.2024	.20339	21
PosNomCB	.2143	.21339	21
PosOrdSP	.4048	.21616	21
PosOrdSB	.3690	.20339	21
PosOrdCP	.3214	.22559	21
PosOrdCB	.4167	.25413	21
PosMetSP	.0595	.15622	21
PosMetSB	.0357	.11952	21
PosMetCP	.0238	.10911	21
PosMetCB	.0357	.11952	21

Appendix F27: Experiment 3, Nominal, Ordinal, Metric ANOVA

Source	<i>df</i>	<i>F</i>	<i>p</i>
Pre/Post	1	0.267	.611
Error	20		
NomOrdMet	2	15.739	.0001
Error	40		
Complexity	1	7.320	.014
Error	20		
Domain	1	4.155	.055
Error	20		
Pre/Post x NomOrdMet	2	0.902	.414
Error	40		
Pre/Post x Complexity	1	1.407	.249
Error	20		
NomOrdMet x Complexity	2	3.258	.049
Error	40		
Pre/Post x NomOrdMet x Complexity	2	3.051	.058
Error	40		
Pre/Post x Domain	1	0.670	.423
Error	20		
NomOrdMet x Domain	2	0.681	.512
Error	40		
Pre/Post x NomOrdMet x Domain	2	0.062	.940
Error	40		
Complexity x Domain	1	0.000	1.000
Error	20		
Pre/Post x Complexity x Domain	1	2.903	.104
Error	20		
NomOrdMet x Complexity x Domain	2	0.448	.642
Error	40		
Pre/Post x NomOrdMet x Complexity x Domain	2	3.223	.050
Error	40		

Appendix F28: Experiment 3 Descriptive statistics Response Time

Condition	Describe/Explain	<i>M</i>	<i>SD</i>	<i>N</i>
PreSimBus	Describe	20.38	13.95	21
PreSimBus	Explain	37.81	20.65	21
PreSimPsy	Describe	17.52	10.50	21
PreSimPsy	Explain	38.10	14.79	21
PreComBus	Describe	19.62	14.05	21
PreComBus	Explain	42.05	19.28	21
PreComPsy	Describe	18.38	11.16	21
PreComPsy	Explain	68.62	28.72	21
PosSimBus	Describe	23.81	11.83	21
PosSimBus	Explain	68.62	28.72	21
PosSimPsy	Describe	26.62	12.48	21
PosSimPsy	Explain	67.43	30.47	21
PosComBus	Describe	24.90	10.36	21
PosComBus	Explain	68.81	30.68	21
PosComPsy	Describe	24.48	14.30	21
PosComPsy	Explain	62.48	26.63	21

Appendix F29: Experiment 3, Response Time ANOVA

Source	<i>df</i>	<i>F</i>	<i>p</i>
Pre/Post	1	40.613	.0001
Error	20		
Complexity	1	0.103	.752
Error	20		
Domain	1	0.577	.456
Error	20		
Task	1	75.713	.0001
Error	20		
Pre/Post x Complexity	1	9.184	.0001
Error	20		
Pre/Post x Domain	1	1.074	.872
Error	20		
Complexity x Domain	1	15.860	.575
Error	20		
Pre/Post x Complexity x Domain	1	1.719	.205
Error	20		
Pre/Post x Task	1	22.682	.0001
Error	20		
Complexity x Task	1	0.597	.449
Error	40		
Pre/Post x Complexity x Task	1	7.041	.015
Error	20		
Domain x Task	1	0.068	.797
Error	20		
Pre/Post x Domain x Task	1	4.395	.049
Error	20		
Complexity x Domain x Task	2	0.448	.972
Error	40		
Pre/Post x Complexity x Domain x Task	1	0.002	.961
Error	20		