Order of Magnitude Markers: An Empirical Study on Large Magnitude Number Detection

Rita Borgo, Joel Dearden, and Mark W. Jones

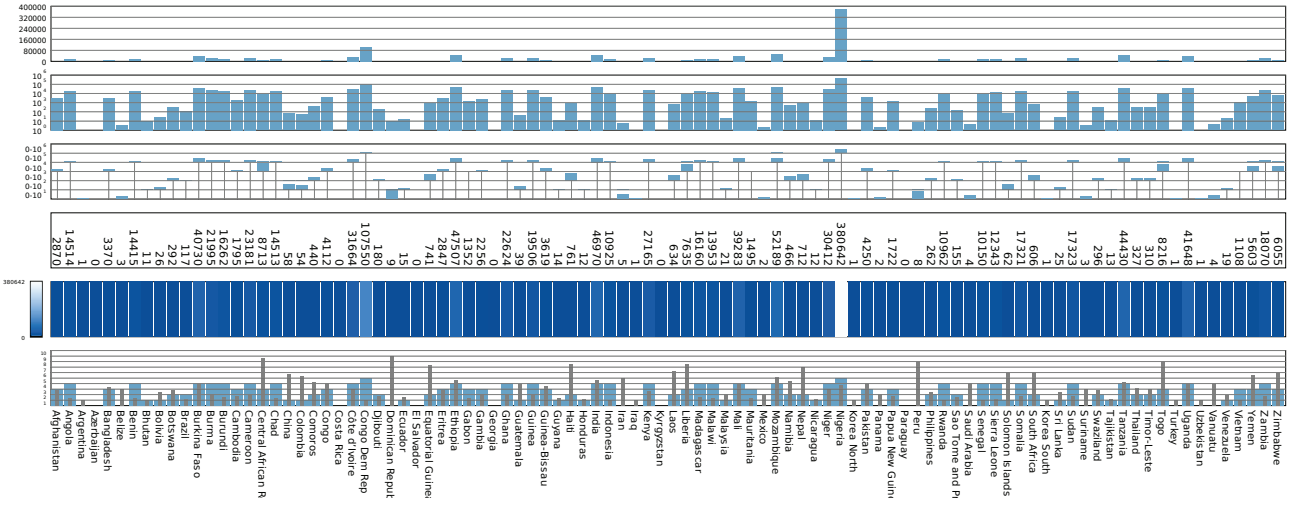


Fig. 1: Malaria cases 2010 [17] using from top to bottom, Linear bar charts, Logarithmic scale, Scale-stack bar charts, Text, Color, and Order Of Magnitude Markers (OOMMs).

Abstract—In this paper we introduce Order of Magnitude Markers (OOMMs) as a new technique for number representation. The motivation for this work is that many data sets require the depiction and comparison of numbers that have varying orders of magnitude. Existing techniques for representation use bar charts, plots and colour on linear or logarithmic scales. These all suffer from related problems. There is a limit to the dynamic range available for plotting numbers, and so the required dynamic range of the plot can exceed that of the depiction method. When that occurs, resolving, comparing and relating values across the display becomes problematical or even impossible for the user. With this in mind, we present an empirical study in which we compare logarithmic, linear, scale-stack bars and our new markers for 11 different stimuli grouped into 4 different tasks across all 8 marker types.

Index Terms—Orders of magnitude, bar charts, logarithmic scale

1 Introduction

When data covers a large range of magnitudes, bar charts require a quantization algorithm to map quantities onto a limited pixel height. Since the dynamic range is far greater than that of the number of pixels available, there can be a large amount of quantization error. Many quantities will be mapped to one pixel, or equivalently, a configuration of the visual representation can represent many quantities. Fig.1 demonstrates this effect. To accommodate the largest value, the linear bar chart representation results in many non-zero values mapping to zero pixels.

Fig.1 also demonstrates the problem with color perception and reading off number values from this single-hue color scale. Vietnam (1108) looks to have the same color as Venezuela (19), when it has over fifty times the number of cases. The remaining charts fare better, where it is possible to see the difference between those two countries. Many visualization tasks involve comparing numbers across a 1D axis as in Fig. 1 or across 2D representations (e.g. data represented according to a color scale in a choropleth map). Asking questions such as *How much larger is one value compared to another* add a whole new complexity to the task.

Data representation using color has been the focus of prior user-studies. The classic work of Cleveland and McGill [5] demonstrated that color was the least effective at presenting data, and we also find that in this work.

The most recent work in the area proposed scale-stack bar charts [10] (also depicted in Fig. 1). A user study was carried out comparing their novel visualization to that offered by linear and logarithmic bar charts. In this paper we propose a set of new visual representations, which we call Order Of Magnitude Markers (OOMMs), that visualise the significand and exponent separately, but within a single marker visualization.

We explore the ability of our new visual encodings to support the significand-exponent separation in the context of high dynamic range values. We analyze performances towards tasks requiring comparison and estimation of individual data values as well as tasks which involve explicit calculation and comparison across more complex displays. Our results show this separation allows a 10× increase in the resolving power of our markers compared to other approaches. We present an empirical study demonstrating this effect and like the work by Hlawatsch et al. [10], we compare it against linear and logarithmic

- *Rita Borgo, Swansea University, E-mail: r.borgo@swansea.ac.uk*
- *Joel Dearden, Swansea University, E-mail: j.r.dearden@swansea.ac.uk*
- *Mark W. Jones, Swansea University, E-mail: m.w.jones@swansea.ac.uk*

- *All authors contributed equally to this work. Dearden was funded by RIVIC and is funded by Leverhulme grant RPG-2013-190.*

Manuscript received 31 Mar. 2014; accepted 1 Aug. 2014. Date of publication 11 Aug. 2014; date of current version 9 Nov. 2014.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.
Digital Object Identifier 10.1109/TVCG.2014.2346428

bar charts. We also compare performance against text, color and other visualization types, and introduce new tasks compared to Hlawatsch et al.

After a study of current techniques and evaluations in Section 2, we introduce our new Order of Magnitude Markers in Section 3. The design and analysis of our user study are provided in Section 4 with a discussion of the findings in Section 5.

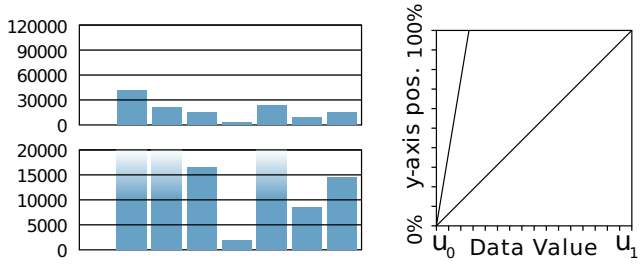
2 LITERATURE

Approaches for presenting data with large dynamic range is an issue discussed in the literature and on newsgroups and the blogosphere. Here we present the current solutions to this problem:

Re-expression of data. Tukey ladder of powers [25](pp88-90) suggests re-expressing the data using powers such as $-1, -\frac{1}{2}, log, +\frac{1}{2}, +1$ or higher. Plotting the data on a logarithmic axis allows small and large magnitude values to be expressed together. Exponential growth (e.g., population growth in the 18^{th} and 19^{th} centuries) can exhibit a linear behaviour when plotted on a logarithmic axis. This is an often used approach for scientific data, but it is difficult to judge values and relative values on a power scale. We explore this aspect using a logarithmic scale in our user study.

Dual scale charts. Isenberg et al. [11] report an empirical study on the use of dual scales within charts. They present an excellent classification of the different types of charts through the consideration of a transformation function (from data space to display space) with an exploration of each type they have identified. From their user study (15 participants) they recommend cut-out charts as the most effective. A cut-out chart consists of a full size context chart with a zoomed section placed below. The study focused on transforming the x-axis rather than our study which gives regard to large magnitude data on the y-axis. We borrow terminology from their work [11] to describe the next two types.

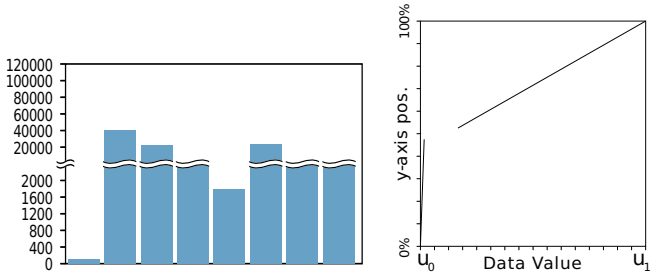
Panel charts. Data is divided into two panels with two different scales [19] (Fig. 2a). One scale is chosen to allow the smaller magnitude data to be compared and the larger data saturates the scale. The other scale allows the depiction and comparison of the larger data both intra-cluster and to the smaller magnitude data. To describe the mapping we borrow the terminology from Isenberg et al. [11] based on the taxonomy by Leung and Apperley [14]. The top panel of (Fig. 2a) displays the data using a linear chart with a scale able to show the whole data range. Fig. 2b shows the transform from input data to available range on the y-axis (line with slope=1). The bottom panel displays a magnified region of the data focused on the smaller values in the data. The steeper transform of Fig. 2b is used. Values that have exceeded the range are depicted in this case using a fade towards the top of the bar. Panel charts still suffer from problems when large magnitude data is used. If large values are clustered closely, it is difficult to distinguish them in the overview panel, and they saturate the focus panel. We do not test panel charts due to this problem.



(a) Panel chart example. (b) Value to chart y-axis transform.

Fig. 2: A panel chart using a small subset of the malaria data (left, country labels omitted to save space). The transform from data value to position on the chart y-axis (right).

Broken axis. Broken axis charts are suitable for situations when the data contains a clusters of high magnitude and low magnitude values that need to be compared intra-cluster. A scale can be chosen that allows all the small values to be displayed. A gap is created in the axis and the large values then appear above this break (Fig. 3a). Fig. 3b shows the transformation from input data to available range on the y-axis. Note the small break on the y-axis of the transform to account for the white-space gap. The method is successful at intra-cluster comparisons when there are distinct clusters at high and low magnitude. It fails in situations where the full range of data is present since values will fall in the cut-out range and not be presented accurately in the chart as in this example. It is also difficult to compare values and relationships across the axis change. Since we are using data that covers the whole range we do not test broken charts.



(a) Broken chart example. (b) Value to chart y-axis transform.

Fig. 3: A broken chart using the same small subset of the malaria data (left, country labels omitted to save space). The transform from data value to position on the chart y-axis (right).

Depiction as area. Using area (or volume) to represent the numbers can allow a wider dynamic range of values to be displayed simultaneously. For example, 2D scatter-plots can depict a third dimension by mapping it to the radius of the plot points (becoming Bubble charts – see figure 2 of the study by Heer and Bostock [9] and Gapminder [20] for example visualizations). User studies [5, 9] demonstrate that the apparent relationship between values is often underestimated. In essence, this approach is largely similar to re-expressing the data using a power and plotting against a 1D scale, often leading to better perception [5]. Since this method has already been evaluated and found to be less effective than using a logarithmic scale, we omit it from the user study.

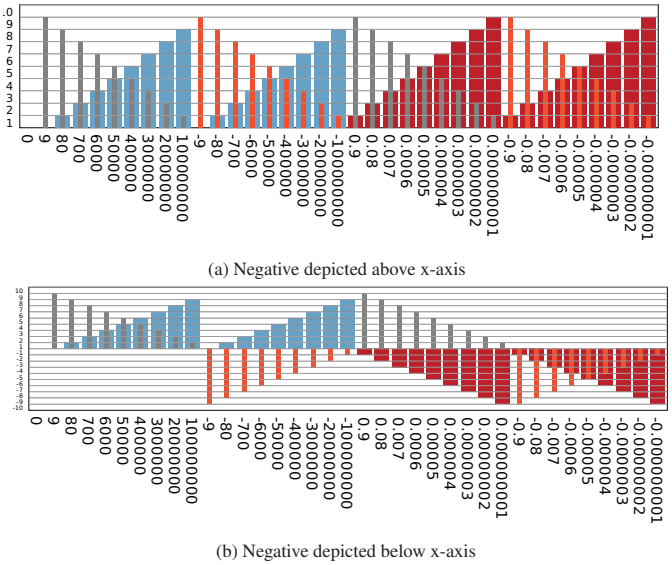
Scale-stack bar charts. The approach related closest to ours [10] represents each number at multiple scales. It differs from panel charts in that it is not limited to two arbitrarily chosen scales, but follows a logical series (power) for each successive scale. Within each scale the number is represented linearly. We compare our work with scale-stack bars in our user study and report our findings in Section 5.

Apart from exploring the question about which approach to follow, there are other ways in which charts or plots can be improved to allow better interpretation of content. These either draw on aesthetic properties or have arisen through user studies trying to discover how people perceive relationships within different representations.

Fink et al. [7] address the problem of selecting aspect ratios for scatter plots by maximizing a measure based on Delaunay triangulation. They derive their properties by comparing to user selected scatter plots in an empirical study. They suggest asking users to solve certain tasks as a future study to determine their effectiveness.

Borkin et al. [4] conduct a user study with Mechanical Turk to discover what makes a visualization memorable. Visualisation types are classified (e.g., Area, Bar, Circle, Map), and a classification appears in their supplementary material.

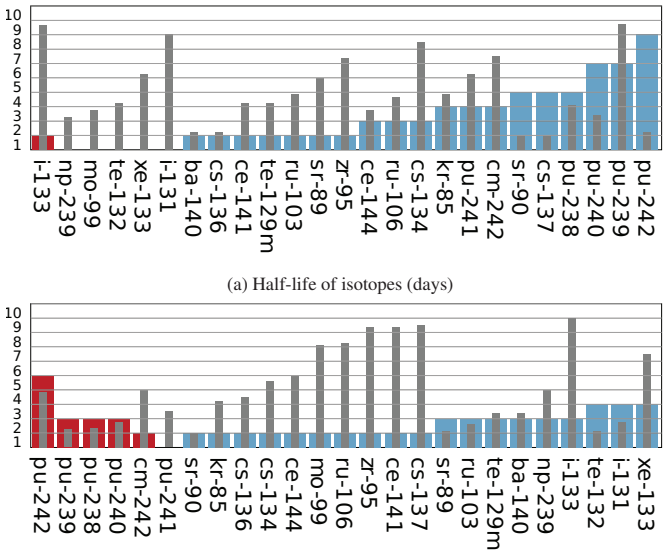
Kong and Agrawala [12] present a system that is able to reverse engineer data sets from existing bar charts and then overlay additional



(a) Negative depicted above x-axis

(b) Negative depicted below x-axis

Fig. 4: Two alternative methods of displaying negative exponents and significands for the OOMM1 marker: (a) displaying sign by colour and (b) displaying sign by colour and direction.



(a) Half-life of isotopes (days)

(b) Activity released of isotopes (PBq)

Fig. 5: Plot of data relating to the Chernobyl accident using the OOMM1 candidate marker: (a) half-life of each isotope and (b) activity released by isotope.

cues to aid reading. Cues include reference structures (e.g., gridlines), redundant encodings (e.g., numerical labels) and summary statistics. The key insight is that users are given control over the additional cues to suit their current stage of processing.

Mackinlay [15], by creating a system to chose automatically good graphical representations for data, highlights expressiveness and effectiveness as important aspects the tool must achieve. Expressiveness requires that all the facts in the set must be encoded, and that no additional information is falsely imparted by the encoding. Effectiveness requires "encoding more important information more effectively". Since this is reliant on the perceiver's capability, some notion of the effectiveness of different visualisation techniques is required. Mackinlay reaches for Cleveland and McGill's [5] work that reports that position and length are the top two representations for accuracy of quantitative perceptual tasks with colour being one of the least accurate. Cleveland and McGill [5] also refer to Weber's law with an example of framed bars to demonstrate they can be useful in distinguishing similar lengths. They include framed rectangle charts for low dynamic range data (a US map with murder rates per state on a 0-16 range). They also report problems of shading the whole states leads to large areas dominating, and false clustering (of large states). By using the framed rectangle charts and exploring quantization they are able to reduce that domination and other clusters become clearer. Speckmann and Verbeek [21] also tackle the problem of map region size obscuring data interpretation by introducing necklace maps. Symbols (e.g., circles) with areas representing the data for each region are located around the map avoiding overcrowding. We use frames around our representations throughout the user study, and it would be interesting additional work to test our visual makers in the context of visualizing geographical data.

There should also be considerations of "chartjunk" [24, 6], in the form of frames, scales and tick marks. Talbot et al. [22] address that problem by demonstrating an improved solution for generating and labeling axes. Our user study is designed to test more functional aspects of the representations such as whether numbers can be perceived and compared correctly, rather than whether aesthetically pleasing representations perform better or worse. Therefore we treat each representation in a similar manner and use scales, tick marks and frames. Our user study does not compare embellished against non-embellished forms. Aesthetically pleasing versions of our visualisations could be designed once effectiveness is established.

3 Marker Design

The motivation for this work came from a need to allow the visual comparison of numbers with a large range of magnitudes across a network. Text was problematical as due to the large range they required great effort to compare. Mapping to colour resulted in a situation where large values would not allow the inspection and intra-cluster comparison of low values. Cleveland and McGill [5] suggest framed bars for such a situation, but using a linear and logarithmic mapping still suffered from the problems detailed above. We deconstructed a number into its significand and exponent constituents and visualised those separately. We found that to be effective in the task of comparing numbers. The effectiveness in this case was our qualitative view on ease of use, speed of use and accuracy. This paper reports our work to determine quantitatively just how effective this new type of visualization is. The tasks (described later) reflect our own usage and generally what is required from such visualizations – comparing numbers across the visualization (ratio), determining values and locating the highest (or lowest).

Our design aims were:

- We wanted a visual marker flexible enough to work in a limited amount of space. This meant it would be usable in both a chart format, where lots of markers would be shown together (see Fig. 1), and separately, for example distributed across a network or map (Task B will demonstrate this).
- It should allow the viewing of all data in a set of numbers regardless of its order of magnitude.
- It should allow the visualization of both positive and negative numbers.
- It should have a greater resolving power than existing techniques.

Like scale-stack bars [10] we consider the significand and exponent of the numbers we wish to represent, but in our design we explicitly attribute them to different elements in the visualization. We use normalised scientific notation $A \times 10^B$ where $1 \leq A < 10$ and $B \in \mathbf{Z}$.

Our markers evolved over time, and ultimately we settled on OOMM1 (see Figs. 1, 4 and 5) as satisfying all of the criteria. For our user study we restrict consideration to positive numbers primarily to enable fair comparison to the logarithmic scale. It also simplifies some of the tasks. For example, if negative numbers are included, some users may make mistakes on the ratio test even if they have suc-

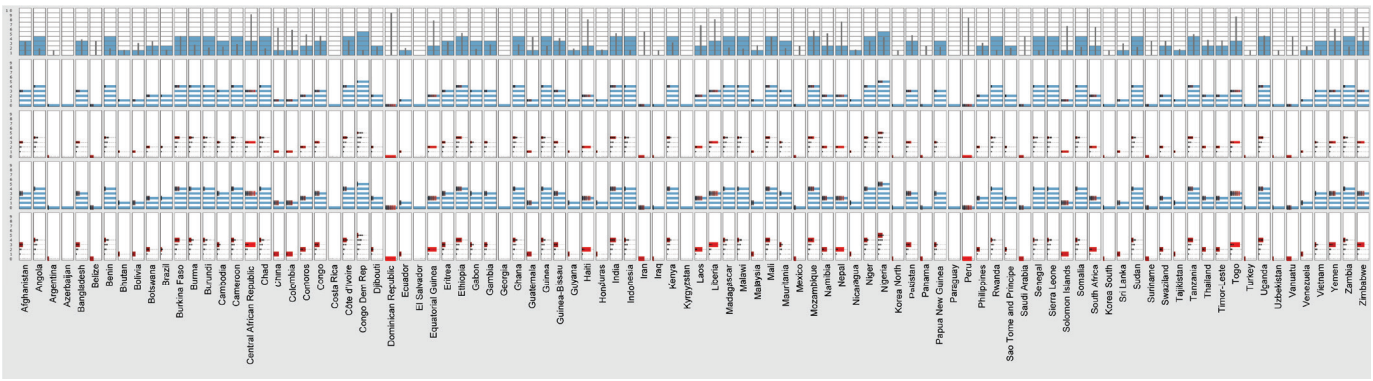


Fig. 6: The five tested markers.

cessfully identified values from their visualization. Therefore, when deciding on markers for the user study (OOMM2-OOMM5) we did not design them with negative significands or exponents in mind, but they could be extended in a similar way to OOMM1. In future user studies we would recommend using OOMM1 only, and testing both variants of representing negative values (Fig. 4).

We developed software to experiment with different marker designs. We privileged designs which favoured pre-attentive processing [8] by either using minimum number of colors (e.g. min 2, max 3), guaranteeing color and/or shape distinctiveness, associating different colors to different shapes. We also favoured designs with low visual complexity, where complexity was defined as the amount of detail and intricacy in the marker visual representation [16], and computed as the number of distinct geometrical features and colors plus the number of overlapping graphical elements [5].

A number of prototype markers were implemented to explore different designs showing the two-dimensional scientific number notation. The final five marker designs developed for the user study are shown in Fig. 6.

The first marker design OOMM1 displays the integer exponent, B, using a stack of B coloured slabs. The real significand, A, is shown using a narrow grey bar which is on a linear scale from 0 to 10 from bottom to top of the marker. The aim is to provide an overview of the size of the number using the wide blue bars and then provide detail on demand with the narrow significand bar. Throughout we follow the big effect/small effect convention so that the exponent which has the biggest effect on the size of the number is represented by the wider blue bars, and the significand by the smaller bars.

The second marker design OOMM2 uses a stack of wide blue slabs, with intra-slab spacing, to represent order of magnitude. The significand is embedded in the top bar as a row of coloured blocks. The total number of blocks in the horizontal row is equal to the significand. If the last segment is not complete then it represents the fractional part of the significand. OOMM4 was derived from this and is identical except that the horizontal row for the significand has taller blocks overlapping the intra-slab spacing.

The third marker design OOMM3 uses the same idea as OOMM2 but the previous slabs that represented the order of magnitude have been replaced by rows of dots. The significand is shown by a horizontal band of colour that overlaps a number of dots equal to the significand. Again fractional parts of a dot overlapped indicate any fractional part of the significand. OOMM5 is derived from OOMM3 and is identical except for the taller band of colour that represents the significand.

The OOMM-type markers can be used to represent both positive and negative significands and positive and negative exponents. This is illustrated by the very large and very small positive and negative numbers shown using two variants of the OOMM1 markers in Fig. 4. Other representations (linear, scale-stack bars) also extend to negative ranges by extending the y-axis downwards. Our representation can also offer a more compact approach by utilising color above the x-axis to indicate negative values. Fig. 4a represents the sign of the exponent and the sign of the significand by colour only, while Fig. 4b uses colour and direction to show the same information. In this case the range of significands that can read off the same chart ranges from -9.99 to +9.99 for all marker types and the range of exponents visible ranges from -10 to +10. This can be helpful when viewing real data sets such as the half-life of isotopes from the Chernobyl accident [26] and the total activity released by those isotopes. Both these data sets contain numbers with a very wide range of orders of magnitude, for example, the half-life dataset ranges from 0.867 days to 137,240,000 days. These two data sets were previously visualised using the scale-stack Bar chart [10] to good effect, though using different units on the y-axis in both cases. Fig. 5 displays this data using OOMM1 markers where negative exponents are indicated by colour and direction. Both marker types allow exploration of the entire data set and comparison between wide ranging values. The OOMM1 provides the entire marker height over which to display the significand and so potentially facilitates more exact comparisons. The OOMM1 also makes clearer the exponent being positive or negative. Examples of all the markers and a detailed description of their construction is provided in the supplementary material.

4 User Study Design

We consulted with researchers in Social Sciences to identify suitable tasks to evaluate our new visual representations. Our aim was to choose a set of tasks that was common during the analytical process of chartered information [23, 1], and that would still address questions of potential interest. We identified four major tasks: target identification, trend detection, ratio and magnitude estimation, with the latter task relevant for analysis involving local values estimation. During the study the four tasks were generically referred to as: Task A (magnitude estimation), B (target identification), C (ratio estimation) and D (trend analysis). For conciseness of writing the same notation is used for the remainder of the document.

Magnitude Estimation - Task A. For a question, the user is presented with one instance of a marker type representing a stimulus. The object of the task is to estimate the quantitative value of the marker. The user enters their estimated value via a text box and clicks on next at which point the time to click is stored. There are two sets of stimuli for this task which are the same for all participants. Therefore there are 16 questions. Questions are presented to users in random order within the task. The stimulus and question order are stored in the output file so the particular study a user experienced could be reconstructed.

Target Identification - Task B. For a question, the user is presented with thirty instances of a marker type (three rows of ten) based on the current stimulus. The object of the task is to pick the markers representing the largest and second largest numbers. A stimulus with a particular marker type is presented to the user at which point a timer is started. The time to first click to select the largest marker is stored. The time to second click and also the interval between clicks is also

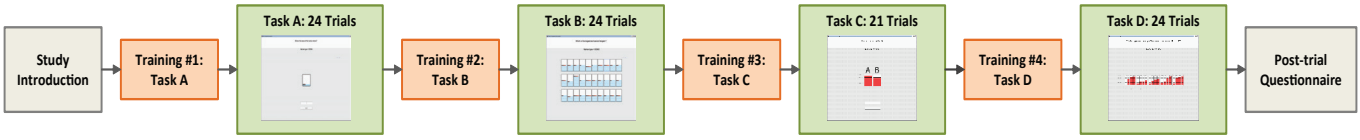


Fig. 7: Test phases flow-chart.

stored. There are three sets of stimuli for this task which are the same for all participants. Therefore there are thirty questions within the first task. Questions are presented to users in random order within the task. Positions of the thirty markers are also randomised. The stimulus, question order and positional order are all stored in the output file so the particular study a user experienced could be reconstructed. The aim of the task is to check whether the marker representation is effective at allowing users to compare numbers across a 2D space.

Ratio Estimation - Task C. For a question, the user is presented with two instances α, β of a marker type ($\alpha > \beta$). The object of the task is to estimate the ratio $\frac{\alpha}{\beta}$. A stimulus with a particular marker type is presented to the user at which point a timer is started. The user enters their estimated ratio via a text box and clicks on next at which point the time to click is stored. There are three sets of stimuli for this task which are the same for all participants. Therefore there are thirty questions. Questions are presented to users in random order within the task. The stimulus and question order are stored in the output file so the particular study a user experienced could be reconstructed.

Trends analysis - Task D. For a question, the user is presented with five company results. Each company result is made up of a chart of four years fictitious profits represented using the marker type. Within each chart, the profits are rising in each year. The object of the task is to determine the company with the highest growth of profits over the entire four years. The user clicks on the company chart that they determine to have the highest growth at which point the time to click is stored. There are three sets of stimuli for this task which are the same for all participants, each consisting of 20 numbers. Therefore there are thirty questions. Questions are presented to users in random order within the task. The company order is also randomised within question. The stimulus, company order and question order are stored in the output file so the particular study a user experienced could be reconstructed.

4.1 Stimuli Design

The stimuli set was designed as follows: for each number $a \times 10^b$ both significand (a) and exponent (b) were created randomly with $0 \leq a, b \leq 10$, $a \times 10^b \in \mathbf{Z}$ and $a \times 10^b > 1$. The integer check was in place to guarantee fairness with text based representation (e.g. avoid the additional complexity of reading floating point numbers). Zero and one were not used so that $log(a \times 10^b) > 0$ and defined. For the target selection task, where participants were asked to choose largest and second largest elements, the largest element was designed as an outstanding outlier, e.g., a number with an exponent a maximum of four times greater in magnitude than the exponent of the distractors. The second largest element, and all the distractors, were within two exponent levels.

4.2 Pilot Study Design

During our pilot study the last three tasks (B, C and D) were tested against all ten markers, namely, linear, logarithm, colour, text, scale-stack bar and OOMM1-OOMM5. We use the following terms consistently in the following. A *stimulus* is a set of numbers to be represented to the user (three for each task, nine in total). A *marker* is one of the ten markers just mentioned. A task is one of the three tasks described in the following paragraphs. A *question* is a triple consisting of stimulus, marker, task (90 in total). Questions are grouped by task. Users had access to a calculator throughout the task if they felt it necessary. Users were told that accuracy was of primary concern and speed would also be measured. We conducted some training using a presentation. Then before each task, there is a set of training questions using each marker type and giving hints as to the correct answer to fully prepare users for the study.

4.3 Pilot Study Analysis

The pilot study was carried out with six participants. Three were the co-authors, two additional were PhD students, and one further was an expert at user studies.

We had one concern that during the task it may become apparent that the same three stimuli were being used for each marker type within a task. The three non-author participants were unaware of this factor. The user study expert participant confirmed this therefore would not be a problem.

4.4 Main Study Design

The pilot study raised several issues that were addressed for the main study. Of primary concern was the duration of the study at typically 60-75 minutes including training, leading to user fatigue. This led us to seek ways to reduce the study time. Our analysis showed that colour performed poorly, which agrees with previous work, e.g., see Mackinlay [15]. It also showed that for the ratio task, it was unnecessary to include the text markers since users were able to read off the two numbers and calculate the ratio (often using the calculator) reliably and quickly. We also removed OOMM2 which at that point was the poorest performing of our new markers. This left eight markers for the trend analysis and target identification tasks (24 questions each), and seven markers for the ratio task (21 questions). We also inserted a fourth task, namely task A, at the start of the study as a measure of how accurately participants understood the new visual representations. This check allowed also for the analysis of possible outliers, if present, and unreliable data. Luckily no data were deemed unreliable, this probably due to the participants selection process.

The final study therefore consisted of fours tasks, eleven stimuli and eight markers. Supplementary material contains the presentation used for participant training.

4.5 Experimental Setting

Participants. A total of 21 participants (2 females, 19 males) took part in this experiment in return for a £10 book voucher. Participants belonged to both the student and academic communities. Pre-requisite to the experiment was a basic knowledge of Calculus and familiarity with concepts such as graphs and logarithmic scale, for this reason recruitment was restricted to the departments of Mathematics, Physics, Computer Science and Engineering, and in the case of students, level 2 and above. Ages ranged from 20 to 33 (Mean=23, SD=3.6). All participants had normal or corrected to normal vision and were not informed about the purpose of the study prior to the beginning of the session.

Apparatus. The visual stimuli and interface were created using custom software written in C♯ with DirectX as the graphics library. Experiments were run using Intel 2.8GHz Quad-Core PCs, 4GB of RAM and Windows 7 Enterprise. The display was 19" LCD at 1440×900 resolution and 32bit sRGB color mode. Each monitor was adjusted to the same brightness and level of contrast. Participants interacted with the software using a standard mouse at a desk in a dimmed experimental room.

Procedure. Fig. 7 illustrates the experimental structure. The experiment began with a brief overview read by the experimenter using a predefined script. Detailed instructions were then given through a self-paced slide presentation. The presentation included a description of the study and also a briefing on how to interpret each of the markers, participants also received a colour copy of the presentation for reference during the study if desired. The experiment was divided into 4 main tasks. Within Task B and D each participant completed a total of 24 trials, Task A featured 16 trials. Task C featured instead a total of 21 trials (as the text marker was removed as discussed in section 4.4). The 4 tasks were always completed in sequential order. Given the nature of the experiment each section assessed a different aspect of the analytical process. Maintaining the same section order for each participant meant that each participant experienced similar experimental conditions. This allowed for a robust analysis of the responses. Randomness was introduced at trial level. Within a task, trials were randomized to avoid learning effects. A training section preceded each task to familiarize the participant with both task and markers. For Task A, B and D and a total of 8 practice trials (one per marker) were completed, for Task C a total of 7 trials (one per marker) were completed. Each training trial included a feedback to the participant regarding the correct answer. Participants were invited to take a short break at the end of each task, if needed. Participants were invited not to take breaks once a task had been commenced. The study was closely monitored and participants abode to the study requirements. When all tasks had been completed each participant completed a short debriefing questionnaire. The purpose of the questionnaire was to obtain comments and recommendations concerning both the experimental procedure and design and usability of the OOMMs. Questionnaires were accompanied by a short post-testing interview. Due to the qualitative nature of the feedback, results were used to support the discussion of quantitative results gathered from the testing phase.

4.6 Main Study Analysis

In our analysis we mainly considered the effect of task vs. marker type. We focused on a comparison of the newly designed markers performances against state of the art markers as this was our primary research question. We made no distinction in terms of the use of markers across different tasks by participant. For Task A and C correct answers were given a 20% error tolerance, therefore in a second phase of analysis we also looked at the performances, in these particular tasks, for varying level of tolerance. Section 4.6.1 describes the overall results. To perform our analysis, as the data is not always normally distributed, we used a non-parametric Friedman test with a standard significance level $\alpha = 0.05$ to determine statistical significance between conditions. Post hoc analysis was performed via separate Wilcoxon signed rank-tests on combinations of related groups for which significance was found. The significance threshold was adjusted using a Bonferroni correction, with corrected significance value of $\alpha = 0.002$. For cases in which both time and error data produced significant results, we performed a correlation analysis over all participants and tasks to see if there was a significant negative correlation, which would indicate the presence of a trade-off effect (e.g. less time led to more errors). When data showed a marked deviation from normality we adopted a non-parametric Spearman correlation measure instead of the traditional Pearson's.

Magnitude Estimation: Task A. Performance in Task A, as a function of marker type, is summarized in Fig. 8. A noticeable variation is visible across markers, the Friedman's test showed a significant main effect in both accuracy ($\chi^2 = 42.75$, $p \ll 0.05$) and response time ($\chi^2 = 50.143$, $p \ll 0.05$). A closer analysis showed:

- Mean Accuracy
  - OOMM1 markers (mean = .88) were significantly more accurate than logarithmic (mean = .52) ($p \ll 0.002$);
  - OOMM3 markers (mean = .88) were significantly more accurate than logarithmic (mean = .52) and scale-stack bar (mean = .69) ($p \ll 0.002$);
  - OOMM4 markers (mean = .88) were significantly more accurate than logarithmic (mean = .52) ($p \ll 0.002$).
  - OOMM5 markers (mean = .84) were significantly more accurate than logarithmic (mean = .52) ($p \ll 0.002$), and scale-stack bar (mean = .69) ($p \ll 0.002$).
- Mean Response Time
  - OOMM1 markers (mean = 19.63) were significantly faster than linear (mean = 34.45) ($p \ll 0.002$) and OOMM4 (mean = 19.14) ($p = 0.002$);
  - OOMM3 markers (mean = 17.8) were significantly faster than linear (mean = 34.45), logarithmic (mean = 25.88), scale-stack bar (mean = 29.49) ($p \ll 0.001$).
  - OOMM4 markers (mean = 19.14) were significantly faster than linear (mean = 34.45) and logarithmic (mean = 25.88) ($p \ll 0.002$). OOMM4 markers were significantly slower than text (mean = 13.48) ($p \ll 0.002$).

No other significant differences were found.

Target Identification: Task B. Performance in Task B, as a function of marker type, are summarized in Fig. 8. A noticeable variation is visible across markers, the Friedman's test showed a significant main effect in both accuracy ($\chi^2 = 82.205$, $p \ll 0.05$) and response time ($\chi^2 = 78.444$, $p \ll 0.05$). A closer analysis showed:

- Mean Accuracy
  - OOMM1 markers (mean = .82) were significantly more accurate than linear (mean = .41), logarithmic (mean = .42) and scale-stack bar (mean = .52) ($p \ll 0.002$);
  - OOMM3 markers (mean = .93) were significantly more accurate than linear (mean = .41), logarithmic (mean = .42), scale-stack bar (mean = .52) and OOMM1 (mean = .82) ($p \ll 0.002$);
  - OOMM4 markers (mean = .93) were significantly more accurate than linear (mean = .41), logarithmic (mean = .42), scale-stack bar (mean = .52) and OOMM1 (mean = .82) ($p \ll 0.002$);
  - OOMM5 markers (mean = .88) were significantly more accurate than linear (mean = .41), logarithmic (mean = .42) and scale-stack bar (mean = .52) ($p \ll 0.002$).
- Mean Response Time
  - OOMM1 markers (mean = 17.95) were significantly slower than linear (mean = 11.22) and significantly faster than scale-stack bar (mean = 24.71) ($p \ll 0.002$). As OOMM1 markers are slower but more accurate than linear, a correlation test of time vs. accuracy was performed to detect any trade-off effect. The correlation result was non-negative (0.043) meaning that faster responses did not led to more errors (and vice-versa);
  - OOMM3 markers (mean = 14.56) were significantly slower than linear (mean = 11.22) ($p \ll 0.002$), and significantly faster than OOMM1 (mean = 17.95), logarithmic (mean = 19.42), scale-stack bar (mean = 24.71), text (mean = 18.98) and OOMM4 (mean = 16.48) ($p \ll 0.002$). As OOMM3 markers are slower but more accurate than linear a correlation test of time vs. accuracy was performed to detected any trade-off effect. The correlation result was non-negative (0.032) meaning that faster responses did not led to more errors (and vice-versa);
  - OOMM4 markers (mean = 16.48) were significantly slower than linear (mean = 11.22) and OOMM3 (mean = 14.56) ($p \ll 0.002$), and significantly faster than scale-stack bar (mean = 24.71) and OOMM5 (mean = 14.19) ($p \ll 0.05$). As OOMM4 markers are slower but more accurate than linear, a correlation test of time vs. accuracy was performed to detect any trade-off effect. The correlation result was non-negative (0.010) meaning that faster responses did not led to more errors (and vice-versa);
  - OOMM5 markers (mean = 14.19) were significantly slower than linear (mean = 11.22) and OOMM4 (mean =

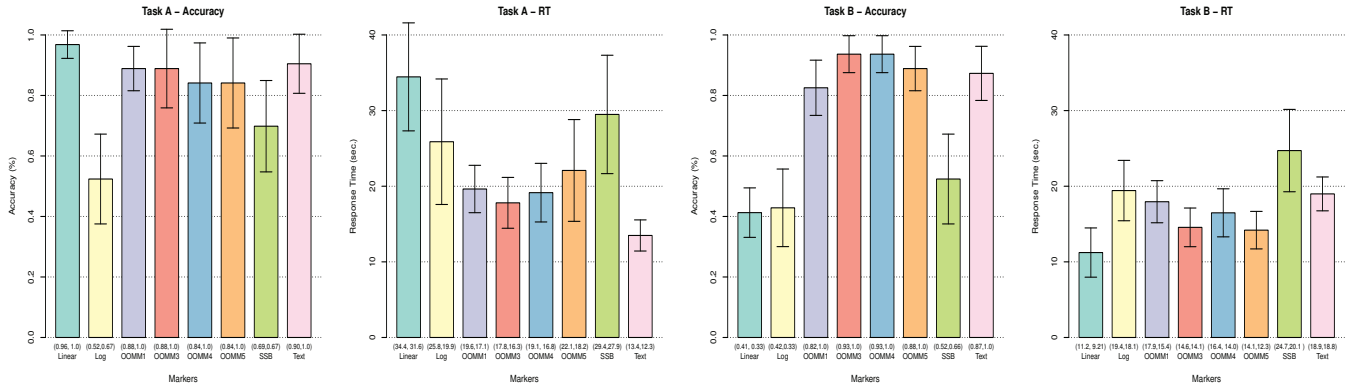


Fig. 8: Analysis of performance results for Tasks A and B, (mean, median) values are indicated below each bar. Error bars show 95% confidence intervals. Bars are colour-coded using RColorBrewer package [18].

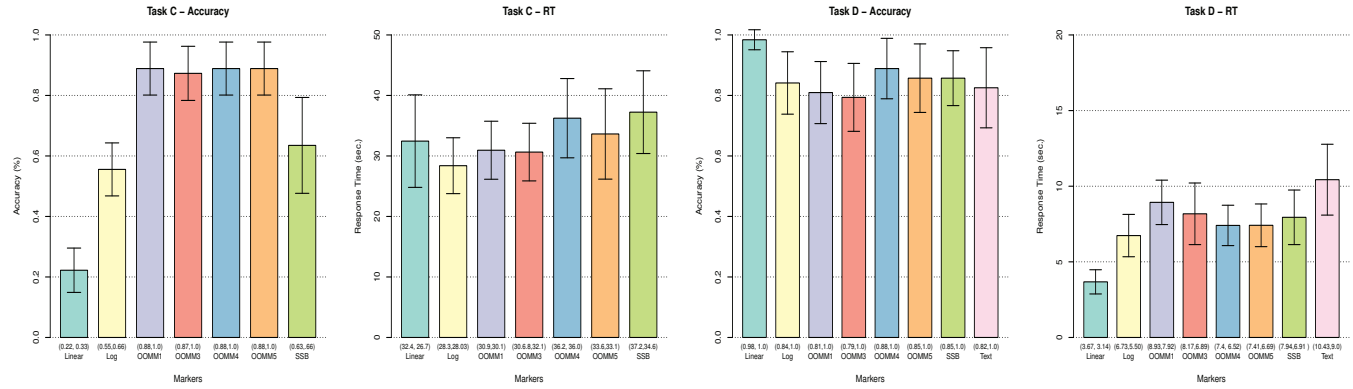


Fig. 9: Analysis of performance results for Task C and D, (mean, median) values are indicated below each bar. Error bars show 95% confidence intervals.

16.48) ($p \ll 0.05$). OOMM5 markers were significantly faster than logarithmic (mean = 19.42), scale-stack bar (mean = 24.71), text (mean = 18.98) and OOMM1 (mean = 17.95) ($p \ll 0.002$). As OOMM5 markers are slower but more accurate than linear, a correlation test of time vs. accuracy was performed to detect any trade-off effect. The correlation result was non-negative (0.436) meaning that faster responses did not led to more errors (and vice-versa);

No other significant differences were found.

Ratio Estimation: Task C. Performance in Task C, as a function of marker type, is summarized in Fig. 9. As aforementioned, due to the nature of the task (computing the ratio of two numbers) performances related to text based stimula were not considered. A noticeable variation is visible across markers, the Friedman's test showed a significant main effect in both accuracy ($\chi^2 = 88.85$, $p \ll 0.05$) and response time ($\chi^2 = 55.111$, $p \ll 0.05$). A closer analysis showed:

- Mean Accuracy
  - OOMM1 markers (mean = .88) were significantly more accurate than linear (mean = .22), logarithmic (mean = .55) and scale-stack bar (mean = .63) ($p \ll 0.002$);
  - OOMM3 markers (mean = .87) were significantly more accurate than linear (mean = .22) and logarithmic (mean = .55). A trend towards significance was found between OOMM3 and scale-stack bar (mean = .63) ($p = 0.018$), further analysis showed a small effect size ($r = .29$);
  - OOMM4 markers (mean = .88) were significantly more accurate than linear (mean = .22) and logarithmic (mean = .55) ($p \ll 0.002$).
  - OOMM5 markers (mean = .88) were significantly more accurate than linear (mean = .22), logarithmic (mean = .55) and scale-stack bar (mean = .63) ($p \ll 0.002$).
- Mean Response Time
  - OOMM1 markers (mean = 30.94) were significantly faster than OOMM4 (mean = 36.24) ($p \ll 0.002$).
  - OOMM3 markers (mean = 30.63) were significantly faster than OOMM4 (mean = 36.24) and scale-stack bar (mean = 37.23) ($p \ll 0.002$);
  - OOMM4 markers (mean = 36.24) were significantly slower than logarithmic (mean = 28.39), OOMM1 and OOMM3 ($p \ll 0.05$). As OOMM4 markers are slower but more accurate than logarithmic, a correlation test of time vs. accuracy was performed to detect any trade-off effect. The correlation result was non-negative (0.271) meaning that faster responses did not led to more errors (and vice-versa).

No other significant differences were found.

Trend Analysis: Task D. Performance in Task D, as a function of marker type, is summarized in Fig. 9. A small variation is visible across markers for accuracy, while a more noticeable variation is visible for response time. The Friedman's test showed a significant main effect in both accuracy ($\chi^2 = 18.667$, $p \ll 0.05$) and response time ($\chi^2 = 61.667$, $p \ll 0.05$). A closer analysis showed:

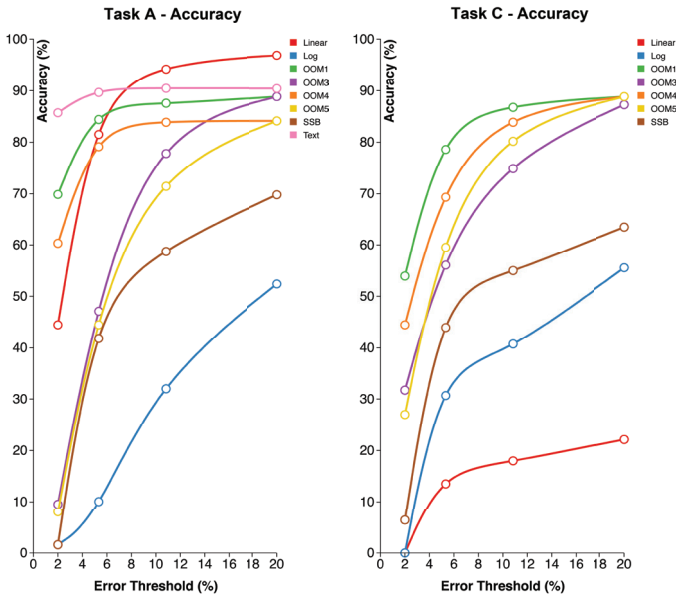


Fig. 10: Task A and C - Comparison of measured accuracy at different levels of error tolerance. Points, from left to right, depict the performances at $2\%, 5\%, 10\%, 20\%$ tolerance values respectively, data are plotted using power trendlines of measured accuracy.

- Mean Accuracy
  - OOMM1 markers (mean = .809) were significantly less accurate than linear (mean = .98) ($p \ll 0.002$);
  - OOMM3 markers (mean = .79) were significantly less accurate than linear (mean = .98) ($p \ll 0.002$);
- Mean Response Time
  - OOMM1 markers (mean = 8.93) were significantly slower than linear (mean = 3.67), logarithmic (mean = 6.73), OOMM4 (mean = 7.4) and OOMM5 (mean = 7.41) ($p \ll 0.002$);
  - OOMM3 markers (mean = 8.17) were significantly slower than linear (mean = 3.67) ($p \ll 0.05$), and significantly faster than text (mean = 10.43) ($p \ll 0.002$);
  - OOMM4 markers (mean = 7.4) were significantly slower than linear (mean = 3.67). OOMM4 markers were significantly faster than text (mean = 10.43) ($p \ll 0.05$);
  - OOMM5 markers (mean = 22.08) were significantly slower than linear (mean = 3.67) ($p \ll 0.002$). OOMM5 markers were significantly faster than text (mean = 10.43) ($p \ll 0.002$).

No other significant differences were found.

4.6.1 Varying Error Tolerance

Measurements of physical quantities are subject to inaccuracies also referred to as uncertainties. Value estimation is likely to deviate from the unknown, true, value of the quantity. Task A and Task C both involved estimation of an unknown value therefore answers were considered correct if falling within a predefined error tolerance (e.g., deviation from the actual value). We empirically chose three levels of tolerance: 20%, 10%, 5% and 2%. We compared accuracy results by varying tolerance level and found significant differences. The Friedman's test showed a significant main effect in accuracy for both Task A and Task C. In particular for Task A: at 2% ($\chi^2 = 96.709$, $p \ll 0.05$), at 5% ($\chi^2 = 80.595$, $p \ll 0.05$), at 10% ($\chi^2 = 58.214$, $p \ll 0.05$). For Task C: at 2% ($\chi^2 = 71.186$, $p \ll 0.05$), at 5% ($\chi^2 = 55.818$, $p \ll 0.05$), at 10% ($\chi^2 = 71.277$, $p \ll 0.05$). A comparison of performance behaviours in terms of accuracy is depicted in Fig. 10. In tables 1 and 2 we report the significant differences between markers. In Task A a trend is visible in which OOMM1 and OOMM4 outperform logarithmic across all four threshold levels, and scale-stack bar for the first three threshold levels. In Task C OOMM1 outperforms linear, logarithmic and scale-stack bar, followed by OOMM4 and OOMM5 which outperform linear and logarithmic across all four threshold levels.

Table 1: Task A results. **p-values** of post-hoc results analysis, using a Bonferroni corrected significance value of $\alpha = 0.002$, for the effects of varying error tolerance level on accuracy. Pairwise significant differences are highlighted in light blue, if the first member of the pair is significantly more accurate than the second member, and red, if the second member of the pair is significantly more accurate than the first. SSB=Scale-Stack Bar.

Task A		Error Tolerance			
		2%	5%	10%	20%
Pairwise Comparisons	OOMM1 vs. Linear	0.13	0.73	0.059	0.024
	OOMM1 vs. Log	$\ll .001$	$\ll .001$	$\ll .001$	$\ll .001$
	OOMM1 vs. SSB	$\ll .001$	$\ll .001$	0.02	0.134
	OOMM1 vs. Text	$\ll .001$	0.20	0. 20	0.317
	OOMM3 vs. Linear	$\ll .001$	$\ll .001$	0.10	0.317
	OOMM3 vs. Log	0.059	$\ll .001$	$\ll .001$	$\ll .001$
	OOMM3 vs. SSB	$\ll .001$	1.0	$\ll .001$	0.002
	OOMM3 vs. Text	$\ll .001$	$\ll .001$	0.52	0.7
	OOMM4 vs. Linear	0.28	1.0	0.1	0.059
	OOMM4 vs. Log	$\ll .001$	$\ll .001$	$\ll .001$	$\ll .001$
	OOMM4 vs. SSB	$\ll .001$	$\ll .001$	$\ll .001$	0.16
	OOMM4 vs. Text	0.01	0.52	0.73	0.73
	OOMM5 vs. Linear	$\ll .001$	$\ll .001$	0.2	0.1
	OOMM5 vs. Log	0.41	$\ll .001$	$\ll .001$	$\ll .001$
	OOMM5 vs. SSB	0.83	0.52	0.002	0.001
	OOMM5 vs. Text	$\ll .001$	$\ll .001$	0.2	0.73
	OOMM1 vs. OOMM3	$\ll .001$	$\ll .001$	1.0	0.48
	OOMM1 vs. OOMM4	0.15	1.0	0.76	1.0
	OOMM1 vs. OOMM5	$\ll .001$	$\ll .001$	0.36	0.73
	OOMM3 vs. OOMM4	$\ll .001$	$\ll .001$	0.65	0.31
	OOMM3 vs. OOMM5	0.41	1.0	0.25	0.31
	OOMM4 vs. OOMM5	$\ll .001$	$\ll .001$	0.25	0.65

5 Findings and Discussion

OOMM markers performances. The overall performance of OOMM markers varied considerably across tasks. When explicit quantitative evaluation (e.g. Task A and C) was required OOMM markers performed consistently more accurately than logarithmic and scale-stack bars (for tolerance 10%, Task A) and linear, logarithmic and scale-stack bars (for tolerances 10% and 2%, Task C), see tables 1 and 2. Task C involved the computation of ratio between two numbers, the increase in the dynamic range introduced by OOMM markers made it easier to the user to provide more accurate answers.

When the task involved target identification (e.g. Task B) OOMM markers performed consistently more accurately than linear, logarithmic and scale-stack bars and consistently faster than logarithmic (OOMM1, OOMM3, OOMM4 and OOMM5) and scale-stack bar (OOMM3, OOMM4, OOMM5, OOMM5). Within the same task OOMM markers were also consistently slower than linear. Perceptual load associated with the increase in visual details and features of the new markers can be one of the reason behind the increase in response time. Cognitive load is also augmented by the learning toll induced by the novelty of the OOMMs visual design, as we recall participants needed to be familiar with concepts such as logarithmic scale and standard charts. It is however interesting to notice how the novelty effect should not be considered when analysing performances against scale-stack bars. It is interesting to note that faster responses did not

Table 2: Task C results. **p-values** of post-hoc results analysis, using a Bonferroni corrected significance value of $\alpha = 0.0027$, for the effects of varying error tolerance level on accuracy. Pairwise significant differences are highlighted in light blue, if the first member of a pair is significantly more accurate than the second member, and red, if the second member of a pair is significantly more accurate than the first.

Task C		Error Tolerance			
		2%	5%	10%	20%
Pairwise Comparisons	OOMM1 vs. Linear	$\ll .001$	$\ll .001$	$\ll .001$	$\ll .001$
	OOMM1 vs. Log	$\ll .001$	$\ll .001$	$\ll .001$	$\ll .001$
	OOMM1 vs. SSB	$\ll .001$	$\ll .001$	$\ll .001$	0.001
	OOMM3 vs. Linear	$\ll .001$	$\ll .001$	$\ll .001$	$\ll .001$
	OOMM3 vs. Log	$\ll .001$	0.07	$\ll .001$	$\ll .001$
	OOMM3 vs. SSB	$\ll .001$	0.8	0.04	0.04
	OOMM4 vs. Linear	$\ll .001$	$\ll .001$	$\ll .001$	$\ll .001$
	OOMM4 vs. Log	$\ll .001$	$\ll .001$	$\ll .001$	$\ll .001$
	OOMM4 vs. SSB	$\ll .001$	0.1	$\ll .001$	0.07
	OOMM5 vs. Linear	$\ll .001$	$\ll .001$	$\ll .001$	$\ll .001$
	OOMM5 vs. Log	$\ll .001$	$\ll .001$	$\ll .001$	$\ll .001$
	OOMM5 vs. SSB	$\ll .001$	0.4	$\ll .001$	0.002
	OOMM1 vs. OOMM3	$\ll .001$	0.01	0.1	0.52
	OOMM1 vs. OOMM4	0.59	0.08	0.56	0.76
	OOMM1 vs. OOMM5	$\ll .001$	0.002	0.24	0.73
	OOMM3 vs. OOMM4	0.24	0.09	0.2	073
	OOMM3 vs. OOMM5	0.15	0.2	0.31	0.7
	OOMM4 vs. OOMM5	0.03	0.24	0.73	1.0

lead to more errors and that slower responses did not imply a loss in accuracy either.

When the task involved both target identification and quantitative evaluation (e.g. Task D) OOMM markers were consistently less accurate and slower than linear. It is interesting to notice how OOMM1 markers were also significantly slower than scale-stack bar, this result could derive from the higher semantic complexity of the OOMM1 markers.

Semantic complexity, lack of familiarity are all elements which could have affected performances of OOMM markers, and, to be fair, scale-stack bars as well. Future investigations should address both aspects by looking at the marker's design features which, if considered in the context of large datasets, experience similar limitations to that of glyph design, and learnability, by assessing the learning curve of users [2, 3, 13]. Semantic complexity and learnability are closely related concepts, simpler representations are easier to learn, the challenge is to find the appropriate trade-off between simplicity and expressiveness.

Text based visualization performances. A full comparison of performances across all representations showed significant differences in accuracy and reaction time of textual representations, versus other visual representations such as linear, logarithmic and scale-stack bars, in tasks involving magnitude estimation or target identification (e.g. Task A and B). For magnitude estimation (Task A), text representations performed significantly faster than linear, logarithmic and scale-stack bars but more accurately only for low threshold levels (e.g. 2% and 5%). For Task A, results are somehow expected since participants only had to read in the value of a number in decimal form.

For target identification (Task B) textual representations performed significantly more accurately than linear, logarithmic and scale-stack bar, behaviour similar to that of OOMM markers, but interestingly not faster. Task B had a similar requirement to Task A: users still had to read in an explicit numerical value, which explains accuracy results. Task B however required to perform visual search within a more complex display than that of Task A: stimuli included 29 distractor elements. This overall behaviour prompts interesting questions on how performances might be affected in scenarios where data aggregation is a necessity. When dealing with extremely large displays the advantage of explicit textual representation is inevitably lost to the lack of available visualization space, also, in terms of visual search, the reaction time to identify a target increases at least linearly with the number of distractors.

Participants' feedback. Participants' feedback was overall extremely positive, they appeared to engage with the new visual representations and keen to see their application in more complex contexts. None of the participants complained about the length of the study and found it easier to interpret the new representations as the test progressed. This last comment in particular suggests that some learning was taking place, further investigation though would be required to support the hypothesis.

Display of numerical magnitude. For linear, logarithmic and scale-stack bars, a height of p pixels will result in at most p quantities represented from the range of the source data. For colour, using a b-bit linear scale we are limited to 2^b different quantities (usually 256, although some monitors only achieve 8 bits through temporal dithering). Text offers the lowest quantization error, being limited by font size and available space. For example, with marker sizes of 150px, 23 digit numbers 99999999999999999999999 are readable, although as our study shows, text markers can be difficult to interpret. Our markers can offer a $10\times$ increase of dynamic range compared to previous markers, thus with a corresponding reduction in quantization error. If s exponents can be represented clearly within the pixel height ($s = 10$ in our study), then we offer $s\times$ the range compared to other markers.

As a specific example, with $s = 10$ and $p = 150$ we can represent 1500 quantities, n, with $0 \leq n < 1 \times 10^{11}$. Scale-stack bars can represent 150 quantities, n, with $0 \leq n < 1 \times 10^{10}$. There is one reduction in magnitude since the absence of a mark on the exponent scale for ours represents $0-10$ whereas scale-stack bars require an explicit $0-10$ scale at the bottom of the marker. If we assume a range of $0 \leq n < 1 \times 10^{10}$ for each other marker apart from ours, we can obtain this example: For an example range of $1,000,000$ to $2,000,000$, on linear, these numbers fall below the first pixel and so are represented with zero pixels rendered. On the logarithmic scale, the 90^{th} (1,000,000) to 94^{th} (1,847,850) pixels cover the range. For SSB, the 92^{nd} (1,333,333) and 93^{rd} (2,000,000). Ours, with the exponent appropriately rendered, the significand is rendered from pixel 15 (1,000,000) to 30 (2,000,000). We can also pick example situations such as $9,000,000 \leq n < 1 \times 10^7$ (on the same example scale), where ours has a range of 15 pixels, and the other markers do not change.

6 Conclusions

In this work we have presented new visual designs to support the display of large value ranges. An empirical study has shown how the increase in expressive power of OOMM markers, and mostly in their numerical dynamic range, outweighs the cognitive load introduced by the novelty of the design. In tasks involving quantitative analysis of large value ranges the OOMM markers outperform state of the art techniques. Our results confirm previous work by Hlawatsch et al. [10] showing that there exist real case study scenarios where markers, which considerably increase the space of representable quantities, make evaluation tasks not only easier but also more accurate. It is of interest to the authors to further the investigation of the OOMM markers performances in terms of lower level cognitive processing such as memorability, learnability and concept grasping and to quantitatively assess their effectiveness in much more complex contexts such as extremely large data displays.

References

[1] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pages 111–117, Oct 2005.

[2] R. Borgo, A. Abdul-Rahman, F. Mohamed, P. W. Grant, I. Reppa, L. Floridi, and M. Chen. An empirical study on using visual embellishments in visualization. *IEEE Trans. Vis. Comput. Graph.*, 18(12):2759–2768, 2012.

[3] R. Borgo, J. Kehrer, D. H. Chung, E. Maguire, R. S. Laramee, H. Hauser, M. Ward, and M. Chen. Glyph-based visualization: Foundations, design guidelines, techniques and applications. In *Eurographics State of the Art Reports*, EG STARs, pages 39–63. Eurographics Association, May 2013. http://diglib.eg.org/EG/DL/conf/EG2013/stars/039-063.pdf.

[4] M. A. Borkin, A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva, and H. Pfister. What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2306–2315, 2013.

[5] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984.

[6] S. Few. The chartjunk debate: A close examination of recent findings. *Visual Business Intelligent Newsletter*, April-June 2011.

[7] M. Fink, J.-H. Haunert, J. Spoerhase, and A. Wolff. Selecting the aspect ratio of a scatter plot based on its delaunay triangulation. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2326–2335, Dec 2013.

[8] C. G. Healey. Perceptual techniques for scientific visualization. *SIGGRAPH 99 Course*, 1999.

[9] J. Heer and M. Bostock. Crowdsourcing graphical perception: Using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 203–212, New York, NY, USA, 2010. ACM.

[10] M. Hlawatsch, F. Sadlo, M. Burch, and D. Weiskopf. Scale-stack bar charts. *Computer Graphics Forum*, 32(3pt2):181–190, 2013.

[11] P. Isenberg, A. Bezerianos, P. Dragicevic, and J. Fekete. A study on dual-scale data charts. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2469–2478, Dec 2011.

[12] N. Kong and M. Agrawala. Graphical overlays: Using layered elements to aid chart reading. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2631–2638, 2012.

[13] P. Legg, D. H. S. Chung, M. L. Parry, M. W. Jones, R. Long, I. W. Griffiths, and M. Chen. Matchpad: Interactive glyph-based visualization for real-time sports performance analysis. *Computer Graphics Forum*, 31(3):1255–1264, 2012.

[14] Y. K. Leung and M. D. Apperley. A review and taxonomy of distortion-oriented presentation techniques. *ACM Trans. Comput.-Hum. Interact.*, 1(2):126–160, June 1994.

[15] J. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Trans. Graph.*, 5(2):110–141, Apr. 1986.

[16] S. J. P. McDougall, M. B. Curry, and O. D. Bruijn. Measuring symbol and icon characteristics: Norms for concreteness, complexity, meaningfulness, familiarity, and semantic distance for 239 symbols. *Behavior Research Methods, Instruments, and Computers*, 31(3):487–519, 1999.

[17] C. Murray, L. Rosenfeld, S. Lim, K. Andrews, K. Foreman, D. Haring, N. Fullman, M. Naghavi, R. Lozano, and A. Lopez. Global malaria mortality between 1980 and 2010: a systematic analysis. *The Lancet*, 379:413–431, Feb. 2012.

[18] E. Neuwirth. *RColorBrewer: ColorBrewer palettes. R package version 1.0-2.*, 2007.

[19] J. Peltier. Excel panel charts with different scales [online]. `http://peltiertech.com/Excel/ChartsHowTo/PanelUnevenScales.html`, Nov. 2011.

[20] H. Rosling. Gapminder [online]. `http://www.gapminder.org/`.

[21] B. Speckmann and K. Verbeek. Necklace maps. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):881–889, Nov 2010.

[22] J. Talbot, S. Lin, and P. Hanrahan. An extension of Wilkinson's algorithm for positioning tick labels on axes. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1036–1043, Nov 2010.

[23] M. Tory and T. Moller. Rethinking visualization: A high-level taxonomy. In *Proceedings of the IEEE Symposium on Information Visualization*, INFOVIS '04, pages 151–158, Washington, DC, USA, 2004. IEEE Computer Society.

[24] E. R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, second edition, 2001.

[25] J. W. Tukey. *Exploratory Data Analysis*. Addison-Wesley, 1977.

[26] UNSCEAR. *Sources and effects of ionizing radiation*, 2008.