Deriving a holistic cognitive fit model for an optimal visualization of data for management decisions.

Indicate submission type: Research-in-Progress

Lisa Falschlunger
University of Applied Sciences Upper Austria
Lisa.falschlunger@fh-steyr.at
Elisabeth Grabmann
University of Applied Sciences Upper Austria
Elisabeth.Grabmann@fh-steyr.at

Othmar Lehner
Said Business School, University of Oxford
othmar.lehner@sbs.ox.ac.uk
Heimo Losbichler
University of Applied Sciences Upper Austria
Heimo.losbichler@fh-steyr.at

Abstract

Research shows that managerial decision making is directly correlated to both, the swift availability, and subsequently the ease of interpretation of the relevant information. Visualizations are already widely used to transform raw data into a more understandable format and to compress the constantly growing amount of information produced. However, research in this area is highly fragmented and results are contradicting. This paper proposes a preliminary model based on an extensive literature review including top current research on cognition theory. Furthermore an early stage validation of this model by experimental research using structural equation modeling is presented. The authors are able to identify task complexity as one of the most important predicting variables for information perception of visual data, however, other influences are significant as well (data density, domain expertise, working memory capacity and subjective visual complexity).

Keywords: *Information visualization, cognitive fit, decision making, PLS modelling.*

1. Introduction

A visual representation of data can be seen as a means to accelerate, as well as to improve cognition and interpretation (Al-Kassab et al., 2014) of such, and thus should in theory improve a rational managerial decision making process. However, such representations are used inconsistently in praxis (Cho et al., 2012; Dilla and Janvrin, 2010) and sometimes either in a way that misses the purpose of informing the reader in an effective and efficient way (Falschlunger et al., 2014) or in a way that may even be misleading or manipulating (Beattie and Jones, 2002).

According to Conati and Maclaren (2008) the success of visualizations should be determined by the ability of users to retrieve relevant information in an effective and efficient way. In this context the theory of cognitive fit, which heavily draws on information processing theory and cognitive load theory, has been used as a research framework in a lot of different empirical investigations, however, results so far are contradicting, often due to the lack of knowledge in relation to the visual perception process (e.g.: Barat, 2007; Dilla et al., 2010; Galletta and Vessey, 1991; Porat et al., 2009; So and Smith, 2004).

Understanding how visualizations affect a user's perception is highly complex as it is influenced by the task and the data at hand as well as by individual factors such as experience, knowledge or culture (Dilla and Janvrin, 2010; Parush et al., 2007; Peck et al., 2012; Yigitbasioglu and Valcu, 2012). So far, research in the area of information visualization lacks a means and method of determining the quality of a graphical representation for such purposes a-priori (Peck et al., 2012; Ziemkiewicz and Kosara, 2010). To advance theory, the authors introduce a new triangulated and validated holistic model based on structural equation modeling. It predicts cognitive effectivity and efficiency, and is adaptive to the individual recipients' differences by incorporating the cognitive burden a graph presents to its reader.

2. Theoretical Background

The model is based on information processing theory and cognitive load theory. Information processing theory provides knowledge on the perceptual process and is divided into three essential stores: sensory store (reception of environmental information for a few seconds), short-term memory (temporarily store that analyses, deconstructs and synthesizes information), and long-term memory (responsible for creating and saving mental constructs or schemas) (Lord and Maher, 1990). Cognitive load theory provides guidelines on how to foster information retrieval and learning by enhancing either germane cognitive load by using standardized or well-known formats or by enhancing extraneous cognitive load by using the right visualization formats and designs (Mostyn, 2012). In order to enhance the extraneous cognitive load cognitive fit theory is important. It states that visualization always needs to fit the task at hand because the problem solver does not need to exert additional cognitive effort to either transform the problem representation to better match the task or to transform their decision processes to better match problem representation (Vessey, 1991). However, as mentioned in the introduction this theory produced contradicting results, indicating that other influences might be relevant as moderator and mediator variables.

Based on an extensive literature review (initial search: 1.952 articles; after abstract sorting based on the content: 237 articles) four essential dimensions of information visualization could be identified and an overview of the model can be seen in figure 1 (red framed variables were alternated):

- Visual Complexity: Visual complexity is the degree of difficulty to transform an image into a consistent verbal description of the designer and the recipient of the visualization (Yigitbasioglu and Valcu, 2012). Ziemkiewicz and Kosara (2010) claim that telling "meaningful stories is the goal of visualization…but stories can rise up purely from differences in shape and arrangement." Two components determine visual complexity, namely the used visualization type and the design or structure of a given visualization (Hill and Milner, 2003; Kuang et al., 2012; Parsons and Sedig, 2014).
- Task Complexity: Task complexity has three variables: task type, task difficulty (or complexity) and task environment (Speier et al., 2003). Two essential classifications can be found for task type: first spatial (determining relationships and making comparisons) and symbolic tasks (use of discrete data values) by Vessey (1991), and second accumulation (acquiring and recalling a single information cue), recognition (recognizing patterns and relationships between 2-3 information cues), estimation (identifying trends between numerous information cues), and projection (making projections to future values) by Hard and Vanecek (1991). Task difficulty can either be calculated objectively or tested subjectively by asking participants, and task environment represents time constraints, task interruptions (Speier, 2006).
- Data Complexity: Data complexity combines data type and data density. To evaluate data type the accompanied dimensions need to be considered (Kuang et al, 2012) and for data density the amount of data compressed in a visualization (Gelman and Unwin, 2013).
- Individual Complexity: Individual complexity can be clustered into three fundamental dimensions: Cognitive traits representing a persons' working memory ability, cognitive states representing situational and emotional influences and experience and

biases (Peck et al., 2012). Situational and emotional states revere to a persons' origin and current state. Of particular interest in conjunction with visual representations are knowledge and expertise, experience, decision making style, gender, and culture for the origin part as well as motivation, concentration and emotional issues that might focus attention (Al-Kassab et al., 2014; Hahn, 2014; Lord and Maher, 1990; Mostyn, 2012).

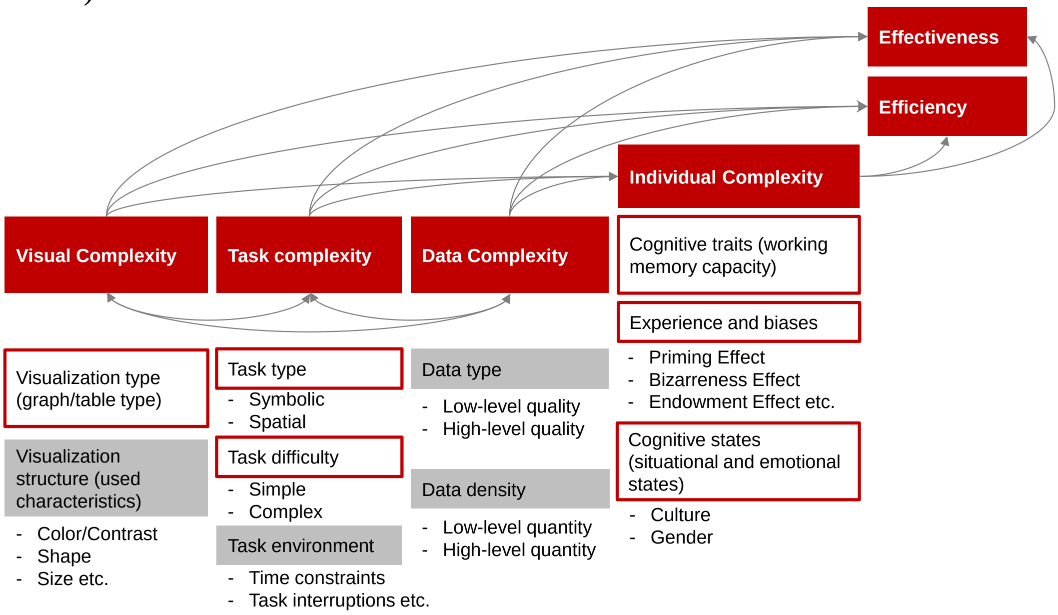

Figure 1: Data Visualization Score (DVS) Model

3. Research Questions and Hypotheses

Despite early individual studies on correlations, there is a lack of research on a more holistic level, bringing the various influences together into a unified model to predict the quality of visualization for managerial decision making. The focus of this paper lies on the examination of the most frequently as well as the strongest influence mentioned in literature – task complexity. A high task complexity is said to hinder effectivity and effectiveness (Speier, 2006). Therefore the hypotheses we are looking out for in particular are:

H1: *A lower task complexity (TC) improves efficiency in cognition.*

H2: *A lower task complexity (TC) positively influences the correct interpretation of the data (effectivity).*

The variable TC was objectively measured by looking at behavioral acts and information cues caused by the used information visualization and its data density (Wood 1986). For evaluating this variable in relation to the other identified influences the authors then came up with a preliminary model based on the main dimensions as laid out before. In a first step the model is transformed into structural equations to cater for the interdependencies and to identify effect sizes.

4. Methodology

We conducted a laboratory experiment in which subjects were given four different tasks while viewing 18 different visualizations, which included different versions of bar charts, column charts, and tables. All of these visualizations showed a company's financial performance and all of the tasks had an optimal solution. Participants were selected from a student population with at least one year background in business administration. During a 30 minute session, tasks were given to subjects by a computer-based decision support system and eye tracking data was recorded. In total, 84 students volunteered to participate in the experiment which resulted in a total of 1,476 data records, since each student completed multiple tasks. Participants were randomly assigned to one of four experimental groups. Using various parametric and nonparametric tests, we found no significant differences across treatments according to gender, age, years in school, major, and prior experience.

A 4x9x2 between-subjects and within-subjects experimental design was used with four levels of task type based on Vessey (1991) and Hard and Vanecek (1991) (accumulation, recognition, estimation, projection), nine levels of information presentation format, and two levels of information types (i.e. time series data and data on the structural split of e.g. revenue within one year on the product mix of the company). Randomization of the tasks within the four groups was used.

Our dependent variables were efficiency and effectiveness. Participants were able to determine the pace of the experiment by independently going through the test by clicking. No time constraints were imposed. Efficiency was measured with the time span between seeing the visualization and stating the answer (net dwell time) and by the total fixation count per participant for each task. The latter was recorded using an eye tracking device (SMI RED with a sampling rate of 120 Hz, a nine point calibration and a 4 point validation). When analyzing eye tracking, data fixations are of particular interest. They are short stops where the eye is able to process information. No information can be processed during a movement of the eye, which is called saccade. Longer fixations are associated with greater visual and/or cognitive complexity. An increased number of fixations can be interpreted as having a negative impact on search efficiency (Goldberg and Helfman, 2014; Renshaw et al., 2003). Effectiveness was measured by the correctness of the stated answers.

Additional factors used in the model include domain expertise, spatial ability, culture, task complexity and visual complexity. Expertise with charts, working experience, years of school education, and visual complexity were measured via self-reported data. Spatial ability was measured by operating span and symmetry span which were collected with an automated test using E-Prime, a software frequently used for psychological tests (Redick, 2012). Cultural dimensions, such as Uncertainty Avoidance and Long Term Orientation, were measured using Hofstede’s scales (Hofstede, 2001) and task complexity was calculated according to Wood (1986).

In order to estimate the parameters of the underlying conceptual model, the authors turn to structural equation modeling (SEM) - in particular to variance-based SEM because of its mild distributional assumptions (Reinartz et al., 2009) and fewer convergence problems. The authors apply a version of partial least square path modelling because of its widely demonstrated capabilities of approximating latent variables using linear composites of observed variables. The computation was supported by the software Smart PLS 3 (Ringle et al, 2015).

5. Early interpretation and Conclusion

H1 asked whether a low task complexity (TC) improves efficiency in cognition and H2 for its influence on the effectivity (correct interpretation). As we can see in figure 2, TC predicts Efficiency and Effectivity at a significance level of $p<0.01$. Therefore we can confirm both hypotheses.

Besides the confirmed factor TC, the main predictors for effectivity and efficiency differ. When analyzing efficiency domain expertise and subjective visual complexity show a high influence. However, individual factors such as cultural dimensions or spatial ability show modest but also significant results as well. When looking at effectiveness data density shows the highest influence. Overall, this preliminary study looks promising as it can be seen as an early endorsement of the created model and the DVS. The highly interlinked model allows the authors to create an optimal data visualization given a specific task for a specific audience, concerning experience and various cultural backgrounds.

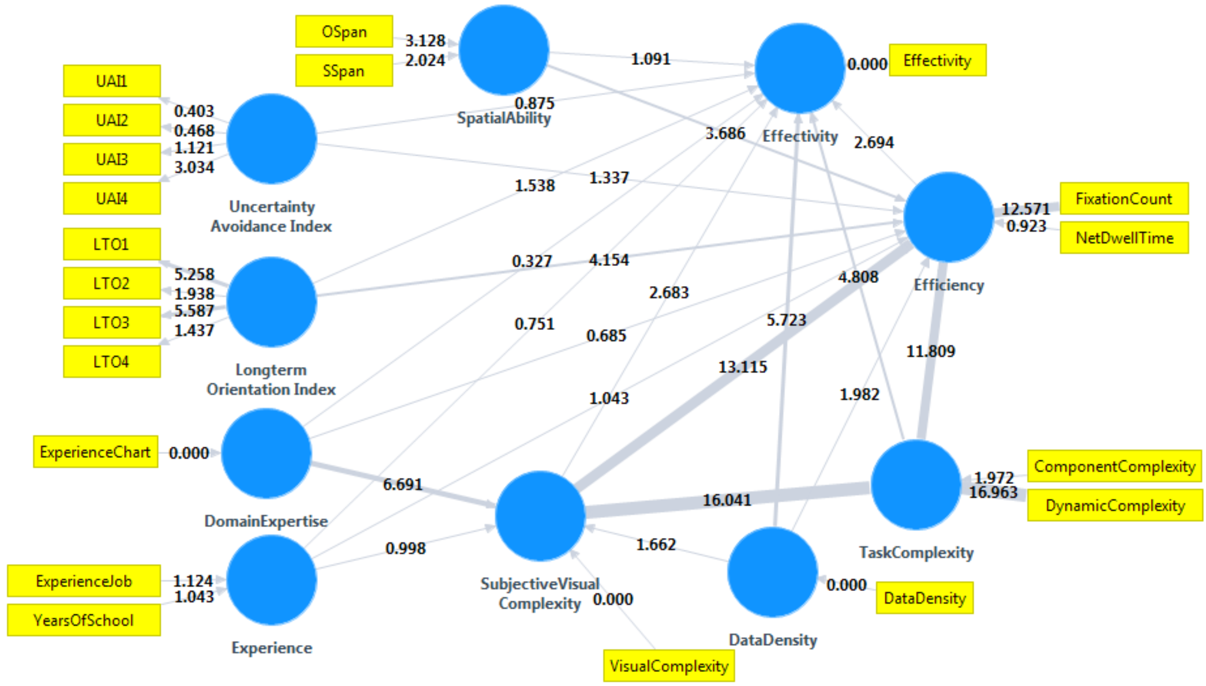

Figure 2: Model for DVS (Consistent PLS after Bootstrapping 1000 subsamples), significant if T score>1,96

6. References

Al-Kassab, J., Ouertani, Z. M., Schiuma, G., Neely, A. (2014). Information visualization to support management decisions, In: International Journal of Information Technology & Decision Making, Vol. 13, No. 2, pp. 407-428

Barat, A. H. (2007). Human perception and knowledge organization: visual imagery, In: Library Hi Tech, Vol. 25, No. 3, pp. 338-351

Beattie, V., Jones, M. J. (2002). Measurement distortion of graphs in corporate reports: an experimental study, In: Accounting, Auditing & Accountability Journal, Vol. 15, No. 4, pp. 546-564

Cho, C. H., Michelon, G., Patten, D. M. (2012). Impression management in sustainability reports: an empiri-cal investitgation of the use of graphs. In: Accounting and the Public Interest, Vol. 12, pp. 16-37

Conati, C. and Maclaren, H. (2008). Exploring the role of individual differences in information visualization, In: Advanced visual Interfaces (Proceedings of AVI'08), ACM, pp. 199-206

Dilla, W. N., Janvrin, D. J. (2010). Voluntary disclosure in annual reports: the association between magnitude and direction of change in corporate financial performance and graph use. In: Accounting Horizons, Vol. 24, No. 2, pp. 257-278

Dilla, W., Janvrin, D. J., and Raschke, R. (2010). Interactive data visualization: new directions for accounting information systems research, In: Journal of Information Systems, Vol. 24, No. 2, pp. 1-37

Dijkstra, T.K., and Henseler, J. (2015). Consistent and asymptotically normal PLS estimators for linear structural equations, In: Computational Statistics and Data Analysis, Vol. 81, pp. 10-23

Falschlunger, L., Eisl, C., Losbichler, H., Greil, A. (2014). Improving information perception of graphical displays – an experimental study on the display of column graphs, In: Full Paper Proceedings of the WSCG 2014 (Conference on Computer Graphics, Visualization and Computer Vision), 2nd -5th June 2014, pp. 19-26

Galletta, D., and Vessey, I. (1991). Cognitive fit: an empirical study of information acquisition, In: Information Systems Research, Vol. 2, No. 1, pp. 63-84

Gelman, A., Unwin, A. (2013). InfoVis and statistical graphics: different goals, different looks, In:

Journal of Computational & Graphical Statistics, Vol. 22, No. 1, pp. 2-28
Hard, N. J., and Vanecek, M. T. (1991). The implications of task and format on the use of financial information, In: Journal of Information Systems, Vol. 5, No. 2, pp. 35-49
Hahn, U. (2014). Experiential limitation in judgment and decision, In: Topics in Cognitive Science, Vol. 6, No. 2, pp. 229-244
Hill, W. Y., and Milner, M. M. (2003). Guidelines for graphical displays in financial reporting, In: Accounting Education, Vol. 12, No. 2, pp. 135-157
Huang, W., and Eades, P. (2005). How people read graphs, In: Proceedings of the APVIS 2005 (Asia Pacific Symposium on Information Visualization – Conference in Research and Practice in Information Technology), Sydney, Vol. 45, pp. 51-58
Kuang, X., Zhang, H. Zhao, S., McGuffin, M.J. (2012). Tracing tuples across dimensions: a comparison of scatterplots and paralell coordinate plots, In: Computer Graphics Forum, Vol. 31, No. 3, pp. 1365-1374
Lord, R. G., and Maher, K. J. (1990). Alternative information-processing models and their implications for theory, research, and practice, In: Academy of Management Review, Vol. 15, No. 1, pp. 9-28
Mostyn, G. R. (2012). Cognitive Load Theory: What it is, why it's important for accounting instruction and research; In: Issues in Accounting Education, Vol. 27, No. 1, pp. 227-245
Parsons, P., and Sedig, K. (2014). Adjustable properties of visual representations: improving the quality of human-information interaction, In: Journal of the Association for Information Science & Technology, Vol. 65, No. 3, pp. 455-482
Parush, A., Hod, A., Shtub, A. (2007). Impact of visualization type and contextual factors on performance with enterprise resource planning systems, In: Computers & Industrial Engineering, Vol. 52, No. 1, pp. 133-142
Peck, E. M., Yuksel, B. R., Harrison, L., Ottley, A., and Remco, C. (2012). Position paper: towards a 3-dimensional model of individual cognitive differences, In: Beyond Time & Errors – Novel Evaluation Methods for Visualization (Proceedings of BELIV 2012), ACM, pp. 1-6
Porat, T., Oron-Gilad, T., and Meyer, J. (2009). Task-dependent processing of tables and graphs, In: Behaviour & Information Technology, Vol. 28, No. 3, pp. 293-307
Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, N. U., Kane, M. J., and Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks, In: European Journal of Psychological Assessment, Vol. 28, No. 3, pp. 164-171
Reinartz,W., Haenlein, M., and Henseler, J. (2009). An empirical comparison of efficacy of covariance-based and variance-based SEM, In: International Journal of Research in Marketing, Vol. 26, No. 4, pp. 332-344
Ringle, C. M., Wende, S., and Becker, J.-M. (2015). "SmartPLS 3." Boenningstedt: SmartPLS GmbH, http://www.smartpls.com.
So, S., and Smith, M. (2004). Multivariate decision accuracy and the presentation of accounting information, In: Accounting Forum, Vol. 28, No. 3, pp. 283-305
Speier, C., Vessey, I., and Valacich, J. S. (2003). The effects of interruptions, task complexity, and computer-supported decision-making performance, In: Decision Sciences, Vol. 34, No. 4, pp. 771-797
Speier, C. (2006). The influence of information presentation formats on complex task decision-making performance, In: International Journal of Human-Computer Studies, Vol. 64, No. 11, pp. 1115-1131
Vessey, Iris (1991). Cognitive fit: a theory-based analysis of the graphs versus tables literature, In: Decision Science, Vol. 22, No. 2, pp. 219-240
Wood, R. E. (1986). Task complexity: definition of the construct, In: Organizational Behavior and Human Decision Processes, Vol. 37, pp. 60-82
Yigitbasioglu, O. M., and Valcu, O. (2012). A review of dashboards in performance management: implications for design and research, In: International Journal of Accounting Information Systems, Vol. 13, No. 1, pp. 41-59
Ziemkiewicz, C, and Kosara, R. (2010). Beyond bertin: seeing the forest despite the trees, In: IEEE Computer Graphics & Application, Vol. 30, No. 5, pp. 7-11.