

**The Impact of Individual Differences on Visualization
Effectiveness and Gaze Behaviour: Informing the Design of
User Adaptive Interventions**

by

Dereck J Toker

B.A., The University of British Columbia, 2010

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

THE FACULTY OF GRADUATE STUDIES
(Department of Computer Science)

The University of British Columbia
(Vancouver)

August 2013

© Dereck J Toker, 2013

Abstract

Research has shown that individual differences can play a role in information visualization effectiveness. Unfortunately, results are limited given that there are so many individual differences that exist, and information visualizations are commonly designed without taking into account these user differences. The aim of this thesis is to investigate the impact of a specific set of individual differences (i.e., user characteristics) in order to identify which of these user differences have an impact on various aspects of information visualization performance. Eye tracking is also employed, in order to see if there is an impact of individual differences on user gaze behavior, both in general and for specific information visualization elements (i.e., legend, labels). In order to gather the necessary data, a user study is conducted, where users are required to complete a series of tasks on two common information visualizations: bar graphs and radar graphs. For each user, the following set of user characteristics are measured: perceptual speed, visual working memory, verbal working memory, and visualization expertise.

Using various statistical models, results indicate that user characteristics do have a significant impact on visualization performance in terms of task completion time, visualization preference, and visualization ease-of-use. Furthermore, it is also found that user characteristics have a significant impact on user gaze behavior, and these individual differences can also influence how a user processes specific elements within a given information visualization.

Preface

The study presented in Chapter 3 was not designed by the author of this Thesis. This includes decisions such as: which user characteristics were selected, which information visualizations were selected, which tasks to use, the domain of the data set, and other study factors such as data distribution and presentation order.

A version of the research reported in Chapter 4 has been published as: Toker, D., Conati, C., Carenini, G., Haraty, M. (2012) Towards Adaptive Information Visualization: On the Influence of User Characteristics. UMAP2012: 274-285.

A version of the research reported in Chapter 5 has been published as: Toker, D., Conati, B., Steichen, Carenini, G. Individual User Characteristics and Information Visualization: Connecting the Dots through Eye Tracking. CHI2013: 295-304.

Table of Contents

Abstract	ii
Preface	iii
Table of Contents	iv
List of Tables	viii
List of Figures	ix
Acknowledgements	xi
Chapter 1 Introduction	1
1.1 Research Goals and Approach	3
1.2 Contributions	4
1.3 Outline	5
Chapter 2 Related Work	6
2.1 Individual Differences in Information Visualization	6
2.2 User Adaptation for Information Visualization	8
2.3 Eye Tracking	9
2.3.1 Eye tracking as low level sensor	9
2.3.2 Eye tracking data analysis techniques	10
Chapter 3 User Study: Bar Graphs & Radar Graphs	12
3.1 Overview	12
3.2 Study Design Contribution	13
3.3 Experimental Tasks	16
3.4 User Characteristics Considered in the Study	17
3.4.1 Perceptual speed	18
3.4.2 Visual working memory	19
3.4.3 Verbal working memory	19
3.4.4 User expertise	20
3.5 Study Design	20

3.5.1	Participants.....	20
3.5.2	Apparatus	20
3.5.3	Procedure	21
3.6	Dependent Measures Collected From Users During Task Execution	22
3.6.1	Objective measures of user performance	23
3.6.2	Subjective measures of user performance	23
3.6.3	Eye tracking data.....	23
Chapter 4 Impact of Individual Differences on User Performance & User Preferences: Analysis & Results.....		
4.1	Summary of User Characteristics Data	24
4.2	Effect on Performance Measures.....	26
4.2.1	Single scenario - model for analysis	26
4.2.2	Single scenario results - main effects	27
4.2.3	Single scenario results - interaction effects	28
4.2.4	Double scenario - significant results	30
4.2.5	Double scenario - marginally significant results.....	30
4.3	Effects on User Preference & Ease of Use	32
4.3.1	Description of statistical model employed - multivariate (MANCOVA).....	33
4.3.2	Subjective measures results - main effects of cognitive abilities	34
4.3.3	Subjective measures results - main effects of expertise	36
4.3.4	Visualization preference direct comparison - follow up analysis	36
4.4	Chapter Summary & Discussion	39
4.4.1	Implications of findings	39
Chapter 5 Eye Tracking: Impact of User Characteristics on Gaze Patterns.....		
5.1	Eye tracking Data & Pre-processing	41
5.1.1	Complete set of eye tracking features computed.....	42

5.1.2	Areas of interest (AOI)	43
5.1.3	Eye tracking feature reduction: principal component analysis	46
5.1.4	Task level family PCA.....	47
5.1.5	AOI transitions family PCA	48
5.1.6	AOI proportionate family PCA.....	49
5.2	Task Difficulty	50
5.2.1	Four indicators of task difficulty	50
5.2.2	Principal component analysis of task difficulty	51
5.3	Mixed Model Analysis	53
5.3.1	Mixed model setup	54
5.4	Mixed Model Results & Discussion - User Characteristics.....	54
5.4.1	Perceptual speed - main effects	55
5.4.2	Perceptual speed - interaction with task difficulty	56
5.4.3	Perceptual speed - interaction with visualization type	58
5.4.4	Verbal working memory - main effects.....	59
5.4.5	User expertise	60
5.4.6	Visual working memory.....	60
5.5	Mixed Model Results - Task Difficulty & Visualization Type	61
5.6	Summary & Discussion	61
5.6.1	Summary of results.....	61
5.6.2	Discussion of perceptual speed.....	62
5.6.3	Discussion of verbal working memory	63
5.6.4	Discussion of user expertise.....	63
5.6.5	Implications for information visualization design.....	64
Chapter 6	Task Difficulty - Follow Up Analysis	65

6.1	Model for Analysis	65
6.2	Results and Discussion	65
6.2.1	Main effect of task difficulty	66
6.2.2	Interaction effect - task difficulty & radar expertise	66
6.3	Discussion	67
Chapter 7	Conclusion and Future Work	68
7.1	Satisfaction of Thesis Goals	68
7.1.1	Impact of user characteristics on visualization performance & preference	68
7.1.2	Additional findings: impact of visualization type on task completion time ..	69
7.1.3	Impact of user characteristics on eye tracking measures	70
7.1.4	Additional contributions - mixed model analysis & task difficulty	71
7.2	Future Work	71
7.2.1	Run a similar study with more difficult tasks	71
7.2.2	Using eye tracking to infer user characteristics	71
7.2.3	Adaptive interventions for a fixed visualization type	72
7.3	Lessons Learned - Design Decisions For Future Studies	72
7.4	Conclusion	74
	Bibliography	75
	Appendix A - Additional Eye Tracking Results	80
	Main Effects of Visualization Type	80
	Main Effect of Task difficulty	81
	Interaction Effects involving Visualization Type & Task Difficulty	82
	Summary	87
	Appendix B - Post Questionnaire	88

List of Tables

Table 3.1: Task set explored in the study (from Amar’s taxonomy).....	17
Table 4.1: Summary of user characteristics collected from the study	25
Table 4.2: Results for direct comparison preference ratings between bar graph and radar graph.....	37
Table 4.3: Summary of statistically significant results from this chapter, *indicates marginally significant	39
Table 5.1: Description of basic eye tracking measures	42
Table 5.2: Task level eye tracking features generated for analysis	43
Table 5.3: AOI level eye tracking features generated for analysis	43
Table 5.4: Generated components for the task-level family. **fixation-rate is the only measure that inversely correlates with the other members of its component	47
Table 5.5: Components generated for the AOI transitions family	48
Table 5.6: Generated components for AOI proportionate measures	49
Table 5.7: Resulting coefficient values of the four measures that were used in principal component analysis for task difficulty	51
Table 5.8: All 28 task combinations from the study, ordered from least difficult to most difficult according to our definition of task difficulty	52
Table 5.9: Main effects of perceptual speed.....	55
Table 5.10: Interaction effects for perceptual speed and task difficulty	56
Table 5.11: Main effects of verbal working memory	59
Table 5.12: Summary of reported results from analysis of eye tracking data	62
Table A.0.1: Main effects of visualization type	80

List of Figures

Figure 3.1: Single scenario example for bar graph used in this study.	14
Figure 3.3: Double scenario example for bar graph used in the study	15
Figure 3.4: Double scenario example for radar graph used in the study	15
Figure 3.5: Sample questions from the perceptual speed test. For each row, participants are required to indicate which item matches the one on the left.....	18
Figure 3.6: The actual Tobii T120 eye tracker used for this study. The circled region shows where the eye tracking cameras are located within the computer screen housing.	21
Figure 3.7: Prompt used to collect user confidence after each task	22
Figure 4.1: Scatter plot with trend line showing task completion time and perceptual speed scores for each participant in the single scenario tasks. The vertical line shows the median split used to report users as high and low	27
Figure 4.2: Chart showing the mean completion times for the interaction effect of perceptual speed and visualization type, users with low perceptual speed are taking even longer for radar graph tasks	28
Figure 4.3: Interaction for visualization type and order. Users always perform faster on the second visualization they use compared to their counterparts	29
Figure 4.4 Task completion time and perceptual speed scores for double scenario tasks with trend line. The vertical line shows the median split used to report users as high and low.....	31
Figure 4.5: Likert-scale data collected for graph preference (left), and ease-of-use ratings (right)	32
Figure 4.6: Histogram showing the responses for radar graph preference based on the median split of visual working memory	34
Figure 4.7: Dot plot showing the visual working memory scores with a vertical line indicating the median split	34
Figure 4.9: Visual working memory scores showing median split	35
Figure 4.10: Direct comparison preference ratings for bar and radar graphs based on verbal working memory	38

Figure 4.11: Direct comparison preference ratings for bar and radar graphs based on expertise for bar graph (left), and radar graph (right)	38
Figure 5.1: Saccade based eye measures	42
Figure 5.2: The five AOI regions defined over bar graph.....	44
Figure 5.3: Five AOI regions defined over radar graph	45
Figure 5.4: Example of experimental data in wide format for GLM (left), and in long format for mixed model (right).	53
Figure 5.5: Interaction between perceptual speed and task difficulty on AOI <i>legend transitions</i>	56
Figure 5.6: Interaction between perceptual speed and task difficulty on AOI <i>label transitions</i>	57
Figure 5.7: Interaction between perceptual speed and visualization type for AOI <i>high transitions</i>	58
Figure 6.1: Shows the interaction effect that radar expertise and task difficulty has on task completion time.....	66
Figure A.0.1: Interaction effect between visualization type and task difficulty for std.dev. path angles component.....	83
Figure A.0.2: Interaction effect between visualization type and task difficulty for AOI legend transitions component	84
Figure A.0.3: Interaction effect between visualization type and task difficulty for label proportion measures component.....	85
Figure A.0.4: Interaction effect between visualization type and task difficulty for AOI high proportion component	86

Acknowledgements

First and foremost, I would like to thank my supervisor, Cristina Conati, for accepting me as a master's student. She is the kind of person that is direct, efficient, and inspiring; all at the same time. These three qualities were exactly what I needed to learn and grow as a burgeoning academic. Thank you.

Next I would like to thank Ben Steichen for being my friend from a country with a strange language, and at the same time being a professional and dependable colleague.

I must also thank Michael Sedlmair, a fellow father from the downstairs apartment, for giving optimistic feedback on my thesis introduction, and positive encouragement in my studies.

Also, thanks to Tamara Munzner for agreeing to be the second reader of this thesis.

Lastly, I must thank my family for their unconditional support. Thank you.

For my wife Eva

Chapter 1 Introduction

Information visualization is a thriving area of research in the study of human/computer communication. Some of the benefits of information visualization include: assisting users in comprehending huge amounts of data, facilitating the better understanding of data, aiding the perception of emerging properties in data, and enabling problems with data to become more immediately apparent [Ware 2012]. The benefits of employing information visualization are especially important given the ever-growing amount of digital information. While visualizations have gained increasingly in terms of general usage and usability, they have in the past typically followed a one-size-fits-all model, focused on the target data set and associated task model, with little consideration for user differences. The primary purpose of this research is to inform the design of user-adaptive visualization systems which *understand* that different users have different visualization needs and abilities, and which can *adapt* to these differences during interaction. However, before adaptation strategies can be effectively specified, the influence of individual differences must be further studied and clarified. The research presented in this thesis will investigate the impact that user characteristics have for users performing information visualization based tasks.

User-adaptation is the idea that an interactive system can adapt the interaction based on specific individual user needs and differences. User-adaptation has been shown to be effective in a variety of applications such as web search, desktop assistance, and e-learning [Jameson 2008], but it is largely unexplored in information visualization. Even so, some initial work has been done that looks at the benefits of user-adaptation by recommending alternative visualizations. In one example, Gotz & Wen [2009] tested a behaviour-driven visualization recommendation system. Their idea was to monitor a user's interaction data in order to define and detect suboptimal usage patterns, and then adapt to these patterns by recommending alternative visualizations to the user. Our research aims to inform similar adaptive visualization systems, but differs significantly in

two ways. First, we¹ believe that user characteristics should be considered as an alternative/additional source to inform the design of user adaptive information visualization systems. Thus, we will investigate how a set of user characteristics affect visualization effectiveness and how this effect is modulated by items such as visualization type and task complexity. Individual differences that exist between users can include both long term user characteristics such as cognitive abilities or expertise, and short term factors such as cognitive load or attention. Studies such as Conati & Maclaren [2008] found that the cognitive ability of perceptual speed impacts visualization effectiveness in terms of which visualization is more suitable for a given user. In another example, Velez et al. [2005] found that a user's spatial abilities, including perceptual speed, were strongly related to visualization comprehension. These studies are an indication that it is important to investigate the possibility of user-adaptive information visualization systems that take into account these individual differences. In particular, a first step towards designing adaptive information visualization systems that can both detect and respond to individual differences, lies in gaining a more fine-grained understanding of the impact that user differences have on visualization processing. Thus, we extend previous work in two ways. First, we look at more cognitive differences and more performance measures. Second, we explore the potential of using gaze data as a source of information to detect individual differences. Eye tracking is an informative and sometimes the only available source of real-time information on visualization processing given that visual scanning and elaboration are both fundamental components of working with a visualization (they are in fact the only components for non-interactive visualizations). Additionally, using eye tracking is promising because current state of the art eye trackers are becoming less invasive given that they no longer require cumbersome head mounted equipment.

¹ Due to the 'we' being so ubiquitous, it is used in this writing, and from here on in signifies that the research presented in this paper is a subset of work involving myself, my supervisor, and fellow researchers on the Advanced Tools for User-Adaptive Visualization (ATUAV) research team.

1.1 Research Goals and Approach

Previously, Conati & Maclaren [2008] studied the impact of several user characteristics (i.e., Perceptual Speed, Visual Memory, Spatial Visualization, Disembodiment, Need for Cognition, and Learning Style) on task accuracy with two different visualizations. Their overall finding was that perceptual speed impacted task accuracy and this effect was mediated by which visualization the user was shown. Our aim is to both confirm that perceptual speed is a user characteristics that does impact visualization interaction, and also to gain a more fine-grained understanding of the impact that other potentially significant user characteristics have on visualization processing, i.e., visual working memory, verbal working memory, visual working memory, and visualization expertise. We will carry out this investigation by conducting a user study² that will examine the impact of these user characteristics on user performance with two common visualizations: bar graphs and radar graphs. Our research will differ from Conati & Maclaren since they only examined the impact of user characteristics on visualization performance in terms of task accuracy, whereas our user study we also include task completion time (that together with task accuracy constitutes our objective measures of performance), as well as visualization preference and visualization ease-of-use (the subjective performance measures). Our study is also different since we are using eye tracking, which will be utilized to investigate the impact of user characteristics on user gaze behavior.

The following are the questions that the research in this thesis addresses:

- Q1.
 - i. Do user characteristics influence (i.e., correlate with) the performance and preference of users while using bar graphs and radar graphs?
 - ii. If there is an effect of user characteristics, how is this effect influenced by visualization type?

² Also mentioned in the preface, the author of this thesis was not involved in the design of this user study.

Q2. Eye Tracking: In order to investigate the impact of user characteristics on user gaze behavior, we analyze eye tracking measures collected from the user study.

- i. Do individual user characteristics influence (i.e., correlate with) a user's eye gaze behavior in a way that is detectable by eye tracking?
- ii. If yes, (a) which gaze features are influenced by which particular user characteristics? (b) Is the effect modulated by task context (e.g., task difficulty or visualization type)?

1.2 Contributions

Our research confirms and extends preliminary existing findings that user characteristics do make a difference in visualization effectiveness, implying that user characteristics should be taken into account when designing user-adaptive information visualization systems. We also show that user characteristics have an impact on a user's gaze behaviours and that this impact is detectable via eye tracking. This finding is important because it provides a better understanding of how user characteristics such as perceptual speed can influence the processing of standard visualization components like legends and labels. Since components such as legends and labels are found in many information visualizations, our findings may generalize to other information visualizations beyond just bar and radar graphs.

The major implication of this research is that it could inform or potentially drive the design of user-adaptive visualizations. Since we demonstrate that user characteristics do in fact have an impact on user performance and that user characteristics also influence eye gaze behavior, then eye tracking could be considered a source, amongst others, of real-time information to be leveraged for detecting user characteristics. This in turn, would provide useful information about the current user that an adaptive information visualization systems could use to generate tailored interventions aimed at improving the visualization system for that user. Examples of some possible interventions could be to either offer an alternative visualization altogether or to provide

support within the current visualization, such as highlighting relevant elements, or s adding explanatory material in order to facilitate visualization processing.

1.3 Outline

The remaining content of this thesis is organized as follows. First, Chapter 2 gives an overview of related work on individual differences in information visualization, user-adaptive information visualization systems, and eye tracking. In Chapter 3 we present an overview of the user study that was conducted in terms of the tasks are were chosen, the user characteristics that are examined, and the experimental measures that were collected from the task interactions. In Chapter 4, we present the data analysis and results on the impact of user characteristics on task performance which is separated into the objective performance measures and subjective preference measures. Chapter 5 presents the data analysis and results obtained on the impact of user characteristics on user gaze behaviour as a result of examining the eye tracking data. Chapter 6 briefly describes a follow up analysis of task difficulty on task completion time. Finally, Chapter 7 concludes with a summary of the research and outlines several directions of future work.

Chapter 2 Related Work

Existing work on identifying the factors that define visualization effectiveness has mostly focused on properties of the data to be visualized or the tasks to be performed, sometimes obtaining inconclusive and conflicting results (see Gotz and Wen [2009] and Nowell [2002], for an overview). Extensive work has been done on comparing the effectiveness of graphical data in terms of accuracy and speed across different visualization types, yet this research has not typically taken into account individual differences (see Cleveland [1984] and Simkin [1987]). In the rest of this chapter, we first look at studies that have begun to explore the impact of individual differences on information visualization. Next, we look at research relating to user-adaptation techniques for information visualization. Finally, since we will be using eye tracking for our study, we look at research that uses eye tracking in order to detect differences in user gaze patterns, as well as various methods employed to analyse this type of data.

2.1 Individual Differences in Information Visualization

Lewandowsky and Spence [1989] examined the effect of expertise on user performance with scatter plots, discovering that high expertise improved accuracy yet decreased completion time. This was an early indication that the impact of individual differences should be investigated further. Recent work examining the impact of individual differences in information visualization include Chen [2000], who found a significant impact of associative memory on search performance using spatial interfaces. He identified that users with high associative memory had higher performance scores. Dillon [2000] argues based on previous studies, that we ought to take into consideration user differences such as knowledge, experience, and behavior patterns when designing information visualizations in order to reduce a user's disorientation during potentially overwhelming scenarios. Velez et al. [2005] explored the link between three spatial abilities (spatial orientation, spatial visualization, and disembedding) and two cognitive factors (visual memory and perceptual speed) on task comprehension (i.e., accuracy & task time) for visualization tasks involving the identification of a 3D object from its

orthogonal projections. They found not only a large diversity in the spatial and cognitive abilities of their study's subjects, but also that these abilities are related to visualization comprehension. For example, they found that all three spatial abilities positively correlated with task accuracy, and the two cognitive traits negatively correlated with response selection time. Our research differs from Velez et al. since we will investigate a more basic set of tasks using two common information visualizations (bar & radar graphs). Furthermore, in addition to investigating the impact of individual differences on both accuracy and task time, we extend our study by exploring the impact of individual differences on visualization preference, visualization ease-of-use, and user eye gaze behavior. Baldonado et al. [2000], a position paper, outlines 8 rules that should be used when designing information visualizations with multiple views. The first rule they define is *Diversity*, arguing that attributes such as user memory, learning abilities, and perceptual abilities can positively or negatively impact information visualization utility, and therefore should be taken into consideration when designing information visualizations. Allen [2000] found that given four different information visualization layouts (e.g., multi-window, world-map) of bibliographic references, perceptual speed significantly impacted which one was better. Users with low perceptual speed had higher recall using a world map interface, whereas users with higher perceptual speed had better recall using any of the other layouts. Perceptual speed was also a significant trait in the study done by Conati and Maclaren [2008]. They looked at how a set of cognitive traits (Visual Memory, Spatial Visualization, Perceptual Speed, Disembodiment, Need for Cognition, and Learning Style) influences two different information visualizations. These visualizations were designed to represent changes in a set variables and consisted of a radar graph and a Multiscale Dimension Visualizer (MDV). The MDV is a visualization that primarily uses color hue and intensity to represent change direction and magnitude [Williams 2004]. Conati and Maclaren found in terms of task accuracy that: (1) a user's perceptual speed was a significant predictor of which of the two visualizations would work better, and (2) that visual memory, perceptual speed, need for cognition, spatial visualization, and some learning styles can each predict performance for certain individual tasks performed with one specific visualization type. The study we describe in this thesis can be seen as an extension of

this previous work in at least three fundamental ways. First, in our study we compare radar graphs with bar graphs, a much more common visualization than MDV. Thus, our findings may potentially have a much stronger impact on adaptive information visualization in general. Second, in addition to only measuring user performance in terms of task accuracy, we include task completion time plus subjective measures such as visualization preference and ease-of-use. Third, we will investigate the impact of user characteristics on user gaze behaviour by using eye tracking.

Recently, there has also been a lot of interest on the impact of personality traits on information visualization effectiveness. Ziemkiewicz et al. [2011] found that the personality trait locus of control significantly impacted response time of correctly answered questions for a set of tasks performed on various tree-structured dataset visualizations. They discovered that users with internal locus of control were slower when correctly answering questions than users with external locus of control. Green and Fisher [2010] also examined personality traits including locus of control, extraversion, and neuroticism. Given two information visualization types (Map-view & Graphical-view), they found that users with internal locus of control completed tasks (both correctly and incorrectly) more quickly, whereas users that were neurotic or extraverted completed tasks more slowly. Our research does not include personality traits, although they could be considered for future studies.

2.2 User Adaptation for Information Visualization

The benefits of user-adaptive interaction have been shown in a variety of tasks and applications such as operation of menu-based interfaces, web search, desktop assistance, and human-learning [Jameson 2003]. However, these ideas have had a minimal application to data visualization, largely due to the limited understanding of which user characteristics are relevant for adaptivity in this domain. Three notable exceptions are the work by Gotz and Wen [2009], Grawemeyer [2006], and Brusilovsky et al. [2006]. Gotz and Wen [2009] propose a technique to automatically detect a user's suboptimal usage patterns based on activity logs collected from interaction with a multi-purpose visualization. The visualization is then adapted accordingly by offering

alternative visualization types such as bar chart, line graph, or scatter plot. Gotz and Wen found that their adaptive system significantly improved task completion time for users compared to a non-adaptive version. Our research, on the other hand, will focus on informing how to adapt the visualization based on user cognitive measures/abilities instead of usage patterns. Grawemeyer [2006] ran a study that required users to perform a series of tasks where the user had to choose which visualization they wanted to use for each task. Visualization options were bar chart, pie chart, plot chart, sector graph, set diagram, and table. A user-adaptive system was tested that adapts by hinting to the user which visualization would be best based on their prior knowledge in the given domain, as well as their performance with the visualizations they selected in previous tasks. Grawemeyer found that the adaptive system significantly improved response accuracy compared to a non-adaptive version where users would choose their visualization with no guidance. Lastly, Brusilovsky et al. [2006], tested a system that adapts the content of a fixed visualization type, based on the user's domain knowledge and current progress in an educational system. The adaptation highlights which information items currently displayed within the visualization are the most relevant for a user to explore next. Brusilovsky et al. found that in general users preferred the adaptive version of the system, and other performance related measures have yet to be tested.

Our research intends to support adaptation that involves both selecting *alternative visualizations* for different users (similar Gotz and Wen, and Grawemeyer), as well as providing *adaptive help* within a given visualization to best accommodate each user's needs (like in Brusilovsky et al.).

2.3 Eye Tracking

2.3.1 Eye tracking as low level sensor

In the field of psychology, the use of eye tracking has long been established as a suitable means for analyzing user attention patterns in information processing tasks [Keith 1995]. Research in this field has also investigated the impact of individual user differences on reading and search tasks [Rayner 1998]. In human-computer interaction

and information visualization, recent research has also started to use eye tracking technology to investigate trends and differences in user attention patterns and cognitive/decision processing. In particular, such research has typically focused on either identifying pattern differences for alternative visualizations [Goldberg 2011], task types [Simola 2008], and activities within a task [Courtemanche 2011], or on explaining differences in user accuracy between alternative visualization interfaces [Plumlee 2006]. While such studies provide valuable insights into differences in gaze behaviours for different tasks and/or activities, they have traditionally ignored individual differences between study participants. To the best of our knowledge, the only research exploring the effect of user traits on gaze behavior is work by Tai et al. [2006], Tang et al. [2012], and Grindinger et al. [2010]. In all three cases, their work focused on a single, domain-specific user characteristic (task-domain expertise), showing that domain experts and novices display significantly different gaze behaviours. The scope of our work is broader, since we investigate a comprehensive array of user characteristics comprising of several cognitive abilities and *visualization expertise*, (i.e., expertise on specific information visualizations). All of these user characteristics are domain-independent, thus our results are more general across different information visualization tasks.

2.3.2 Eye tracking data analysis techniques

Several researchers have looked at how to process raw eye tracking data in order to detect attention patterns that distinguish different types of users, e.g. novices vs. experts [Grindinger 2010], or to distinguish information goals [Fukuhara and Nakano 2011], [Muldner 2010]. One possible way to analyze eye tracking data is to apply data mining techniques such as Hidden Markov Models [Courtemanche 2011], Scan-Path clustering [Goldberg 2010], or explicitly defined unsupervised algorithms [Eivazi and Bednarik 2011], [Grindinger 2010]. While data mining methods can quickly identify clusters of similar attention patterns during visualization tasks, the results they return are often difficult to interpret, since unsupervised algorithms are typically applied as black-boxes. By contrast, although traditional human-guided statistical analyses can be more time-consuming, its findings tend to be more transparent and easier to interpret. Our investigation of eye tracking presents such a human-guided analysis of how gaze behavior relates to both user characteristics, and visualization properties. In particular,

we use several statistical models in order to provide fine-grained insights on how individual user characteristics interact with visualization components and task difficulty, and how these effects can impact user gaze patterns.

Chapter 3 User Study: Bar Graphs & Radar Graphs

3.1 Overview

This chapter contains the details and design of the user study that was conducted to investigate the impact of individual differences. The overall goal of this work is to A) identify what specific user characteristics influence visualization effectiveness, and B) explore the relationship of these user characteristics with a user's gaze behavior. As case studies, two basic visualization techniques are used: bar graphs and radar graphs. Bar graphs were chosen because they are one of the most popular and effective visualization techniques. Radar graphs were chosen because they were among the simpler of the two visualizations studied in Conati & Maclaren [2008], and though it has been argued that bar graphs are superior to radar graphs on common information seeking tasks [Few 2005], the reality is that radar graphs are still being discussed and used in various applications³. One notable study worth mentioning is the work done by Diehl et al. [2010] on uncovering the strengths and weakness of radial visualizations (e.g., radar graph, pie chart) versus non-radial visualizations (e.g., bar chart, parallel coordinates). Their work mentions that empirical evaluations between radial and non-radial visualizations are rare, therefore the fact that their study compares the effectiveness of one radial visualization versus one non-radial visualization is very similar to our study except they do not consider individual user differences. Diehl et al. found that overall, the non-radial visualization (i.e., rectangular visualization in a Cartesian coordinate system) was significantly better in terms of speed and accuracy for a short set of 8 simple tasks. As we will see in our results, we too find that the non-radial (i.e., bar graph) visualization is also better than the radial, but only for the simple visualization tasks we examine. However, in Diehl et al., they acknowledge that the validity of their findings may not generalize to more complex visualizations, and in our study we find that there is in fact no longer a significant difference between the two visualization types for more complex tasks.

³ A Google search on "Radar Chart" as of January 1, 2013, produced 40,000 unique results that were all less than 10 years old.

3.2 Study Design Contribution

An important point to make clear before diving into the specifics of the study, is that the author of this thesis was not directly involved in the design of the study described in this chapter (namely which visualizations, tasks, and cognitive traits to use). The reason for this is that the author became a member of the ATUAV research group right at the moment where the study was being administered. Nevertheless, while the author was not able to influence the choice of which information visualizations and tasks to use, as well as which cognitive traits to test for, it was still necessary to read and summarize all of the relevant research in order to generate the rationale and arguments for why all the choices had been made. Furthermore, the contribution of statistical and data analyses presented in Chapter 4, and Chapter 5 were not a substantial part of the study design (i.e., not clearly defined a priori) and are therefore another significant contribution of the author of this thesis.

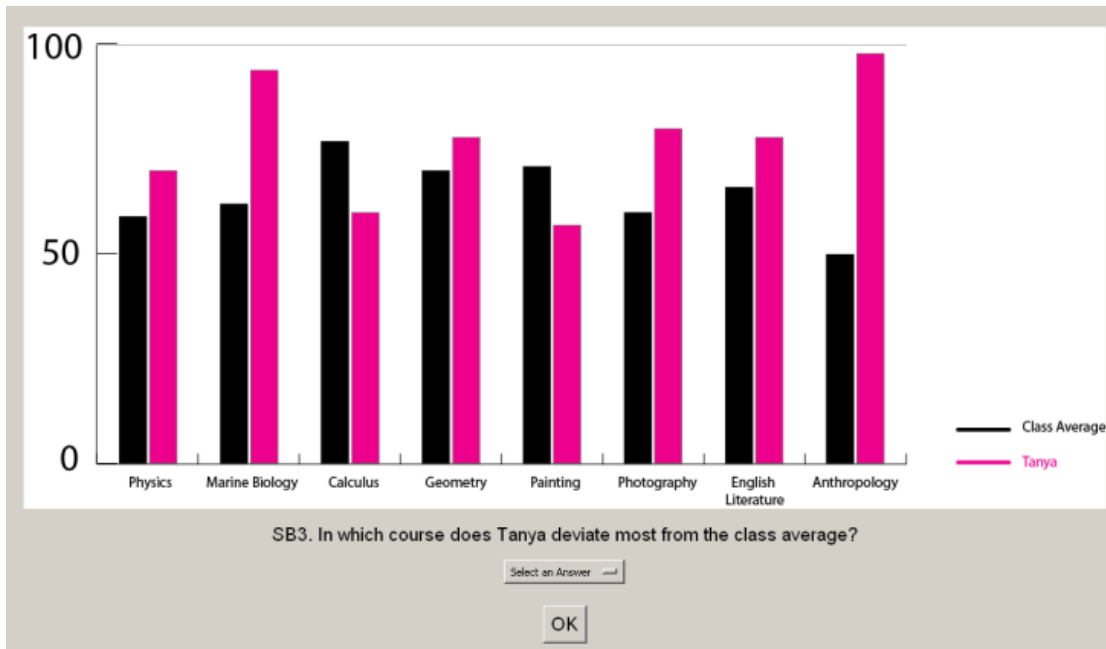


Figure 3.1: Single scenario example for bar graph used in this study.

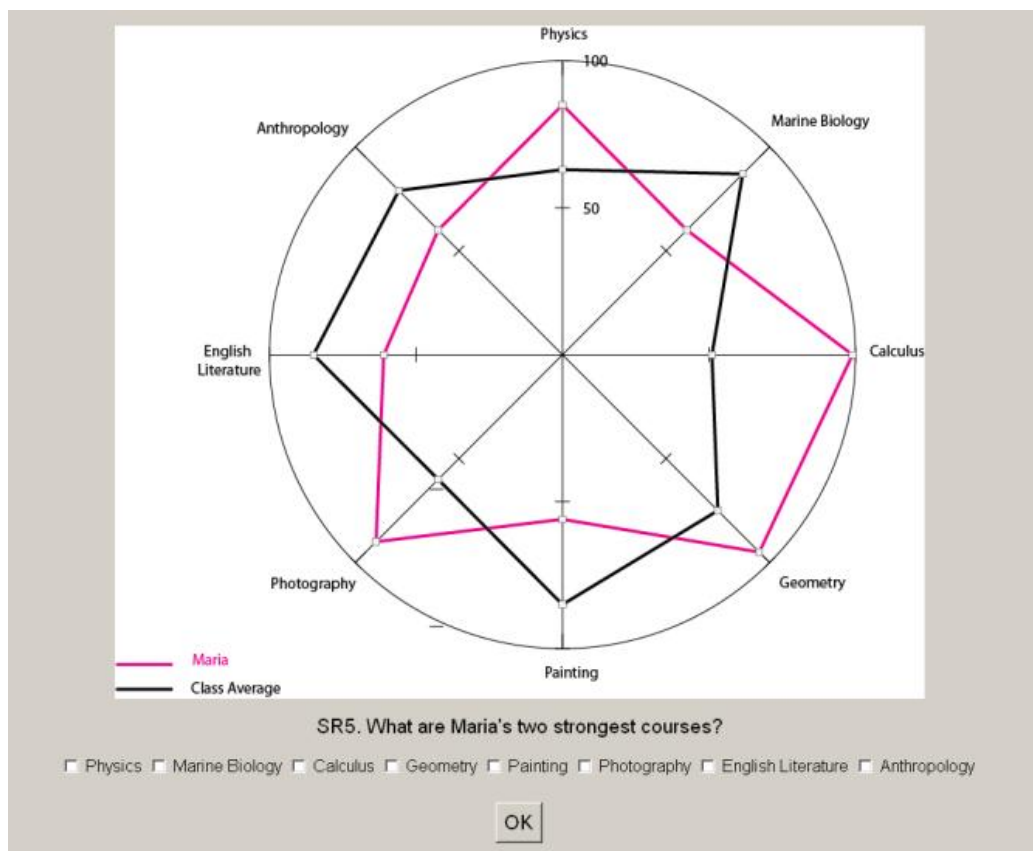


Figure 3.2: Single scenario example for radar graph used in the study

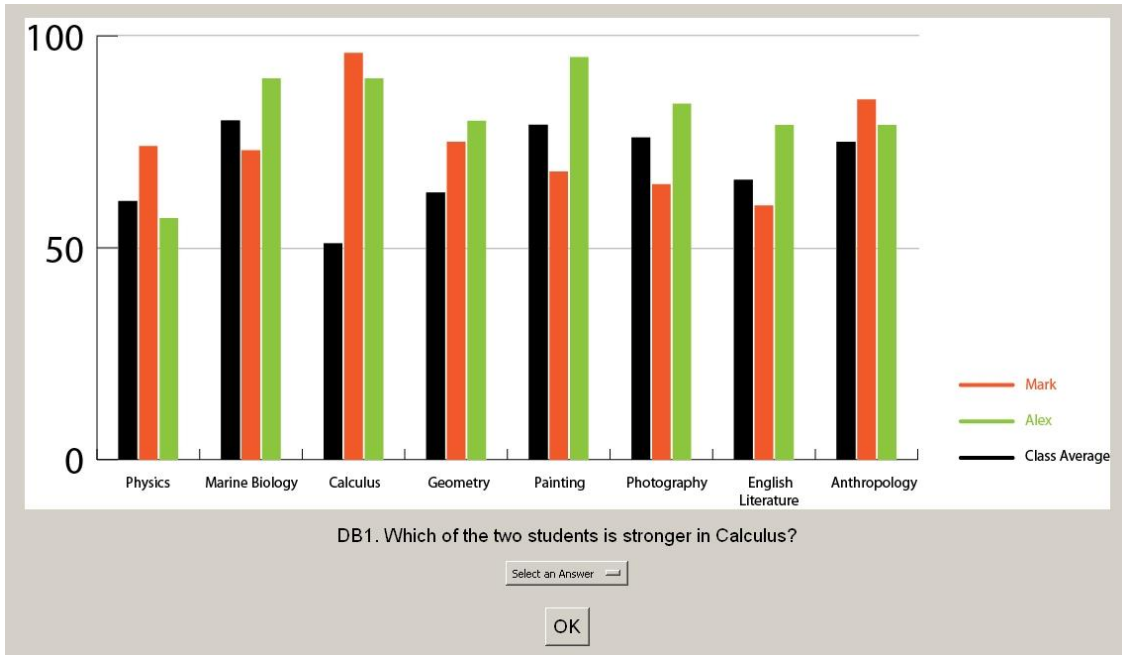


Figure 3.3: Double scenario example for bar graph used in the study

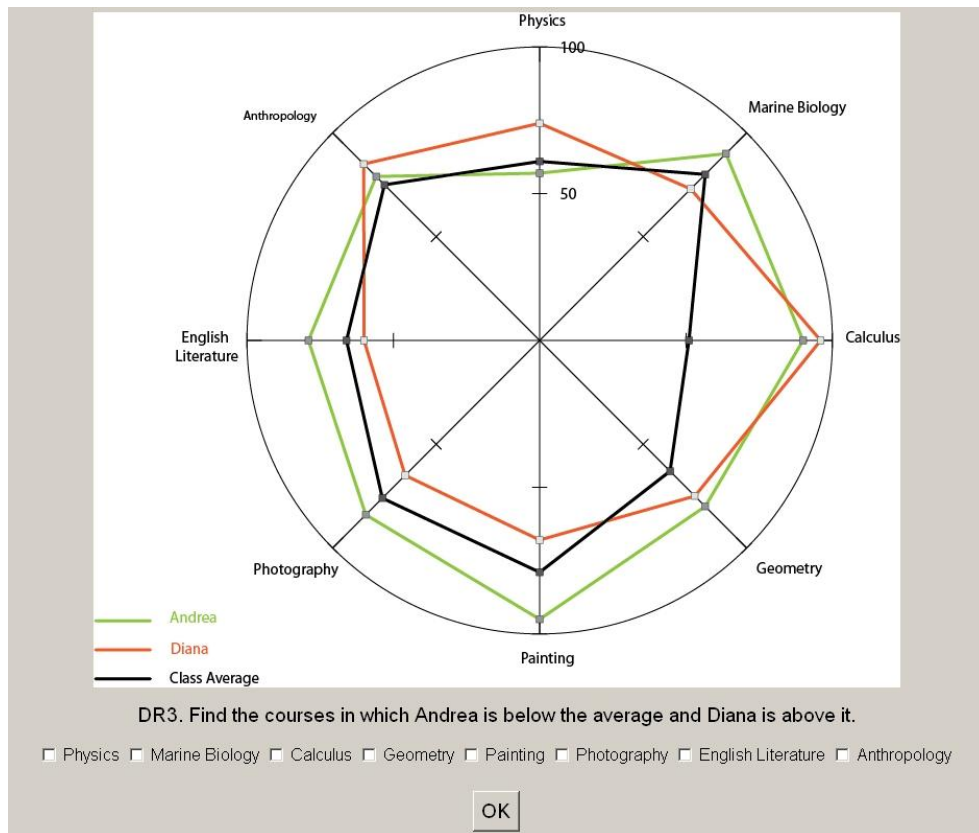


Figure 3.4: Double scenario example for radar graph used in the study

3.3 Experimental Tasks

We constructed a fictional data set of 8 university courses containing student grades. One reason for choosing this data set was to avoid any confusion due to misunderstanding of the data, we wanted to have a data set that was familiar to most of our participants whom were all university students.

The tasks were based on a set of low-level analysis tasks that Amar et al. [2005] identified as largely capturing people's activities while employing information visualization, (see Table 3.1 for a list of the tasks). The tasks were chosen so that each of the two target visualizations would be suitable to support them. A first battery of tasks involved 5 questions comparing the performance of one student with the class average for 8 courses, e.g., "In how many courses is Maria below the class average?" (**single scenario tasks** from now on, since they involve a single student, refer to Figure 3.1 and Figure 3.2). A second battery of tasks involved 4 questions comparing the performance of two students and the class average in order to increase task complexity, e.g., "*Find the courses in which Andrea is below the class average and Diana is above it?*", (these are labelled the **double scenario tasks** since they pertain to two students, refer to Figure 3.3 and Figure 3.4). Participants repeated each of the 5 tasks in the single scenario with two different datasets that varied in terms of skewness of the value distribution to account for a possible effect of distribution type on visualization effectiveness. Specifically, this was meant to compare a spiky distribution against a close-to-uniform distribution, where the spiky distribution was created by alternating student grades between high and low for some of the courses displayed in adjacent positions. No variations on distribution were used for the double scenario tasks in order to keep the experiment's length under one hour, as it is generally recommended for studies involving visual attention [Goldberg 2010]. The student names used in the task dataset were also changed between each round of tasks in order to prevent knowledge transfer.

Single Scenario Tasks (involving one student)

Task Type	Sample Question
Compute Derived Value	In how many courses is <i>[Mary]</i> below the class average?
Retrieve Value	In which course does <i>[Mary]</i> deviate most from the class average?
Find Extremum	In which courses is <i>[Mary]</i> above the class average?
Retrieve Value	Did <i>[Mary]</i> receive a higher mark in <i>[marine biology]</i> or <i>[painting]</i> ?
Sort	What are <i>[Mary's]</i> two strongest courses?

Double Scenario Tasks (involving two students)

Task Type	Sample Question
Retrieve Value	Which of the two students is stronger in <i>[Calculus]</i> ?
Filter	Find the courses in which <i>[Mark]</i> and <i>[Alex]</i> are both above the class average?
Filter	Find the courses in which <i>[Alex]</i> is below the average and <i>[Mark]</i> is above it?
Compute Derived Value	Which student has performed better than the class average in a greater number of courses?

Table 3.1: Task set explored in the study (from Amar's taxonomy)

3.4 User Characteristics Considered in the Study

The individual characteristics investigated in this study include three cognitive abilities: perceptual speed, visual working memory, and verbal working memory, as well as two measures of user visualization expertise, one for each of the two visualizations prior to the study. These user characteristics will be among the independent measures considered in the statistical analyses that are conducted in the following chapters.

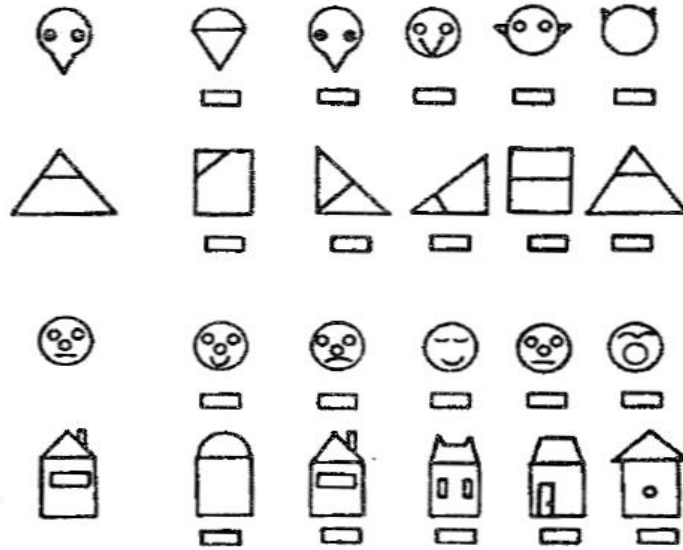


Figure 3.5: Sample questions from the perceptual speed test. For each row, participants are required to indicate which item matches the one on the left

3.4.1 Perceptual speed

Perceptual speed is a measure of speed when performing simple perceptual tasks (e.g., finding figures or making comparison between letters, numbers, objects, pictures, or patterns) such that everyone would score perfect if there were no limitation on time [Salthouse 2000]. Perceptual was selected because it was part of the original set of cognitive measures related to perceptual abilities that were explored by Velez et al. [Velez 2005]. Perceptual speed was also the only measure in the set for which [Conati 2008] found a significant interaction effect with visualization type when comparing the effectiveness of radar graphs and Multiscale Dimension Visualizer (MDV). In order to measure perceptual speed, we used Identical Figures Test (P-3) [Ekstrom 1976]. P-3 is a paper-based test that measures a participant's speed in matching a figure on the left side of each row to one of five possible figures appearing on the right side (see Figure 3.5). There are two trials, each consisting of 48 rows, and subjects are given 90 seconds per trial to correctly match as many rows as possible. The final test score is the number of correctly answered rows, ranging from 0 to 96.

3.4.2 Visual working memory

Visual working memory is part of the working memory responsible for temporary storage and manipulation of visual and spatial information [Logie 1995]. Visual working memory was also selected because it part of the original set of cognitive measures related to perceptual abilities that were explored by Velez et al. [Velez 2005]. Conati [2008] also found a significant result for visual working memory on user accuracy of certain filter tasks. Several tests are available for measuring visual working memory, but they each measure slightly different constructs (e.g., the ability to remember sequences of visual elements or the ability to integrate information from long-term memory). The test most appropriate for the nature of the tasks in this study is the one proposed and tested by Fukuda & Vogel [Fukuda 2009] which measures the ability to recall the location of basic colored squares over short intervals and most closely matched the task structure in our study. We were not specifically interested in looking at sequential or long-term visual memory since each of our tasks was static and did not require any visual information to be remembered from one task to the next. The Fukuda & Vogel test is computer based, and consists of 120 trials. In each trial an array of either 4, 6, or 8 different colored squares flash on the screen for 500ms. After 200ms, one of the colored squares re-appears in its original location, and participants must select whether the square has the same or a different color than when it was first flashed. The final score for the test measures the percentage of correct answers.

3.4.3 Verbal working memory

Verbal working memory is part of the working memory responsible for temporary maintenance and manipulation of verbal information [Baddely 1986]. Verbal working memory was selected because we hypothesized that this cognitive trait may affect performance in processing textual components of a visualization, which, in our study include legends, labels, and task descriptions. We chose to use the Operation-word span test (OSPAN) for verbal working memory to measure this user characteristic. This test accounts for both the storage and the processing components of this trait and has shown the highest correlation with reading comprehension when compared to other tests [Turner 1989]. The test is computer-based, and shows a series of arithmetic equations that the participant must classify as True or False. Once a selection is made

for a given equation, one word appears on the screen for 800ms and then the next arithmetic equation appears. At the end of a series, participants are required to recall the words in correct order. This process is repeated several times with each series varying in length from 2 to 6 words. The insertion of secondary tasks between memory items (here, solving math examples) means that subjects are required to recall information that is periodically unattended and vulnerable to proactive interference [Kane 2002]. The final test score is a measure of the average word span in working memory, ranging from 0 to 6.

3.4.4 User expertise

There are numerous examples where domain expertise has been shown to impact both accuracy and speed with information visualizations (see Dillon [2000]). However, in our study, we choose to assess the impact of expertise with respect to each visualization, independent of the task domain. Participants in our study self-reported their expertise in using each of the two visualizations prior to the study as follows: for each visualization they answered the question: “*Had you ever used radar(bar) graphs before this study?*”. Then, users expressed their agreement with the statement “*I am an expert in using radar(bar) graph,*” on a Likert-scale from 1 to 5.

3.5 Study Design

3.5.1 Participants

Thirty-five subjects (18 were female), ranging in age from 19 to 35, participated in the experiment over a period of 3 weeks. Participants were recruited via advertising at the university, as is the standard practice. Ten participants were computer science students, while the rest came from a variety of backgrounds, including microbiology, economics, classical archaeology, and film production. Participants were each given \$20 cash as compensation for participating in the study.

3.5.2 Apparatus

The experiment was conducted on a Pentium 4, 3.2GHz, with 2GB of RAM, and a Tobii T120 eye tracker as the main display. Tobii T120 is a remote eye tracker embedded in a

17" display, providing unobtrusive eye tracking (see Figure 3.6). The experimental software that we used, was fully automated and coded in the Python programming language.⁴

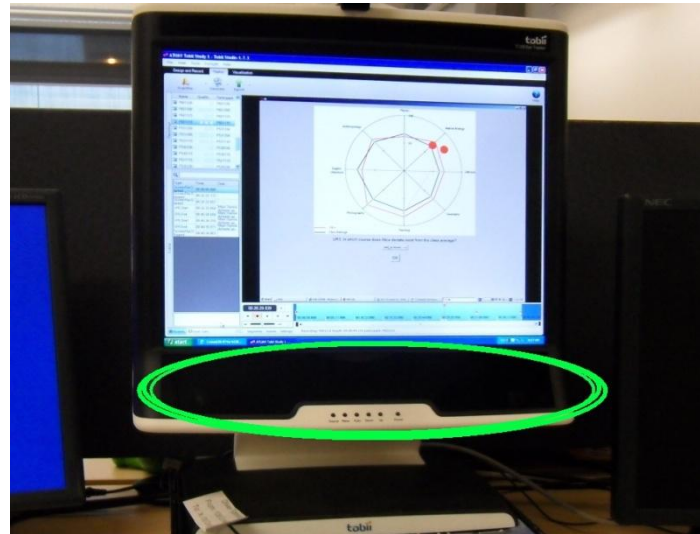


Figure 3.6: The actual Tobii T120 eye tracker used for this study. The circled region shows where the eye tracking cameras are located within the computer screen housing.

3.5.3 Procedure

The experiment was a within-subjects study, designed and pilot-tested to fit in a single session lasting at most one hour. Participants began by completing tests for the three cognitive measures: first the paper-based perceptual speed test (3 minutes long), then the computer-based verbal working memory test (lasting between 7 and 12 minutes), and finally the computer-based test for visual working memory (10 minutes long). Next was a calibration step for the eye tracker software, then training on each of the two visualizations, followed by the main barrage of tasks for the study. Participants were trained on each of the two visualizations which lasted about 10 minutes. For training, participants were required to correctly identify several components of the bar and radar graphs (e.g., 'point to the course names', 'point to where a grade of 0 would be', 'point to the class average', etc.), and were then required to answer some sample task questions for each visualization. Each participant was offered a chance to re-do the training step if

⁴ This programming work was done by Nicholas FitzGerald at the Department of Computer Science in the LCI laboratory under the ATUAV project.

they felt they needed more practice. No participants re-trained on the bar graph and two participants requested one extra round of training for the radar graph. As for the main portion of the study, each participant performed the 14 tasks described in Section 3.3 twice, once with each of the two target visualizations, for a total of 28 trials per participant. The presentation order with respect to visualization type was fully counterbalanced across subjects. Each task consisted of presenting the participant with a radar/bar graph displaying the relevant data, along with a textual question (see Figure 3.1 and Figure 3.2). Participants would then select their answer from a drop-down list or set of radio boxes, and click OK to advance to the next task. Before seeing the next task, participants were also presented with a screen asking them to rate their confidence with their submitted answer on a Likert scale from 1 to 5 (see Figure 3.7). Finally, at the end of the study, participants were asked to complete a questionnaire in order to ascertain age, gender, background, visualization expertise, etc., (see Appendix B - Post Questionnaire).

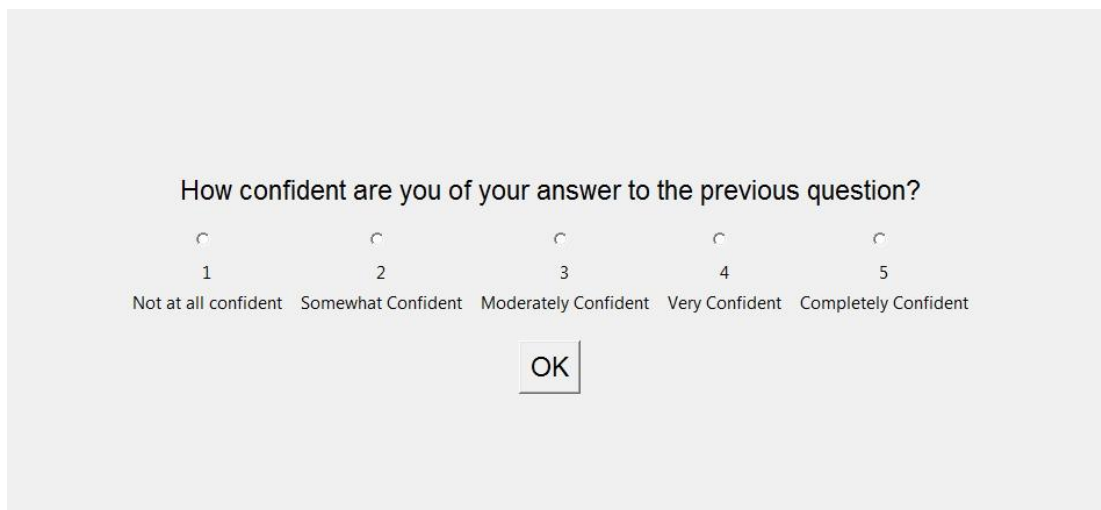
The image shows a screenshot of a user interface prompt. At the top, the text reads "How confident are you of your answer to the previous question?". Below this text, there are five radio button options arranged horizontally, labeled 1 through 5. Underneath each number is a corresponding label: "1 Not at all confident", "2 Somewhat Confident", "3 Moderately Confident", "4 Very Confident", and "5 Completely Confident". At the bottom center of the prompt area, there is a rectangular button with the text "OK" inside it.

Figure 3.7: Prompt used to collect user confidence after each task

3.6 Dependent Measures Collected From Users During Task Execution

Several measures were collected regarding task performance. There were two objective measures: completion time and task accuracy, and two subjective measures: bar/radar graph ease-of-use, and bar/radar graph user preference. These measures, which are

described next, acted as the dependent variables in the statistical analyses that are performed and are explained in detail in Chapter 4. Eye tracking data was also recorded for each user, and many features are generated from this data set which will be explained in greater detail in Chapter 5.

3.6.1 Objective measures of user performance

We considered two objective measures of user performance. First, we measured *task accuracy*, a measure of how well users performed in terms of providing the correct answer for each task. Second, we measured *completion time*: the experimental software recorded the total amount of time in milliseconds that participants spent on each task.

3.6.2 Subjective measures of user performance

There are two subjective measures we collected. First, preference ratings for each of the two visualizations were collected in the post-questionnaire via the two statements "*I prefer to use bar graph for answering the questions*" and "*I prefer to use radar graph for answering the questions*", rated on a Likert scale from 1 to 5. Secondly, an assessment of overall ease-of-use of each visualization was collected in the post-questionnaire by asking participants to rate on a Likert scale from 1 to 5 the two statements: "*In general, radar graph was easy to understand*," and "*In general, bar graph was easy to understand*." The questions were worded in such a way that "easy to understand" was chosen over "easy to use" since the visualizations in the study were not interactive, and thus it was more natural to express usability in terms of understandability.

3.6.3 Eye tracking data

Each user's gaze data was recorded using eye tracking, and a detailed explanation of all of the collected measures, as well as the processing and analysis of this data is described in Chapter 5.

Chapter 4 Impact of Individual Differences on User Performance & User Preferences: Analysis & Results

The goal of this section is to investigate if user characteristics influence the effectiveness of two common information visualizations: bar graphs and radar graphs. In order to answer these questions, several statistical analyses are performed on the following objective measure and subjective measures that were discussed in the previous chapter: *task completion time*, *visualization preference*, and *visualization ease-of-use*.

First we look at the descriptive statistics of the characteristics collected from the users in this study, and then briefly compare the results to other studies. Next we analyse and discuss results for task completion, and then do the same for the two subjective measures: preference and ease-of-use. Lastly we present a summary of the findings and discuss the implications of our results for user-adaptive information visualization design.

4.1 Summary of User Characteristics Data

Table 4.1 (next page) summarizes the descriptive statistics of the user characteristics that were measured from the participants in our study. The rather large variances for most measures indicate that the study succeeded in collecting a pool of participants that was quite diverse.

Measure	N	Min	Max	Mean	Std. Dev
Percep. Speed	35	54.00	96.00	85.70	11.64
Visual WM	35	0.3	5.4	2.72	1.36
Verbal WM	35	2	6	4.58	1.10
Bar Expert	35	2	5	4.15	0.87
Radar Expert	35	1	5	2.12	1.22

Table 4.1: Summary of user characteristics collected from the study

The user characteristic measurements we collected were also compared against measurements from other studies. For visual working memory, a different study of 117 university students⁵ had a mean of 2.51 which is quite comparable to ours (see second row in Table 4.1), and this other study had a standard deviation of 0.7, which is not very similar to our study. We are not sure why our study would have a considerably higher standard deviation for visual working memory. Next, for verbal working memory, a different study of 70 university students reported by Lin [2007], had a mean of 4.97 and standard deviation of 0.92, which are both fairly comparable to results measured for this study. Lastly, for perceptual speed, a different study involving 83 army enlistees [Ekstrom 1976] reported a mean score of 68.6 which is considerably lower than the mean measured from this study of 85.7. One possibility for this difference arises from the fact that the perceptual speed test is paper based, and was designed to be graded using a test scoring machine. This type of test typically requires the answer-bubble for each question to be completely shaded in with a pencil. For our study we graded the tests by hand since we did not have access to a scoring machine, therefore the study participants were informed that a simple checkmark or 'X' would suffice when indicating their answer to the test questions. Since the perceptual speed test consisted of 96 questions, and shading in the answer-bubbles takes a bit more time than just making a checkmark, it is possible that these little savings in time made a big difference over the course of the whole test and was the cause for our mean score to be higher. Given that

⁵ This was a study performed at the UBC Department of Psychology, and the values were reported to us via email by faculty member James T. Enns.

the other study's standard deviation for perceptual speed was 9.8 versus 11.6 for our study, this similarity in standard deviations is a good indicator that the range in perceptual abilities of both user populations were quite similar, despite the mean being skewed to the higher side in our study.

4.2 Effect on Performance Measures

Unfortunately there was ceiling effect on accuracy for this experiment, and nearly all users received perfect scores on the set of tasks. This may be due to the fact that subjects could take as much time as they wanted to generate an answer. Therefore, the objective measures analyzed in this chapter will consist solely of task completion time. The analysis of completion time has been separated into single scenario tasks and double scenario tasks because in the single scenario phase there is an additional between-subject control for order of distribution⁶, and also because the statistical model (General Linear Model) used was not able to account for the unequal quantity of tasks administered to users between the single and double scenario. The single scenario set of tasks comprised of 10 different tasks, repeated once per visualization type, and the double scenario comprised of 4 different tasks, repeated once per visualization type. Statistical significance is reported at the 0.05 level, as well as partial eta squared (η_p^2) for effect size, where .01 is a small effect, .09 is a medium effect, and .25 is a large effect [Field 2003].

4.2.1 Single scenario - model for analysis

For the single scenario phase, the statistical model we used was a repeated-measures 2 (visualization type) by 2 (distribution type) by 5 (task) general linear model with visualization-type order and distribution-type order as between-subject factors, and the user characteristics as the covariates. The sphericity assumption was verified for this data set using Mauchly's test.

⁶ The effectiveness of a particular visualization can depend on the distribution of data visualized. Therefore, we considered two different distributions, spiky and uniform, to test the effect of distribution on effectiveness of the visualization. For creating a spiky distribution, we alternated student grades between high and low values in neighbouring courses.

4.2.2 Single scenario results - main effects

There was a large significant effect of **visualization type** (bar vs. radar), $F(1, 20) = 8.06$, $p = .01$, $\eta_p^2 = 0.29$. Completion time was faster with bar graphs ($M = 14.25s$, $SE = 0.6s$), than with radar graphs ($M = 19.0s$, $SE = 0.76s$). This results supports the findings in Diehl et al. [20120] that for simple tasks, bar graphs are better than radar graphs in terms of performance since users are always faster with bar graphs and there is no loss of accuracy (users were on average 99% correct for all single scenario tasks). There was also a large significant main effect of **perceptual speed**, $F(1,20) = 7.61$, $p = .01$, $\eta_p^2 = 0.28$, indicating that the higher the perceptual speed, the faster the completion time for both visualizations. The mean completion time for participants with low vs. high perceptual speed was 18 and 16 seconds respectively (where high/low is defined based on the median split of perceptual speed values). A plot of the perceptual speed scores is shown in Figure 4.1. This result confirms previous findings that differences in cognitive measures can sufficiently impact general visualization effectiveness and, like in [Conati 2008], it singles out perceptual speed as a relevant measure.

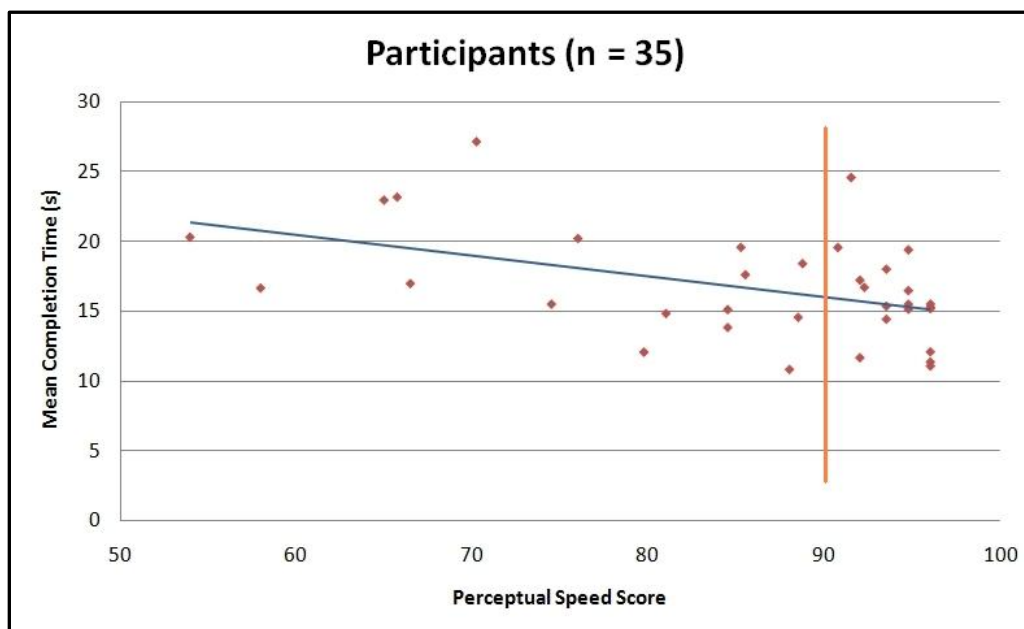


Figure 4.1: Scatter plot with trend line showing task completion time and perceptual speed scores for each participant in the single scenario tasks. The vertical line shows the median split used to report users as high and low

4.2.3 Single scenario results - interaction effects

There was a medium-large significant interaction effect between **visualization type and perceptual speed**, $F(1,20) = 4.49$, $p < .05$, $\eta_p^2 = 0.18$. Even though completion time is always faster with the bar graph, the difference in time performance between bar and radar decreases as a user's perceptual speed increases (See Figure 4.2). This result is important because it confirms the finding in [Conati 2008] that perceptual speed is a cognitive measure that can sufficiently impact the *compared effectiveness* of two different visualizations, at least when one of them is a radar graph.

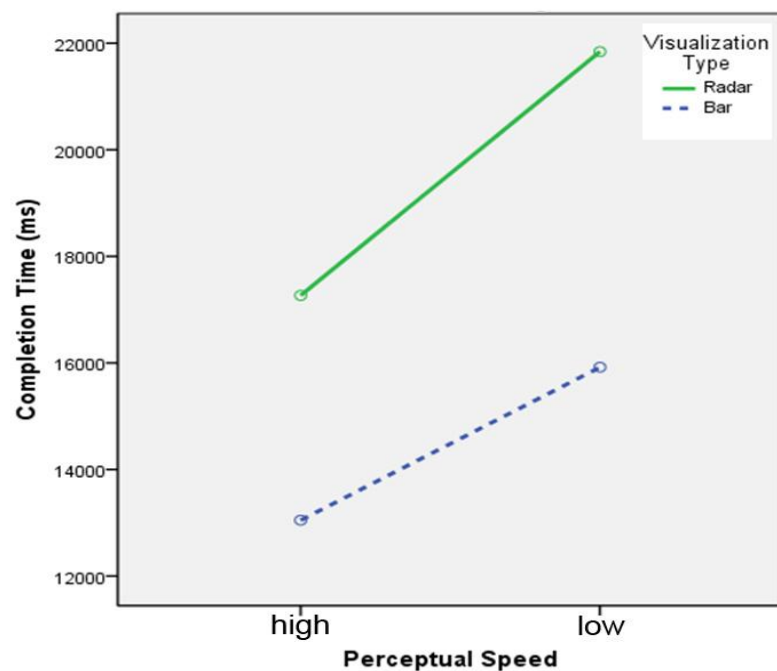


Figure 4.2: Chart showing the mean completion times for the interaction effect of perceptual speed and visualization type, users with low perceptual speed are taking even longer for radar graph tasks

There was also a large significant interaction effect between **visualization type and visualization order**, $F(1,20) = 8.66$, $p < .01$, $\eta_p^2 = 0.30$. Subjects that saw radar graphs first, proceeded to perform better with bar graphs than those who saw bar graphs first. Conversely, subjects who saw bar graphs first, proceeded to perform better on radar graphs than those who saw radar graphs first (see Figure 4.3). Thus, it appears that there is a *training effect* between visualizations, despite the fact that task details are changed from the first to the second visualization provided. What is likely happening is that the user is becoming familiar with the general task context/domain (e.g., the fact that the user is looking for values of school courses) after seeing it with the first visualization provided, which facilitates task performance with the second visualization.

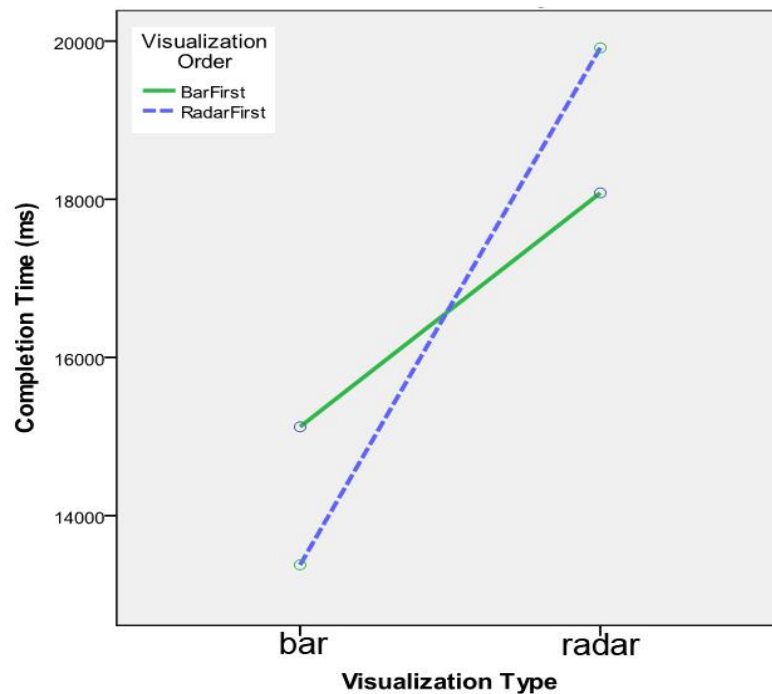


Figure 4.3: Interaction for visualization type and order. Users always perform faster on the second visualization they use compared to their counterparts

No significant findings were found for distribution type, so we do not consider this factor in subsequent analyses. There was also a lack of effect for visualization expertise for the single scenario tasks which may be due to a training effect. Given that the single scenario tasks were fairly simple and each user performed 20 of these tasks, all users became relative experts for this set of tasks by the end of the study.

4.2.4 Double scenario - significant results

The double scenario comprised of 4 different tasks, repeated once per visualization type. These tasks required the user to answer questions that related to two students along with the class average, instead of just one student. For the double scenario, the statistical model we used was a repeated-measures 2 (visualization type) by 4 (task) general linear model, with visualization order as a between-subject factor, along with the user characteristics as covariates.

The only significant effect found was a medium-sized effect of **task**, $F(2, 50) = 4.32$, $p < .05$, $\eta_p^2 = 0.14$. This effect suggests that in this phase there is a larger spread of difficulty across tasks as compared to the single scenario phase, resulting in a significant impact of the double scenario tasks on completion time.

The *lack of a significant effect* and very low effect size of **visualization type** is interesting ($p = .465$, $\eta_p^2 = 0.02$), because it opens the possibility to challenge claims in the literature that bar graphs are generally superior to radar graphs (e.g., [Few 2005], [Simkin 1987], [Diehl et al. 2010]). Given the low effect size for visualization type, the lack of significant effect may be due to a training effect generated by the participants' interactions with the two visualizations in phase one, which managed to eliminate the effect of visualization type detected in phase one. An alternative explanation is that radar graphs are as good as bar graphs for the types of comparison tasks covered in the double scenario phase. While we do not have data to reliably choose between these two explanations, the fact remains that this is a scenario in which radar graphs are as effective as bar graphs, which presumably is a unique finding.

4.2.5 Double scenario - marginally significant results

There were also two marginally significant results worth mentioning here because their medium-large effect size.

First, **perceptual speed** has a marginally significant main effect, $F(1,26) = 3.87$, $p = .06$, $\eta_p^2 = 0.13$, which reflects the correlation of this cognitive measure with visualization effectiveness, similar to what was detected in the single scenario phase. A plot of the perceptual speed scores is shown in Figure 4.4. We also calculated the experimental power for this result which was 0.6. This is below the typical recommended value for

experimental power of 0.80 [Field & Hole 2003]. An additional 20 participants would be required in order to increase the experimental power from 0.60 to 0.80, which means we would have a 20% better chance of detecting a significant result given the medium-large effect size.

Second, **radar expertise** has a marginally significant main effect, $F(1,26) = 4.01$, $p = .055$, $\eta_p^2 = 0.14$. In this case, the experimental power was 0.63, and 16 more participants would be required in order to increase the power to 0.80. Recall that there was no effect of expertise in the single scenario, suggesting that for the simpler single scenario tasks, the training provided to participants as part of the experimental setup managed to remove differences due to existing expertise. For the double scenario tasks, it appears that expertise starts having an effect, possibly due to the fact the tasks in the double scenario are more difficult. Furthermore, the effect of radar expertise is in relation to completion time for both visualization types, which means that radar expertise is linked to both radar graph and bar graph performance.

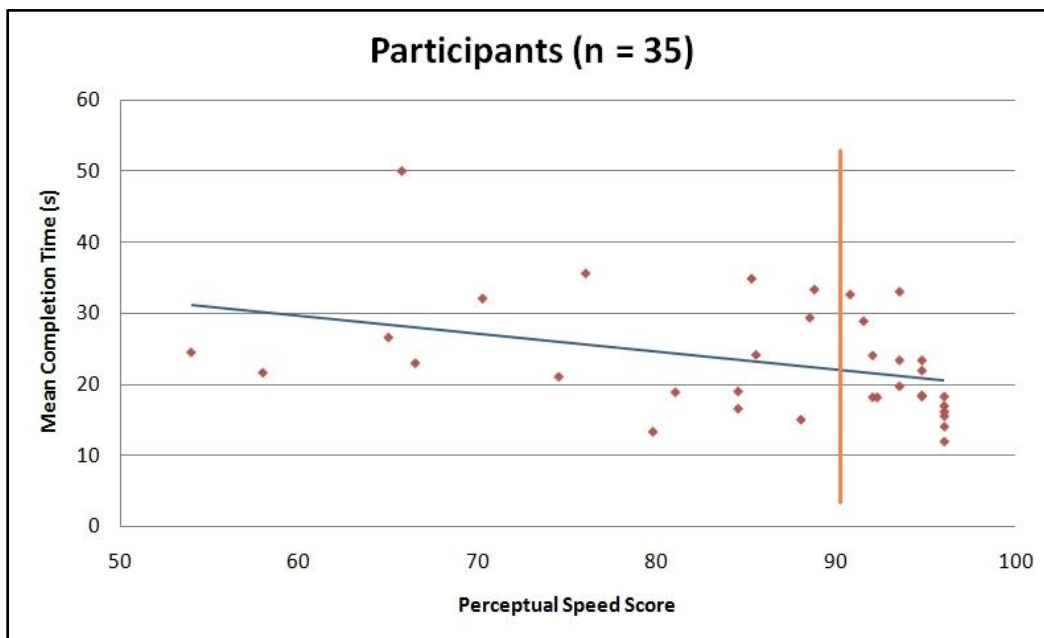


Figure 4.4 Task completion time and perceptual speed scores for double scenario tasks with trend line. The vertical line shows the median split used to report users as high and low

4.3 Effects on User Preference & Ease of Use

In this section, the data from the study is no longer analyzed per task, but rather with respect to the whole interaction for each user. This is because the questions we asked in order to assess visualization preference and ease-of-use were given at the end of the experiment in the post test (see Appendix B - Post Questionnaire), and apply to the overall user experience with radar and bar graph, as opposed to specific tasks. Figure 4.5-left shows the distribution of preference ratings for bar and radar graph that were collected from the users of this study. The distribution of ratings for the bar graph is skewed towards high values, whereas it is more uniformly distributed for the radar graph, indicating a higher variance in user preferences for the radar graph visualization. Figure 4.5-right shows ease-of-use ratings for bar and radar graph. As was the case for preference ratings, more users give their highest rating to the bar graph compared to the radar graph with a higher variance for the radar graph ratings. It is worth noting however that in Figure 4.5-right, both the radar and bar graph are generally skewed towards high values, indicating that neither visualization is particularly difficult to understand.

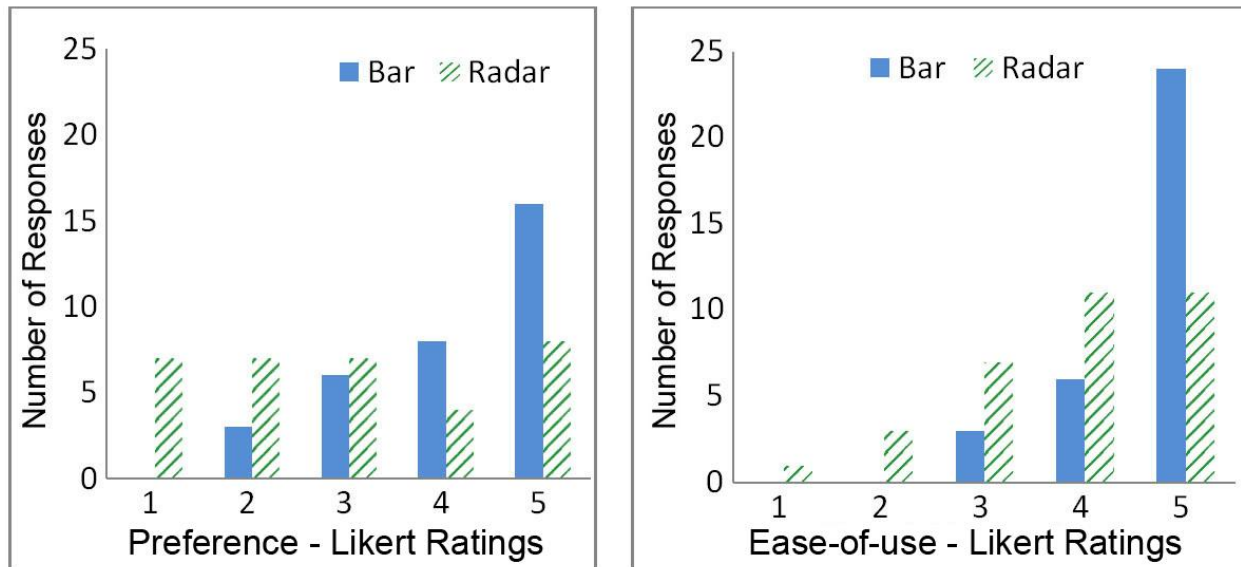


Figure 4.5: Likert-scale data collected for graph preference (left), and ease-of-use ratings (right)

4.3.1 Description of statistical model employed - multivariate (MANCOVA)

Recall that preference and ease-of-use data was collected using a standard 5 point Likert-scale, and as such is not necessarily suitable for standard parametric analysis since ordinal data is not invariant to monotone transformations [Kaptein 2010]. Therefore, the Aligned Rank Transformation (ART) was applied using the web-based ART-Tool [Wobbrock 2011] which transforms the Likert ratings for Radar Preference, Bar Preference, Radar Ease-of-Use, and Bar Ease-of-Use into a form that is consistent over monotone transformations, which can then be correctly analyzed using standard parametric analysis. The statistical model we used was a multivariate analysis, with preference and ease of use ratings as the dependent variables, and the user characteristics as covariates.

4.3.2 Subjective measures results - main effects of cognitive abilities

We found the following two effects of cognitive abilities on subjective measures:

- A large significant effect of **visual working memory** on *radar graph preference*, $F(1, 26) = 10.65$, $p < .01$, $\eta_p^2 = 0.29$. In general, users with higher visual working memory had higher preference ratings for radar graphs (see Figure 4.6 and Figure 4.7). Because radar graphs display extra visual clutter due to more intersecting colours, shapes, and overlapping lines, our results suggest that users with a higher visual working memory capacity prefer a more visually dense or stimulating visualization type.

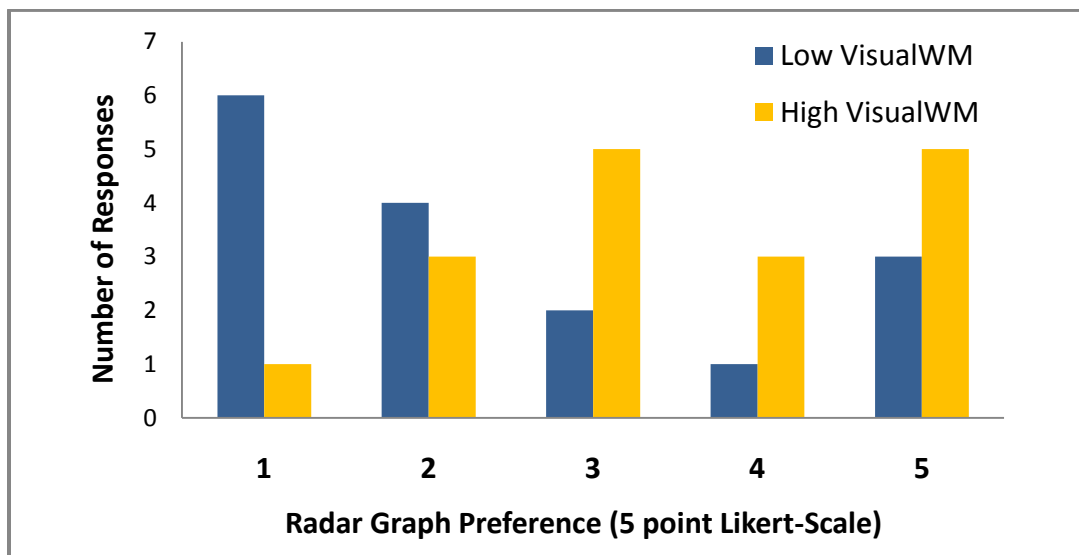


Figure 4.6: Histogram showing the responses for radar graph preference based on the median split of visual working memory

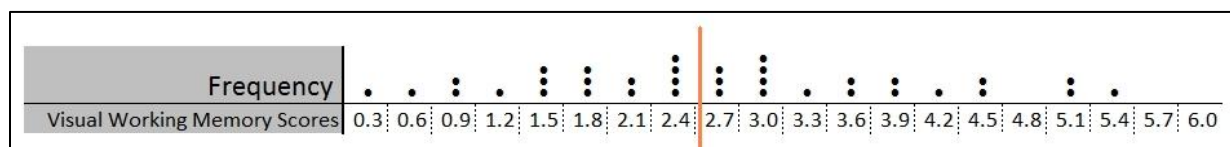


Figure 4.7: Dot plot showing the visual working memory scores with a vertical line indicating the median split

- We also found a large significant effect of **verbal working memory** on *bar graph ease-of-use*, $F(1, 26) = 9.69$, $p < .01$, $\eta_p^2 = 0.27$. In general, users with lower verbal working memory had a higher ease-of-use rating for bar graphs (see Figure 4.8 and Figure 4.9). It is difficult to interpret the link between verbal working memory and bar graph ease-of-use, especially given that users with higher verbal working memory are rating bar graphs less easy to use. Generally, we would expect having increased cognitive abilities would mean a visualization would be easier to use. Verbal working memory however, is not related to visual processing so it is not surprising that it does not facilitate it. Further investigation however is necessary to understand why it interferes with user visualization ease-of-use.

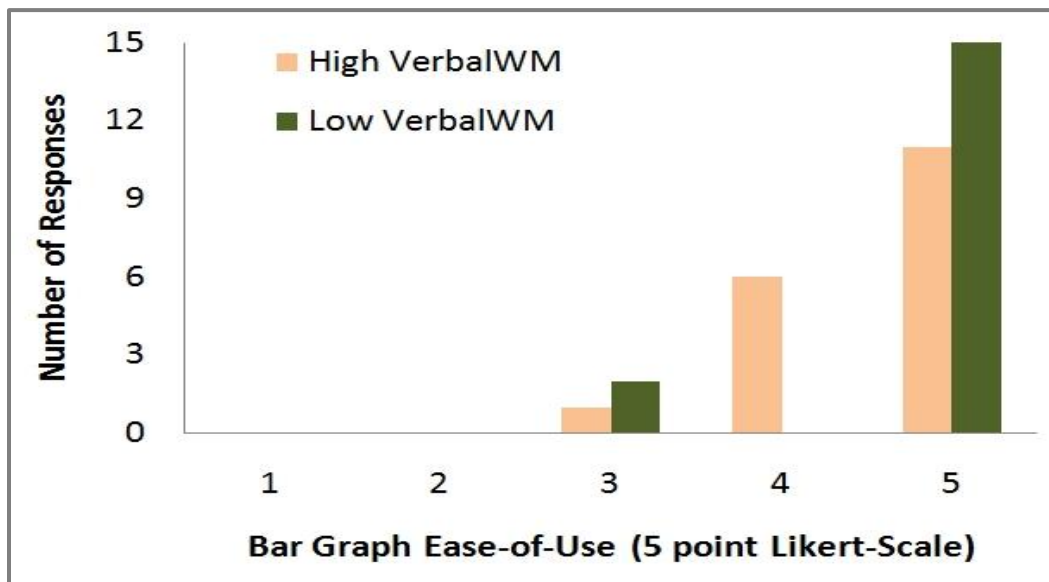


Figure 4.8: Histogram showing the responses for bar graph ease-of-use based on the median split of verbal working memory

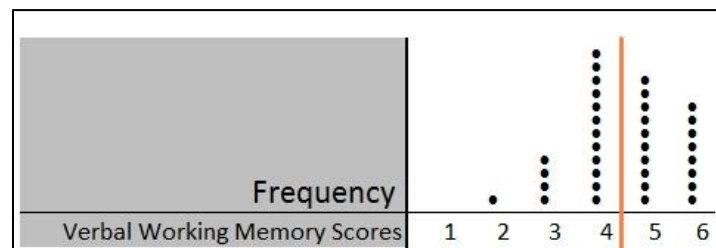


Figure 4.9: Visual working memory scores showing median split

These findings relating cognitive abilities to subjective measures are extremely interesting for two reasons. First, they are further evidence that user characteristics in general can affect a user's experience with visualizations. Second, they indicate that different characteristics may influence different factors that contribute to the user's overall experience with a visualization. In the case of our study, perceptual speed correlated with actual performance (completion time), whereas visual working memory and verbal working memory correlated with non-performance related subjective measures such as preference and ease-of-use, respectively.

4.3.3 Subjective measures results - main effects of expertise

We also found two significant results relating to user expertise:

- A very large significant effect of **Radar Expertise** on both *i) Radar preference*, $F(1, 26) = 45.80$, $p < .001$, $\eta^2 = 0.64$, and *ii) Radar ease-of-use*, $F(1, 26) = 19.6$, $p < .001$, $\eta_p^2 = 0.43$. We found that users with higher radar expertise had a higher preference and ease-of-use ratings for radar graphs.
- We also found a very high significant effect of **Bar Expertise** on *Radar ease-of-use*, $F(1, 26) = 931.86$, $p < .001$, $\eta_p^2 = 0.97$. Users with higher bar expertise had a higher rated ease-of-use for radar graphs.

Whereas it is quite intuitive that visualization expertise should correlate with the degree of preference and perceived ease-of-use of that visualization, it is interesting that in our study, visualization expertise only correlates with subjective measures, and not actual performance.

4.3.4 Visualization preference direct comparison - follow up analysis

In this subsection, we present a short follow up analysis that was run on visualization preference as a direct comparison between the two visualizations. We opted to add this section for three reasons:

- Our survey questions did not include a question that asked users to explicitly elicit their preference for bar graph and radar graph as a head-to-head

comparison, which could have been useful to verify user preferences between the two visualizations.

- We report results regarding preferences for bar and radar graph independent of one another, whereas the concept of having a preference may imply that an innate comparison is made between items, especially given the binary nature of the number of items to prefer in our study.
- Direct comparison preference ratings can still be calculated by subtracting the preference ratings of one item from the other (in our case, a positive score would mean a user prefers bar over radar, a negative score indicates a user prefers radar over bar, and a neutral score means the user is indifferent).

We computed direct comparison preference ratings by subtracting the radar graph preference from the bar graph preference for each user. We then used the Aligned Rank Transformation (see Section 4.3.1), and ran a univariate general linear model with the direct comparison preference ratings as the dependent measure, visualization order as a fixed factor, and the user characteristics as the covariates. Three significant results were found and are reported below in Table 4.2.

User Trait	F-Ratio	Effect Size	Sig. Value
<i>Verbal Working Memory</i>	F(1,27) = 631.93	$\eta^2 = 0.959$	p < 0.001
<i>Bar Graph Expertise</i>	F(1,27) = 998.59	$\eta^2 = 0.974$	p < 0.001
<i>Radar Graph Expertise</i>	F(1,27) = 430.0	$\eta^2 = 0.941$	p < 0.001

Table 4.2: Results for direct comparison preference ratings between bar graph and radar graph

We found that verbal working memory had a significant effect on the compared preference, where on average, users with higher verbal working memory had a stronger preference for bar graph over radar graph (see Figure 4.10).

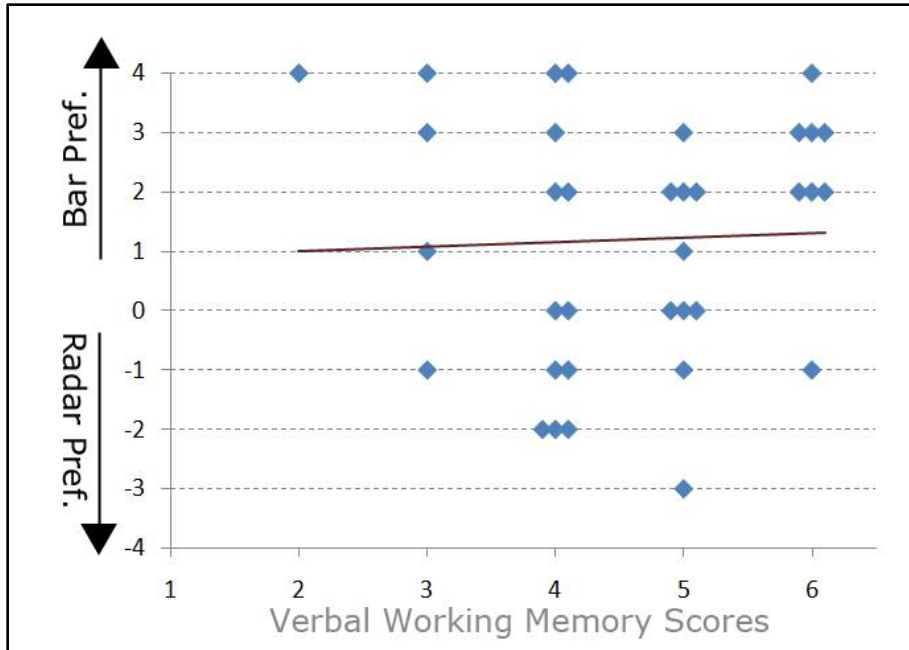


Figure 4.10: Direct comparison preference ratings for bar and radar graphs based on verbal working memory

Next, both bar graph expertise and radar graph expertise had a significant effect on compared preference, and in both cases, increased expertise correlated with a increase in preference for radar graphs over bar graphs (see Figure 4.11).

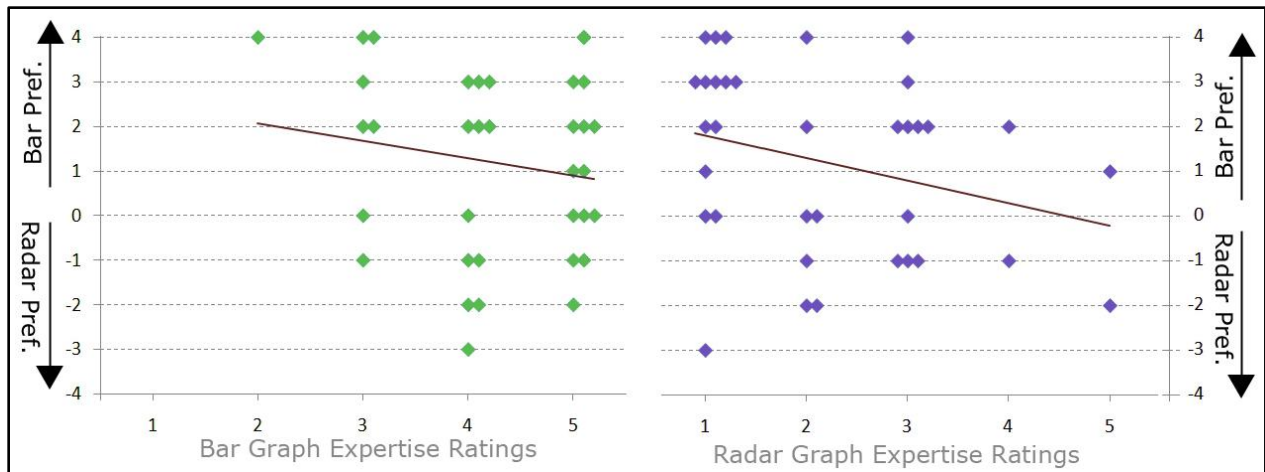


Figure 4.11: Direct comparison preference ratings for bar and radar graphs based on expertise for bar graph (left), and radar graph (right)

4.4 Chapter Summary & Discussion

The results of the analyses conducted in this chapter confirm and extend preliminary existing findings that individual user characteristics do have an impact on visualization effectiveness. A summary of our findings is shown below in Table 4.3.

Factor	Measure
Visualization Type	single scenario - <i>completion time</i> - (main effect & interaction effect)
Perceptual Speed	single scenario - <i>completion time</i> - (main & interaction effect) double scenario - <i>completion time</i> - (main effect, $p = .06$)*
Verbal Working Memory	subjective measure - <i>bar graph Ease-of-use</i> - (main effect)
Visual Working Memory	subjective measure - <i>radar graph Preference</i> - (main effect)
Bar Expertise	subjective measure - <i>radar graph Ease-of-use</i> - (main effect)
Radar Expertise	double scenario - <i>completion time</i> - (main effect, $p = .055$)* subjective measure - <i>radar graph Preference</i> - (main effect) subjective measure - <i>radar graph Ease-of-use</i> - (main effect)

Table 4.3: Summary of statistically significant results from this chapter, *indicates marginally significant

4.4.1 Implications of findings

For the specific comparison between bar graphs and radar graphs, it was discovered that while bar graphs are more effective than radar graphs in terms of completion time on simple information seeking tasks (i.e., single scenario tasks), the difference in performance with radar graphs is mediated by perceptual speed, decreasing for users with high perceptual speed. Yet, the two visualizations seem to be equivalent on more complex tasks (double scenario tasks). It is an open question to verify which of the two visualizations would be more effective on a set of tasks more complex than the ones considered in this study. As for impact of user characteristics on visualization effectiveness, in addition to the abovementioned interaction between perceptual speed and visualization type, there were also effects of other user characteristics such as

expertise, and visual/verbal working memory on the subjective measures of visualization preference and ease-of-use.

We envision two possible forms of adaptation based on the results in this chapter. The first is an adaptive system that would select between different visualizations for different users based on their user characteristics (i.e., which visualization is best for a given user). The second form of adaptation would be to provide some users with additional support by also offering interventions within a given visualization. To illustrate, assume that an adaptive system has a model of the current user that specifies values for their characteristics. If the target visualization is intended to support simple, single-scenario-like tasks, bar charts should be the default choice. However, if the user is low on perceptual speed, they may benefit from adaptive interventions, such as highlighting or arrow pointing to portions of the visualization relevant to the task. In contrast, if the visualization is intended to support more complex, double-scenario-like tasks, the adaptation may consist of selecting a different visualization for different user groups. For instance, users with high visual working memory or high radar graph expertise would likely prefer a radar graph, while users with none of these features would be more effective with bar graph.

Chapter 5 Eye Tracking: Impact of User Characteristics on Gaze Patterns

The goal of the analysis described in this chapter is twofold. First, we want to see if user characteristics influence a user's eye gaze behavior in a way that is detectable by eye tracking. Second, if there is an effect of user characteristics, which characteristics impact which eye gaze features, and how are these effects mediated by task difficulty and visualization type? The rest of this chapter will describe how we utilized the eye tracking measures collected from the study we ran (see Chapter 3) in order to address our goals for this chapter. First, eye tracking is defined and explained in terms of the eye gaze data we collected, and also how we pre-processed this data and defined areas of interest. Next, we define a novel way to define and assess task difficulty, which will be included as a model parameter for the statistical analyses executed in this chapter. We then provide a description of the statistical models used for analysis, followed by the experimental results. Finally, we end the chapter with a summary and discussion of our findings.

5.1 Eye tracking Data & Pre-processing

An eye tracker captures gaze information in terms of *fixations* (i.e., maintaining gaze at one point on the screen) and *saccades* (i.e., a quick movement of gaze from one fixation point to another), which can then be analyzed to derive a viewer's attention patterns. For this analysis, several eye tracking measures are used that belong to the set of basic eye tracking features described by Goldberg and Helfman [2010] as the building blocks for comprehensive gaze processing. These features are built by calculating a variety of statistics upon the basic eye tracking measures that are described in Table 5.1.

Measure	Description
Fixation rate	Rate of eye fixations per milliseconds
Number of Fixations	Number of eye fixations detected during an interval of interest
Fixation Duration	Time duration of an individual fixation
Saccade Length	Distance between two fixations delimiting a saccade (d in Figure 5.1)
Relative Saccade Angles	The angle between the two consecutive saccades (angle y in Figure 5.1)
Absolute Saccade Angles	The angle between a saccade and the horizontal (angle x in Figure 5.1)

Table 5.1: Description of basic eye tracking measures

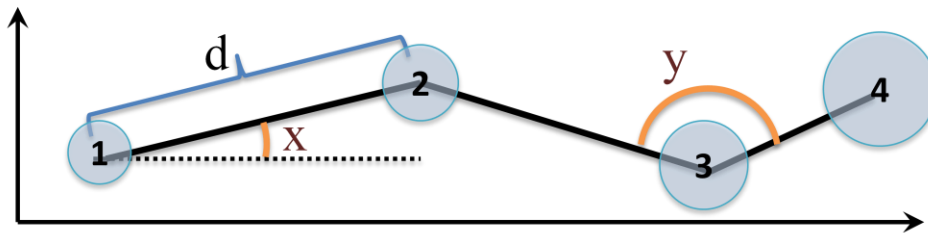


Figure 5.1: Saccade based eye measures

Among the measures described in Table 5.1, fixation rate, number of fixations, and fixation duration are widely used. In addition, saccade length, relative saccade angle, and absolute saccade angle are included, as suggested by Goldberg and Helfman [2010] because these measures are useful to summarize trends in user attention patterns within a specific interaction window (e.g., if the user's gaze seems to follow a planned sequence as opposed to being scattered). In order to extract individual eye tracking features, the raw gaze data from the Tobii eye tracker was processed using customized Python scripts⁷.

5.1.1 Complete set of eye tracking features computed

The total set of gaze features used for this analysis was obtained by computing statistics such as sum, average, and standard deviation over the measures shown in Table 5.1, at two levels of granularity. At the *Task Level*, features are computed over each task as a whole (see Table 5.2). At the *AOI level*, features are computed based on gaze activity within a specific region of the screen, or *Area Of Interest* (see Table 5.3). Included at the AOI level are transitions between pairs of defined AOIs (five in this

⁷ This programming work was done by Nicholas FitzGerald at the Department of Computer Science in the LCI laboratory under the ATUAV project.

analysis, described in the next sub-section). In order to keep the analysis at a reasonable number of features at the AOI level, only proportionate features were calculated and did not include features related to path angles (note that each added AOI feature actually increases the total number of AOI features by a factor of 5 since there are 5 AOIs). In total, there were 49 different features computed for this analysis (14 Task-level and 35 AOI-level).

Task level measures
Overall Fixation rate
Total Number of Fixations
Sum, Mean, and Std. Dev. of Fixation Durations
Sum, Mean, and Std. Dev. of Saccade Length
Sum, Mean, and Std. Dev. of Relative Saccade Angles
Sum, Mean, and Std. Dev. of Absolute Saccade Angles

Table 5.2: Task level eye tracking features generated for analysis

AOI level measures
Proportion of Fixation Durations
Proportion Total Number of Fixations
Number of Transitions from High AOI to each other AOI (5 separate measures)
Number of Transitions from Low AOI to each other AOI (5 separate measures)
Number of Transitions from Labels AOI to each other AOI (5 separate measures)
Number of Transitions from Question AOI to each other AOI (5 separate measures)
Number of Transitions from Legend AOI to each other AOI (5 separate measures)

Table 5.3: AOI level eye tracking features generated for analysis

5.1.2 Areas of interest (AOI)

A total of five AOIs were defined for each of the two visualizations. These regions were selected in order to capture the distinctive and typical components of these two information visualizations. Figure 5.2 and Figure 5.3 show how these AOIs map onto the bar graph and radar graph visualizations. The selection of these five AOIs is the result of a trade-off between having detailed information on user attention by measuring very specific areas that are salient for task execution, versus keeping the number of AOIs manageable for data interpretation and analysis. For example, initially the *High* and *Low* AOIs (described next), each consisted of 8 separate regions (i.e., one for each course), which resulted in 16 more AOI's instead of just 2. We did this because for transition analysis, which looks at how often a user transitioned from a given AOI to any

other AOI, we wanted to have very detailed transition features, but unfortunately the number of transitions to be included grows n-squared for each AOI we add. Therefore we opted to include fewer AOIs while still maintaining a granularity that captured distinctive and meaningful areas of an information visualization. The following is a description each of the five AOIs we chose to include:

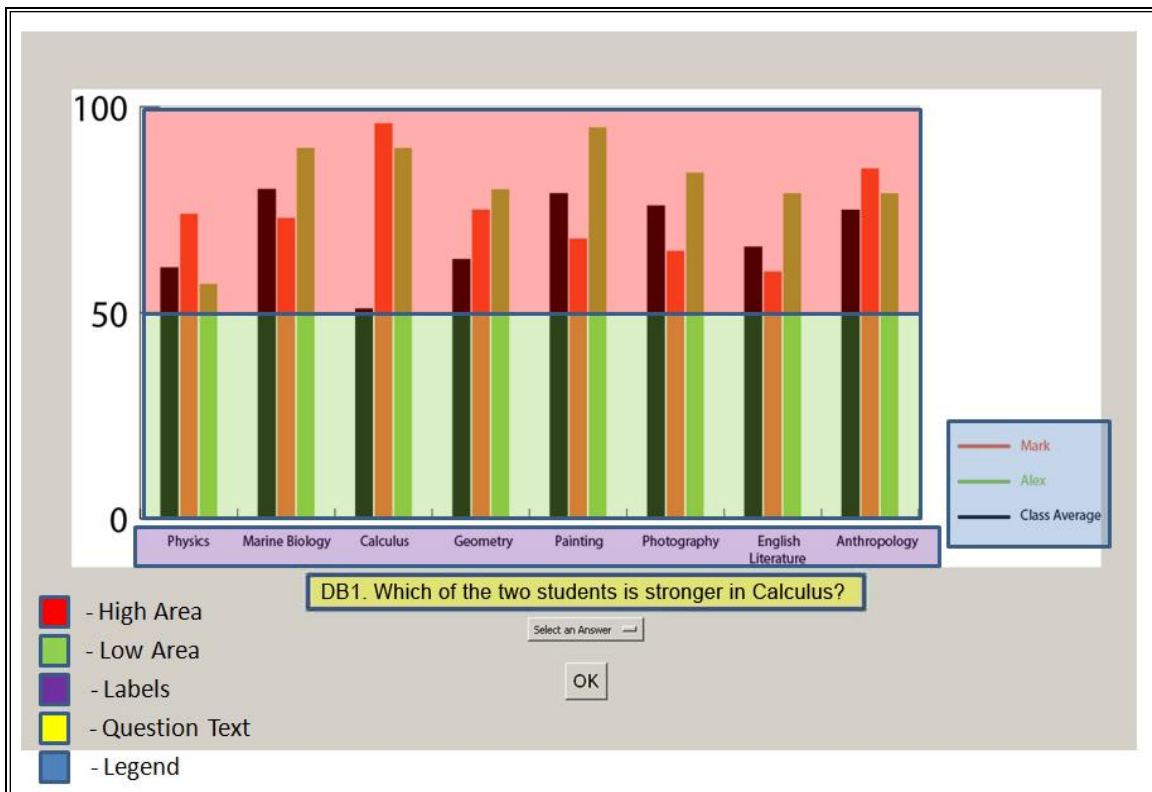


Figure 5.2: The five AOI regions defined over bar graph

- **High Area:** covers the upper half of the data elements of each visualization. This area is the graphical portion of the information visualization that contains the relevant data values. On the bar graph, it corresponds to a rectangle over the top half of the vertical bars (region between 50 and 100, see Figure 5.2). For the radar graph, it corresponds to the combined area of the 8 trapezoidal regions covering the data lines between 50 and 100 (see Figure 5.3).
- **Low Area:** is the graphical portion of the information visualization that contains the least useful data values. For our analysis, it covers the lower half of the data elements for each visualization (area between 0 and 50).

- **Labels Area:** covers all the data labels in each graph. For the bar graph, this AOI corresponds to the rectangle covering the labels just below the graph (see Figure 5.2). For the radar graph, it corresponds to the combined areas of the individual rectangles each covering a different label around the outside of the graph circle (see Figure 5.3).
- **Question Text Area:** covers the question text describing the task to be performed.
- **Legend Area:** covers the legend showing the mapping between each student and the color of the visualization elements that represent that student's performance in the visualization.

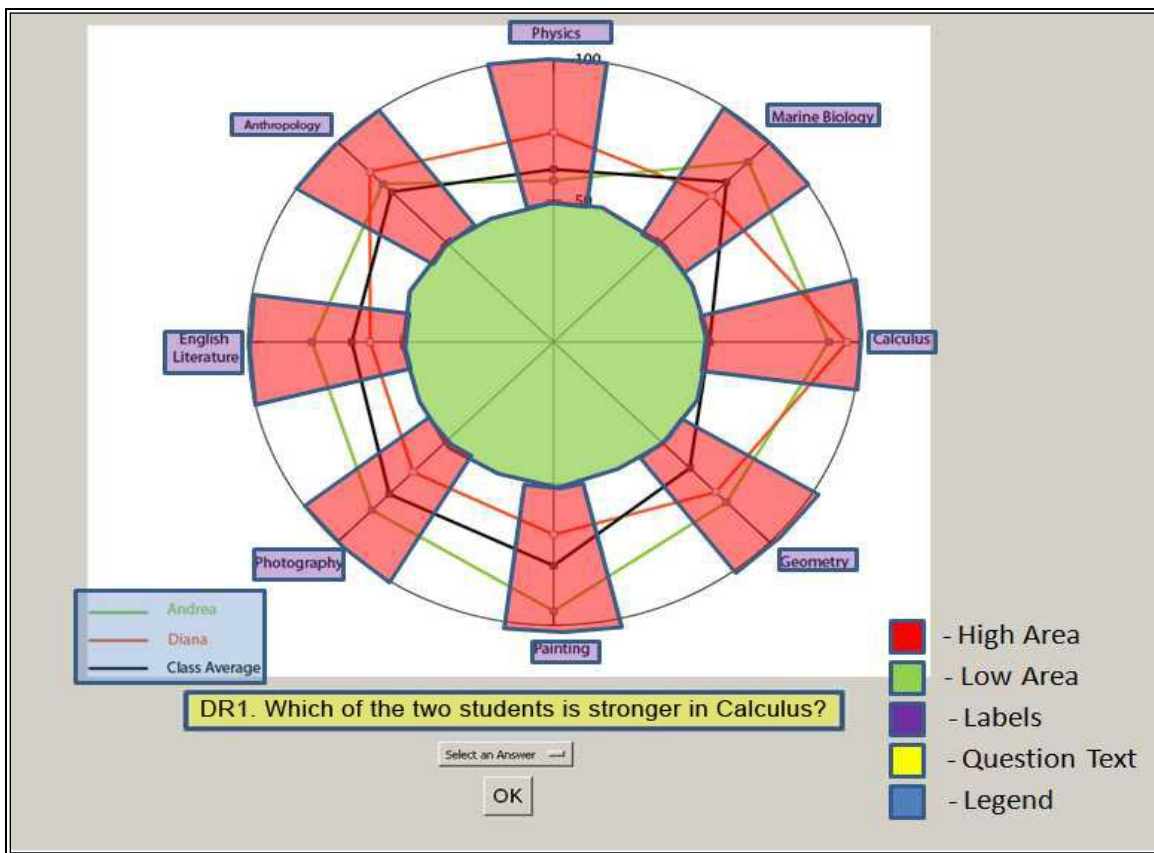


Figure 5.3: Five AOI regions defined over radar graph

5.1.3 Eye tracking feature reduction: principal component analysis

To account for possible correlations among eye tracking features, three principal component analyses (PCA) are performed on the set of 49 gaze features. A principal component analysis is a form of dimensional reduction that allows one to reduce a number of variables into potentially fewer components. This is useful because it allows one to both identify and combine groups of inter-related variables in order to identify their relationship with one-another, and to simplify the data by reducing the number of overall variables [Field 2009]. The 49 gaze features are grouped into three non-overlapping *families* according to how the measures are intuitively related, namely **(i)** task-level features (e.g., *fixation-rate*), **(ii)** AOI transitions (e.g., *number of transitions* from one AOI to another), and **(iii)** AOI proportionate features (e.g., *proportion of fixation durations* in a given AOI). One principal component analysis was performed for each of these three families and the resulting components will allow us to discuss any subsequent findings in terms of high-level related gaze components rather than many low-level features.

5.1.4 Task level family PCA

The task level family consisted of 14 gaze features. Five components were generated⁸ using PCA ($\chi^2 = 15434.49$, $df = 91$, $p < .001$), explaining 86.69% of the variance. Table 5.4 shows the breakdown of the original 14 features into the five components. The table also shows the coefficient values for each measure which gives some insight into the weighting of each measure within its respective component. Note that, for most components, it is quite easy to identify intuitive commonalities among its features, as reflected in the components' names (e.g., all features of component 1 are based on sums, all features in component 2 relate to fixation measures, etc.). The same is true for the components resulting from PCA on the other two families of features.

Measure	Coefficient	Component	Name
num_fixations	.995	1	Sum Measures
sumrelpathangles	.986	1	
sumabspathangles	.985	1	
sumpathdistance	.944	1	
sumfixationduration	.929	1	
meanfixationduration	.975	2	Fixation Rate
fixation_rate **	-.914	2	
stddevfixationduration	.900	2	
meanpathdistance	.919	3	Path Dist.
stddevpathdistance	.919	3	
meanrelpathangles	.790	4	Std.Dev. Path Angles
stddevrelpathangles	.704	4	
stddevabspathangles	.695	4	
meanabspathangles	.941	5	Abs. Path Mean

Table 5.4: Generated components for the task-level family. **fixation-rate is the only measure that inversely correlates with the other members of its component

⁸ The number of components generated is either determined by using the Catell scree test, or if this test results in an ambiguous scree plot, then Kaiser's criterion is checked to select only components with eigenvalues greater than 1 [Field 2009].

5.1.5 AOI transitions family PCA

The AOI transitions family consisted of 25 features and the PCA generated four components (see Table 5.5), ($\chi^2 = 8506.86$, $df = 45$, $p < .001$), which explained 45.24% of the variance. The PCA proved to be especially useful for reducing the many AOI transition features down to a small set of meaningful components, each including features mostly related to a specific AOI.

Measure	Coefficient	Component	Name
legend_to_legend	.781	1	Legend Transitions
legend_to_high	.768	1	
high_to_legend	.763	1	
text_to_legend	.690	1	
legend_to_text	.608	1	
legend_to_low	.606	1	
text_to_text	.595	1	
low_to_legend	.467	1	
legend_to_labels	.434	1	
labels_to_legend	.406	1	
high_to_low	.813	2	Low Trans.
low_to_high	.805	2	
text_to_low	.536	2	
low_to_low	.460	2	
low_to_text	.421	2	
labels_to_labels	.716	3	Label Trans.
labels_to_low	.697	3	
low_to_labels	.635	3	
text_to_labels	.590	3	
labels_to_text	.446	3	
high_to_labels	.776	4	High Trans.
labels_to_high	.737	4	
high_to_text	.462	4	
high_to_high	.409	4	

Table 5.5: Components generated for the AOI transitions family

5.1.6 AOI proportionate family PCA

The AOI proportionate family consisted of 10 gaze features. Five components (see Table 5.6) were produced from the PCA ($\chi^2 = 5706.32$, $df = 300$, $p < .001$) and explained 97.13% of the variance. The components generated indicate that both the amount of time and the number of fixations in each AOI is correlated, and as such the features have been grouped accordingly into 5 pairs, one for each AOI.

Measure	Coefficient	Component	Name
low_prop_num_fix	.981	1	Low
low_prop_time	.980	1	Prop
labels_prop_num_fix	.976	2	Label
labels_prop_time	.972	2	Prop
legend_prop_num_fix	.973	3	Legend
legend_prop_time	.968	3	Prop
text_prop_time	.975	4	Text
text_prop_num_fix	.975	4	Prop
high_prop_num_fix	.953	5	High
high_prop_time	.945	5	Prop

Table 5.6: Generated components for AOI proportionate measures

5.2 Task Difficulty

We are also interested in looking at effects that task difficulty has when user characteristics are taken into consideration. Defining tasks as being easy or difficult a priori is challenging, since difficulty depends upon user expertise and perceptual abilities, which were varied on purpose in this study. Therefore, we define task difficulty a posteriori, based on four different measures (two objective and two subjective) that are then aggregated using a principal component analysis [Field 2009].

5.2.1 Four indicators of task difficulty

Because there was a ceiling effect on task correctness, the first objective measure of task difficulty is *task completion time* (assuming that in general, more time is needed for more difficult tasks). However, longer completion times may also simply be an indication of a task requiring more time while not necessarily being more difficult. Therefore the second objective measure of difficulty is the *standard deviation of completion time* for each task, across all users. A high value of this metric indicates a high variability among completion times, meaning that the task may be difficult or confusing for some users.

The two chosen subjective measures of task difficulty are based on the users' reported confidence of their performance, which was elicited after each task. The first subjective measure is the *average confidence* reported by users on each task. Intuitively, less difficult tasks would have higher values for this average. However, we also want to take into account that some users may tend to be more confident overall than other users. Therefore, the second subjective measure is the *average deviation of confidence* for each task across all users, and is computed as follows. For each user, the average confidence across all of their tasks is used. Then, for each task, the *deviation of confidence* is computed as the difference between the user's reported confidence for that task and that user's average confidence across all tasks. Finally, for each task, the deviation of confidence across all users is averaged. This average indicates for which tasks users were giving confidence ratings that were above or below their typical input.

5.2.2 Principal component analysis of task difficulty

The four measures of task difficulty described in the previous subsection are each used as input to four principal component analyses respectively. Bartlett's test of sphericity ($\chi^2 = 73.35$, $df = 6$, $p < .001$), indicated that the principal component analysis was appropriate. Kaiser's sampling adequacy was 0.55 and all variables showed a communality > 0.52 which is above the acceptable limit of 0.5 [Field 2009]. One component was generated, and had an eigenvalue above Kaiser's criterion of 1, and explained 62.22% of the variance. Our measure of task difficulty is therefore the output component that is generated by this PCA. This measure will be used as an independent factor in the proceeding statistical analyses. Table 5.8 shows a detailed breakdown of the experiment tasks, ordered by task difficulty. What is interesting is that many of the hardest tasks have uniform distributions, whereas the easier tasks have spiky distributions. Even though no significant results were obtained in Chapter 4 regarding distribution type, this table does shed some light on the fact that uniform data distributions tend to be more difficult, likely due to the differences in adjacent data values being less pronounced, and therefore requires more effort to discern.

Measure	Coefficient	Component	Name
task_completion_time	0.693	1	Task Difficulty
standard_deviation_completion_time	0.550	1	
average_confidence	-0.905	1	
average_deviation_from_confidence	0.941	1	

Table 5.7: Resulting coefficient values of the four measures that were used in principal component analysis for task difficulty

Task Type (Amar)	Visualization Type	Task Scenario	Distribution Type	Task Difficulty Component Value
Retrieve Value	Bar	single scenario	spiky	-1.255
Sort	Bar	single scenario	spiky	-1.147
Retrieve Value	Bar	double scenario	n/a	-1.069
Find Extremum	Bar	single scenario	spiky	-0.96
Retrieve Value	Radar	double scenario	n/a	-0.793
Filter	Bar	single scenario	spiky	-0.775
Filter	Bar	single scenario	uniform	-0.718
Sort	Bar	single scenario	uniform	-0.641
Filter	Bar	double scenario	n/a	-0.634
Retrieve Value	Radar	single scenario	spiky	-0.537
Sort	Radar	single scenario	spiky	-0.441
Compute Derived Value	Bar	single scenario	spiky	-0.438
Retrieve Value	Radar	single scenario	uniform	-0.436
Compute Derived Value	Radar	single scenario	spiky	-0.361
Compute Derived Value	Bar	single scenario	uniform	-0.322
Compute Derived Value	Radar	single scenario	uniform	-0.237
Filter	Radar	single scenario	uniform	-0.189
Filter	Radar	single scenario	spiky	-0.005
Find Extremum	Radar	single scenario	spiky	0.01
Filter	Radar	double scenario	n/a	0.216
Compute Derived Value	Radar	double scenario	n/a	0.301
Compute Derived Value	Bar	double scenario	n/a	0.654
Filter	Radar	double scenario	n/a	0.872
Find Extremum	Radar	single scenario	uniform	1.05
Filter	Radar	double scenario	n/a	1.643
Sort	Radar	single scenario	uniform	1.725
Retrieve Value	Radar	single scenario	uniform	1.989
Find Extremum	Bar	single scenario	uniform	2.497

Table 5.8: All 28 task combinations from the study, ordered from least difficult to most difficult according to our definition of task difficulty.

5.3 Mixed Model Analysis

Since the study data involved repeated measures (e.g., each subject performed the same task type with each of the two different visualizations), a suitable means for analysis is a *Mixed Model* [7]. Mixed models can handle both repeated measures as well as the mix of categorical and continuous independent measures, which are both present in this study. An alternative model commonly used for repeated-measures analysis is a General Linear Model Repeated Measures analysis (GLM for short) [Field 2009]. GLM, however, is less suitable than a Mixed Model for eye tracking analysis, because it is less resilient to missing data. This issue is due to the fact that GLM requires data to be in *wide format* (Figure 5.4-left), where all repeated measures (trials) for each participant are listed in one data entry row. When there is an invalid trial, GLM is forced to discard the entire data for that participant. This can be costly in an experiment with several invalid trials, as is often the case when using unobtrusive eye trackers that do not constrain subjects' movements. By contrast, a Mixed Model uses data in *long format* (Figure 5.4-right), listing each trial as a different data entry, and discarded invalid trials do not interfere with valid ones. Thus, a Mixed Model analysis is able to leverage at best potentially noisy eye tracking data.

User	Trial A	Trial B	Trial C
P1	1.2	1.9	1.3
P2	2.3	miss	2.4

User	Trial	Score
P1	A	1.2
P1	B	1.9
P1	C	1.3
P2	A	2.3
P2	C	2.4

Figure 5.4: Example of experimental data in wide format for GLM (left), and in long format for mixed model (right).

5.3.1 Mixed model setup

For each of the three families of gaze features (i.e., *Task Level*, *AOI Proportionate*, and *AOI transition*) a mixed model is ran over each of the generated PCA components within that family. Each mixed model is a 2 (visualization type) by 2 (visualization order) model, with the user characteristics and task difficulty as the model's covariates. Mixed Models are univariate analyses (ANCOVA) and do not support multivariate analysis, i.e., having more than one depended measure per model which is why we run a single mixed model for each dependent measure. The models must therefore be adjusted for family-wise error, (i.e., the problem that the more dependent measures that are tested, the higher the chance of finding something significant), by applying the Bonferroni adjustment to each family of results. Since there are three families of eye tracking measures, each model will be adjusted by applying the Bonferroni correction according to the number of components within that family. We report statistical significance at the 0.05 level, and effect sizes are reported as small for $r = 0.1$, medium for $r = 0.3$, and large as $r = 0.5$ [Field & Hole 2003]. The next section describes the most salient results of the analysis. When going over the results involving directionality, keep in mind that the dependent measures are PCA components, each consisting of a single value that represents a much larger collection of underlying measures. Each component is generated by (i) calculating the weighted values of its underlying members; (ii) aggregating and scaling these values into one number typically ranging from -1 to +1. If an underlying member is positively correlated to its corresponding component the directionality will be the same, otherwise it will be opposite.

5.4 Mixed Model Results & Discussion - User Characteristics

In this section, results are presented to address the research question: which individual user characteristics influence a user's eye gaze behavior? And is the effect modulated by visualization type and task difficulty? The results that follow are discussed per user characteristic.

5.4.1 Perceptual speed - main effects

There was a main effects of perceptual speed on three PCA components (see Table 5.9).

Family	Component	F-Ratio	Effect Size	Sig. Value
Task level	<i>Fixation Measures</i>	$F(1,27) = 8.9$	$r = 0.37$	$p = 0.03$
AOI	<i>Legend Proportion</i>	$F(1,21) = 25.2$	$r = 0.21$	$p < 0.001$
	<i>Legend Transitions</i>	$F(1,26) = 10.25$	$r = 0.16$	$p = 0.016$

Table 5.9: Main effects of perceptual speed

One main effect of perceptual speed was at the task level (first row in Table 5.9), showing that high perceptual speed users had lower values of *Fixation Measures* than low perceptual speed users. An analysis of the underlying members of this component further showed that users with high perceptual speed had a higher fixation-rate than low perceptual speed users, indicating that they were able to scan the screen more quickly. High perceptual speed users also had lower average and lower standard deviation of fixation durations, i.e. shorter and more consistently timed fixations. These combined findings closely match the definition of perceptual speed, and are interesting because they show that individual differences for this cognitive ability may be captured via eye tracking measures that are not related to information on specific elements of the visualization.

The other two main effects of perceptual speed are at the AOI level (see bottom 2 rows of Table 5.9), showing that this cognitive ability also affects eye gaze measures relating to specific visualizations elements, namely the legend, in terms of *Legend Proportion* and *Legend Transitions*. We found that low perceptual speed users spent more of their time in the legend AOI and also transitioned to it more often compared to high perceptual speed users. This result indicates that users with low perceptual speed are taking more time to process/store legend-related information and are looking at the legend more frequently (possibly because they tended to forget the mapping of the information contained in the legend).

5.4.2 Perceptual speed - interaction with task difficulty

We found two significant interactions of task difficulty and perceptual speed; one on the *Legend Transitions* and one on the *Label Transitions* component (see Table 5.10).

Family	Component	F-Ratio	Effect Size	Sig. Value
AOI	<i>Legend Transitions</i>	F(1,686)=6.85	r = 0.10	p < 0.05
	<i>Label Transitions</i>	F(1,676)=7.97	r = 0.11	p = 0.02

Table 5.10: Interaction effects for perceptual speed and task difficulty

For *Legend Transitions*, all users generated more legend-related transitions with difficult tasks than with easy tasks (see Figure 5.5), likely due to the fact that an increased difficulty increases cognitive load and causes users to more easily forget some of the information in the legend. This effect, however, is higher for users with low perceptual speed.

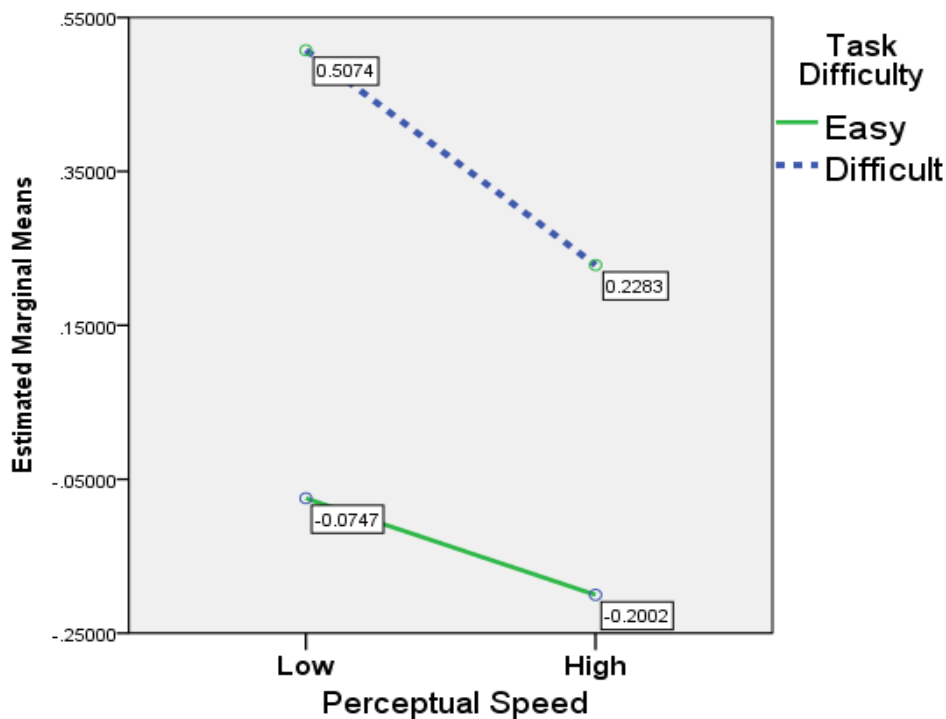


Figure 5.5: Interaction between perceptual speed and task difficulty on AOI legend transitions

As for the interaction between perceptual speed and task difficulty on *Label Transitions*, Figure 5.6 illustrates this finding. What we found is that all users have more label-related transitions for easy tasks compared to difficult tasks, but the difference is much higher for low perceptual speed users. This effect is not as intuitive as the previous one found on *Legend Transitions*, but whatever it is that causes users to have more label-related transitions for easy tasks, it seems to be affecting low perceptual speed users the most.

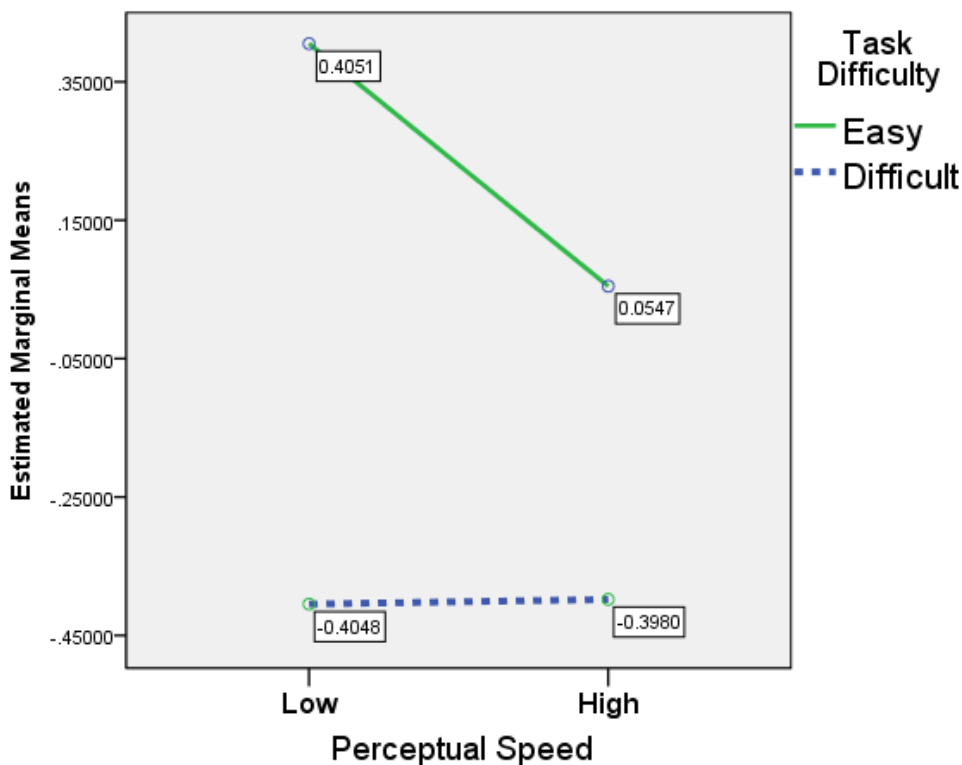


Figure 5.6: Interaction between perceptual speed and task difficulty on AOI *label transitions*

5.4.3 Perceptual speed - interaction with visualization type

There was one significant interaction effect between perceptual speed and visualization type in terms of the AOI *High Transitions* component, $F(1,680)=22.2$, $r=0.18$, $p < 0.001$, (see Figure 5.7). All users showed more High AOI related transitions with the radar graph than with the bar graphs, but the difference is much greater for low perceptual speed users. Given that the High AOI is the graphical portion of the information visualization that contains the relevant data values, this effect suggests that low-perceptual speed users are more affected than high perceptual speed users by alternative forms of visualizing data.

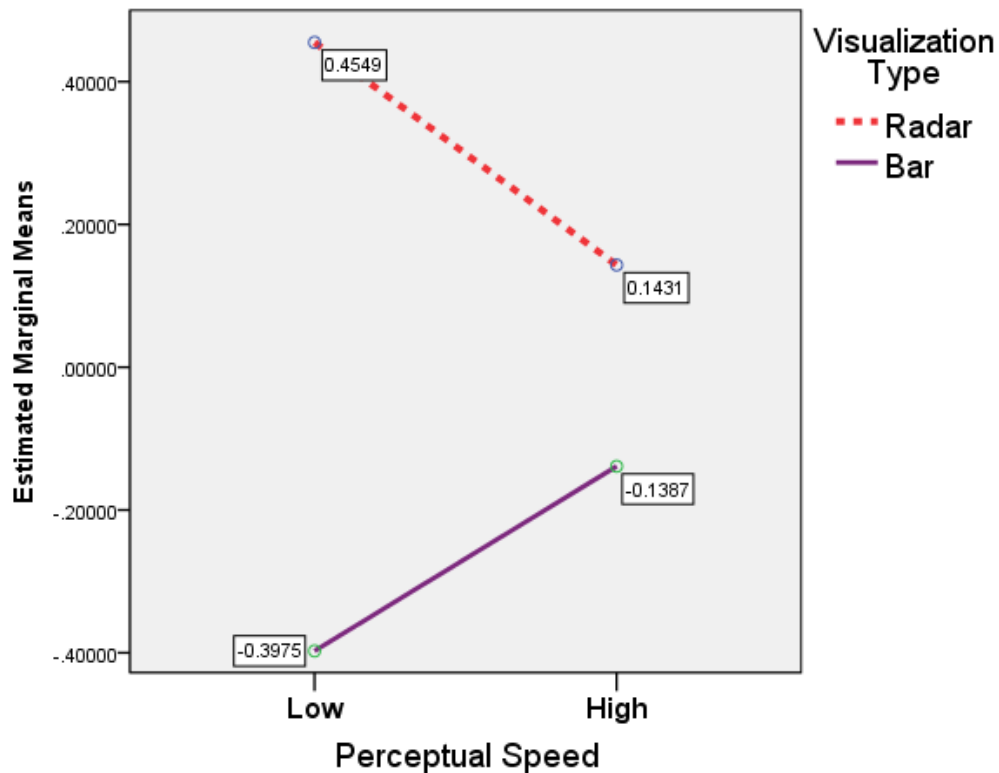


Figure 5.7: Interaction between perceptual speed and visualization type for AOI *high transitions*

5.4.4 Verbal working memory - main effects

We found two main effects of verbal working memory: one on the *Text Proportion* component and one on *Standard Deviation of Path Angles* (see Table 5.11).

Family	Component	F-Ratio	Effect Size	Sig. Value
AOI	<i>Text Proportion</i>	F(1,28) = 7.24	r = 0.36	p = 0.04
Task Level	<i>Std.Dev. PathAngles</i>	F(1,25) = 8.06	r = 0.32	P < 0.05

Table 5.11: Main effects of verbal working memory

The *Text AOI* relates to the most textual element of both visualizations, namely the question text. An analysis of the members of the Text Proportion component shows that the proportionate amount of time users spent looking at the text and the number of fixations within this AOI is significantly lower for users with high verbal working memory. This effect indicates that high verbal working memory users refer to the task question less often than their low verbal working memory counterparts, which is consistent with the definition of verbal working memory as a measure of storage and manipulation capacity of verbal information. This result is interesting because it shows that differences in users' verbal working memory can be directly captured by eye tracking features related to the primary textual element of a visualization.

The second main effect of verbal working memory is on the *Standard Deviation of Path Angles*. This component essentially captures the consistency of a user's gaze patterns during a visualization task, because it is built upon features related to measuring the deviation of angles between consecutive saccades. We found that users with low verbal working memory had higher values for *Std.Dev. Path Angles* than users with high verbal working memory. Where these values are higher, it indicates that a user is frequently looking across different areas of the screen, rather than following more planned or consistent path directions. Therefore, the finding that users with low verbal working memory had higher values for *Std.Dev. Path Angles* is consistent with the finding that low verbal working memory users referred back to the question text more often. Additionally, Goldberg and Helfman [2010] attribute a higher sum of relative angles with uncertainty of the task. This suggests that users with low verbal working memory may

be more uncertain than users with high verbal working memory. This is an interesting finding because it provides evidence for a user characteristic that can impact a user's level of uncertainty independent of specific tasks or task characteristics.

5.4.5 User expertise

We found two non-significant main effects of both bar graph and radar graph expertise, which are discussed because of their large effect sizes.

There was a main effect of Bar Graph Expertise for the *AOI Label Proportion* component, $F(1,21) = 6.042$, $r = 0.80$, $p = 0.1$, showing that users with high bar expertise spent a greater proportion of their time looking at labels compared to non-experts. We also found a main effect of Radar Graph Expertise on the *AOI Legend Proportion* component, $F(1,21) = 5.732$, $r = 0.78$, $p = 0.129$, with radar experts spending less time looking at the legend compared to non-experts. The discrepancy between strong effect sizes for these two expertise-related measures and the lack of statistical significance is most plausibly due to a lack of statistical power. The power for Bar Graph expertise on Label AOI Prop is 0.67, and the power for Radar Graph expertise on Legend AOI Prop is 0.64. The recommended value of power to aim for is 0.8 [Field 2009], and we would therefore have to add 17% (or 6) more users to reach this value.

It may seem surprising that we did not find stronger influences of visualization expertise on gaze patterns. This result however, is consistent with findings in the previous chapter, which showed that bar and radar graph expertise may only have significant effects on user visualization preference, but not on performance, provided that preference has little to do with gaze behaviour. These findings suggest that there may not be easily detectable differences in the visualization processing behaviours of experts and novices, as defined by our self-rated measures of expertise.

5.4.6 Visual working memory

There were no significant effects to report for visual working memory. This lack of findings may be due to the fact that the study tasks were relatively easy and that the visualizations were static in nature. It is thus likely that users were not required to reach their maximum visual memory capacity, especially since they could easily get an overview of the whole graph in a single look. Moreover, individual tasks were

independent of each other, thus users were not required to store any successive visual information (one of the functions affected by visual working memory).

5.5 Mixed Model Results - Task Difficulty & Visualization Type

We also found results relating exclusively to main effects of visualization type, main effects of task difficulty, and interactions between these two independent measures. These findings do not relate to user characteristics, and therefore were not part of the main focus of this analysis. Nevertheless, these results may have some useful implications for information visualization design and also the understanding of visualization processing. We have opted to include and briefly discuss these findings in Appendix A - Additional Eye Tracking Results.

5.6 Summary & Discussion

In this chapter we presented research aimed at investigating the relationship between individual user characteristics, task difficulty, and user attention patterns when using different visualization techniques. The goal of the analysis on the eye tracking data was to investigate i) if user characteristics impact gaze patterns during visualization processing, and if the impact can be detected through eye tracking; ii) which eye gaze measures are influenced by which user characteristics, and if/how is this mediated by task difficulty and visualization type. The analyses we conducted revealed that there does in fact exist a set of user characteristics that sufficiently influence user gaze behavior, which is detectable through a variety of eye tracking metrics. Furthermore, the influence is sometimes mediated by non-user characteristic factors such as visualization type and task difficulty.

5.6.1 Summary of results

Given the user characteristics we investigated, Table 5.12 shows a full summary of the reported effects that were found (either statistically significant or with large effect sizes) on various eye gaze components which serves to confirm that user characteristics do in fact impact user gaze behaviour. We will next discuss selected results pertaining to

specific user characteristics that have an impact on gaze features and how these results could inform user-adaptive information visualization systems.

User Characteristic	Eye tracking Component
Perceptual Speed	<i>Fixation Measures</i> (main effect) <i>Legend Proportion</i> (main effect) <i>Legend Transitions</i> (main & interaction effect) <i>Label Transitions</i> (interaction effect) <i>High AOI Transitions</i> (interaction effect)
Verbal WM	<i>Std. Dev. Path Angles</i> (main effect) <i>Text Proportion</i> (main effect)
Bar Expertise	<i>Label Proportion</i> ($p > 0.05$, but large effect size)
Radar Expertise	<i>Legend</i> ($p > 0.05$, but large effect size)
Visual WM	<i>None</i>

Table 5.12: Summary of reported results from analysis of eye tracking data

5.6.2 Discussion of perceptual speed

Perceptual speed is the cognitive ability with the highest number of effects. This finding provides encouraging evidence that perceptual speed is a user characteristic could be reliably detected in real time using gaze information. This result is particularly important for the long term goal of our research (i.e., informing the design user-adaptive visualizations), given that we already know that low perceptual speed can negatively affect task performance in terms of accuracy [Conati 2008] and task completion time (Chapter 4). In particular, we identified that perceptual speed influences AOI-specific gaze measures relating to the legend, labels, and High AOIs. These findings suggest that adaptive interventions could be particularly useful if they support the access and/or processing of these three AOIs for users with low perceptual speed given their lower performance. In addition, the interaction effects found for perceptual speed suggest that task difficulty and visualization type should also be taken into account, if known, when providing adaptive interventions. For instance, it was found that low perceptual speed users tended to access the visualization legend more than high perceptual speed users,

suggesting that they should be specifically supported in terms of legend processing. However, we also found that this effect is exacerbated in the presence of difficult tasks. Thus, while it may not be worthwhile disrupting a low speed user with a legend-related intervention for tasks known to be easy, it may be important to do so as task difficulty increases.

5.6.3 Discussion of verbal working memory

The results on verbal working memory indicate, intuitively, that this cognitive ability correlates with eye-tracking features related to the main textual element of a visualization, and thus may be detectable in real time by tracking these features. For our user study, the textual element was the question text, but in other settings this could be the visualization caption or the portion of text in which the visualization is embedded (e.g., possibly providing verbal descriptions of the displayed data).

In terms of user-adaptation, it is plausible that users with low verbal working memory may benefit if textual elements of a visualization were given more emphasis than the purely graphical elements. However, because in Chapter 4 no results were obtained on verbal working memory and performance during information processing, it remains a topic for future research to investigate if and how adaptive interventions would impact visualization effectiveness for users with different levels of verbal working memory.

5.6.4 Discussion of user expertise

We reported two non-significant main effects of the expertise-related user characteristics that were included because of their large effect sizes. Bar expertise had a large effect on label access, while radar expertise had a large effect on legend access. These findings may be an indicator that we need to explore adaptive interventions relating to the legend and labels in order guide non-experts by encouraging them to access these elements in a way that is more similar to expert gaze behaviour (even though we are not necessarily able to link this to performance given our set of tasks). We would therefore need to run additional studies with more reliable, objective measures of expertise (the ones used in this study were self-reported) and relate them to performance before we can make any reliable decision on how to provide adaptive support for novice users.

5.6.5 Implications for information visualization design

In summary, we have identified a set of user cognitive abilities that sufficiently impact gaze measures both in general and in relation to specific AOIs of a visualization. We discussed how adaptive interventions could be informed by certain user characteristics by targeting specific AOIs in order to influence a user's experience with a given visualization. While our analysis has only investigated two simple visualization techniques, several of our results may be generalized to a wider array of information visualization designs because they involve components that are common to most types of information visualizations (i.e., graph labels or legend). Additionally, these findings may further generalize since many of our results are effects that are independent of a specific information visualization type.

Chapter 6 Task Difficulty - Follow Up Analysis

The purpose of this chapter is to present results from a short follow up analysis done in order to include task difficulty in the analysis of the impact of user characteristics on user performance and preference. At the time that the analysis of the impact of user characteristics on user performance and preference measures (Chapter 4) was done, task difficulty had not yet been considered. It was during the design of the analysis for the eye tracking data (Chapter 5) when the idea to define and test task difficulty arose. Since there were significant results (both main effects and interactions with user characteristics) relating to task difficulty for the eye tracking data, we decided to see if task difficulty would also produce any significant results with respect to user performance.

It should be noted that it is not possible to analyze the effect of task difficulty on the subjective measures (user preference and ease-of-use). The reason for this is that the measure of task difficulty is defined for each task (repeated measures), whereas the subjective measures are given for the experiment as a whole. It is therefore not possible to run a statistical model that investigates task specific independent measures when the dependent measure is not task specific.

6.1 Model for Analysis

The goal of this analysis is to test whether task difficulty (see section 5.2 for an explanation of how task difficulty is derived) has any impact on user performance, and if this effect is mediated by user characteristics. In order to do this, a 2 (visualization type) by 2 (visualization order) mixed model was used, with the user characteristics and task difficulty as the model's covariates. As usual, significance is reported at the 0.05 level.

6.2 Results and Discussion

There were two results of task difficulty on user performance as follows.

6.2.1 Main effect of task difficulty

There is a main effect of task difficulty on task completion time, $F(1,883) = 9.389$, $r = 0.09$, $p < 0.01$. This result shows that more difficult tasks take longer to complete. This is not too surprising, considering that task difficulty is a component generated measure that is partly made-up of completion time. The fact that this result was found is a good sanity-check.

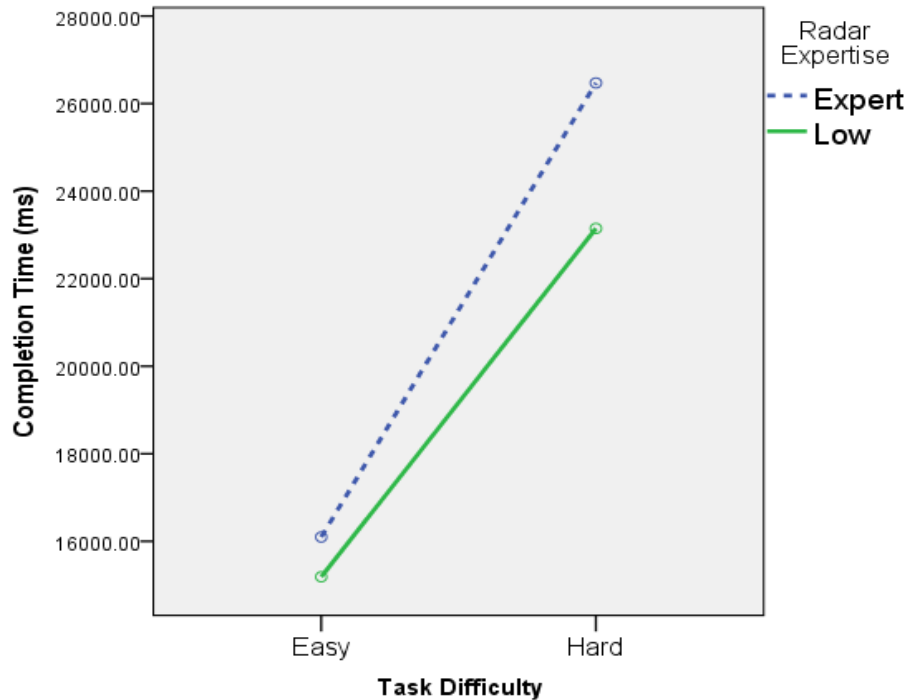


Figure 6.1: Shows the interaction effect that radar expertise and task difficulty has on task completion time

6.2.2 Interaction effect - task difficulty & radar expertise

There is a significant interaction effect between task difficulty and radar expertise on task completion time, $F(1,883) = 7.66$, $r = 0.06$, $p < 0.01$. As Figure 6.1 shows, radar experts are always taking longer to finish each task than users with low expertise. What this result suggests is that experts are perhaps more careful when providing answers since they are more concerned with not making a mistake, and they are even more careful as task difficulty increases. This explanation would be consistent with Lewandowsky & Spence [1989], where they found that domain experts completed a set of information visualization tasks more accurately but more slowly than non-experts.

However, given that there was a ceiling effect on task accuracy in our study, we are unable to confirm this conclusion without conducting a future experiment with more complex or difficult tasks.

6.3 Discussion

The findings in the chapter show that there is in fact a significant correlation of task difficulty with completion time, which is an intuitive result given that a more difficult task should generally take more time to complete. The more interesting finding though is that this effect is mediated by user visualization (radar) expertise.

As a further point of discussion, recall that we previously found an interaction effect of task difficulty and perceptual speed on *legend transitions* (Section 5.4.2). Given that perceptual speed (see Table 4.3) and now task difficulty have both been shown to relate significantly to task completion time, it is likely the case that increased legend transitions (see Figure 5.5) is a direct indicator of longer task time. This effect would be further mediated by perceptual speed (i.e., low perceptual speed users are transitioning even more and taking even longer to finish their tasks) and also task difficulty (i.e., users generate more legend transitions and take longer to complete more difficult tasks). A correlation between *legend transitions* and *task completion time* confirms this argument ($r = 0.455$, $p < 0.001$), showing that there is a significant correlation with a large effect size.

It was also found that perceptual speed and task difficulty impact *label transitions* (see Figure 5.6), and we can make a similar argument as the one just made for legend transitions, only in this case increased label transitions is an indicator of shorter task completion time and this effect is similarly mediated by both perceptual speed and task difficulty. Although the effect size is not as large, a correlation between *label transitions* and *task completion time* also confirms this argument ($r = -0.088$, $p < 0.01$), showing that there is a significant correlation with a small effect size.

Chapter 7 Conclusion and Future Work

We presented and discussed a user study that investigates the impact of five different user characteristics (perceptual speed, visual/verbal working memory, radar/bar graph expertise) on the effectiveness of two common data visualization techniques: bar graph and radar graph. We also investigated how these factors influence a user's gaze behavior during the visualization tasks. The results of in this study confirm preliminary existing findings that individual differences do make a difference in visualization effectiveness, and thus should be taken into account in selecting suitable visualization support for each specific viewer. Furthermore, we have shown that user characteristics are detectable using eye tracking, suggesting that this technology could be used as a source for detecting user characteristics in real time. All of these findings offer insights into the design of user-adaptive intervention visualization systems.

7.1 Satisfaction of Thesis Goals

7.1.1 Impact of user characteristics on visualization performance & preference

This objective entailed investigating the impact of user characteristics on task completion time, task accuracy, and user preferences for the two given information visualizations examined in this study. We employed the general linear model and multivariate statistical analyses in order to answer the following research questions:

i) Which user characteristics influence (i.e., correlate with) the performance and preference of users while using bar graphs and radar graphs?

Results confirmed that perceptual speed is a user characteristic that correlates with completion time. Users with higher perceptual speed were significantly faster than those with lower perceptual speed. We also found results relating user characteristics to user visualization preference. Users with higher visual working memory gave higher preference ratings to radar graphs, as did users with higher radar graph expertise. These results are interesting because they show that while certain users characteristics

(i.e., perceptual speed) can influence task performance, other user differences such as expertise and visual working memory abilities impact user preferences.

ii) If there is an effect of user characteristics, how is this effect influenced by visualization type?

We also found that perceptual speed was mediated by visualization type with respect to its impact on task completion time. Users with low perceptual speed performed even more poorly with radar graphs and this supports the conclusion that perceptual speed is a cognitive ability that can impact the effectiveness of alternative visualizations.

7.1.2 Additional findings: impact of visualization type on task completion time

Even though it was not the primary focus of our investigation, we did find an impact of visualization type on task completion time. Our analysis indicated that users were faster on bar graphs for single scenario tasks, but this effect was not detected on the double scenario tasks. For single scenario tasks, questions involved one student and the class average, whereas double scenario tasks involved answering questions for two students and the class average. What this means is that visualization effectiveness can depend a great deal on the complexity or amount of amount of data presented. As data complexity is increased, certain visualizations such as radar graphs, can become just as effective as others.

In terms of preferences, we found that more users gave a higher preference rating to bar graphs, whereas user preference for radar graphs was more evenly distributed across all preference values. The implication of this result is that it is important to consider the intended purpose for a given information visualization. Is it more important to design a system that users prefer, or one where they perform faster, or a balance of the two? Knowing that these questions are mediated by user characteristics can inform better design choices.

7.1.3 Impact of user characteristics on eye tracking measures

This objective aimed to investigate the impact of user characteristics on user gaze behavior captured via eye tracking. Using principal component analyses and mixed models, we are able to answer the following two questions that were posed at the outset:

i) Do individual user characteristics influence (i.e., correlate with) a user's eye gaze behavior in a way that is detectable by eye tracking?

User characteristics do in fact influence user gaze behaviour, and many examples of correlations were detected using eye tracking. Some of the most interesting results relate to perceptual speed, and verbal working memory which are discussed next.

ii) Which gaze features are influenced by which particular user characteristics? Is the effect modulated by task context (e.g., task difficulty or visualization type)?

Several instances of user characteristics influencing gaze features were found. For example, perceptual speed was linked to eye gaze features such as fixation rate, and legend use. These are both interesting findings because they give promising insight into how future systems could begin to detect user characteristics since we have identified eye tracking features that are potential candidates to monitor in real time. Additionally, because we also found that perceptual speed correlates with task completion time, features such as increased legend use or lower fixation-rate could also be monitored in real-time as indicators of poor performance. We also found that verbal working memory was linked to the amount of time users spent reading the task question. This too is an important finding since it shows that different user characteristics influence different elements of visualization processing, a finding that provides better understanding for both visualization design and user-adaptive interventions, especially within a fixed visualization type.

Additionally, task difficulty and visualization type both had an interaction effect with perceptual speed for user gaze features relating to bar/radar graph legend and labels.

These results highlight that it is important to consider not only the impact of user characteristics, but also how task context can impact different users in conjunction with their individual differences.

7.1.4 Additional contributions - mixed model analysis & task difficulty

An additional contribution of this thesis is showing that a mixed model analysis can be used as a suitable method for analyzing potentially noisy eye tracking data. Our analysis generated detailed results, linking a given set of user characteristics and factors to numerous eye tracking features. We also provide a novel way of defining task difficulty a posteriori, based on the aggregation of objective performance measures and subjective user confidence. Task difficulty also generates significant results, indicating that in addition to user characteristics, task difficulty can also impact both completion time and user gaze behaviour.

7.2 Future Work

There are several interesting ways in which this work could be extended.

7.2.1 Run a similar study with more difficult tasks

Possible future work would involve running a similar study but with more complex information visualization tasks. This could be achieved by some combination of harder questions, increasing the amount of data being visualized, and showing other types of visualizations beyond just bar and radar graphs. Our hypothesis is twofold: first, we would expect the impact of individual differences on time-based performance to be even more pronounced than what we found in this study; and second, because of the complexity of the associated tasks, we would expect to see effects on task accuracy in addition to just completion time

7.2.2 Using eye tracking to infer user characteristics

In order to apply our results to a user-adaptive information visualization system, the system must be able to acquire a model of the user characteristics. The next step would be to investigate how relevant user characteristics could be detected and classified in

real-time, in order to directly drive adaptive interventions. To this end, a future direction of research would be to investigate various machine learning techniques in order to infer user characteristics based on that user's eye tracking data. Some promising results have already been shown by Steichen et al. [2012].

7.2.3 Adaptive interventions for a fixed visualization type

Suitable adaptation strategies must also be devised in order for a system to know when to adapt and what sort of adaptation to deliver. We are currently designing an experiment that will test the impact of overlay type interventions, which are interventions that do not change the ordering or configuration of the visualization. These interventions will be in the form of visual cues, such as highlighting or adding arrows, with the intention of addressing the following questions. Which overlay interventions are best suited for which type of task? When is the best time to intervene? How are these interventions mediated by user characteristics?

7.3 Lessons Learned - Design Decisions For Future Studies

The goal of this section is to highlight the 'lessons learned' regarding the experimental study design presented in this thesis. We present several design decisions that we would change if we had the ability to go back and re-design our study. These post hoc insights are intended to inform both other studies in general if applicable, as well as our own future studies relating to eye tracking, adaptive interventions, and individual differences in information visualization.

Study design decisions worth changing:

- Carefully and concisely document the rationale behind all design choices as soon as each choice is made, in order to have a clear reference for later reporting & publications.
- Perform a power analysis a priori in order to estimate the minimum number of required participants necessary to achieve the desired power for a given effect size threshold. If the effect size of the experimental treatment is unknown (as it

was the case in this experiment), then a power analysis can be performed a priori for low, medium, and large effect sizes to ascertain how many participants would be needed for the desired power at each of these effect sizes. For example, if only a few extra participants are needed to obtain optimal experimental power for both large and small effects then it may be advantageous to simply add the extra participants. Alternatively, if many more participants are needed to achieve desired power for low effect sizes, then it may be advantageous not to run extra participants beyond the lowest amount needed since there would be little gain for lower effect sizes.

- In the questionnaire, when asking users to rate preferences for individual items (i.e., visualization types), ensure to include additional questions that require participants to rank the items against each other, either by rating how much they prefer one item over the other using a Likert scale, or by asking them to place the items in order of preference (which is ideal if there are more than two items to compare).
- When measuring expertise via a questionnaire, ensure to administer this prior to the user performing the main portion of the study. Furthermore, instead of (or possibly in addition to) measuring subjective self-reported expertise, be sure to ask more quantitative type questions regarding expertise, (i.e., units of time per month, per week, per day, etc.).
- When requiring input from an answer box, ensure that all input methods are the same since we want to define meaningful AOIs. For example, having all the input boxes as radio boxes (as opposed to some being drop down menus) is better for eye tracking analysis because we can then define a static region over this area.
- Avoid potential confounds resulting from grapheme placement. In our study, the legend was on the right for bar graphs, and on the left for radar graphs. This is not ideal since results relating to legend use and comparisons between visualization types are prone to this confound. In the future, ensure the legend is in the same location for all visualization types if possible.

7.4 Conclusion

In summary, we have identified a set of user characteristics that have a strong impact on task completion time, visualization preference, and user gaze behaviour. Our research suggests that eye tracking could be used as a source for enabling user-adaptive systems to detect user characteristics in real-time, and then provide interventions related to the user characteristics. These interventions could either suggest a more suitable visualization for the user, or provide support to improve user performance or experience with the current visualization. Our research has provided a rich set of results that can be used to inform the design of user-adaptive systems, by showing which user characteristics are worth considering and which visualization elements impact visualization processing.

Bibliography

- [1] Allen, B. Individual differences and the conundrums of user-centered design: Two experiments. *Journal of the American Society for Information Science* 51, (2000), 508-520.
- [2] Amar, R.A., Eagan J., & Stasko, J.T. Low-Level Components of Analytic Activity in Information Visualization. In 16th IEEE Information Visualization Conference, (2005), 15-21.
- [3] Baddeley, A. D. Working memory. New York: Clarendon Press/Oxford University Press, (1986).
- [4] Baldonado, M.Q.W., Woodruff, A., & Kuchinsky, A. 2000. Guidelines for using multiple views in information visualization. *Proceedings of the working conference on Advanced visual interfaces*, ACM (2000), 110-119.
- [5] Brusilovsky, P., Ahn, J., Dumitriu, T., & Yudelson, M. Adaptive Knowledge-Based Visualization for Accessing Educational Examples., in: *Proceedings of Information Visualization*, (2006) 142-150.
- [6] Canham, M., & Hegarty, M. Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*, (2010), 155-166.
- [7] Chen, C. Individual differences in a spatial-semantic virtual environment. *Journal of the American Society for Information Science*, 51(6): (2000), 529–542.
- [8] Cleveland, W. S., & McGill, R. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association* , Vol. 79, No. 387 (1984), 531-554.
- [9] Conati, C., & Maclaren, H. Exploring the Role of Individual Differences in Information Visualization. In *Proceedings of the working conference on Advanced visual interfaces (AVI '08)*, (2008), 199-206.
- [10] Courtemanche, F., Aïmeur, E., Dufresne, A., Najjar, M., & Eboa, F.H.M. Activity recognition using eye-gaze movements and traditional interactions. In *Proceedings of Interacting with Computers*, (2011), 202-213.

- [11] Diehl, S., Beck, F., & Burch, M. Uncovering Strengths and Weaknesses of Radial Visualizations - an Empirical Approach. *IEEE Transactions on Visualization and Computer Graphics*, (2010), 935-942.
- [12] Dillon, A. Spatial-semantics: how users derive shape from information space. *Journal of the American Society for Information Science*, 51(6), (2000), 521-528.
- [13] Dillon, A. & Shaap, D. Expertise and the perception of structure in discourse. *Journal of the American Society for Information Science*, 47(10), (1996), 786-788.
- [14] Eivazi, S., & Bednarik, R. Predicting Problem-Solving Behavior and Performance Levels from Visual Attention Data. In the proceedings of 2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction at IUI 2011, (2011), 9-16.
- [15] Ekstrom, R.B., French, J.W., Harman, H.H. Kit of factor-referenced cognitive tests, Educational Testing Service, Princeton, NJ (1976).
- [16] Few, S. Keep Radar Graphs Below the Radar - Far Below: *Information Management Magazine*, (2005), 48.
- [17] Field, A. *Discovering Statistics Using SPSS*, Third Edition. Sage Publications, London, (2009).
- [18] Field, A., & Hole, G., *How to Design and Report Experiments*. Sage Publications, London, (2003).
- [19] Fukuda, K., & Vogel, E.K. Human variation in overriding attentional capture. *Journal of Neuroscience*. (2009), 8726-8733.
- [20] Fukuhara, Y., & Nakano, Y. Gaze and Conversation Dominance in Multiparty Interaction. 2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction. (2011).
- [21] Goldberg, J.H., & Helfman, J.I. Comparing Information Graphics: A Critical Look at Eye Tracking, *BELIV '10*, (2010), 182-195.
- [22] Goldberg, J.H., & Helfman, J.I. Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. *Information visualization* 10(3), (2011), 182-195.

- [23] Gotz & Wen. Behavior-Driven Visualization Recommendation. Proceedings of the 14th international conference on Intelligent user interfaces, (2009), 315-324.
- [24] Grawemeyer, B. Evaluation of ERST – an external representation selection tutor. In Proceedings of the 4th international conference on Diagrammatic Representation and Inference (Diagrams'06), (2006), 154-167.
- [25] Green, T. M., & Fisher, B. Towards the personal equation of interaction: The impact of personality factors on visual analytics interface interaction. In IEEE Visual Analytics Science and Technology (VAST), (2010), 203-210.
- [26] Grindinger, T., Duchowski, A., & Sawyer, M. Group-wise similarity and classification of aggregate scanpaths. In Proceedings of the 2010 Symposium on Eye-Tracking Research 38, (2010), 101-104.
- [27] Jameson, A. "Adaptive Interfaces and Agents" in Human-Computer Interface Handbook, eds J.A. Jacko and A. Sears, (2003), 305-330.
- [28] Kane, M. J., & Engle, R. W. The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, 9(4), (2002), 637-671.
- [29] Kaptein, M.C., Nass, C., & Markopoulos, P. Powerful and Consistent Analysis of Likert-Type Rating Scales. Proceedings of the 28th international conference on Human factors in computing systems, CHI 2010, (2010), 2391-2394.
- [30] Keith, R. Eye movements and cognitive processes in reading, visual search, and scene perception. *Eye Movement Research Mechanisms, Processes, and Applications*. North-Holland Press, (1995).
- [31] Lewandowsky, S., & Spence, I. The perception of statistical graphs. *Sociological Methods and Research*, 18, (1989), 200-242.
- [32] Lin, T. Cognitive Trait Model for Adaptive Learning Environments. Doctoral Dissertation, Massey University, Palmerston North, New Zealand. (2007).
- [33] Logie, R. H. Visuo-spatial working memory. Psychology Press, United Kingdom, (1995).
- [34] Nowell, L., Schulman, R., & Hix, D. Graphical encoding for information visualization: an empirical study. IEEE Symposium on Information Visualization, INFOVIS 2002, (2002), 43–50.

- [35] Plumlee, M.D., & Ware, C. Zooming versus multiple window interfaces: Cognitive costs of visual comparisons. *ACM Trans. Comput.-Hum. Interact*, 13, 2, (2006), 179-209.
- [36] Rayner, K. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124, (1998), 372-422.
- [37] Salthouse, T. A. Aging and measures of processing speed. *Biological Psychology*, 54, (2000), 35-54.
- [38] Simkin, D., & Hastie, R. An Information-Processing Analysis of Graph Perception. *Journal of the American Statistical Association*, Vol. 82, No. 398 (1987), 454-465.
- [39] Simola, J., Salojarvi, J., & Kojo, I. Using hidden Markov model to uncover processing states from eye movements in information search tasks. *Cognitive Systems Research*. 9, (2008), 237-251.
- [40] Steichen, B., Carenini, G., & Conati, C. Adaptive Information Visualization - Predicting user characteristics and task context from eye gaze. *UMAP Workshops*, volume 872 of Workshop Proceedings, CEUR-WS.org, (2012).
- [41] Tai, R.H., Loehr, J.F., & Brigham, F.J. An Exploration of the Use of Eye-Gaze Tracking to Study Problem-Solving on Standardized Science Assessments. *International Journal of Research & Method in Education* 29, 2, (2006), 185-208.
- [42] Tang, H., Topczewski, J.J., Topczewski, A.M., & Pienta, N.J. Permutation test for groups of scanpaths using normalized Levenshtein distances and application in NMR questions. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, (2012), 169-172.
- [43] Turner, M. L., & Engle, Randall W. Is working memory capacity task dependent? *Journal of Memory and Language*, 28(2) (1989), 127-154.
- [44] Velez, M.C., Silver, D., & Tremaine, M. Understanding visualization through spatial ability differences. *Proceedings of Visualization*, (2005), 511-518.
- [45] Ware, C. *Information Visualization: Perception for Design*, Third Edition. Morgan Kaufmann, Waltham, MA, (2012).
- [46] Williams, M., & Munzner, T. Steerable, Progressive Multidimensional Scaling. *Proceedings of Information Visualization*, (2004), 57-64.

- [47] Wobbrock, J., Findlater, L., Gergle, D., & Higgins, J. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only ANOVA Procedures. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2011, (2011), 143-146.
- [48] Ziemkiewicz, C., Crouser, R.J., Yauilla, A.R., Su, S.L., Ribarsky, W., & Chang, R. How Locus of Control Influences Compatibility with Visualization Style. Proceedings of IEEE (VAST), (2011), 81-90.

Appendix A - Additional Eye Tracking Results

We found many significant results on the main effects that both visualization type and task difficulty have on gaze features, independent of the user characteristics. The aim of this section is to show the impact that the two information visualization types examined in this study can have on gaze patterns. We also shed some insight toward the effectiveness of these visualizations when task difficulty is decreased or increased. The results presented here result from the same data set and analysis that is described in section 5.3.

Main Effects of Visualization Type

We found two main effects of visualization type (bar/radar), which are shown below in Table A.0.1. For each result, the directionality will be reported, along with a short discussion.

Family	Component	F-Ratio	Effect Size	Sig. Value
AOI	Text Proportion	$F(1,678) = 7.20$	$r = 0.09$	$p < 0.05$
	High Transitions	$F(1,681) = 21.56$	$r = 0.18$	$p < 0.001$

Table A.0.1: Main effects of visualization type

AOI Text Proportion: The amount of time a user spent reading the main textual components on each task changes significantly based on the visualization type. In our study, we found that the proportionate amount of time users took to read the question text and also the quantity of fixations in the text AOI, were both lower when using radar graphs and both greater for bar graphs. Since the question text was in the same area for all tasks, it seems that there is an effect of information visualization layout on the amount of time a user spends reading the question text. This finding could be an indication that unfamiliar or more cluttered visualization layouts (i.e., radar graph) distract the user and may be one of the reasons that they spend less time reading the question text.

AOI High Transitions: We found that the total number of high transitions was greater when using radar graphs compared to bar graphs (bottom row of Table A.0.1). One possibility for this may be due to the circular shape of the radar graph which inherently requires more fixations to do comparisons (i.e., going back and forth between items multiple times), versus a more flat or horizontal layout of data values on the bar graph which requires fewer fixations to perform similar comparisons.

Main Effect of Task difficulty

There was one significant main effect of task difficulty for the *AOI Label Transitions*, $F(1,681) = 6.37$, $r = 0.10$, $p < 0.05$. For this finding, users' fixations transitioned over the label area more often on easy tasks than for difficult tasks. This is a possible indication that certain interventions should pertain to the labels of a graph, based on the difficulty of that task. For example, since we know that difficult tasks take more time to complete (see results in Chapter 6), and users transitioned over the labels more often for the less difficult tasks, then perhaps intervening on the labels in order to encourage users to notice them more could help reduce the time it takes to complete difficult tasks.

Interaction Effects involving Visualization Type & Task Difficulty

This section looks at the interaction effects between task difficulty and visualization type. Table A.0.2 summarizes all of the significant results for these interaction effects. In general these findings are interesting because they all show that the amount of time a user spends in each of these different AOIs can differ significantly based on the interaction effect between visualization type and task difficulty.

Family	Component	F-Ratio	Effect Size	Sig. Value
Task	Std.Dev. Path Angles	F(1,681) = 13.95	r = 0.14	p < 0.001
AOI	Legend Transitions	F(1,682) = 8.02	r = 0.11	p = 0.02
AOI	Legend Proportion	F(1,694) = 9.27	r = 0.12	p = 0.01
	Labels Proportion	F(1,670) = 7.53	r = 0.11	p = 0.03
	High Proportion	F(1,678) = 10.65	r = 0.12	p < 0.01

Table A.0.2: Results for interactions between visualization type and task difficulty.

Std.Dev. Path Angles: As reported in section 5.4.4, increased standard deviation of angles may be an indicator of uncertainty. In Figure A.0.1, we see that for all tasks, the standard deviation of angles is lower for radar graphs compared to bar graphs. This finding may be linked to the fact that the circular layout of the data elements within the radar graph visualization affords a more consistent visual work flow in terms of how a user's gaze traverses the data elements. Most notably, for difficult tasks, the standard deviation of path angles decreases for radar graphs but increases for bar graphs, indicating that radar graphs may be more suitable when tasks become more complex. This is consistent with the results we saw in Section 4.2.4, where radar graphs are potentially just as good as bar graphs for more complex tasks.

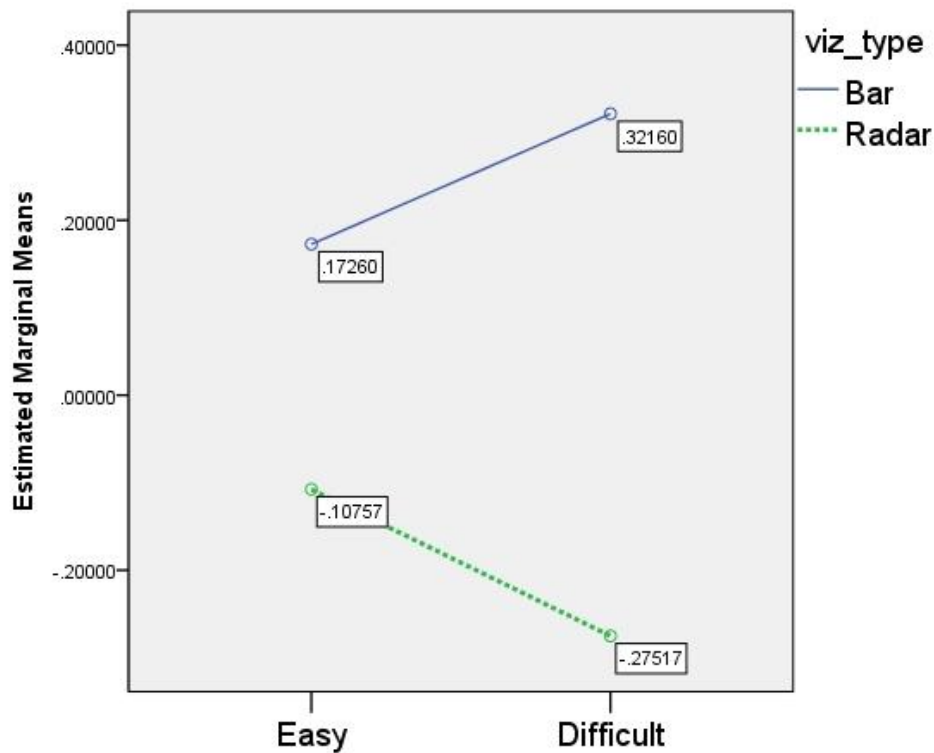


Figure A.0.1: Interaction effect between visualization type and task difficulty for std.dev. path angles component

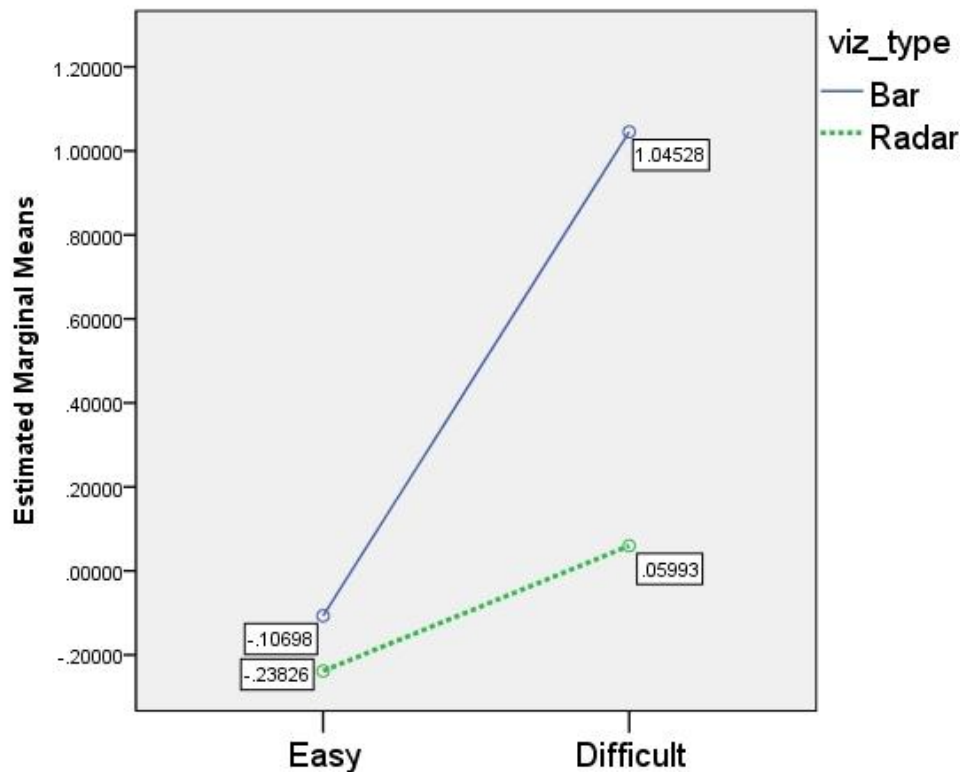


Figure A.0.2: Interaction effect between visualization type and task difficulty for AOI legend transitions component

Legend Proportion & Legend Transitions: Both of these components generated similar directionalities and the interaction effect is shown in Figure A.0.2 for legend transitions which is representative of the directionality for both results. For easy tasks, users essentially spent the same proportion of time, and transitioned just as often regardless of the visualization type. The major difference occurs when using the bar graph on difficult tasks versus the radar graphs. In these cases, users are devoting a great deal more attention in terms of time spent and gaze transitions to the legend when they are looking at the bar graph. This may be another indication that with difficult tasks, bar graphs require some extra processing that makes them comparable in time with radar graphs (see Section 4.2.4). It should be noted however, that this result may be subject to confound since the placement of the legend in our study differs between the two visualization types (as shown in Figure 5.2 and Figure 5.3).

Label Proportion: As shown in Figure A.0.3, users always spend less time looking at labels for radar graphs compared to bar graphs. This may be an indicator that radar graphs are more effective at coding label information. Goldman and Helfman [2010] had a similar finding, and attributed it to the fact that bar graphs present a mass of parallel lines which may overload a user's visual system, thus making it more difficult for a user to associate labels to their values. As for the interaction effect, we also see that the difference in time spent looking at the labels between bar and radar is larger with easy tasks.

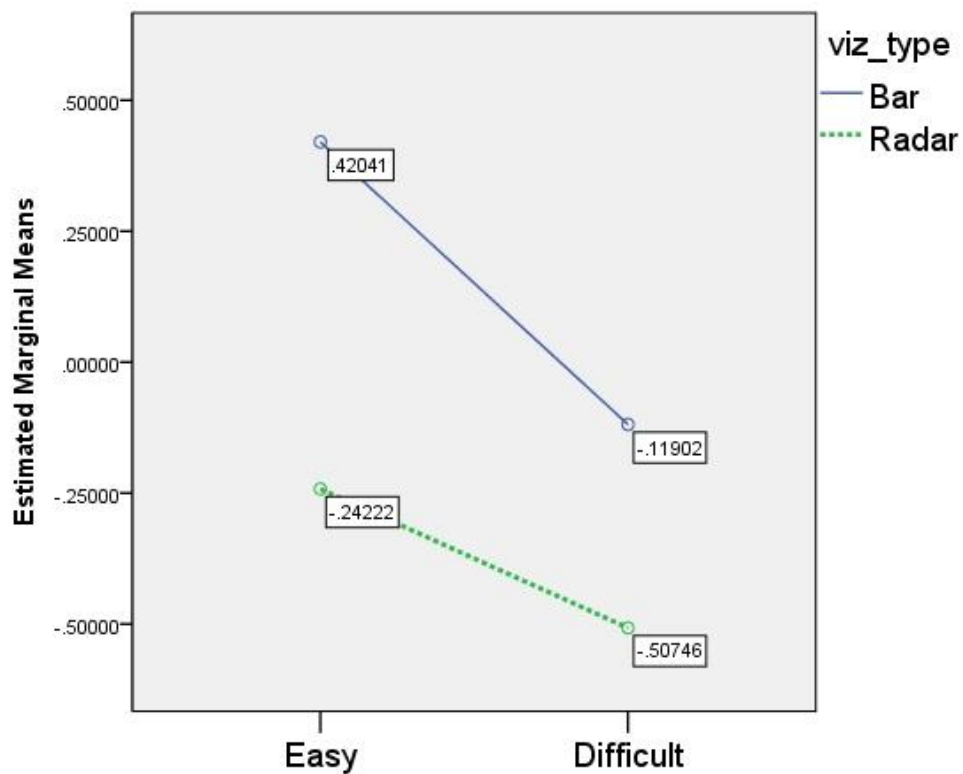


Figure A.0.3: Interaction effect between visualization type and task difficulty for label proportion measures component

High AOI Proportion: Figure A.0.4 shows the interaction effect for this finding. Users always spent more of their time looking at the *High* AOI on bar graphs, compared to radar graphs. But, as task difficulty increases, the proportion of time spent in the *High* area of interest increases even more for bar graph. In other words, increased task difficulty impacts how much time a user spends looking at the main region where the data values are displayed within a visualization, and this changes significantly depending on which information visualization type is used.

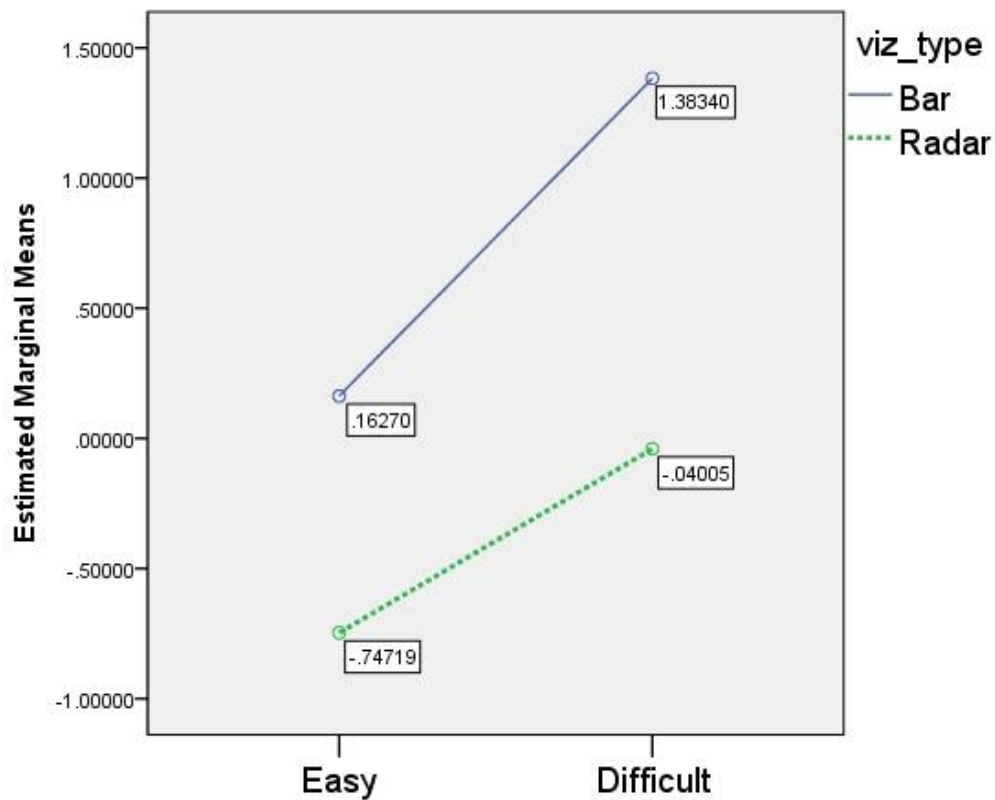


Figure A.0.4: Interaction effect between visualization type and task difficulty for AOI high proportion component

Summary

This section has shown that there are significant effects that both visualization type and task difficulty can have on a user's eye gaze behaviour. We saw that users spend less time looking at the question text on radar graphs, and that this illustrates that there may be an impact from other elements within a given visualization. We also saw that task difficulty can significantly affect how much time and transitions a user devotes to specific areas of an information visualization. These findings are useful since they offer a more detailed insight into visualization processing related solely to elements of visualization type, layout, and task difficulty.

Appendix B - Post Questionnaire

Notification Experiment

Name:

Participation Number

Age

Gender

How did you hear about this study?

Field of Study

Had you ever used bar graph before this study?

- Yes
 No

Had you ever used radar graph before this study?

- Yes
 No

I am an expert in using bar graph.

1 2 3 4 5

Strongly disagree Strongly agree

I am an expert in using radar graph.

1 2 3 4 5

Strongly disagree Strongly agree

1. In general, bar graph was easy to understand?

1 2 3 4 5

Strongly Disagree Strongly Agree

2. In general, radar graph was easy to understand?

1 2 3 4 5

Strongly Disagree Strongly Agree

3. I prefer to use bar graph for answering the questions.

1 2 3 4 5

Strongly Disagree Strongly Agree

4. I prefer to use radar graph for answering the questions

1 2 3 4 5

Strongly Disagree Strongly Agree

5. Bar graph was easy to use when finding the courses in which the student's grade was below/above the class average?

e.g. "In which courses is Maria below the class average?" OR "In how many courses is Maria above the class average?"

1 2 3 4 5

Strongly Disagree Strongly Agree

6. Radar graph was easy to use when finding the courses in which the student's grade was below/above the class average?

e.g. "In which courses is Maria below the class average?" OR "In how many courses is Maria above the class average?"

1 2 3 4 5

Strongly Disagree Strongly Agree

7. Bar graph was easy to use when finding the student's strongest course between 2 courses?

e.g. "Did Tanya receive a higher mark in Geometry or Physics?"

1 2 3 4 5

Strongly Disagree Strongly Agree

8. Radar graph was easy to use when finding the student's strongest course between 2 courses?

e.g. "Did Tanya receive a higher mark in Geometry or Physics?"

1 2 3 4 5

Strongly Disagree Strongly Agree

9. Bar graph was easy to use when finding the course in which the student's grade deviated most from the class average?

e.g. "In which course does Maria deviate the most from the class average?"

1 2 3 4 5

Strongly Disagree Strongly Agree

10. Radar graph was easy to use when finding the course in which the student's grade deviated most from the class average?

e.g. "In which course does Maria deviate the most from the class average?"

1 2 3 4 5

Strongly Disagree Strongly Agree

11. Bar graph was easy to use when finding the student's 2 strongest courses?

e.g. "What are Maria's two strongest courses?"

1 2 3 4 5

Strongly Disagree Strongly Agree

12. Radar graph was easy to use when finding the student's 2 strongest courses?

e.g. "What are Maria's two strongest courses?"

1 2 3 4 5

Strongly Disagree Strongly Agree

13. Bar graph was easy to use when comparing two students' grades in one course?

e.g. "Which of the two students is stronger in Calculus?"

1 2 3 4 5

Strongly Disagree Strongly Agree

14. Radar graph was easy to use when comparing two students' grades in one course?

e.g. "Which of the two students is stronger in Calculus?"

1 2 3 4 5

Strongly Disagree Strongly Agree

15. Bar graph was easy to use when comparing two students' grades with the class average?

e.g. "Find the courses in which Andrea and Diana are both above the average." OR "Find the courses in which Diana is below the average and Andrea is above it."

1 2 3 4 5

Strongly Disagree Strongly Agree

16. Radar graph was easy to use when comparing two students' grades with the class average?

e.g. "Find the courses in which Andrea and Diana are both above the average." OR "Find the courses in which Diana is below the average and Andrea is above it."

1 2 3 4 5

Strongly Disagree Strongly Agree

17. Bar graph was easy to use when finding the student that had a greater number of courses, performing better than the class average?

e.g. "Which student has performed better than the class average in a greater number of courses?"

1 2 3 4 5

Strongly Disagree Strongly Agree

18. Radar graph was easy to use when finding the student that had a greater number of courses, performing better than the class average?

e.g. "Which student has performed better than the class average in a greater number of courses?"

1 2 3 4 5

Strongly Disagree Strongly Agree

Submit

Powered by [Google Docs](#)

[Report Abuse](#) - [Terms of Service](#) - [Additional Terms](#)