ABSTRACT
Due to the fact that the quality of decisions is linked to the availability of information and to the ability of the human brain to process this in an effective and efficient way, its selection and representation are of major importance in business communication. Graphs and tables are widely used to transform raw data into a more understandable format, but there are not any empirically tested guidelines that consider the cognition and perception abilities of humans. This paper therefore explores how specific visual designs applied to column graphs influence effectiveness and efficiency by applying the technique of eye-tracking to make an accurate assessment of what the observer is looking at. The tested design elements show significant results and allow the deduction of the following design guidelines for column graphs: do not use a 3D view for depicting two dimensional data, do not use non-zero or broken axes, do show label values, do not use horizontal gridlines or the label axis when showing label values and do align the label values depending on the available space (either horizontally or vertically).

Keywords
Reporting Design, Information Visualization, Information Perception, Graphs, Column Graphs, Eye Tracking

1. Introduction
The constantly expanding amount of available information within and outside a company necessitates the importance of selecting the right information as well as depicting it in a way that reduces cognitive load [May03a] [Muk07a]. One way to highlight and emphasize important information is to visualize the data in the form of graphs, because they tend to draw attention and when properly designed also reduce complexity while enhancing the ability of the human brain to process and remember information [Con08a] [Lur07a] [Pen09a] [Ren03a]. Graphs are seen to be better in supporting the comprehension of large amounts of quantitative information as well as being more effective for detecting trends, patterns or time sequences in comparison to textual and tabular information [Bea08a] [Get12a].

How graphical information effects the cognitive load can be directly influenced by the designer of a graph [Bea08a]. Germane Cognitive Load can be reduced by applying standardized display formats, thus relieving the short term memory which represents a bottleneck in the human information processing [And11a] [Mos12a] [Pec12a]. Additionally, Extraneous Cognitive Load can be reduced with the use of design guidelines that enhance effectiveness and efficiency of graphs by supporting the thinking and working processes of the human brain [And11a] [Pec12a]. Empirically tested and widely distributed design guidelines for graphical displays therefore would not only help to enhance effectiveness and efficiency but would also help to standardize business reports when applied by numerous people responsible for creating reports [Hil03a]. Guidelines in this context are rules about how to design particular elements of a graph.

A way to observe Cognitive Load and therefore to evaluate the effectiveness and efficiency of guidelines is to apply the technique of eye-tracking. Recently, a concentration of research in the field of business communications and in particular in the design of graphical displays in combination with cognitive abilities of users can be found [Con08a] [Eis13b]. Eye tracking methodology is supposed to provide more insights into visual behavior than previous methods used to evaluate graphical displays. This is due to the fact that scanpaths with their saccades and fixations can help to examine and understand the perceptual process in more detail [Gol14a] [Ren03a].
Therefore, this paper sets out to formulate such guidelines for effective and efficient graphical displays under the consideration of the cognitive abilities of humans by applying the technology of eye tracking.

The area of research will refer to the field of Reporting Design, while Reporting Design is defined as the perception-optimized information visualization of graphs and tables in business reports (e.g. management and annual reports) [Eis13a]. For a better understanding, the following paragraphs discuss previous research in this area.

2. Visual Design of Graphical Display

There are many ways to depict data, but there are in particular four types of graphs that are used in business communications – bar, pie, column and line graphs [Bea08a]. All four display formats have been tested in a larger study but due to spatial constraints only column graphs will be discussed in this paper, which also represent the most common graph type found in financial reports.

Column graphs usually have rectangular bars of lengths that are proportional to the magnitudes or frequencies of what they represent. They are primarily used to show changes in data over a period of time whereas the horizontal or x-axis represents the time (shown from left to right) and the y-axis the depicted financial variable (such as sales or earnings) [Bea08a]. Their common design elements can be inferred from figure 1 [Col97a]

![Figure 1: design elements of a column graph](image)

Most of the literature and experiments on graphs and graph types compare graphical and tabular displays. However, there is little literature on the design of graphs and their elements. Furthermore, most of it does not take the ability of the human brain into account. A visual stimulus that does not consider human cognitive architecture is likely to be random in its effectiveness [Gol14a]. Therefore, findings of former studies from other areas and those from studies using other methods have to be analyzed and considered when designing the experiment. The following bullet points represent the most important findings in conjunction with the purpose of this paper and are the basis of the derived hypotheses:

- A well-known problem in this field of research, especially in the communication between a company and external stakeholders, is that graphical displays can be distorted. This means that graphs are designed to give a wrong impression to the observer (intentionally or unintentionally) [Pei02a]. A distortion can be achieved by the selection of different graph types, colors, scales, emphasis, size or other modifications like applying 3D effects [Bea08a] [Pen08a]. Two distortions are said to have the biggest impact on the ability to process information in the right way: mislabeled value axes and three-dimensional displays. Firstly, mislabeled value axes, because they exaggerate the variation of the displayed data [Bea02a]. Secondly, the use of a third dimension for two-dimensional data because it can create visual illusions that can cause biased comparison judgments [Ame10a]. This findings lead to the following two zero-hypotheses:

  H1: A mislabeled value axis does not affect effectiveness or efficiency of information perception.
  H2: The use of a three-dimensional display when depicting two-dimensional data does not affect effectiveness or efficiency of information perception.

- Studies on Working Memory in the context of Information Load state that it is better for the decision making process to display relevant information in close proximity because of limited resources in short-term memory [Par13a]. Findings suggest that labels and figures should be placed in juxtaposition to their data series to emphasize their relationship and reduce cognitive load [Vin93a]. Transferred to the design of column graphs this result suggests marking label values as close as possible to their data marker, which leads to the following hypothesis:

  H3: The distance between the data marker and the data value does not affect effectiveness or efficiency of information perception.

- Invented by Edward Tufte the data ink-ratio suggests that performance is maximal when the ratio is maximal (equal 1) [Inb12a] [Tuf83a]. Applied, this finding suggests that all additional ink that is not crucial to display relevant information should be omitted. However, findings in conjunction with line graphs suggest that when choosing the depiction of the label axis instead of label values, gridlines help with estimating data points (Poggendorff Illusion) [Ren03a]. These two contradicting findings in
literature provide the basis for the following hypotheses:

**H4:** The use of additional ink (label axis and gridlines) when displaying label values does not affect effectiveness or efficiency of information perception.

**H5:** The use of gridlines, either with label values or with a label axis, does not affect effectiveness or efficiency of information perception.

- As humans in the western world are trained since pre-school to read text horizontally, the brain is supported better when data values are aligned horizontally [Laa04a]. However, due to spatial constraints, graphs often show value labels aligned vertically. According to the literature, this should lead to increased process time.

**H6:** The alignment of the data values does not affect effectiveness or efficiency of the information perception process.

These stated findings of related studies are used to conduct experiments which will be explained in section 4.

3. Methodology

Eye Tracking is already used in a wide range of research areas such as marketing, psychology or human-computer interaction [Kör08a] [Mau07a] [Tsi09a]. This is the case because more in depth analyses and objective observations can be conducted through the use of this technology [Gol14a]. In the last few years the field of information visualization also applied this technique and gained new insights [Hua05a] [Rat08a] [Ste13a]. Goldberg and Helfman [Gol11a] stated in their paper that understanding differences in these sequential strategies between various design alternatives is valuable for improving designs to help maximize effectiveness and efficiency, which is the purpose of this paper.

When analyzing eye tracking data fixations, saccades and scanpaths are of particular interest. Fixations are short dwells where the eye can process information, whereas longer fixations are associated with greater visual and/or cognitive complexity and an increased number of fixations can be interpreted to have a negative impact on search efficiency [Gol14a] [Ren03a]. Saccades are quick movements from one fixation to another, which can be used to derive a participant’s attention pattern [Tok13a] and scanpaths represent a string of related fixations and saccades. For analyses, an unduly long scanpath is believed to indicate a non-meaningful representation or a poor layout [Ren03a].

Commonly used measures for the evaluation of eye tracking data are error rates (perception of the information is correct or incorrect = effectiveness) and response time (perception of the information is fast or slow = efficiency) [Hol12a]. These measures cannot only be used to analyze the whole graph but also for special areas of interest (AOIs). AOIs can be designed for each visual target independently by dragging rectangles or other arbitrary shapes [Gol14a]. For statistical analysis the Mann-Whitney U test for effectiveness, the ANOVA for efficiency and Student’s t-test for estimation analysis is recommended and used.

Additionally, error analysis can help with understanding differences and sequential strategies between error and non-error. Furthermore, random guesses can be identified and sorted out. These insights can provide diagnostic information to a designer that exceeds the information provided by the analysis based solely on response time and error rate [Gol11a].

For this paper the process model of Eisl et al. [Eis13b] that explains the phases conception, execution and analyses in detail, is used.

4. Experimental Design

The data presented in this paper is a portion of the collected data from an ongoing larger study [also see Eis13a, Eis13b and Fal14]. Three different experiments are conducted to test the stated hypotheses of section 2. General comments on the used experimental design are stated here, while specific comments can be found in sections below.

In conformity with the used process model participants are asked to read a question on the top of the screen. The questions used are defined in accordance with the literature and verified by a focus group. After reading, participants are supposed to answer the question and then move on to the next slide by clicking. No time constraints are imposed since the effectiveness is more important than the efficiency. This is due to the fact that answering fast but incorrectly is worse than needing more time to find the right information. Randomization of the displays is used and participants are full- and part-time students in business administration.

The study is conducted in a pervasive lab and the height of and distance to the eye tracker is the same for each participant. A headrest is used to ensure minimum head movement. The eye tracking hardware is a binocular eye tracking system with a sampling rate of 120 Hz by Interactive Minds. A nine point calibration and NYAN 2.0 software are used.

In order to evaluate solely the time that is needed to answer the stated questions, AOIs are defined. All results shown in this paper do not include reading
time for the task but allow the experimentee to read the task again during execution due to cognitive constraints in process working memory.

Experimental design - hypothesis 1 (broken axis)
Hypotheses 1 is tested using two displays (see figure 2 & 3). The question used with figure 2&3 is: “Which division generates more return in 2014?”

30 participants (14 female and 16 male; average age: 25) are tested individually in a 15-minute session and divided into two groups. Each participant views either the example with the broken axis or the one with the normal axis to eliminate remembering.

Experimental design - hypotheses 2-5 (3D, value labels, value axis and gridlines)
Hypotheses 2-5 are tested in one coherent experiment. All displays are depicted in figures 4-8 and these two questions are asked to evaluate effectiveness: Q1 “Does the actual sales volume exceed the budget in June?” and Q2 “How high is the actual sales volume in June?”

84 participants (45 female and 39 male; average age: 24) are tested individually in a 30-minute period with a recalibration after 15 minutes. Participants are divided into three groups which leads to maximal two one displays per participant. To reduce remembering when viewing the same data twice, one display is shown before and one after recalibration. Each of the used displays can be compared with each other to calculate the effects on information perception for the stated hypotheses. According to
previous research, the column graph using a three-dimensional display for two-dimensional data should lead to the worst result in effectiveness and efficiency. On the other hand, the display using direct labels and no gridlines or value axis should be best for information perception (low cognitive load because of close proximity of information while maintaining a high data-ink ratio). Direct markings and the depiction of a value axis should lead to a time delay and when no values are shown gridlines should help in estimating the right value.

Experimental design - hypothesis 6 (alignment of label values)

Experiment 3 tests the alignment of label values. Figure 9 and 10 show the two used examples in the experiment. The stated question is: “How high was the turnover in 2013?”

56 participants (32 female and 24 male, average age: 25) are tested individually in a 15-minute session and divided into two groups. Due to the fact, that people are used to reading text horizontally, better results should be achieved when using a horizontal alignment of label values.

5. Results and Discussion

Results - hypothesis 1 (broken axis)

Table 1 shows the results on effectiveness for displays using a non-zero axis compared to one using a zero axis.

<table>
<thead>
<tr>
<th>Effectiveness</th>
<th>Error Rate</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>zero axis</td>
<td>6.7%</td>
<td>&gt; 0.1</td>
</tr>
<tr>
<td>broken (non-zero) axis</td>
<td>26.7%</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: error rates – broken axis

The use of a broken axis does influence information perception negatively when comparing information in two coherent graphs, however, no significance can be found. Individual scanpath analysis shows that participants tend to look at the value axis on the left side more carefully than the right side. Wrong answers result from not looking at the value axis of the right graph.

Table 2 shows the results on efficiency. Almost no differences in response time can be detected.

<table>
<thead>
<tr>
<th>Efficiency</th>
<th>Mean</th>
<th>Median</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>zero axis</td>
<td>15.65</td>
<td>15.83</td>
<td>&gt; 0.1</td>
</tr>
<tr>
<td>broken (non-zero) axis</td>
<td>15.72</td>
<td>14.92</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: response time – broken axis

H1, that a mislabeled value axis does not impact effectiveness or efficiency of information perception, can be confirmed. However, strong indications of a negative effect of a broken axis can be found. Further research in this matter is necessary.

Results - hypothesis 2 (3D)

Error rates for the experiment using the 3D format are calculated for the first question and show significance at a p<0.1 level. The second question is analyzed by estimation error (see table 4).

<table>
<thead>
<tr>
<th>Effectiveness</th>
<th>Error Rate</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D effect</td>
<td>0.0%</td>
<td>&lt; 0.1</td>
</tr>
<tr>
<td>3D effect</td>
<td>10.7%</td>
<td>&lt; 0.1</td>
</tr>
</tbody>
</table>

Table 3: error rates – 2D vs. 3D

Estimations when using a two-dimensional display show an overall-range from 100 to 200 (mean 179) and when using a three-dimensional display from 80 to 210 (mean 148). When calculating the differences of the estimations to the correct answer, the following results for variance and standard deviation can be found:

<table>
<thead>
<tr>
<th>Estimation errors</th>
<th>Variance</th>
<th>Std dev.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D effect</td>
<td>407.57</td>
<td>20.19</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>3D effect</td>
<td>2,652.11</td>
<td>51.50</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: estimation errors – 2D vs. 3D

The average over and underestimation is higher with the use of a three-dimensional display and is significant at a p<0.01 level. This correlation can be extracted from figure 6.

Additionally, high significance (p<0.01) can be found when analyzing response time on two- vs. three-dimensional displays. Response time is 46.5% longer for 3D than for 2D.

Table 5: response time – 2D vs. 3D

H2, that the use of a three-dimensional display when depicting two-dimensional data does not affect effectiveness or efficiency of information perception, can be rejected.

Results - hypotheses 3 and 4 (value labels and additional ink)

Almost no errors are made when giving the answer to the first question (if actual or budget is higher). Scanpath analysis shows that people who answered wrongly confused columns for budget and actual.
When analyzing the second question (the actual height of one column), those graphs without label values show significant higher variations in results on estimations than those with label values (p<0.01). Additionally, significance for task completion time can be derived at a p<0.05 level. The mean task completion time varies between 8.05 and 10.44 seconds or 29.7%.

Results - hypothesis 6 (alignment of value labels)
Results are very similar when using horizontal and vertical aligned label values and do not show significance, neither by evaluating error rates nor by evaluation of response time.

H6, that the alignment of the data values does not impact effectiveness or efficiency of the information perception process can be confirmed.

6. Conclusion and further research
The literature shows that guidelines can help to reduce Extraneous Cognitive Load and Germane Cognitive Load and therefore enhance perceptual speed and accuracy [And11a] [Pec12a] [Ren03a]. The goal of this study is to identify specific design guidelines for column graphs to enhance effectiveness and efficiency in the process of information perception. Three experiments using the technology of eye tracking are conducted to test the following elements: display of a value axis, display of gridlines, display of value labels, 3D effect for two-dimensional data and the alignment of value labels.

In accordance with the literature it can be concluded that graphical distortions have an impact on the information perception process. Non-zero axes inhibit the ability of the human brain to process information correctly, thereby increasing error rates (but not significantly). Further tests using a bigger sample size should be conducted in this area. The use of a three-dimensional display for two-dimensional data increases error rates by 10% and response time by about 50% and shows significance at a p<0.01 at a p<0.001 level respectively.

As the theory of Cognitive Load suggests, data values should be placed in close proximity to the data
series. Placing label values in juxtaposition to its data markers has the highest influence on the process of information perception. If values cannot be placed within the graph due to spatial problems, gridlines do not influence effectiveness or efficiency. These results contradict studies concerning line graphs [Ame10a] [Pre06a] which state that the display of gridlines reduces estimation errors.

It can be concluded that if values are stated above or within the label marker it is best not to use gridlines or display the value axis because it delays information processing time (but not significantly) and decreases the data-ink ratio. The alignment of the values does not have an influence and therefore should be placed in accordance with the available space but preferably horizontally due to reading habits.

Further research on the design of variants of commonly used column graphs (segmented column graphs and grouped column graphs) is necessary. Additionally, the design of other types of frequently used graphical display (bar, pie and line graphs) has to be explored in conjunction with the cognitive abilities of the human brain.

7. References


[Inb07a] Inbar, O., Tractinsky, N., and Meyer, J. (2007): Minimalism in Information Visualization – Attitudes towards Maximizing the Data-Ink


[Mel10a] Meili, H., and Qian, F. (2010): An Evaluation Research on Usability of Taobao’s Homepage and Main Search Engine Based on Eye Tracking; In: the Proceedings of ICSEM,


