CONSTRUCTING KNOWLEDGE FROM INTERACTIVE VISUALIZATIONS.

HOW DATA VIEWING STRATEGIES INFLUENCE

COMPREHENSION OF COMPLEX RELATIONS.

A Thesis

Presented

to the Faculty of

California State University, Chico

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

in

Interdisciplinary Studies – International Cognitive Visualization

by

© Ulrich Ludewig 2015

Summer 2015
CONSTRUCTING KNOWLEDGE FROM INTERACTIVE VISUALIZATIONS.
HOW DATA VIEWING STRATEGIES INFLUENCE
COMPREHENSION OF COMPLEX RELATIONS.

A Thesis
by
Ulrich Ludewig
Summer 2015

APPROVED BY THE ACTING DEAN OF THE GRADUATE SCHOOL

APPROVED BY THE GRADUATE ADVISORY COMMITTEE:

Neil H. Schwartz, Ph.D.
Graduate Coordinator

Erica de Vries, Ph.D., Chair

Neil H. Schwartz, Ph.D.

Wolfgang Schnotz, Ph.D.
PUBLICATION RIGHTS

No portion of this thesis may be reprinted or reproduced in any manner unacceptable to the usual copyright restrictions without the written permission of the author.
ACKNOWLEDGEMENTS

I owe thanks to Dr. Wolfgang Schnitz, Dr. Erica de Vries, and Dr. Neil Schwartz for introducing me to the exiting field of cognition and learning, for countless compelling discussions, and for preparing me for an academic career. This work would not have been possible without the practical support of the Learning, Instruction & Cognition lab, and research assistants. Finally, I owe gratitude to my colleagues in ICV3 for a great and productive time in Landau, Grenoble and Chico.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publication Rights</td>
<td>iii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iv</td>
</tr>
<tr>
<td>List of Figures</td>
<td>vii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>viii</td>
</tr>
<tr>
<td>Acronyms</td>
<td>ix</td>
</tr>
<tr>
<td>Abstract</td>
<td>x</td>
</tr>
<tr>
<td><strong>CHAPTER</strong></td>
<td></td>
</tr>
<tr>
<td>I. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Graph Comprehension</td>
<td>3</td>
</tr>
<tr>
<td>Interaction with External Displays</td>
<td>5</td>
</tr>
<tr>
<td>Representational Choice</td>
<td>7</td>
</tr>
<tr>
<td>Top-Down Effects on Allocation of Attention</td>
<td>8</td>
</tr>
<tr>
<td>Present Study</td>
<td>9</td>
</tr>
<tr>
<td>II. Literature Review</td>
<td>12</td>
</tr>
<tr>
<td>Theories of Graph Comprehension</td>
<td>12</td>
</tr>
<tr>
<td>Effects of Knowledge</td>
<td>17</td>
</tr>
<tr>
<td>Information Reduction Hypothesis</td>
<td>17</td>
</tr>
<tr>
<td>Representational Choice</td>
<td>19</td>
</tr>
<tr>
<td>Cognitive Functions of Representations</td>
<td></td>
</tr>
<tr>
<td>III. Method</td>
<td>22</td>
</tr>
<tr>
<td>Design</td>
<td>22</td>
</tr>
<tr>
<td>Experimental Material</td>
<td>22</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1. Diagram of the different representations (indicated by boxes) and processes (indicated by arrows) involved in understanding a visual display (Hegarty, 2011). ........... 3

Figure 2. Essay Task Display. First sub question in the surface condition, and the overview initial display. In the blank condition no data lines are displayed. ............... 24

Figure 3. Example graph reading task in the overview condition. In the blank condition, no data lines are displayed. ................................................................. 26

Figure 4. Irrelevant information tolerance .......................................................... 33

Figure 5. Performance (Speed) in correctly answered graph reading tasks............. 35

Figure 6. Performance (Accuracy)...................................................................... 37

Figure 7. Essay Quality..................................................................................... 38

Figure 8. Essay length...................................................................................... 39

Figure 9. Interactions with the display during essay writing. .............................. 40
LIST OF TABLES

Table 1. Essay sub questions for content and surface conditions..........................25
Table 2. Definition and example for each statement type in the essays .....................31
Table 3. Analysis of Variance for Average Number of Irrelevant Data Lines Displayed in Correct Items..........................................................33
Table 4. Mann-Whitney-U Test for Average Number of Irrelevant Data Lines Displayed in Correct Tasks ...............................................................34
Table 5. Analysis of Variance for Average Time-on-Task after Graph Selection ........35
Table 6. Analysis of Variance for Correct Items .....................................................37
Table 7. Analysis of Variance for Essay Quality Score ............................................38
Table 8. Analysis of Variance for Number of Words Written in the Essays ............39
Table 9. Analysis of Variance for Number of Interactions during Essay Task ..........41
Table 10. List of all graph reading tasks.................................................................51
## ACRONYMS

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>ET</td>
<td>Essay Task</td>
</tr>
<tr>
<td>ER</td>
<td>External Representation</td>
</tr>
<tr>
<td>CI</td>
<td>Construction-Integration</td>
</tr>
<tr>
<td>IDV</td>
<td>Interactive Data Visualization</td>
</tr>
<tr>
<td>ID</td>
<td>Initial Display</td>
</tr>
<tr>
<td>IHD</td>
<td>Indicator of Human Development</td>
</tr>
</tbody>
</table>
This study investigated how differences in understanding of an interactive data visualization influenced representational choice and performance on a graph reading task. One-hundred-three participants received a training exercise where they learned about historical events and indicators of human development. Participants were then asked to view an interactive data visualization (IDV) and required to interpret the data either within the historical context (content group), or to describe only how the data was displayed (surface group). Participants then solved 12 graph reading tasks using one of two IDVs; an overview initial display that required users to gradually deselect the amount of task-irrelevant information or a blank initial display that required participants to select task-relevant information. Participants that interpreted the data were hypothesized to
focus more on task-relevant information; ignore task-irrelevant information and, therefore, deselect less task-irrelevant information from the display. Results indicated that the content group deselected less irrelevant information, $d(94) = .67, p < .01$, than the surface groups while performing as well as the surface group. The content group appeared to have a higher tolerance for task-irrelevant information, which is consistent with the information reduction hypothesis. Our results suggest that users’ tendency to choose representations with task-irrelevant information was moderated by their level of understanding of the specific external display.
CHAPTER I

INTRODUCTION

According to Hegarty (2013, p. 3) “Information technology, [and] graphical displays are becoming increasingly complex and interactive. Nowadays, users have the flexibility to choose and design their own displays.” Indeed, interactive data visualizations (IDV) have been available for use for a long time, but recently they have become publicly available, with large data sets being transformed into visualizations that make it easy to understand for the general public. For instance, applications like “GapMinder”¹ have made complex issues related to history meaningfully accessible to a wide audience of interested novices.

IDVs have been described by Card, Schneiderman, and McKinley (1999, p. 6) as visualizations that represent data to amplify cognition within computer-supported environments. And, there is common agreement on the potential value of scientific visualizations to help students learn (Card et al., 1999; Zhang & Norman, 1993). However, even though user-visualization interactivity is emphasized by Card et al. (1999), few studies have investigated the role of interaction with IDV when learners are attempting to derive meaning from visualizations—particularly when those visualizations are graphs. We should account for interaction with the visualization in cognitive models of graph comprehension to develop effective instructions and training for graphical

¹ http://www.gapminder.org/
literacy. However, to date, interactivity per se has been absent from those models. Moreover, while outcomes from user-visualization interaction have been examined (Moreno & Mayer, 2007), few if any of these studies have investigated the cognitive processes responsible for specific interaction behaviors and how these behaviors affect performance.

Interacting with an external display is a top-down process influenced by a user’s meta-representational competence (diSessa, 2004) and the user’s momentary internal representation of the display and its referent (Trafton, Trickett, & Mintz, 2005). We wanted to investigate how top-down processes influence the interaction with external displays within cognitive models of graph comprehension. While these displays can contain more or less task-irrelevant information, research in representational choice has shown that experts tend to choose displays with task-irrelevant information, rather than novices (Hegarry, Smallman, Stull, & Canham, 2009). The experts also show a greater ability to ignore these task-irrelevant information (Canham & Hegarty, 2010; Haider & Frensch, 1999). The problem is that since experts tend to choose representations with task-irrelevant information and ignore those task-irrelevant information more efficiently, we were interested in knowing whether all users choose representations with more or less task-irrelevant information based on their internal representation of the display—an internal representations based on their understanding of that display.

In the following sections, we will cover a cognitive model of graph comprehension and explain how interaction with external displays fits into this model. Moreover, we will examine how expertise influences the interaction with external display
and representational choice. Finally, we will discuss how top-down effects on allocation of attention could influence representational choice.

Graph Comprehension

Graph comprehension is a complex interaction of bottom-up perceptual processes of encoding and top-down processes of applying graph schema and domain-knowledge (Trickett & Trafton, 2006; Freedman & Shah, 2002; Hegarty, 2001). These processes take place between three stages of graph comprehension. Each stage results in its own internal representation. First, users have to encode the visual features of the external display. The visual features are encoded depending on a goal that directs the allocation of attention. When the encoding is successful it results in an internal representation of the display. Secondly, users have to map these internal representations of the display onto the conceptual relationship it conveys by applying graph schema (Pinker, 1990; Hegarty, 2011 uses the term “display schema”). Finally, users must relate this conceptual relationship to its referent by applying domain knowledge to these

![Diagram of the different representations (indicated by boxes) and processes (indicated by arrows) involved in understanding a visual display (Hegarty, 2011).](image)

Figure 1.
conceptual relationships. This process of mapping from a conceptual relationship to its referent results in an *internal representation of the referent*. The internal representation of the referent can contain more information than explicitly presented in the external display, if inferences are made based on domain knowledge (See Figure 1). Schnott (2014) emphasizes that meaningful learning from external displays requires a coordinated process of selecting information, organizing information, activating prior knowledge resulting in coherence formation. The mapping of the internal representation of a display to its referent, can be considered as a coherence formation process. A good understanding of the display requires coherence mapping between the internal representation of the display and its referent. The coherence formation process is an active process that involves interaction with the external display. In Hegarty’s (2011) model, interaction with the external displays can be the allocation of attention to specific information (Arrow A, see Figure 1) or the manipulation of the external display (Arrow B, see Figure 1). Interaction in terms of allocation of attention to specific information can be identified by eye-movement patterns (Gegenfurtner, Lehtinen, & Säljö, 2011), while the manipulation of the external display is a sequence of representational choices a user performs (Trafton et al., 2005). While the active processing of forming coherence during task solving is researched with eye-movement patterns (Gegenfurtner et al., 2011; Jarodzka, Scheiter, Gerjets, & Gog, 2010; Schnott, Ludewig, Ullrich, McElvany, Horz, & Baumert, 2014), few studies have investigated the active process of manipulating an external display to form coherence during task solving.
Interaction with External Displays

The manipulation of an external display is limited to its functions. Each function of the external display has the potential to facilitate a cognitive process. For instance, filtering task-irrelevant information from a very cluttered graph might increase the accuracy for a graph reading task, but the function of the external display facilitates a cognitive process relative to a task. This filtering function may not affect the conceptual interpretation of the graph. For the graph reading task, filtering data points results in a computational offload. It is clear that computational offload facilitates the cognitive process (Ainsworth, 2006) demands of the task, but it is unclear whether users take advantage of computational offload, if they have the choice.

Multiple top-down processes influence the interaction of the user with the IDV. The interaction with the IDV during graph comprehension is a sequence of representational choices. The representational choices depends on meta-representational competence (diSessa, 2004), if the decision is made before the graph comprehension process. The manipulation of the external display during the comprehension process depends additionally on users’ internal representation of the display and its referent (Trafton et al., 2005). During the comprehension process users tend to follow a processing mode. Schnottz et al. (2014) distinguished task-driven selective processing from coherence-oriented general processing. Task-driven selective processing puts the emphasis on rather top-down goal-driven selective processing, and coherence-oriented general processing puts the emphasis on a bottom-up global understanding of the content. For instance, a user that mostly processes information task-driven selective might choose an external display that only displays task-relevant information. Schnottz et al. (2014)
investigated these processing modes with eye-movements on text-picture integration tasks. The processing mode should also effect the manipulation of the external display, since eye-movements and manipulations of the external display are active processes to form coherence. We need to measure users’ behavior during task-driven selective processing to investigate the top-down effect of a user’s understanding on the user’s representational choice.

Experts are users with a good understanding of the external displays they are working with, due to their domain knowledge and familiarity with the display. An expert compared to a novice’s behavior indicates how domain knowledge and familiarity with the external display affect representational choice, in a top-down fashion (Hegarty, 2013). Trafton et al. (2005) investigated how scientists from different domains (meteorology & physics) interact with external displays during their work, by analyzing verbal protocols. The external displays were IDV. The authors described the scientist’s activity as an alternating process of manipulating an external display to align the display with their internal representation and align their internal representation to the external display, using spatial transformations. In a different study, Trafton, Trickett, Stitzlein, Saner, Schunn, & Kirschenbaum (2006) demonstrated that experts compared to journeyman use more spatial transformations during comprehension of graphs, while the journeyman group interacted more with the external display. This indicates that experts with domain knowledge and familiarity with the external display use more spatial transformations to form a coherent mapping between internal representation of the display and its referent. Their domain knowledge and familiarity might enable them to internally transform their internal representation instead of manipulating the external display. This raises the
question whether users with a good understanding of the external display (like experts) use less manipulations of the external display to solve a task.

While Trafton et al. (2006) observed whole sequences of graph manipulation, the following sections focuses on the representations users chose relative to their task, expertise and domain knowledge.

Representational Choice

Representational choice research investigates whether users choose the most efficient representation for a specific task. The most efficient representational choice should follow the parsimony principle. The parsimony principle states that one should not display more information than needed for a specific task (Bertin, 1967). Hegarty, Smallman, Stull, and Canham (2009) investigated whether novices and experts chose the external displays that follow the parsimony principle. Hegarty at el. (2009) found that novices picked 71% of the trials in the simpler map with only task-relevant information. Experts chose in only half of the trials in the map with only task-relevant information. The authors argued that this might be due to the familiarity with maps that contain political borders and information about the train. A different reason for this effect is misconception about perception (Smallman & Cook, 2011). Smallman et al., (2011) argue that users believe that perception is simple and always in accuracy representations, even though their actual representation is inaccurate. The tendency to choose displays with task-irrelevant information is consistent with user’s preference for more realistic displays (Smallman et al., 2011) and 3-D over 2-D graphs (Zacks, Levy, Tversky, & Schiano, 1998). However, realism and 3-D effects do not add any interpretable information, while experts tend in the study of Hegarty et al. (2009) to choose displays
with more interpretable information. It is unclear how this tendency is influenced by the user’s understanding of the external display, for example, which top-down processes possibly mediate this effect and how users will actually perform on the representations they choose.

Moreover, it is unknown whether the same effect occurs when users choose the external display during solving a task instead of before solving the task. Users that manipulate the external display during the solving task might choose differently. The representational choice is in this case less due to the conception or misconception about perception, but to the experience of the success of the perceptual process. Experts may choose display with task-irrelevant information, because they have a greater ability to ignore task-irrelevant information (Haider et al., 1999). The allocation of attention to task-relevant information might influence the representational choice during the interaction with the external display. In the following section we will cover how top-down processes influence allocation of attention to task-relevant information of external display.

**Top-Down Effects on Allocation of Attention**

Top-down effects on allocation of attention to task-relevant information in graph comprehension are researched in the context of expertise (Gegenfurther et al., 2011). For instance, the information reduction hypothesis explains expert’s perceptual ability with a learned selectivity of information processing (Haider et al., 1999). Experts learn to neglect task-irrelevant information and to actively focus on task-relevant information, which they accomplish by strategic allocation of attention. In eye-tracking research, this perceptual ability to ignore irrelevant information results in fewer fixations.
on irrelevant information and more fixations on task-relevant information. In the original experiment, participants learn to ignore task-irrelevant information over the course of training in repeated trials of judgment tasks of alphanumeric character strings (Haider et al., 1999). Ho, many studies showed that the information reduction hypothesis can be applied to more complex stimuli (Charness, Reingold, Pomplun, & Stampe, 2001).

Canham et al. (2010) investigated if the information reduction hypothesis can be generalized to comprehension of weather maps and the effect of domain-relevant instructions on accuracy and proportion of views on task-relevant information. They compared the proportion of fixations on task-relevant areas before and after users participated in a tutorial in which they received domain-relevant instruction. They found that both accuracy and the proportion of view time on task-relevant information, increased after instruction. This finding indicates that the allocation of attention it not only influenced over the course of gradually developed expertise, but also by relatively short domain-relevant instructions of a rather conceptual more than procedural nature. A domain-relevant tutorial affected user’s allocation of attention in a way that they received a higher “tolerance” for task-irrelevant information, which facilitated their ability to focus on task-relevant information. This is consistent with the moderating effect of visualizations on expertise difference in performance (Gegenfurter et al., 2011). The tolerance for task-irrelevant information should affect the representational choice when users’ gradually reduce the amount of task-irrelevant information for an external display.

Present Study

This study investigates whether coherence forming orientation and initial display of an IDV affects users’ irrelevant information tolerance and graph reading task
performance. Coherence forming orientation is manipulated through an essay task with two possible orientations: content and surface. A content orientation, as opposed to a surface orientation, stimulates the coherence formation between the internal representation of the display and its referent. This means, user in the content orientation interpret the data using domain-relevant information from the text and user in the surface orientation only describe how the data is displayed. Whether users need to deselect irrelevant information or, on the contrary, select relevant information to solve graph reading task is manipulated through the initial display with two possible settings: overview and blank respectively. These two factors might affect users’ representational choice and in particular the degree to which they allow irrelevant information to be displayed in graph reading.

The external display allows users to externally reduce task-irrelevant information, which in turn results in computational offloading (Hegarty, 2011; Ainsworth, 2006). Users that formed coherence between their internal representation of the display and its referent will have a higher tolerance for irrelevant information, because they allocate their attention to task-relevant information. These users with a higher tolerance for task-irrelevant information, even if they can reduce the amount by using the IDV, should leave more task-irrelevant information on the display when solving a graph reading task.

We hypothesize that the content group has a higher tolerance for task-irrelevant information, so that they deselect less task-irrelevant information from the overview display in the graph reading tasks (Hypothesis 1). The amount of irrelevant information should not negatively influence performance. The performance in the content
group should be at least as good as the performance in the surface groups, when they
deselect task-irrelevant information from an overview graph (Hypothesis 2). In the blank
initial display condition, the content group and the surface group should select only task-
relevant information (Hypothesis 3), and the performance of the content group should not
suffer in case they selected irrelevant information (Hypothesis 4).
CHAPTER II

LITERATURE REVIEW

In the following literature review I will, first, give a summary of contemporary theories of graph comprehension, second, elaborate on top-down effects on graph comprehension.

Theories of Graph Comprehension

Graph comprehension has been investigated for decades. Steven Pinker proposed a popular model in graph comprehension in 1990. He considers visual features of the display, gestalt processes, and the graph schema as factors that allow the user to extract the conceptual message of a graph. His model of graph comprehension can be summarized in seven steps: (1) The user has a goal to extract a specific piece of information, (2) the user looks at the graph, while activating graph schema and gestalt processes, (3) based on gestalt principles the user encodes salient features of the graph, (4) the user now knows which cognitive/interpretative strategies to use, the user then extracts goal-directed visual chunks, (6) the user may compare two or more visual chunks and (7) finally, the user extracts the relevant information to satisfy the goal.

Eric Freedman and Priti Shah (2002) developed a model that focuses on the prior knowledge that is necessary for graph comprehension. They criticized most empirical research on quantitative graphs because they only required fact retrieval. This
leaves the question of how one might apply fact retrieval to a real-life context, such as making a decision based on complex data, developing a scientific theory from the data, or making a judgment about a theory. It is fundamental to identify the characteristics of graph comprehension in order to describe the interpretation process. The graph format and types of data sets influence the interpretation process. Freedman et al. (2002) focused mostly on higher levels of information extraction, to learn more about the real-world situations of graph interpretation. Graph interpretation requires making inferences and solving problems. A graph can only be interpreted in the context of relevant prior knowledge. The authors describe multiple types of prior knowledge: domain knowledge, graphical skills, and explanatory skills. The emphasis of domain knowledge is novel compared to Pinker’s model that focused on graph schema as a prior knowledge factor. Freedman et al. (2011) claims that subjects chunk the visual features of a graph during the initial processing, while prior knowledge guides the processing of visual features.

Freedman et al. (2011) proposed “A Construction-Integration Model” in order to explain the graph comprehension process, which is analogous to Kintsch’s (1988) Construction-Integration (CI) Model of text comprehension. The CI model consists of two phases: The construction phase and a comprehension phase. Graph and text comprehension have both a construction and integration phase, which takes place in alternating cycles. This was shown in studies using eye-fixations. Graph and text comprehension involve serial and incremental processing. During the construction phase relevant prior knowledge is constructed. The reader attempts to construct a coherent representation of available information. This process is influenced by superficial aspects
of the text or graph. The construction processes is also influenced by prior knowledge and task demands during the initial processing of text or graph information.

In the integration phase of the CI Model, disparate knowledge is combined into a coherent representation. This integration process is effortful when readers have to make inferences in order to form a coherent representation. Comprehension is effortless, when relevant information is explicitly represented in the visual features of the graph. When the visual features are identified, they can be easily linked to prior knowledge. The graph comprehension process shares many characteristics with text comprehension.

The CI model of graph comprehension considers three pools of units: Visual features, domain knowledge, and interpretation propositions. Visual feature units represent the properties of the graph.

*Visual features* such as format and color, have an effect on low-level perceptual aspects of graph comprehension as well as on high-level cognitive processes. For instance, the way information is grouped on the display is an important display characteristic. It influences chunking during graph comprehension. The interpretation of relevant quantitative information is most sufficient when relevant quantitative information is directly represented in visual chunks. When relevant information must be derived by mentally transforming data to make inferences about relationships or facts, some viewers struggle to comprehend that information. “Graphical displays are most useful when they make quantitative information perceptually obvious” (Freedman et al., 2002, p.22).

*Domain knowledge* units represent viewer’s prior knowledge about quantitative relationships. Proposition units represent possible interpretations of the
information in a graph. The visual features and domain knowledge units are important factors that correspond to a viewer’s graphical reading skills. A good representation has strong links between visual features and interpretations. The CI model suggests that visual features of a graph are likely to interact with a viewer’s prior knowledge. Individuals with more graphical skills will chunk specific types of information that are relevant together, compared to individual with poor graphical skills. Domain knowledge includes any mental representation of the content of the graph. This domain knowledge is activated under a number of factors and has to be activated by the graph in order to interpret the graph. The activation depends on how accessible the domain knowledge is to the user. The second factor is salience; salience indicates how meaningful and relevant the particular explanation is in a specific context.

Content familiarity has an influence on the viewers’ comprehension goal. It helps viewers to keep track of information and serves viewers when they are making mental computations required in inference generation. Furthermore, it helps the viewer to identify possible errors. Therefore, viewers with well-developed abilities to mentally manipulate information in the graph will be less impacted by the presentation format.

Shah, Freedman, and Vekiri (2005) draw a distinction between perceptual and conceptual processes. Perceptual processes are “bottom-up encoding mechanisms” which focus on the visual features of the display. Conceptual processes equate to “top-down encoding processes” which influence interpretation. Perceptual processes account for performance on simple, fact retrieval tasks.

Trickett and Trafton (2006) remarked that although graphs depend on spatial arrays, the processes by which information is extracted are largely perceptual. They claim
that other models such as Pinker’s Theory of Graph Comprehension and Freedman and Shah’s Construction-Integration model, have not incorporated spatial processes, because most of the task-graph combinations used in psychology laboratories are very simple and can be addressed using perceptual processes. They showed with research in complex domains with complex tasks that many spatial processes are involved. This is especially the case if the information is not explicitly represented in the graph and if simple perceptual processes are inadequate to extract that implicit information.

Work of Trickett et al. (2006) focuses on situations in which subjects are unable to directly extract the information they need. In hours of in-depth analysis of several hours of verbal protocol they found that the use of spatial transformations are more frequently used than any other strategy to generate information. Spatial transformations can be observed, particularly, in complex domains for which complex visualizations are required. These results were consistent: The verbal protocols show that subjects use a great deal of spatial processing to extract and use information from data visualizations (Trafton, Kirschenbaum, Tsui, Miyamoto, Ballas, & Raymond, 2000). Additionally, experts use far more spatial processing than novices (Trafton et al., 2006).

The term spatial process was established by Baddeley (1999), who made the distinction between verbal and spatial processes. Trickett et al. (2006) operationally defined spatial processing in two ways. First, spatial processing involves maintaining spatial information in working memory, and secondly, spatial processing can also be identified via the use of mental spatial transformations, which occur when a spatial object is transformed from a mental state into another mental state. We will refer to these as simple spatial transformation. There are different types of mental transformation:
Creating a mental image, modifying that mental image by adding or deleting features, mentally rotating features, mentally moving an object, animating a static image, making comparisons between views, and any other mental operation which transforms a spatial object from one state into another. Trickett et al. (2006) make a distinction between spatial processing and purely perceptual processing, in which graph users are able to make direct or explicit comparisons from the graph itself, without the need to hold spatial information in working memory.

Trafton et al. (2005) compared the number of spatial transformations with the number of physical transformations scientists performed on visualizations. The authors were interested in the use of mental imagery comprehension of scientific visualizations. They found that scientists performed significantly more spatial transformations than physical transformations.

Effects of Knowledge

Freedman and Shah (2002) emphasized that graph comprehension is influenced by prior knowledge. However, it is still unclear on which stage knowledge affects graph comprehension. Different types of prior knowledge might affect different processes of graph comprehension. Domain knowledge affects the interpretation of graphs, but also the perception of the graph (Canham et al., 2010). Domain knowledge has the potential to affect which locations and visual features users focus on and consequently encode.

Information Reduction Hypothesis.

The effect of prior knowledge and training in comprehension of visualizations can be explained by the information-reduction hypothesis (Haider et al. 1999). This
hypothesis focuses on learned selectivity of information processing. Experts gain the ability to neglect task-irrelevant information and to actively focus on task-relevant information, which they accomplish by strategic allocation of attention. Irrelevant information is perceptually ignored whenever this is possible, due to the learners training. In eye-tracking research, this perceptual ability to ignore task-irrelevant information should result in fewer fixations on task-irrelevant information and more fixations on task-relevant information.

Canham et al. (2010) investigated whether the information reduction hypothesis by Haider et al. (1999) can be generalized to the comprehension of weather maps. They investigated the effect of domain-relevant instructions on accuracy and proportions of viewing task-relevant information. Canham et al. (2010) were able to show that both accuracy and the proportion of view time on task-relevant information increased after instruction. This demonstrated that domain-relevant instruction affects the process of information selection form complex graphical displays, the processes of interpretation, and making inferences from this selected information. Several studies have shown that experts pay more attention to task-relevant aspects of graphical displays, but expertise reflects 10 or more years of experience in a domain (Ericsson & Charness, 1994). This study demonstrated that attention to visual displays could change significantly within 10-15 minutes of instructions that are more conceptual than procedural. These findings are consistent with the information reduction hypothesis, and demonstrated that it applies to more ecologically valid tasks.

The information reduction hypothesis can be generalized to tasks with complex and highly visual stimuli. Experts gain a perceptual ability that also effects the
inferences they make based upon the perceived information (Jarodzka, Scheiter, Gerjets, & van Gog, 2010). This effect was shown in number of domains such as fish locomotion (Jarodzka et al., 2010) or radiology (Cooper, Gale, Darker, Toms, & Saada, 2009). The information reduction hypothesis is strongly supported by a meta-analysis of effect sizes in eye-tracking research about differences between novices and experts. The meta-analysis included a variety of studies in different domains, using different visualizations, and different task demands. The experts fixated more on task-relevant areas ($r=0.53$) and fewer on task-irrelevant areas ($r=-0.31$) compared to novices (Gegenfurtner et al., 2011).

Representational Choice

The representational choice research investigates whether users choose the most efficient representation for a specific task. Researches have suggested many principles about what an efficient representation might be. For instance, the parsimony principle states that one should not display more information than is needed for a specific task (Bertin, 1967). The principle works relative to the task demands. Wickens and Carwell (1995) proposed that focused tasks are facilitated by simpler displays that include only task-relevant variables, whereas integrative tasks are facilitated by more complex displays with several superimposed variables. Hegarty, Smallman, & Stull (2012) found that the parsimony principle applies to the use of weather maps by systematically vitiating the number of displayed variables in a map. The search time of participants was effected by the number of irrelevant variables represented on the displayed. In a different experiment, Hegarty et al. (2009) investigated whether subjects choose the most efficient representation. They found that about half of the novices and
about 70% of the experienced participants chose representations that displayed more than the relevant amount of information.

A different principle was chosen by Zacks and Tversky (1999); they asked their participants which representation they prefer. The representations depicted either discrete comparisons between data points or depicted trends, in either bar or line graphs.

Cognitive Functions of Representations

Graphical displays can “augment” cognition in different ways (Hegarty, 2011). They have different functions. For instance, an appropriate display can augment cognition as an external storage of information, by organizing information, by offloading cognition on perception and by offloading cognition on action.

- Visual displays are external representations that externally store information and free up working memory resources for other aspects of thinking (Card et al., 1999).
- Grouping of information represents a close relationship in the represented world.
- Nonvisual data can be mapped onto visual variables, so that visual patterns emerge that can then be interpreted.
- In interactive displays, people can offload internal mental computations on external manipulations of the display itself (Card et al., 1999).

Similar functions of external representations in the context of learning from multiple representations are proposed in the DeFT framework by Sharon Ainsworth (2006). She suggests that an external representation (ER) can reduce the amount of cognitive effort required to solve an equivalent problem. The more appropriate representation provides computational offload. ER can also represent the same abstract
structure with a different representational format. This *Re-representation* can have positive effects on reasoning and problem solving. An ER is *graphically constraining* if it limits the range of inferences that can be made about the represented concept.
CHAPTER III

METHOD

Design

The study follows a 2 x 2 (Essay Task x Initial Display) between-subjects design with two dependent variables. The factor, essay task has two levels: Content versus surface. The factor, Initial Display has two levels: Overview versus blank. Participants were randomly assigned to their conditions. The experiment had three phases: Pre-training, coherence forming and a testing phase. In the pre-training phase, participants were familiarized with the IDV and gained relevant prior knowledge. In the coherence forming phase participants wrote essays about the content of the material (content groups) or about the function and design of the IDV (surface group). The IDV initially displayed either all of the data lines in the initial display condition, overview, or no data lines in the initial display condition, blank. In the testing phase, subjects answered 12 graph reading tasks which were counterbalanced by a Latin square in sets of four items.

Experimental Material

All experimental materials and instructions were built into a website. The website was presented on square screens in a lab room. Participants clicked a “next” button to proceed through the instructions, but there was no possibility to navigate backwards or to skip pages.
Pre-Training and Background Information

As a background story in the study, the participants were asked to imagine that they were working for the “World Health Organization” in the research department. Their task was to interpret data to make suggestions about the issues of overpopulation.

The pre-training included a static picture that explained the organization of the interface, followed by two animations that demonstrated the functions of the IDV. The animations demonstrated the changes that occur on the display when data lines are selected and deselected. The first animation explained the changes that occur when countries are selected and deselected. The second animation demonstrated the changes that occur when indicators are selected and deselected. The initial animation shown to each participant depended on the Initial Display condition the participant was in; either an overview or a blank display. The example IDV showed artificial data that demonstrated the function of the IDV analogous to the animations.
Interactive Data Visualization (IDV)

The interface of the IDV consisted of a control button sidebar on the left and a data display in the middle. The instructional text was presented to the right side of the screen.

Figure 2 shows the IDV in the overview initial display condition, all three countries and all three indicators were initially displayed. In the blank initial display condition, the graph initially displayed no data. In both conditions, participants were able to select and deselect as many data lines as needed. Only the initial display varied between the overview and blank initial display condition.

Figure 2. Essay Task Display. First sub question in the surface condition, and the overview initial display. In the blank condition no data lines are displayed.
Essay Task

The essay tasks for content and surface conditions are displayed in Table 1. In the content condition, participants were asked to write an essay about three sub questions. These sub questions demanded complex information extraction, mapping historical events to the graph, and building inferences based on the text. In the surface condition, participants were also asked to write an essay about three sub questions on the function of the IDV and design of the display.

Table 1. Essay sub questions for content and surface conditions.

| Content #1 | “How do population, fertility rate, and GDP per capita relate to each other?” |
| Content #2 | “How are the indicators influenced by historical events in each country?” |
| Content #3 | “What can you say about the issue of overpopulation based on this data?” |
| Surface #1 | “Explain to someone who is unfamiliar with this interface how it works. Use the relations between the indicators as an example.” |
| Surface #2 | “For each indicator, describe how the graphs are changing when you select and deselect countries?” |
| Surface #3 | “Argue about the advantages and disadvantages of interactive graphs that changes scale?” |

Graph Reading Tasks

Participants were exposed to 12 graph reading tasks. The data display reset after every task back to the participant’s Initial Display condition.

Figure 3 shows the initial display of the overview condition. Each task starts in the overview initial display condition with all data lines and in the blank initial display condition without any data lines. The task format was multiple choice questions. Each task had one correct answer and three distractors. The number of words per item ranged from 20 to 109 with an average number of $M = 49.25$ ($SD = 27.10$). For each task two data lines were relevant. Participants in the blank and overview initial display condition
had to, respectively, select or deselect three checkboxes to display the relevant data lines and those lines only. For example, some tasks required a participant to select (or deselect) two countries and one indicator. Other tasks required a participant to select (or deselect) one country and two indicators. Participants in the overview initial display condition were expected to deselect as many irrelevant data lines as needed until they could solve the graph reading task. The tasks could be solved by using an overview graph, but it was easier to solve the task if less irrelevant data lines were displayed. Participants in the blank initial display condition had to select the relevant data lines to solve the task. If they did not select the relevant data lines they could not solve the task.

Figure 3. Example graph reading task in the overview condition. In the blank condition, no data lines are displayed.
Instructional Text

The instructional text described historical events and indicators of human development between the years 1950-2000 in China, Germany and South Korea. The countries, historical events, and indicators were selected, because they allow for inferences relevant to the interpretation of the issues of overpopulation. The text had 335 words in total. The text structure is shown in Figure 3.

Demographic Questionnaire & Debriefing

The participants were asked to answer questions about their age, sex, year of study, major, and prior experience with the topic (“Have you taken any courses or had any experience (friends, voluntary activities) in the past that have helped you solve these kinds of tasks?”). Finally, participants were debriefed.

Procedure

The procedure was divided into three phases: The pre-training phase, the coherence forming phase and the testing phase.

During the pre-training phase participants were introduced to the scenario, and were given background information. The background information involved information about the organization and function of the IDV. Additionally, they were pre-trained with an IDV with example data. Participants then read the instructional text about historical events and the indicators of human development. The participants reading time was not restricted. All pre-training activities required passing a series of control questions. There were two control questions about background information and scenario, four control questions about pre-training and five control questions about the
instructional text. Participants that gave incorrect answers were sent back to the related instructional material. The number of tries to answer the control questions was not restricted.

The coherence forming phase required participants to write three short essays about either the content of the IDV and the instructional text or the function of the IDV. The essay-writing task was limited to six minutes. We chose this time restriction, because in a pre-study without time restriction on the same essays, participants with good answers took about six minutes per essay. The website proceeded automatically after the time ran out.

Next, in the testing phase, participants were exposed to the IDV and the instructional text with 12 graph reading tasks. Depending on the Initial Display, the IDV started with either an overview or a blank display. The graph reading tasks were presented in a counterbalanced order using Latin squares in batteries of four items. After selecting one answer in the graph reading task, subjects had to click “next” to proceed to the next task. Finally, participants answered a demographic questionnaire and read a debriefing document once they had finished the testing phase.

Sample

One hundred and three undergraduates in psychology and economics volunteered for participation (72 percent female, 28 percent male) and were randomly assigned to one of four conditions. The median age was $M = 21$, 11 percent were 25 and older. Sixty-seven percent were majors in psychology and 33 percent were business or business related majors. Fourteen percent of the participants were freshmen, ten percent juniors, 42 percent sophomore, and 32 percent seniors. We excluded eight participants,
because their number of correct answers to the graph reading tasks was below the guess rate (25%) or they worked for less than 90 seconds in total on all 12 graph reading tasks.

Variables

We used tolerance (for task-irrelevant) and performance as dependent measures. Tolerance for task-irrelevant information shows the extent to which participants display task-irrelevant information to solve the graph reading tasks. The performance is the time-on-task after graph selection. These dependent variables are averaged over the 12 items of the testing phase.

In addition to the dependent variables, we calculated a number of control variables, as manipulation checks, such as the total accuracy of graph reading task, the quality and number of words of the essay, and number of interactions with the IDV during essay writing.

Dependent Variables

The tolerance for task-irrelevant was operationalized as the average amount of irrelevant data lines displayed on the graph that were used to solve a graph reading task for correct items. This measure can vary between zero (only relevant data lines displayed) to 7 (all irrelevant data lines displayed). A participant with a score of one, displayed an average of one irrelevant data line per correct item. The tolerance for task-irrelevant score is low when the number of irrelevant data lines is low and high when the number of irrelevant data lines is high. Items in which the relevant data lines were not displayed are excluded from the score, because if data lines were not displayed and they were relevant data lines they should not be able to solve the task. Therefore, they are represented by a score of zero.
The second dependent variable is *performance (speed)*, operationalized as the average time-on-task after graph selection on correct items. We expect no differences in accuracy between the groups, due to the low difficulty. Therefore, differences in performance should be represented by the time participants needed to solve the task correctly. Only the time after graph selection was counted to avoid influences from interaction behavior. A short time-on-task on correct items refers to a high performance. Items in which relevant data lines were not displayed were excluded to match the same set of items used in the tolerance for irrelevant information score.

**Control Variables**

We used total accuracy on graph reading tasks, essay quality and number of words in essays and number of interactions during essay task as control variables.

The total accuracy is the sum of the number of correct graph reading tasks. Total accuracy can potentially range from 5 to 12. High total accuracy refers to high performance.

The essay quality was rated by allocating each statement in the essay to one of 5 categories: *Wrong statement, description of the display, simple data extraction, or complex data extraction*. Table 2. shows the definition and an example for each category. The sub questions were rated separately in a random order. Only simple extractions and complex extractions counted in the essay quality score. Simple extractions account for one point and complex extractions account for two points in the final essay quality score.

The number of words was the sum of the words from all three sub questions. Letter combinations without meaning were excluded from the word count (e.g. asdf, tttttt, jdnfnrjd).
The number of interactions during essay task was the sum of the interactions with the IDV during the three sub questions of the essay writing. An interaction was any manipulation of the display during the essay writing.

Table 2. Definition and example for each essay statement type.

<table>
<thead>
<tr>
<th>Category</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wrong statement</td>
<td>Statement inconsistent with the displayed data within the essay or wrong data extraction.</td>
<td>&quot;GDP was at a constant rate during the great leap of china&quot;</td>
</tr>
<tr>
<td>Description of the display</td>
<td>Statement that describes the data display, without referring to the quantity of an indicator.</td>
<td>&quot;The x axis represents time and the y represents GDP. The three lines stand for each of the three countries- china, germany and south korea.&quot;</td>
</tr>
<tr>
<td>Simple data extraction</td>
<td>A simple statement extracted from the data</td>
<td>&quot;Germany has the highest per capita&quot;</td>
</tr>
<tr>
<td>Complex data extraction</td>
<td>A complex statement extracted from the data</td>
<td>&quot;For germany and south korea the GDP per capita increased but the populations stayed the same over the years&quot;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&quot;GDP per capital was increase, but fertility rate is increased during 1960-1970 the period of baby-boom, but after 1960-1970 was decrease&quot;</td>
</tr>
</tbody>
</table>
CHAPTER IV

RESULTS

Analysis of Variance for Dependent Variables.

We performed two separate ANOVA’s to investigate the effect of Essay Task and Initial Display on irrelevant information tolerance and performance as measured by speed in the graph reading tasks. The Levene’s test for equal variance is not significant, if no indicated difference.

Tolerance for Irrelevant Information

Tolerance for irrelevant information operationalized as the average number of irrelevant data lines displayed was analyzed as a function of Essay Task (content vs. surface) and Initial Display (overview vs. blank). The descriptive statistics are displayed in Figure 4 and the ANOVA results are displayed in Table 3.

The Initial Display (ID) had a large statistically significant effect, $F(1, 91) = 16.175$, $\eta^2 = .151$, $p < .001$, regarding the number of task-irrelevant data lines displayed. Participants in the overview condition displayed more irrelevant information, $M = 1.39$ ($SD = 2.14$), than participants in the blank condition, $M = 0.21$ ($SD = 1.8$). The Essay Task (ET) was also significant and yielded a medium effect size, $F(1, 91) = 5.888$, $\eta^2 = .06$, $p < .05$. The content group displayed more task-irrelevant information $M= 1.15$ ($SD = 2.01$), than the surface group, $M = 0.45$ ($SD = 1.01$). Finally, the interaction between ID x ET was significant and yielded a small to medium effect, $F(1, 91) = 5.027$, $\eta^2 = .06$, $p < .05$ (See Table 5). The simple main effects showed that content, $M = 2.10$ ($SD = 2.57$)
and surface groups, \( M = 0.71 \) (SD = 1.40) were only different within the overview initial display condition, \( d(94) = .67, p < .01 \). These results indicate that the content group deselected less task-irrelevant information than the surface group in the overview condition.

Table 3. Analysis of Variance for Average Number of Irrelevant Data Lines Displayed in Correct Items.

<table>
<thead>
<tr>
<th>Effect</th>
<th>( df_1 )</th>
<th>( df_2 )</th>
<th>( F )</th>
<th>( \eta^2 )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Display (ID)</td>
<td>1</td>
<td>91</td>
<td>16.175</td>
<td>.15**</td>
<td>.000</td>
</tr>
<tr>
<td>Essay Task (ET)</td>
<td>1</td>
<td>91</td>
<td>5.888</td>
<td>.06*</td>
<td>.017</td>
</tr>
<tr>
<td>ID x ET</td>
<td>1</td>
<td>91</td>
<td>5.027</td>
<td>.05*</td>
<td>.027</td>
</tr>
</tbody>
</table>

\( \text{Note: *p < .05, **p < .01. N = 95. R}^2 = .23. \)

In this ANOVA we found a significant Levene’s test. Therefore, we investigated the difference between the conditions using the Mann-Whitney-U test for
non-parametric measures. The Initial Display still indicated a medium effect, $U(95) = 513.50, r = -0.047, p < .01$, while the effect of essay task was not reproduced with the non-parametric method (See Table 4).

Table 4. Mann-Whitney-$U$ Test for Average Number of Irrelevant Data Lines Displayed in Correct Tasks.

<table>
<thead>
<tr>
<th>Effect</th>
<th>$N$</th>
<th>$U$</th>
<th>$r$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Display (ID)</td>
<td>95</td>
<td>513.50</td>
<td>-0.47**</td>
<td>.000</td>
</tr>
<tr>
<td>Essay Task (ET)</td>
<td>95</td>
<td>994.00</td>
<td>-0.1</td>
<td>.316</td>
</tr>
<tr>
<td>Initial Display (ID) in content</td>
<td>47</td>
<td>117.00</td>
<td>-0.5**</td>
<td>.001</td>
</tr>
<tr>
<td>Initial Display (ID) in surface</td>
<td>48</td>
<td>136.50</td>
<td>-0.45**</td>
<td>.002</td>
</tr>
<tr>
<td>Essay Task (ET)   in blank</td>
<td>48</td>
<td>264.50</td>
<td>-0.07</td>
<td>.623</td>
</tr>
<tr>
<td>Essay Task (ET)   in overview</td>
<td>47</td>
<td>205.50</td>
<td>-0.22</td>
<td>.132</td>
</tr>
</tbody>
</table>

*Note:* *p* < .05, **p** < .01.
Performance (Speed)

Performance (Speed) measured in average time-on-task after graph selection was analyzed as a function of Essay Task (content vs. surface) and Initial Display (overview vs. blank). The descriptive statistics are displayed in Figure 5 and the results of the ANOVA are displayed in Table 5.

A medium interaction effect between Initial Display and Essay Task was found, $F(1, 91) = 5.723, \eta^2 = .06, p < .05$ and further investigated by simple main effects tests. The difference between the content group, $M = 28.62$ seconds ($SD = 9.63$) and the surface group, $M = 21.82$ seconds ($SD = 9.50$) was significant within the blank initial display condition, $d(94) = .10, p < .05$.

![Figure 5. Performance (Speed) in correctly answered graph reading tasks. Y-axis represents the average time-on-task after graph selection in correct tasks in seconds. Error bars represent 95%-confidence interval. N = 95. R² = .08.](image)

**Table 5.** Analysis of Variance for Average Time-on-Task after Graph Selection.

<table>
<thead>
<tr>
<th>Effect</th>
<th>df1</th>
<th>df2</th>
<th>F</th>
<th>$\eta^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Display (ID)</td>
<td>1</td>
<td>91</td>
<td>0.343</td>
<td>.00</td>
<td>.559</td>
</tr>
<tr>
<td>Essay Task (ET)</td>
<td>1</td>
<td>91</td>
<td>1.716</td>
<td>.02</td>
<td>.194</td>
</tr>
<tr>
<td>ID x ET</td>
<td>1</td>
<td>91</td>
<td>5.723</td>
<td>.06*</td>
<td>.019</td>
</tr>
</tbody>
</table>

*Note: *$p < .05$, **$p < .01$. N = 95. R² = .08.*
Analysis of Control Variables

The control variables performance accuracy, essay quality, essay length and interaction during essay task were investigated with separate ANOVAs using Initial Display and Essay Task. The Levene’s test for all following ANOVAs was not significant. Additionally, we performance overall correlates within and between dependent and control variables.

Performance (Accuracy)

Performance (Accuracy) measured in number of correctly answered graph reading tasks was analyzed as a function of Essay Task (content vs. surface) and Initial Display (overview vs. blank). The descriptive statistics are displayed in and results of the ANOVA are displayed in Table 6.

![Performance (Accuracy)](image)

Figure 6. Performance (Accuracy). Y-axis represents the number of correct items. Potential range from 0-12. Error bars represent 95%-confidence interval. Differences are not significant. \( N = 95. R^2 = .04. \)
Table 6. Analysis of Variance for Correct Items.

<table>
<thead>
<tr>
<th>Effects</th>
<th>df₁</th>
<th>df₂</th>
<th>F</th>
<th>$\eta^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Display (ID)</td>
<td>1</td>
<td>94</td>
<td>0.155</td>
<td>0.00</td>
<td>0.695</td>
</tr>
<tr>
<td>Essay Task (ET)</td>
<td>1</td>
<td>94</td>
<td>0.116</td>
<td>0.00</td>
<td>0.734</td>
</tr>
<tr>
<td>ID x ET</td>
<td>1</td>
<td>94</td>
<td>3.270</td>
<td>0.03</td>
<td>0.074</td>
</tr>
</tbody>
</table>

Note: No significant differences were found. N = 95. $R^2 = .04$

The total average over all conditions was $M = 8.7$ ($SD = 1.66$) of 12 possible correct items. The analysis of variance displayed in Table 6 indicated no significant differences.
Essay Quality

We analyzed the word count and the essay quality as a function of Essay Task (content vs. surface) and Initial Display (overview vs. blank). A large main effect for Essay Task was found, $F(1, 94) = 81.197$, $\eta^2 = .47$, $p = .000$. The essay quality of the content groups, $M = 32.75$ ($SD = 13.35$) was much higher the essay quality of the surface group, $M = 9.71$ ($SD = 13.02$). This suggest that the manipulation was successful.

Table 7. Analysis of Variance for Essay Quality Score.

<table>
<thead>
<tr>
<th>Effect</th>
<th>$df_1$</th>
<th>$df_2$</th>
<th>$F$</th>
<th>$\eta^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Display (ID)</td>
<td>1</td>
<td>91</td>
<td>0.013</td>
<td>.00</td>
<td>.911</td>
</tr>
<tr>
<td>Essay Task (ET)</td>
<td>1</td>
<td>91</td>
<td>81.197</td>
<td>.47**</td>
<td>.000</td>
</tr>
<tr>
<td>IDD x ET</td>
<td>1</td>
<td>91</td>
<td>1.583</td>
<td>.02</td>
<td>.212</td>
</tr>
</tbody>
</table>

Note: *$p < .05$, **$p < .01$. $N = 95$, $R^2 = .48$. 

Figure 7. Essay Quality. Y-axis represents the essay quality score. Error bars represents a 95%-confidence interval. $N = 95$. $R^2 = .48$. 
Essay Length

We analyzed the number of word as a function of Initial Display and Essay Task. Number of words written in the essays differed significantly between the overview initial display condition and the blank initial display condition, $F(1, 94) = 7.041, \eta^2 = .07, p < .05$. Participants in the blank initial display condition wrote an average $M = 190.83 (SD = 63.89)$ words and in the overview initial display condition $M = 231.47 (SD = 84.65)$. Subjects in the blank initial display condition engaged less in the essay writing (See Figure 8). A blank initial display seems to inhibit the essay writing process.

![Essay length](image)

**Figure 8.** Essay length. Y-axis represents the number of words written in the essays. Error bars represents a 95%-confidence interval. $N = 95$. $R^2 = .09$.

**Table 8.** Analysis of Variance for Number of Words Written in the Essays.

<table>
<thead>
<tr>
<th>Effect</th>
<th>$df_1$</th>
<th>$df_2$</th>
<th>$F$</th>
<th>$\eta^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Display (ID)</td>
<td>1</td>
<td>91</td>
<td>7.041</td>
<td>.07*</td>
<td>.009</td>
</tr>
<tr>
<td>Essay Task (ET)</td>
<td>1</td>
<td>91</td>
<td>0.193</td>
<td>.00</td>
<td>.662</td>
</tr>
<tr>
<td>ID x ET</td>
<td>1</td>
<td>91</td>
<td>1.305</td>
<td>.01</td>
<td>.256</td>
</tr>
</tbody>
</table>

*Note:* *p<.05, **p<.01. $N = 95$. $R^2 = .09$. 
Number of Interactions during Essay Task  Finally, we analyzed the number of interactions with the display during the essay task as a function of Essay Task (content vs. surface) and Initial Display (overview vs. blank). Means are presented in Figure 9 and the results of the ANOVA are presented in Table 8. Results indicated a significant medium effect for Initial Display, $F(1, 91) = 80.729$, $\eta^2 = .47, p < .000$. Participants in the overview initial display condition interacted on average only $M = 18.19$ (SD =10.16), while participants in the blank initial display condition interacted on average, $M = 40.69$ (SD = 14.12) times with the display. Moreover, the Essay Task had a marginally significant effect on the number of interactions, $F(1,91) = 3.329$, $\eta^2 = .04$, $p = .071$.

![Interactions during essay task](image)

**Figure 9. Interactions with the display during essay writing. Y-axis represents the number of words written in the essays. Error bars represents a 95%-confidence interval. N = 95. $R^2 = .48$.**
Table 8. Analysis of Variance for Number of Interactions during Essay Task.

<table>
<thead>
<tr>
<th>Effect</th>
<th>df1</th>
<th>df2</th>
<th>F</th>
<th>$\eta^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial display (ID)</td>
<td>1</td>
<td>91</td>
<td>80.729</td>
<td>.47**</td>
<td>.000</td>
</tr>
<tr>
<td>Essay Task (ET)</td>
<td>1</td>
<td>91</td>
<td>3.329</td>
<td>.04</td>
<td>.071</td>
</tr>
<tr>
<td>IDD x ET</td>
<td>1</td>
<td>91</td>
<td>0.148</td>
<td>.00</td>
<td>.701</td>
</tr>
</tbody>
</table>

*Note: *$p < .05$, **$p < .01$. $N = 95$. $R^2 = .48$.*

Correlations

In addition to the ANOVAs, we investigated the correlation between and within the dependent and control variables. Important to mention, the two dependent measures time-on-task after graph selection and irrelevant data lines displayed were not correlated, $r(95) = .13$, $p = .18$. The performance (accuracy) correlated with tolerance of irrelevant information, $r(95) = -.31$, $p < .01$. Meaning that participants with a higher tolerance for irrelevant information answer less items correctly.
CHAPTER V

DISCUSSION

This investigation provides insight into how top-down processes affect the representational choice for graph reading tasks. We manipulated users’ understanding of an external display with essay tasks that stimulated different coherence forming orientations. The coherence forming orientations prompted the content group to interpret the display using domain-relevant information and prompted the surface group to only describe how data is displayed.

In the overview initial display condition, participants had to gradually reduce the amount of task-irrelevant information displayed until they were able to solve the graph reading tasks. Consistent with our hypothesis, the content group showed a higher tolerance for task-irrelevant information than the surface group (Hypothesis 1). In addition, the content group performed as well as the surface group (Hypothesis 2). We argue that the content group had a higher tolerance for task-irrelevant information because they were better able to focus on task-relevant information, due to their previous efforts to interpret the data during the essay task.

These findings contribute to the body of research on graph comprehension, which can be applied to inform the design of training for graphical literacy with IDV and enhance instruction to scaffold learning with IDV. Specifically, our results demonstrate that coherence forming orientation influences an individual’s representational choice. Many studies have shown that users tend to choose representations with task-irrelevant
information (Hegarty, 2009; Smallman et al., 2011; Zacks, Levy, Tversky, & Schiano, 1998). This is can be explained by users’ meta-cognitive concepts, specifically “misconceptions about perception” (Smallman et al., 2011). Misconception about perception is the belief that perception is simple, and results in accurate representations. Our findings suggest that the tendency to choose representations with task-irrelevant information is moderated by a user’s understanding of the external display. We argue that a good understanding of the display enables users to focus their attention to task-relevant information, which makes users with a good understanding choose representations with more task-irrelevant information.

Moreover, we find it remarkable that the conceptual process of writing an essay over the course of 15 minutes significantly affected the representational choice for a rather perceptual process of solving a graph reading task.

The reported effects on the tolerance for task-irrelevant information are consistent with the information reduction hypothesis (Haider et al., 1999). Although it was originally proposed in the context of skill acquisition for judgment tasks of alphanumeric character strings, this study suggest that a user’s understanding helps them to ignore task-irrelevant information, which influences their representational choice. This shows that the perceptual processes described by the information reduction hypotheses affect behavior in more complex situations, more similar to those in the classroom or work place.

Our results further suggest that, with regards to Hegarty’s Model of Visual-Spatial Displays (2011), the active processes of allocating attention and manipulating an
external display are not independent from each other. The process that is used for coherence forming is moderated by understanding of the display.

We found that in the blank initial display condition, participants tended to select task-relevant information. The results within this condition were partly inconsistent with our hypothesis. However, the content and surface group did select only task-relevant information (Hypothesis 3), this group’s performance was contrary to our prediction. The surface groups was faster than the content group (Hypothesis 4). We speculate that this difference was a result of a higher number of interactions in the blank initial display condition and the surface essay task. The surface essay task stimulates the building of an internal representation of the display without interpreting the data. This focus on the internal representation of the display, in addition with the higher number of interactions within the blank initial display condition, might have supported the process of solving the graph reading task.

Limitations

We have argued that time is an appropriate performance indicator, because the individual item difficulty is very low. Despite high accuracy in a pilot study, the number of incorrect items were relatively high (about 25%) in the actual sample. This might indicate low engagement of the participants in the actual sample. Moreover, the time-on-task was positively correlated with the accuracy. Not surprisingly, participants that took longer were more accurate. This may indicate that time-on-task in correct items could be biased by the number participates that got answers correct by guessing. Participants that guess the correct answer are more likely to be faster than participants that actually solve
the graph reading task. However, this should not influence the between groups
comparison, because accuracy between the groups is not different.

The reported difference in number of irrelevant data lines displayed within the
overview initial display condition were not reproduced with the Mann-Whitney-U test.
This suggests that the difference between the content and surface groups is due to a
distribution different from a normal distribution. We argue that the non-normal
distribution of number of irrelevant data lines is due to the construction of the IDV. In the
overview initial display conditions participants can deselect one, two or three button in
the side bar. They start with the overview display with nine data lines. One click brings
them to seven data lines. The second click displays between four and three data lines. The
third click brings them to two data lines. Assuming participants navigating linearly, the
behavior of deselecting data lines is automatically transformed in a non-metric scale.

Future Work

We recommend that future studies combine our indirect measure of tolerance
for irrelevant information with a direct measure (e.g. eye-tracking). Additionally, one
should evaluate a user’s belief as to whether the graph reading task was answered
accurately. This may allow researchers to collect more evidence for the interaction
between perceptual process and meta-cognitive processes of representational choice.
Further, we suggest the use of IDV with more data line to increase possible variance for
representational choice, and to systematically vary the difficulty of the graph reading
tasks. In addition to task difficulty, one could also vary the level of cognitive process
required by the tasks (e.g. form perceptual to conceptual). The interaction between meta-
cognitive, cognitive, and perceptual process for representational choice and graph comprehension require more investigation.
REFERENCES


APPENDIX

Table 9. *List of all graph reading tasks.*

<table>
<thead>
<tr>
<th>Q1</th>
<th>During The Great Leap of China, which declined first GDP per capita or fertility rate?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A. GDP per capita.</td>
</tr>
<tr>
<td></td>
<td>B. Fertility rate.</td>
</tr>
<tr>
<td></td>
<td>C. GDP per captia and fertility rate declined simultaneously.</td>
</tr>
<tr>
<td></td>
<td>D. GDP per captia and fertility rate did not decline during The Great Leap.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q2</th>
<th>Did the decrease of fertility rate in China significantly reduce the population growth rate?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A. The fertility rate did not decrease.</td>
</tr>
<tr>
<td></td>
<td>B. Population growth does not slow down while the fertility rate decreases.</td>
</tr>
<tr>
<td></td>
<td>C. Population growth slows down while fertility rate decreases.</td>
</tr>
<tr>
<td></td>
<td>D. Population is not growing.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q3</th>
<th>What is the difference in fertility rate between China and South Korea when The Miracle on the Han River started?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A. About 6</td>
</tr>
<tr>
<td></td>
<td>B. About 4</td>
</tr>
<tr>
<td></td>
<td>C. About 1</td>
</tr>
<tr>
<td></td>
<td>D. There is not difference.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q4</th>
<th>In Germany, did the fertility rate always decrease while the economy was growing?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A. The fertility rate did not decrease over the whole time.</td>
</tr>
<tr>
<td></td>
<td>B. The fertility rate reached its climax around 2.5 in in the 60s, and then decreased.</td>
</tr>
<tr>
<td></td>
<td>C. The economic growth of Germany had many regression periods. The fertility rate always decreased.</td>
</tr>
<tr>
<td></td>
<td>D. The fertility rate reached its climax around 4.7 in the 80s, and then decreased.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q5</th>
<th>What is similar in South Korea's and Germany's fertility rate and GDP per capita?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A. A constant economic growth goes along with constant increase in fertility rate.</td>
</tr>
<tr>
<td></td>
<td>B. Both countries enforced a strict birth control policy. Due to this policy, fertility rate decreased constant.</td>
</tr>
<tr>
<td></td>
<td>C. A constant economic growth goes along with an increase followed by a decrease in fertility rate.</td>
</tr>
<tr>
<td></td>
<td>D. Both countries enforced a strict birth control policy. Due to this policy, fertility rate decreased constant.</td>
</tr>
</tbody>
</table>
Q6 Did Germany's or South Korea's population grow more over the whole time period?
   A. South Korea's population grew by 15% while Germany grew by 50%.
   B. Germany grew by 50% while South Korea more than doubled it's population.
   C. South Korea more than doubled it's population while Germany grew by about 15%.
   D. Germany more than doubled it's population while South Korea grew by about 25%.

Q7 In Korea, how did the fertility rate change while the economy was growing?
   A. The fertility rate decreased for short time while economy grow slow, then fertility rate decreased slower while the economy grow exponential.
   B. The fertility rate increased for short time while economy was in regression, then fertility rate decreased while the economy grow back to the old level.
   C. The fertility rate decreased for short time while economy was in regression, then fertility rate increased while the economy grow back to the old level.
   D. The fertility rate increased for short time while economy grow slow, then fertility rate decreased while the economy grow exponential.

Q8 How many times did Korea and China have the same fertility rate?
   A. 8
   B. 4
   C. 6
   D. 0

Q9 In which decade are Germany's and Korea's fertility rate the closest in number?
   A. 90s.
   B. 50s.
   C. 80s.
   D. 70s.

Q10 In Korea, which showed a greater change, the expansion in GDP per capita or the decrease in fertility rate?
   A. Fertility rate.
   B. GDP per capita.
   C. Both change in the same way.
   D. GDP per capita is decreasing.

Q11 Did China's or Germany's population increase faster after 1990?
   A. China.
   B. Germany.
   C. Both do not significantly grow.
   D. Germany's population decreases.
Q12 Does Korea or China have more regression periods in which the GDP per capita was lower for a short period of time?
   A. China.
   B. Korea.
   C. Both the same.
   D. Korea never had regression.