# A Psychophysical Investigation of Size as a Physical Variable

Yvonne Jansen and Kasper Hornbæk

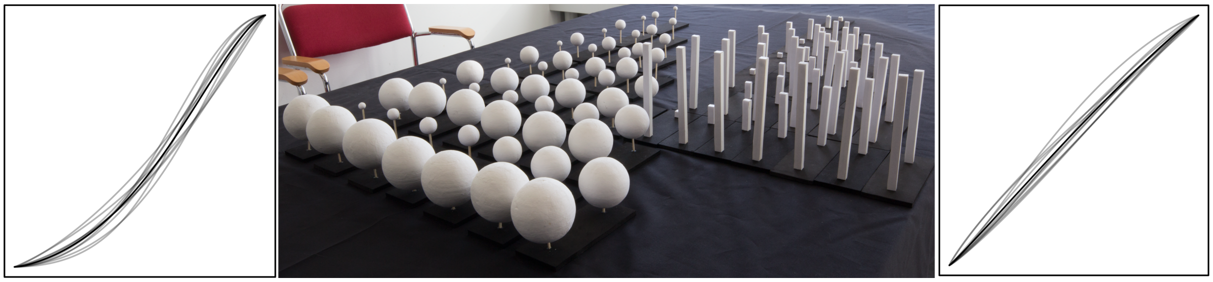

Fig. 1. Middle: the pairs of physical marks used during the experiment. Left and right: fitted psychophysical response curves for spheres and bars respectively, overplotted to show individual curves for all participants.

**Abstract**—Physical visualizations, or data physicalizations, encode data in attributes of physical shapes. Despite a considerable body of work on visual variables, "physical variables" remain poorly understood. One of them is physical size. A difficulty for solid elements is that "size" is ambiguous – it can refer to either length/diameter, surface, or volume. Thus, it is unclear for designers of physicalizations how to effectively encode quantities in physical size. To investigate, we ran an experiment where participants estimated ratios between quantities represented by solid bars and spheres. Our results suggest that solid bars are compared based on their length, consistent with previous findings for 2D and 3D bars on flat media. But for spheres, participants' estimates are rather proportional to their surface. Depending on the estimation method used, judgments are rather consistent across participants, thus the use of perceptually-optimized size scales seems possible. We conclude by discussing implications for the design of data physicalizations and the need for more empirical studies on physical variables.

**Index Terms**—Data physicalization, physical visualization, psychophysics, experiment, physical variable

## 1 Introduction

Long before the crafting of the first visualizations, humanity has encoded and manipulated data in physical form [15]. While paper and later computer screens took over as the preferred way of presenting data, physical data representations – or *data physicalizations* [28] – are again becoming prevalent. Data physicalizations have been a popular visualization medium for artists and designers for the past 20 years [62, 15], and recently academia started to formally investigate their merits. Multiple articles have reported promising findings on the suitability of data physicalizations for information retrieval [27], for education purposes [63], for public data communication [38, 59], and for motivational purposes [55, 31]. Work on data physicalizations is gaining momentum through emerging technologies such as digital fabrication (for static physicalizations) and shape-changing interfaces (for dynamic physicalizations).

Data physicalizations have been put to many uses. While some find data physicalization most appropriate for casual visualization [45, 24], others hold that with advancing technology, data physicalization could offer an equal level of interactivity and resolution to that of current desktop systems [28]. In both cases, a key design question is how to encode data using physical variables and, on a more abstract level, what the available physical variables are and how they differ from purely visual variables. For instance, if we want to show rainfall data physically, should we use physical bar charts (using laser cut or mechanized bars)? How accurately will people perceive the data? Would it be more or less efficient to show the data encoded as spheres (3D printed or inflatable)?

For visual variables, there is a large body of previous work investigating the perceptual efficacy of visual encodings to guide a designer's choices. The studies by Cleveland and McGill are probably the best known [10, 11] but many others exist [14, 41, 7, 17, 19, 50, 25, 58]. In contrast, little is known about the effectiveness of physical data encodings. Psychologists have studied the perception of volumes, and most report nonlinear response curves and a considerable amount of variation between people [1, 16]. Since physicalizations necessarily encode information in volumetric shapes – *all* physical objects have some volume – it could be possible that the perception of physicalizations would need to be assumed unreliable and that they were only suitable for special visualization needs where accurate perception of the data is not pertinent. However, Teghtsoonian [60] demonstrated that depending on whether people are asked to estimate how large a stimulus *looks* or how large a stimulus *is*, they respond differently and provide accurate estimates for how large the stimulus *is*. It is currently unclear how people interpret sizes of physical stimuli when they appear in the context of data representations, and whether the variability of judgments between people is sufficiently low so that a scale correction would lead to consistently more accurate estimates.

To shed light on these questions, we present a psychophysical investigation of two physical marks, bars and spheres (Figure 1). We designed an experiment such that participants were instructed to *estimate the quantities represented by the physical marks*. We describe our analysis process to determine how bars and spheres can be used for data physicalization, what accuracy can be achieved, and in how far perceptually-optimized scales might be necessary. Our findings inform designers about how size as a physical variable can be used in data physicalizations. We conclude by discussing methodology for the further study of physical variables.

• *Yvonne Jansen is with University of Copenhagen. E-mail: jansen.yv@gmail.com.*
• *Kasper Hornbæk is with University of Copenhagen. E-mail: kash@di.ku.dk.*

*Manuscript received 31 Mar. 2015; accepted 1 Aug. 2015; date of publication xx Aug. 2015; date of current version 25 Oct. 2015. For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.*


Fig. 2. Three examples of the use of spheres to represent quantitative information: worry (prayer) beads showing deaths through terrorist attacks (photo courtesy Loren Madsen); a physicalization of the location and magnitudes of earth quakes created as a school project [37]; and number of internet users created with the handmade visualization toolkit by Jose Duarte.

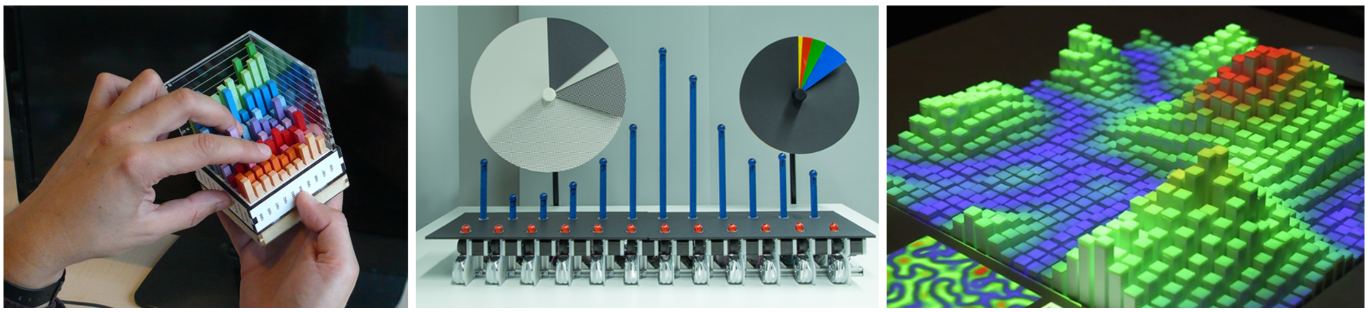

Fig. 3. Three examples of the use of bars to represent quantitative information: physical bar charts by Jansen et al. [27]; physical charts by Microsoft Research [59]; and Tangible CityScape (photo courtesy MIT Tangible Media Group) [21].

## 2 Background

The background for this paper is data physicalization, which we cover first. We then discuss visual variables and contrast the extensive knowledge about such variables with the rather meager knowledge on physical variables. We briefly review size as one particular important parameter for physical variables and discuss the psychophysical methods used to model the perception of visual variables.

### 2.1 Data Physicalization

Information has been put into physical form for millennia; the list curated by Dragicevic and Jansen [15] contains hundreds of examples, including the 7500 years old Mesopotamian clay tokens to support counting, representations created by physicists, biologists, and chemists in the 19th and 20th century, tactile graphics for the visually impaired, and data sculptures created by artists and designers. A recent review of the research problems that need to be tackled about physical representations of data coined the term *data physicalization* to mean how "computer-supported physical representations of data (i.e., physicalizations) can support cognition, communication, learning, problem solving, and decision making" [28].

Visualizations encode data in attributes of graphical marks such as points, lines, and areas. These attributes, for instance, size, color, or orientation are called *visual variables*. Similarly, physicalizations encode data in attributes of *physical marks*, and thus we call these attributes *physical variables*. The effectiveness of physical variables to represent data has so far not been studied empirically, and we can only make conjectures based on prior findings about the perception of physical objects in general (e.g., [1]).

The benefits of turning data into physical form are many and include increased use of non-visual perception and active touch, support for thinking through physical manipulation, and increased engagement. These and other benefits have been previously reviewed [62, 28] and are related to benefits of shape-changing interfaces [46] and of tangible interfaces [49]. The benefits of data physicalization have been corroborated in recent empirical studies. For instance, direct active manipulation facilitates information retrieval from physical representations in comparison to on-screen 3D visualizations [27], physicalizations can improve memorability of data compared to paper visualizations [54], and physical representations of physical activity such as running can motivate behavior [55, 31].

Producing physical representations of data has been a key challenge, but in recent years it has become much cheaper and faster. Key technologies have been 3D printing (e.g., [38]), laser cutting (e.g., [56]), and mechanical actuation (e.g., [57]). In particular, recent work on shape-changing technology has helped make physicalizations dynamic, for example, as in the systems inFORM [21] and EMERGE [57].

The combination of benefits and improved technologies has lead to an increase of work on physical data representations, including work for casual use [55], work on data sculptures for education [63], and work on communicating numerical information [27]. The present paper focuses on how to encode information effectively.

### 2.2 Visual Variables

To appreciate the aims of this paper, it is necessary to contrast the situation in data physicalization to that of visualization with respect to the variables used to encode data. For visualization, a large body of work has catalogued the visual variables that can be used to encode information and studied their effectiveness. The French cartographer Jacques Bertin was the first to formalize a language of graphics [3]. He identified three types of visual marks (points, lines, and areas), and eight variables to vary these marks such that they encode information in two-dimensional space: the two dimensions of the plane (x/y position), size (length, area, or repetition), value (gray scale saturation), texture (grain), color (hue), orientation, and shape. This work has since been extended and validated, most notably by Cleveland and McGill [10, 11, 12] and Mackinlay [34].

Even long before Bertin, psychologists, statisticians, and cartographers studied the effectiveness of various visual encodings. In the 1920's a debate started on the relative effectiveness of bars and circles to encode quantitative information [33]. Of these early studies, the one by Croxton and Stein is best known today [14], investigating two-dimensional bars, squares, circles, and cubes. Further, many studies have charted the relative effectiveness of visual variables [41, 11, 34, 50]. Position of a graphical mark is widely accepted as a more effective way of showing quantitative data than the size of the mark [11, 34]; shape is better at showing nominal data than at showing quantitative data [34]; extraneous dimensions of visual marks can diminish accuracy [65] though conflicting evidence exists [41, 50]. This information has proven useful for designers of visualizations because it helps select the appropriate visual variables for particular data and to avoid making visualizations that are misleading [47].

### 2.3 Physical Variables

In contrast, physical variables are much less well understood. We take *physical variable* to mean modifications in attributes of physical marks. For instance, a sphere is the physical equivalent of a graphical mark and its attributes such as position, size, color, or texture can be modified to encode information. Figures 2 and 3 show examples of physicalizations using the physical marks spheres and bars which encode information mainly in their size.

Candidates for physical variables have been catalogued in the literature on shape-change [46, 40]. Among others, they include variables from the visual variables literature, such as orientation or texture. But these variables are physical, and thus not just perceived visually [28]. Furthermore, the percept transformation [26] likely plays a greater role due to a more active perception, leading, for example, to frequently changing viewing angles, which have already been shown to play a role in the effectiveness of visual variables on large displays [4].

Little is known, about how physical variables perform in relation to each other. We could make some assumptions based on findings from psychophysical studies on the perception of volumetric shapes [16, 1, 51]. But much of this prior work refer by "volumes" to two-dimensional perspective drawings of various shapes. In the few studies using actual physical stimuli, participants' heads were usually fixed [1]. Also, Teghtsoonian found that size judgments depend on whether participants were asked to estimate how large a stimulus *looks* or how large a stimulus *is* [60]. Such context effects are relevant to consider when studying the effectiveness of variables for visualization or physicalization. It thus seems pertinent to frame experimental tasks for studies on graph perception in the appropriate context: the marks to be estimated represent quantities and the task is to make judgments about the *represented quantities* and not about the marks themselves. Furthermore, Schneider and Bissett found that depending on whether participants are asked to make ratio or difference judgments, some interpret the task differently which affects their response scale and thus might explain the observed variety of responses in previous studies [48].

### 2.4 Why Study Size?

In this first study on physical variables we focus on size. Size is a key visual variable and often used in static [27] as well as dynamic data physicalizations [21, 57] (see also Figures 2 and 3). Depending on which physical mark is used, size can be interpreted differently since the attributes of the mark can vary in one, two, or all three dimensions. For example, bars vary only in one dimension – height – whereas spheres vary in all three dimensions. Still, the dimensions of a sphere can be expressed in a single number, its diameter. More commonly however, the size of a sphere is indicated by its volume. Figure 4 illustrates the different growth rates for bars and spheres depending on whether one indicates their size in volume or surface area.

The variance in experimental findings on the size perception of three dimensional objects [1] suggests that mapping data directly proportional to volumetric units is often problematic. For example, if we encounter a physicalization that uses spheres as physical marks, how would people interpret the ratio between two spheres with diameters of 5cm and 2.5cm? And do people agree on an interpretation or do estimates vary considerably between people? If people were to agree on an interpretation, then this could be taken into account when encoding data in the size of spheres. In this article, we investigate these two questions for the size perception of bars and spheres.

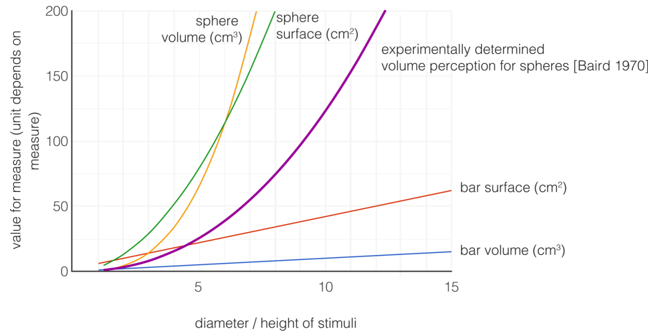


Fig. 4. Comparison of growth for different measures of bars and spheres.

### 2.5 Modeling Perception of Visual and Physical Variables

The present paper approaches the study of size through models and methods from psychophysics. There, human (visual) perception is commonly modeled through a power law $P = bS^a$ with $P$ being the perceived quantity, $S$ the true quantity (in some unit), $a$ an exponent, and $b$ a scaling factor. A power law relationship was first described by Plateau [43] and is known today as 'Stevens' power law' [52]. Most of the psychophysical literature reports estimates for the parameters of this model. The advantage of the model is that it helps compare response curves across studies and for different types of stimuli. However, the model has been criticized for a variety of reasons, most notably because parameter estimates can vary considerably across studies (see for example Green and Luce [22]). Poulton [44] discusses in detail possible biases introduced by psychophysical methods.

Cleveland and McGill contributed a ranking of visual variables, based their ranking on an analysis of absolute errors of estimates [11], arguing that variables which exhibit low errors are good choices for effective encodings of data. While such a classification is clearly useful to choose effective encodings, it does not further investigate badly performing variables to determine if and why they should indeed be generally avoided. For example, Cleveland and McGill [12] report that color (saturation and hue) is estimated most unreliably of all variables tested. However, this does not mean that visualization designers should never use color but that they need to consider the limitations of this variable and use it appropriately. Research into perceptual color scales provides such guidance (for example [47]).

While the investigation of absolute errors helps to identify effective variables, it is not a sufficient tool to examine what type of errors participants made. For example, absolute errors hide whether errors are generally high but spread around 0 or whether people systematically over- or underestimate a variable. One might argue that such a variable should just be generally avoided – although this is not always possible for aesthetic or practical purposes (for example, graduated circles have received much attention by cartographers [18, 19]). Following the observation that circle sizes are commonly underestimated, Flannery proposed a perceptually-optimized circle scaling [19]. While Flannery's proposed circle scaling was later challenged by Cleveland et al. [9], in general, a scale correction has the purpose of equalizing the direction of errors such that overestimations become as common as underestimations.

A theoretical scale correction is not always sufficient to generally minimize errors in judgment. As Green and Luce pointed out [22], variation (between people) can be extremely high and Stevens' exponents in the literature can fall in a wide range (for example, for loudness between 0.15 and 0.6). It is thus important to identify variables which are always unreliable from those which can be corrected. Two important factors for assessing the *potential* effectiveness of a visual variable are the *variability between* people and the variability *within* a single person (cf. Cleveland et al. who studied these factors for circles [9]). Variables which are low on both accounts, such as position and length, are already well accepted as good encodings and widely used[1]. Still, a variable can show little variability (between and within people) but high absolute errors. These errors must then be systematic such that a perceptual scale correction could be applied. If variability was only high *between* people, then an individual correction would be possible although impractical as the resulting encoding would only be perceived accurately by the one person to which it was adapted.

## 3 Study Rationale

We detailed in the background section several factors due to which previous psychophysical studies on the judgment of size are not sufficient to inform the design of data physicalizations. In this section we give a detailed account of the study rationale to further motivate our choices in the experimental design and to facilitate the design of future studies on other physical variables. As far as possible, the design of our experiment is guided by previous experiments notably the classic studies by Cleveland and McGill [10, 11, 12] and Spence [50]. All mentions of Cleveland and McGill or Spence in this section refer to these studies.

### 3.1 Methodology

Experimental methodologies for the study of graph perception are inspired by those common in psychophysics (see Stevens [53] for a discussion of the different methods and their respective strengths and shortcomings). A basic distinction is between *estimation* and *production* methods: estimation requires participants to indicate a judgment whereas production requires participants to recreate a presented stimuli with an adjustable version of the stimuli. For graph perception, estimation methods are most common. Cleveland and McGill used the ratio estimation method (RE) whereas Spence used the constant sum method (CS) [13]. The main difference between these two methods is the domain in which the judgment task can be performed: RE is a cross-modality matching task requiring to indicate a ratio between stimuli as a number (percentage) whereas a CS judgment requires to indicate on a visual scale the ratio between two stimuli. Consequently, RE requires a conversion from the visual domain, the experimental stimulus, into the numeric answering domain, whereas CS remains in the visual domain but requires a conversion from one type of shape into another. Spence argued [50] that the latter is a more ecologically valid task as the reading of graphs commonly includes the visual comparison of different quantities and that transforming the percept into a different modality – numbers – might bias the response.

We include both methods, ratio estimation and constant sum in our experiment to investigate possible differences between the two methods. Method is thus a factor in our experiment design.

### 3.2 Choice of Stimuli

#### 3.2.1 Physical Marks

A wide range of different physical marks is possible as experimental stimuli such as bars, spheres, cubes, cylinders, ellipsoids, or toruses. We decided to include bars since they are already commonly used [15], and to investigate whether physical, three dimensional bars exhibit the same linear response curves as their two dimensional counterparts. Additionally, we were interested in physical marks which vary in all three dimensions. We opted for spheres since they have no edges which people might use to apply simple heuristics; spheres have also been used in existing data physicalizations [15] (Figure 2). We only studied these two marks to keep the number of factors and the duration of the experiment manageable.

#### 3.2.2 Sizes

In ratio estimation experiments, either a 'complete' set of comparisons between all sizes is possible, or few 'standard stimuli' are chosen against which the other sizes are compared. Both Cleveland and McGill, and Spence opted for the latter and included two or three standards in their experiments. They found only small effects for the size of the standard. A potential side effect of this choice is that participants are likely to judge stimuli in light of their previous choices for the same standard, that is, they can build a mental model of the scale across stimuli presentations. However, when reading graphs, people are likely to compare different subsets of the graphs thus the 'standard' of comparison constantly changes. Therefore, a more ecologically valid approach is to use a 'complete' set of sizes where the 'standard' stimulus varies between consecutive judgments.

We chose to include eight stimuli sizes of which two are presented per judgment, and included all possible combinations resulting in $\frac{n!}{(n-r)!(r!)} = \frac{8!}{(8-2)!(2!)} = \frac{40320}{720\times 2} = 28$ pairs (with n being the number of sizes and r being the number of sizes presented concurrently).

A comparison of the findings of previous studies demonstrates possible *range effects* [53, 44]. Depending on the choice of the absolute difference between the maximum and the minimum stimulus, response curves can vary, so that experimenters report different exponents for the power law between actual and perceived size. We account for range effects with our choice of a complete set of comparisons, that is participants judge stimuli from different ranges. Also, we indicate angular stimulus sizes such that future studies can more easily contrast their findings with ours.

Still, physicalizations can be scaled up in size more easily than on-screen visualizations. Thus, room-sized or even building-sized physicalizations are possible. Such scales are out of the scope of this article, and we only consider sizes suitable for hand-sized to table-sized physicalizations.

#### 3.2.3 Material

We expect that the choice of material can affect the percept of physical marks. For the bars, an important criterion was that the edges of the bar were salient to ensure that it was perceived as a three dimensional shape and not as a flat bar.

For the spheres, it was important that the material exhibited a light texture such that the mark had unmistakably a spherical shape and was not perceived as a flat disk. At the same time the material should not lead to specular reflections as these have been shown to be beneficial, at least for shape detection [20]. We chose spheres made of matte surface cotton pulp (see Figure 5 for a close-up of the texture).

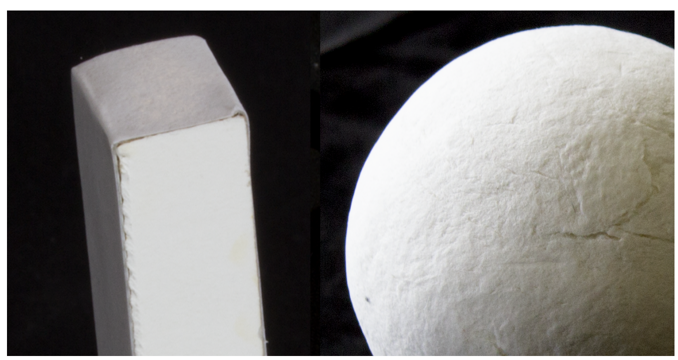

Fig. 5. Close up of the bars and spheres showing their edge saliency and surface texture.

### 3.3 Task Framing & Level of Specificity

Teghtsoonian demonstrated previously that size judgments depend on whether participants are asked to estimate how large a *is* or how large it *looks* [60]. We are interested in the *perception of physical variables* (and not the perception of physical stimuli in general), that is in how physical shapes are interpreted if they *occur as marks in the context of physicalization*. We therefore framed the experimental task accordingly: the initial instructions for the experiment therefore specified that the participant was about to see a series of shapes, and that she should think about them as *representing certain quantities*[2]. Then an

[1] In an analysis of encodings for a study on the memorability of visualizations, Borkin et al. report that 45.3% of 2070 analyzed visualizations used either bars, lines, or points [5], i.e., position or length.

[2] This task framing is closest to how we expect people would interpret physicalizations they might encounter, for example, in a museum.

example was given of two spheres of different sizes labeled with the names of two countries[3]. Participants were then told that throughout the experiment they will be asked to judge the relative difference, i.e., ratio between two shapes, and that they should make “quick visual judgments and not try to make precise measurements, either mentally or with a physical object such as a pencil or a finger.” Furthermore, we specified that there was no time limit but suggested that each judgment should not take them more than 10 seconds.

### 3.4 Sensory Modality

Some recent work has suggested that for haptic perception, a shape’s surface area is a fitting predictor for its perceived size [30]. Work into cross-modal integration of sensory information shows that the haptic sense is more sensitive to lower spatial frequencies and the visual sense to higher spatial frequencies [42]. To avoid possible cross-modal effects [29], we focus in this article on the *visual* perception of physical marks only. While we believe that active touch likely plays an important role when exploring data physicalizations [27, 28], we first need to collect empirical data on the visual perception of physical marks[4]. The benefits are twofold: (i) our findings can be compared to previous work on the visual perception of graphical marks on two-dimensional surfaces (specifically Cleveland and McGill’s [11] and Spence’s [50] findings), (ii) later studies including both vision and touch can contrast their results against both Kahrimanovic et al.’s findings for haptic exploration [29, 30] and ours for vision.

## 4 Experiment

We now describe the design of our experiment. Further details are available on the project’s website: yvonnejansen.me/size.

### 4.1 Design

The experiment included two factors: physical mark (bars and spheres), and estimation method (RE: ratio estimation and CS: constant sum). Conditions were blocked as follows:
first method (could be either ratio estimation or constant sum):
→ 28 ratios of shape A (could be either bars or spheres)
→ 28 ratios of shape B
second method:
→ 28 ratios of shape A
→ 28 ratios of shape B
Bars and spheres were therefore blocked, but always alternated to minimize possible transfer effects. Each participant made in total 112 estimates.

### 4.2 Stimuli

We included bars and spheres of eight different sizes respectively (see Table 1). All possible pairs of solids of different sizes were created resulting in a total of 28 displays of bars and 28 displays of spheres. Each pair of marks was mounted on a black foam pad of 15x6cm such that the centers of all solids were 10cm apart. Cleveland and McGill spaced their marks 65mm, 130mm, and 195mm. They reported no discernible difference between the two smaller distances. Talbot et al. [58] reported an effect of separation for bar charts, particularly for small reference bars. Our choice might therefore introduce some noise in our data due to a separation effect dependent on the size of the comparison standard. However, correcting for this effect, that is reducing the separation for smaller stimuli, would have provided participants with additional information about stimuli sizes. While constant center distances and mounting heights can also provide additional insight into sphere sizes, we followed Spence’s stimulus design [50] and kept the center distance equal for all stimuli and both physical marks. We include post-experiment self-reports from participants about their answer strategies.

[3]Our experiment only included physical stimuli, but the example in the written instructions was presented through an annotated photograph.

[4]While Kahrimanovic et al. [29] study the perception of spheres for both vision and touch, they only asked participants to judge whether the volume of two different shapes, e.g., a sphere and a cube, have the same volume.

| | stimuli sizes | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **mark/measure** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| bars | | | | | | | | |
| height (cm) | 1 | 3 | 5 | 7 | 9 | 11 | 13 | 15 |
| angular (deg) | 0.5 | 1.4 | 2.2 | 3.1 | 3.9 | 4.7 | 5.4 | 6.1 |
| spheres | | | | | | | | |
| diameter (cm) | 1.2 | 1.8 | 2.5 | 3 | 4 | 5 | 6 | 7 |
| angular (deg) | 0.6 | 0.8 | 1.1 | 1.4 | 1.8 | 2.2 | 2.7 | 3.1 |

Table 1. Stimuli sizes in metric units and angular size (angular sizes are based on the initial seating position but participants were free to move their head closer or sideways).

### 4.3 Task & Instructions

Teeghtsonian previously demonstrated that instructions can have a considerable effect on how participants make their judgments [60]. Since our motivation for this experiment is to inform the design of data physicalizations, we framed the experimental task accordingly: the instructions stated that all shapes should be thought of as representing a quantity. Specifically we instructed participants to either “indicate the percentage of the quantity represented by the smaller shape relative to the larger shape” (ratio estimation method) or to “divide the line such that the left part represents the quantity represented by the left shape and the right part represents the quantity of the right shape” (constant sum method).

### 4.4 Answer Interfaces for the Two Estimation Methods

Participants entered their estimations using an iPad 2. For the ratio estimation method, the tablet displayed a number pad to enter the number estimate. For the constant sum method, a 12cm long line was initially displayed. A divider indicator was only displayed once the participant touched the line. It could be adjusted by sliding the finger sideways. In contrast to Spence’s implementation of the constant sum method [50], the line had no scale indicators (see Figure 6 left). We chose this design to ensure that the task could only be solved in the visual domain, that is by judging the ratio between the two line segments without any mental computations. The instructions for using this interface were the same as those reported by Spence [50].

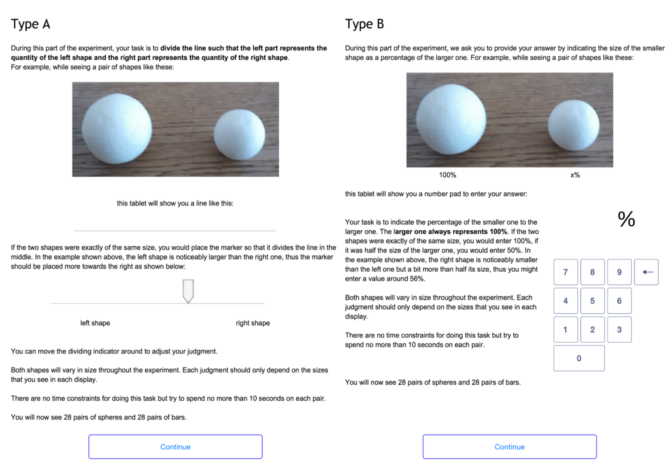


Fig. 6. Instructions shown on the iPad for the two estimation methods, including the priming to interpret marks as representing data.

### 4.5 Procedure

Participants were welcomed and offered a refreshment (water or coffee) while reviewing the consent form. They were then asked to adjust the height of their seat such that the top of their head aligned with a line drawn on a whiteboard on the wall (140cm from the floor). They were then introduced to the experimental setup and asked to read the instructions presented on an iPad. All participants thus received the same written instructions. After participants finished reading the instructions, the experimenter placed a first training stimuli at a distance of 50cm from the table's edge, and participants started to familiarize themselves with the iPad interface to provide their answer. Each experimental condition was preceded by two training stimuli. After participants confirmed that they were ready to start, the experimenter placed in succession 28 stimuli in front of them. All participants performed all four conditions. After finishing participants were debriefed and completed a short questionnaire to assess their numeracy skills[5] [32]. The experiment lasted between 40 and 50 min.

### 4.6 Experimental Setup

Participants were seated at a table covered by a black cloth. The cloth covered the entire table such that it provided a uniform, contrasted background to the white experimental stimuli (see Figure 7). The room's windows were covered with blinds to avoid change in lighting between participants due to the weather. The room was lit with fluorescent ceiling lights.

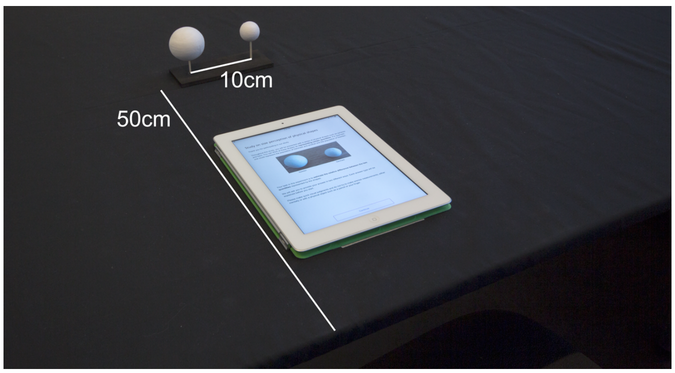


Fig. 7. Setup of the experiment.

### 4.7 Participants

Ten participants were recruited through a mailing list. Previous studies demonstrated negligible variation for factors such as age [2] or level of education [11], thus we collected no further information on participants' demographics. Participants were compensated for their time with a gift voucher of about 20€.

### 4.8 Dependent Measures

We collected two measures: ratio estimation, and time-on-task. Estimates made using the constant sum method were converted to be comparable to ratio estimations using the following formula:

For constant sum, the relative size of the larger mark $cs_1$ depends on the smaller one $cs_2$:

$$cs_1 = 100 - cs_2$$

For ratio estimation, the larger shape always represents 100:

$$re_1 = 100$$

To convert, we simply express the smaller shape as a percentage of the larger one:

$$re_2 = \frac{cs_2}{cs_1} * 100$$

### 4.9 Hypotheses

We expected that response curves for bars would follow a similarly linear scale as established for two-dimensional bars on paper or screens [11, 50]. For spheres, we expected that response curves would vary considerably across participants as suggested by previous work on volumetric shapes [1]. Spence [50] reported lower error rates using the constant sum method than Cleveland and McGill who used the ratio estimation method [11]. We therefore expected that the constant sum method would reduce variation across participants.

[5]Since the ratio estimation method required to provide answers in the form of percentages, we assessed numeracy as a covariate (only reported online).

## 5 Results

We report our results[6] in this section following the analyses performed for previous studies, particularly Cleveland and McGill's [10, 12] and Spence's [50]. Since our main interest is to inform the design of data physicalizations and to give estimates for predicting the accuracy for the perception of quantities, we report uncertainty measures (95% bootstrapped confidence intervals) and investigate biases (systematic over- and under estimations) [8].

Unless otherwise noted, the results reported in this section are based on height for bars and diameter for spheres. This choice allows us to get a better understanding of the dimensionality of the percepts. For example, an exponent of 2 indicates a perceived size proportional to area, and an exponent of 3 indicates a perceived size proportional to volume. Our numbers can however easily be compared to previous work since $v^a \propto (d \times d \times d)^a = d^{a \times 3}$ (with v: volume, d: diameter).

### 5.1 Main Factors

Figure 8 shows all data points for the four combinations of estimation method and physical mark. To estimate the Stevens' exponents, we fitted a nonlinear regression model with one parameter: estimate $= (\frac{\text{size}_1}{\text{size}_2})^a$. Figure 8 shows the fitted curves and indicates the exponent, a, together with 95% confidence intervals. Furthermore, variability (residual standard error) is indicated, also with CI. All individual data points are plotted with 0.3 transparency such that gray scale value encodes density of data points.

Estimates of bars follow a more linear response curve (a = 0.941) than estimates of spheres (a = 1.6). Estimation method had a considerable effect on answer variability: estimates given using the constant sum method vary about twice as much as estimates made using ratio estimation contrary to our hypothesis. This is true for both marks, bars and spheres. Accordingly, participants are more consistent and similar in their judgments when using ratio estimation. We investigate this question further later in this section. However, as is common with ratio estimation (cf. [10, 58, 44]), we observed in some participants a tendency to round their estimate to the nearest multiple of 5.

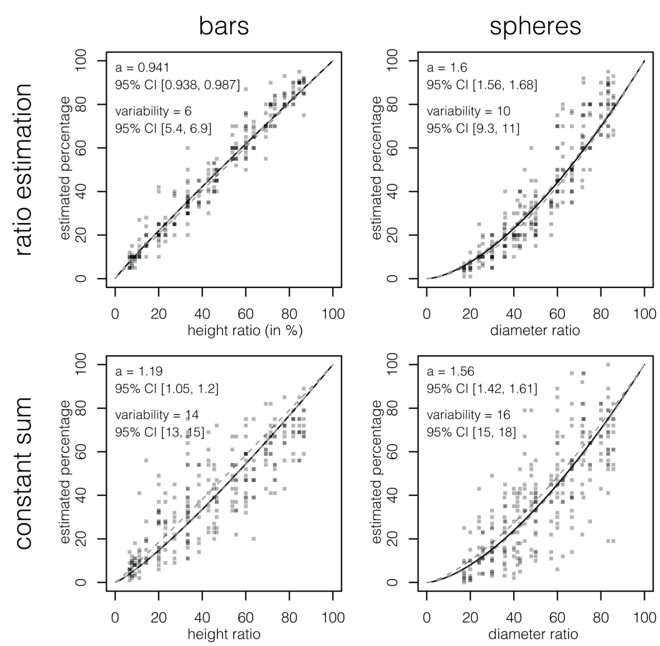


Fig. 8. Scatterplots and fitted nonlinear regressions for all factor combinations including variability (residual standard error), and bootstrapped 95% confidence intervals for regression exponent and variability.

[6]Raw data, R scripts, and additional figures are available at yvonnejansen.me/size.

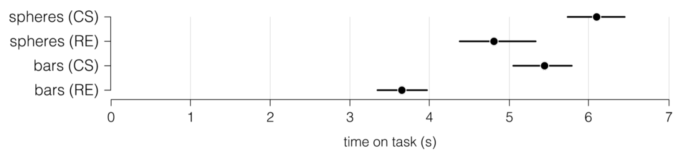


Fig. 9. Time on task for each condition.

Estimates of bars using ratio estimation are in line with previous studies for the size perception of 2D bars. Spence also reports exponents of 0.95, that is, the visual perception of physical bars can in practice be considered as linear.

For spheres, we found an exponent of 1.6. Previously reported exponents for volumetric shapes vary around 0.7 [1] $\frac{1.6}{3} = 0.533$ which corresponds to 2.1 if ratios are based on diameters. However, it should be noted that many studies did not use physical stimuli but rather two-dimensional drawings with various depth cues [1], and that studies using physical solids at the time commonly fixated participants heads.

We measured how much time people spent on each instance. However, this can only be interpreted as rough estimates since the exact presentation time with physical stimuli is less controlled than when presented on a computer display. Nonetheless, Figure 9 shows bootstrapped median estimates for the four conditions. The data suggests, that participants had more difficulty entering their estimates when using the constant sum method, and that estimating bars was faster than estimating spheres. Absolute size of stimuli had no noticeable effect on point estimates or confidence intervals for either shape.

### 5.2 Individual Response Curves

Figure 8 shows considerable spread of data points around the fitted model. We now investigate individual response curves to examine whether the variability, found especially with the constant sum method, is within or between participants. Our results for the constant sum method are peculiar in light of the high accuracy previously reported by Spence when using this method.

Figure 10 shows an overplot of all individually fit responses. For the ratio estimation method, we find only little variation of curves between participants, both for bars and spheres. This indicates that much of the visible spread around the regression line in Figure 8 is indeed due to inaccuracy of estimates.

In contrast, for the constant sum method we find much spread, not only in the raw data but also between the individually fit regression lines. This finding is contrary to our hypothesis. A possible explanation is that participants had different mental models of the task. Indeed we find a variety of individual Stevens' exponents between about 1 (proportional to diameter) to 3.5 which is slightly higher than volumetric growth and indicates that the estimates made by this participant are proportional to the ratio between the actual volumes of the spheres. Note though that the participant did not make estimates proportional to volume when using the ratio estimation method.

### 5.3 Accuracy

To further investigate the observed spread, we now analyze the accuracy of judgments. Cleveland and McGill [8, 11] as well as Spence [50] indicated accuracy by analyzing absolute errors between actual sizes and estimates. To compare our findings to those for two-dimensional graphical marks, we also compute the absolute error of estimates (estimated value minus an objective measure). The choice of objective measure affects now the error metric for spheres as illustrated by Figure 11 across the different measures volume, diameter, surface, and fitted regression (a=1.6). As a comparison, we include the absolute errors reported by Spence [50] which are slightly lower than those reported by Cleveland and McGill. While bars as physical marks are comparable in accuracy to two-dimensional graphical marks, the errors for spheres – if volume encodes the data – are unacceptably large. However, by encoding data in other measures of spheres such as their surface or by using further fitted functions, accuracy can be immensely improved such that spheres as physical marks could become a reasonable choice.

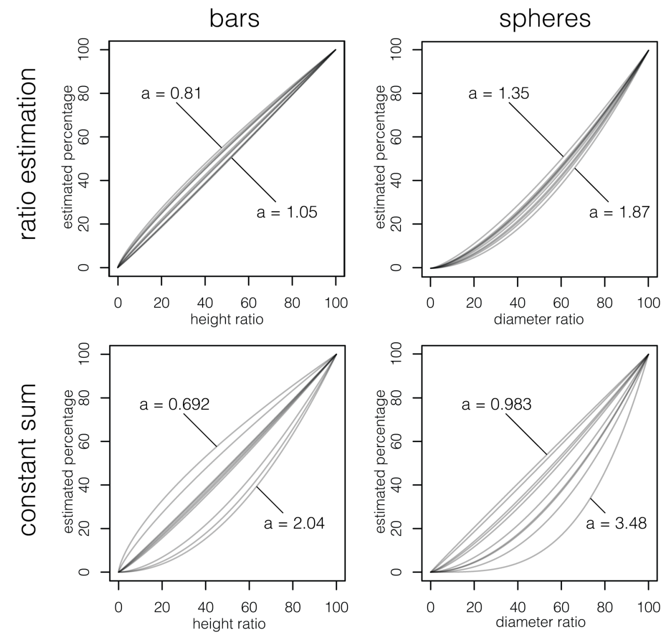


Fig. 10. Spread of individual regression curves with indicated exponents for the respective extremes.

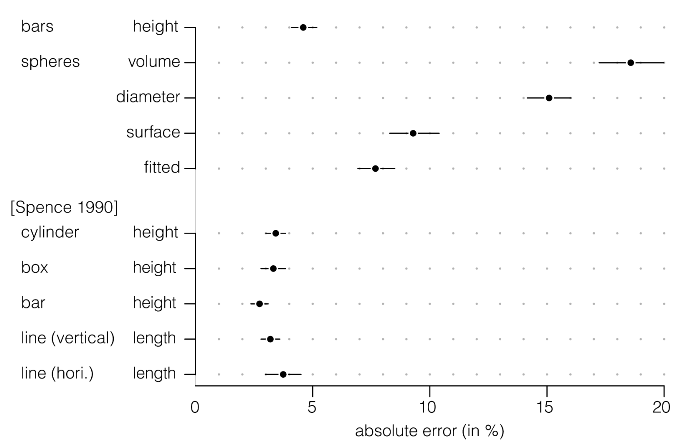


Fig. 11. Dot plot of absolute errors for bars and spheres computed for different objective measures. Error bars indicate 95% bootstrapped confidence intervals. Data for Spence taken from [50].

### 5.4 Bias

In addition to average accuracy, it is important to investigate bias (direction of error) in estimates, particularly to test whether these are dependent on the ratio to judge, that is, whether the bias differs over the range of different sizes. Figure 12 shows the residuals for the four combinations of mark and method against height or diameter ratio. Bars judged with ratio estimation show only little bias to overestimation. Spheres judged with the same method are underestimated for large differences between marks (small ratios), reaching the maximum error at ratios of around 50%. The constant sum method introduces an additional bias towards underestimations.

The curved spread of data points for spheres (ratio estimation) confirms that the diameter ratio is not a suitable indicator for perceived size. Together with the observation from Figure 10 that individual curves vary only moderately (at least when using ratio estimation), a perceptually corrected scale for spheres as physical marks seems possible.

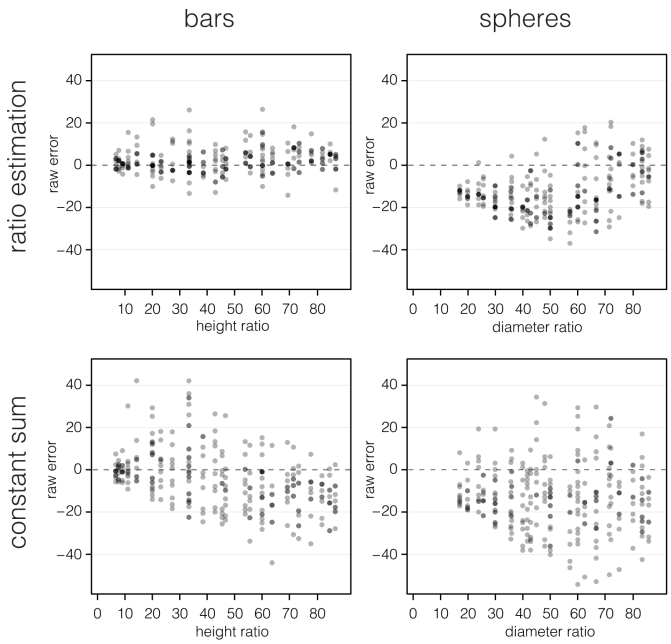


Fig. 12. Residual differences of individual data points from the height or diameter ratio of the bars and spheres.

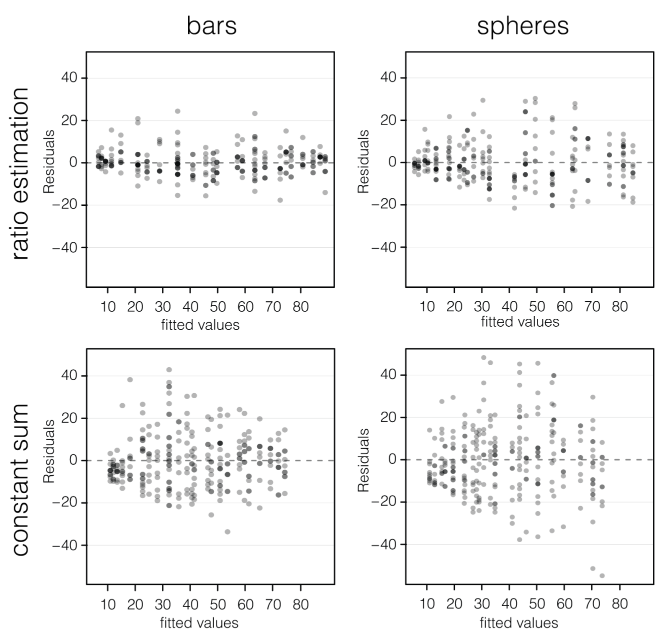


Fig. 13. Residuals after scale correction with our fitted model.

If we apply the model derived from our data $P = S^{1.6}$, then a plot of the residuals normalizes around 0 (see Figures 13). It is interesting to note that the perceived size of spheres is even lower than the growth rate of areas, that is the exponent is $< 2$. However, our experimentally derived exponent is an intrinsically good fit for our data. Previous work on various volumetric shapes report exponents around 0.7 which translates to 2.1 for sphere diameters. It is thus possible that the actual perceived size of spheres is proportional to their surface area; a predictor that has also previously been proposed for the haptic perception of shapes [30].

### 5.5 Informal Participant Feedback

During the post-experiment debriefing, we inquired whether participants consciously employed any strategies for estimating the ratio between the spheres such as basing their judgments on the distance between the spheres. No one reported any consistent strategy but three indicated that for large differences they imagined filling the larger sphere with multiples of the smaller one. All participants expressed low confidence in their sphere judgments and three even apologized for not being "very good at this".

## 6 Discussion

We have presented a study of the perception of size for two physical marks, bars and spheres, using the two most popular estimation methods to assess graph perception *ratio estimation* (used by [8, 11]) and *constant sum* (used by [50, 39]).

### 6.1 Bars

Our results show that participants perceive ratios between bars quite accurately, at least when they are presented isolated; this is consistent with prior work on the size perception of plates [51]. Talbot et al. [58] showed for two-dimensional bar charts that separation and distractors can have an effect. Further studies are therefore necessary to evaluate possible context effects of distractor bars. Furthermore, we only tested bars arranged perpendicular to the line of sight of the observer. For bars that vary in their placement along the line of sight, other factor such as size constancy play a role. While a considerable amount of work exists on size constancy and the perception of distances (e.g., [35, 61, 36, 23]), it is currently unclear how these findings apply to the perception of physical marks due to the many factors playing together in an actual physicalization with more than two data points such as size, distance, distractors, possible occlusions between marks, or the visibility of a baseline for comparison.

### 6.2 Spheres

For spheres our results show that data would be wrongly interpreted if it was encoded in the volume of a sphere. We found underestimations of in average 18.6% $\pm$ 1.4% assuming a volume-scale encoding. However, we also found that if a different scale is applied for encoding, such as surface-based, the accuracy of sphere perception can be immensely improved to an error rate of 9.3% $\pm$ 1.1% for a surface-scale and, somewhat unsurprisingly, an experimentally derived perceptual scaling can make the error rates drop further to 7.7% $\pm$ 0.8%. Since no prior work exists on how the presence of distractors or separation for spheres, it is difficult to predict the size of possible effect. More studies are required to address this question.

### 6.3 Other Physical Marks

We only studied bars and spheres to represent two classes of physical marks: bars for marks which only vary in one dimension, and spheres for marks which vary in all three dimensions. A next step is to test whether these two marks are indeed representative for these two classes of physical marks. Previous work on the size perception of volumes [1] reports overall similar values for different types of shapes. Recent work on the haptic perception of cubes, spheres, and pyramids also found surface area to be the best predictor [30]. We can thus hypothesize similar findings for other (regular) marks.

### 6.4 Alternative Models

We used a simple nonlinear model to fit our data: $P = S^a$. It is based on Stevens' law and provides a reasonably good fit for data sets with multiple participants (see Figures 8 and 10). While visually checking the fit for individual regressions, we noticed that some people showed response curves with a tendency to overestimations at one end of the range and underestimations at the opposite range. Such response behavior can be best described with a logistic curve (an S-shaped curve). Figure 14 illustrates an overplot of logistic curves fitted per participant. While the curves for the ratio estimation method are rather similar to the ones we found with the simpler model, the curves for the constant sum method are more informative than what the simpler model

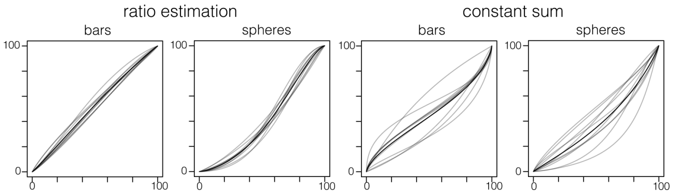


Fig. 14. Overplotted individually fit regressions for a logistic curve model.

showed – especially for bars. While the simpler model indicated an almost linear response curve for 5 of the 10 participants, the logistic curves fit the data better and show that for large differences between the two stimuli (small ratio), 7/10 participants were prone to overestimations whereas small differences led to underestimations for 9/10. The same effect can be observed for spheres where the response bias for large differences is less pronounced with the constant sum method then with the ratio estimation method.

While the logistic curve model can be useful for an exploratory analysis of data, it is less descriptive and parsimonious [64] than the simpler model. The simplicity of Stevens' law supports an intuitive interpretation of the exponent: $a < 1$ indicates a bias to overestimations, and $a > 1$ indicates the opposite. For a = 1, we see a linear response curve. Such an intuitive interpretation is not possible with the logistic curve model where two parameters interact in a more complex fashion.

### 6.5 Estimation Methods

The two estimation methods produced notably different results: while each participant made consistent judgments with both methods, judgments *between* participants varied immensely for the constant sum method (see Figure 8). Spence [50] however reported that he observed a high level of accuracy using this method (higher than Cleveland and McGill [11]). There are multiple possible explanations for our findings. Spence reports that the experimenter observed participants during 10 practice trials and asked them to repeat the practice trials if they did not perform the task correctly. No information is included about how often this was the case and which rationale was used to determine whether participants performed the task correctly.

In contrast to the procedure used by Spence, we only showed a simple line to the participants without any scale indicators whereas Spence displayed tick marks on the line and, for half of his participants, he provided also a number scale next to the tick marks. Thus it is possible that his participants performed the task by estimating a number and then searching for it on the scale.

Either way, the constant sum method is of interest for the study of graph perception as it is a purely visual method whereas the ratio estimation method is a cross-modality matching task [53]. We refrain therefore from discouraging the use of the constant sum method due to our findings but recommend verifying that all participants have adopted the same mental model of the task. This could be accomplished for example by asking them to estimate marks which are not included in the experiment and by providing them with feedback. Alternatively, during the training trials, the corresponding ratio estimate number could be displayed when moving the line divider such that participants can build the correct mental model of the line. They would continue to practice until their error rate falls below a certain threshold.

### 6.6 Implications for the Design of Physicalizations

If a physicalization designer encodes data in the height of a bar, she can expect that observers will perceive the data almost as accurately as when using a two-dimensional bar.

The case is different for spheres though. One of the potential benefits of spheres is that wide ranges of data can be encoded. Let us assume a designer needs to encode values, for example, between 1 and 1000. By encoding data in the *volume* of a sphere, the range could be encoded in sphere diameters of about [1.2, 12]. This would lead though to unacceptable errors of about 20% in average. The use of diameter would lead to slightly lower errors, which are still around 15%, but lead to huge differences in diameter between spheres. Now if she instead uses the *surface* of the spheres for her encoding scale, then the range of sizes remains within acceptable ranges: [0.6, 18]. From a practical standpoint, the use of the surface area of a sphere (or possibly even any physical mark's surface [30]) seems a reasonable heuristic to reduce error rates immensely.

Several participants felt the need to apologize during the experiment that they are "really bad at estimating spheres". While this is only an informal observation, and we did not explicitly ask participants to indicate their confidence for each estimate, it is worth studying further. If we can identify physical marks (or graphical marks) for which participants' judgments are within acceptable error margins but for which participants feel little confidence in their estimates, then such marks could be suitable candidates to encode uncertainty (as has been previously proposed for sketchiness as a visual variable [6]).

## 7 Conclusion

We presented the first study of the perception of size as a physical variable for two physical marks, bars and spheres. Our results showed that physical bars achieved almost the same levels of accuracy as two-dimensional bars. For spheres, the use of a volumetric scale leads to large error rates of 18.6%. However, the surface area of a sphere approximates the perceived size of this physical mark reasonably well such that error rates go down to 9.3%.

Additionally, we presented a series of analysis steps to determine the suitability of a physical variable to encode data. Most of this analysis relies on a combination of fitting (already widely accepted) models and visually assessing the variability between subjects, their accuracy, and finally their estimation biases. If multiple objective measures exist to describe the physical variable being tested, then all of these should be tested as possible predictors for the perception of the variable. If one is found that shows little variation across participants, error rates within acceptable limits, and no change of bias across the range of values of the variable, then this measure is a good candidate for a scale for the physical variable being tested.

Many more possible physical variables need to be tested before we can establish a similar ranking for them as exists for visual variables [11]. To support this effort, all data and R scripts used for the analysis of our data will be available on the project website.

## 8 Acknowledgments

We thank Pierre Dragicevic and our anonymous reviewers for their helpful comments to improve this article. This work has been supported by the EC within the 7th framework programme through the FET Open schemes GHOST project (grant #309191) and and by the Danish Council for Strategic Research (grant #10-092316).

## References

[1] J. C. Baird. *Psychophysical Analysis of Visual Space*. International Series of Monographs in Experimental Psychology. Elsevier, 1970.
[2] A. Bartholomew, J. F. Norman, J. Swindle, A. Boswell, and H. Norman. Aging and the use of implicit standards in the visual perception of length. *Journal of Vision*, 10(7):485–485, 2 Aug. 2010.
[3] J. Bertin. Semiology of graphics: diagrams, networks, maps. 1983.
[4] A. Bezerianos and P. Isenberg. Perception of visual variables on tiled wall-sized displays for information visualization applications. *TVCG*, 18(12):2516–2525, 2012.
[5] M. A. Borkin, A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva, and H. Pfister. What makes a visualization memorable? *IEEE Trans. Vis. Comput. Graph.*, 19(12):2306–2315, Dec. 2013.
[6] N. Boukhelifa, A. Bezerianos, T. Isenberg, and J.-D. Fekete. Evaluating sketchiness as a visual variable for the depiction of qualitative uncertainty. *TVCG*, 18(12):2769–2778, 2012.
[7] J. I. Clarke. Statistical Map-Reading. *Geography*, 44(2):96–104, 1 Apr. 1959.
[8] W. S. Cleveland. Graphs in scientific publications. *Am. Stat.*, 38(4):261–269, 1 Nov. 1984.

[9] W. S. Cleveland, C. S. Harris, and R. McGill. Judgments of circle sizes on statistical maps. *J. Am. Stat. Assoc.*, 77(379):541–547, 1 Sept. 1982.
[10] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *J. Am. Stat. Assoc.*, 1984.
[11] W. S. Cleveland and R. McGill. An experiment in graphical perception. *Int. J. Man. Mach. Stud.*, 25(5):491–500, Nov. 1986.
[12] W. S. Cleveland and R. McGill. Graphical perception: The visual decoding of quantitative information on graphical displays of data. *J. R. Stat. Soc. Ser. A*, 150(3):192–229, 1 Jan. 1987.
[13] A. L. Comrey. A proposed method for absolute ratio scaling. *Psychometrika*, 15(3):317–325, Sept. 1950.
[14] F. E. Croxton and H. Stein. Graphic comparisons by bars, squares, circles, and cubes. *J. Am. Stat. Assoc.*, 27(177):54–60, 1932.
[15] P. Dragicevic and Y. Jansen. List of physical visualizations. dataphys.org/list, 2012.
[16] G. Ekman and K. Junge. Psychophysical relations in visual perception of length, area and volume. *Scand. J. Psychol.*, 2(1):1–10, 1 Mar. 1961.
[17] G. Ekman, R. Lindman, and W. Willlam-Olsson. A Psychological Study of Cartographic Symbols. *Percept. Mot. Skills*, 13(3):355–368, 1961.
[18] J. J. Flannery. *The graduated circle: A description, analysis, and evaluation of a quantitative map symbol*. University of Wisconsin–Madison, 1956.
[19] J. J. Flannery. The relative effectiveness of some common graduated point symbols in the presentation of quantitative data. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 8(2):96–109, 1971.
[20] R. W. Fleming, A. Torralba, and E. H. Adelson. Specular reflections and the perception of shape. *J. Vis.*, 4(9):798–820, 23 Sept. 2004.
[21] S. Follmer, D. Leithinger, A. Olwal, A. Hogge, and H. Ishii. inform: Dynamic physical affordances and constraints through shape and object actuation. In *Proc. UIST'13*, pages 417–426. ACM.
[22] D. M. Green and R. Duncan Luce. Variability of magnitude estimates: A timing theory analysis. *Percept. Psychophys.*, 15(2):291–300, 1974.
[23] R. N. Haber and C. A. Levin. The independence of size perception and distance perception. *Percept. Psychophys.*, 63(7):1140–1152, Oct. 2001.
[24] T. Hogan and E. Hornecker. How does representation modality affect User-Experience of data artifacts? In *Haptic and Audio Interaction Design*, Lecture Notes in Computer Science, pages 141–151. Springer Berlin Heidelberg, 1 Jan. 2012.
[25] J. G. Hollands and I. Spence. Judgments of change and proportion in graphical perception. *Hum. Factors*, 34(3):313–334, June 1992.
[26] Y. Jansen and P. Dragicevic. An interaction model for visualizations beyond the desktop. *TVCG*, 19(12):2396–2405, 2013.
[27] Y. Jansen, P. Dragicevic, and J.-D. Fekete. Evaluating the efficiency of physical visualizations. In *Proc. CHI'13*, pages 2593–2602. ACM.
[28] Y. Jansen, P. Dragicevic, P. Isenberg, J. Alexander, A. Karnik, J. Kildal, S. Subramanian, and K. Hornbaek. Opportunities and challenges in data physicalization. In *Proc. of CHI'15*, 2015.
[29] M. Kahrimanovic, W. M. Bergmann Tiest, and A. M. L. Kappers. Seeing and feeling volumes: The influence of shape on volume perception. *Acta Psychol.*, 134(3):385–390, July 2010.
[30] M. Kahrimanovic, W. M. B. Tiest, and A. M. L. Kappers. Haptic perception of volume and surface area of 3-D objects. *Atten. Percept. Psychophys.*, 72(2):517–527, Feb. 2010.
[31] R. A. Khot, L. Hjorth, and F. F. Mueller. Understanding physical activity through 3D printed material artifacts. In *Proc. CHI'14*, pages 3835–3844. ACM.
[32] I. M. Lipkus, G. Samsa, and B. K. Rimer. General performance on a numeracy scale among highly educated samples. *Med. Decis. Making*, 21(1):37–44, Jan. 2001.
[33] M. Macdonald-Ross. How numbers are shown. *AVCR*, 25(4):359–409, 1 Dec. 1977.
[34] J. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Trans. Graph.*, 5(2):110–141, Apr. 1986.
[35] G. Martius. Über die scheinbare Größe der Gegenstände und ihre Beziehung zur Größe der Netzhautbilder. *Philos. Stud.*, 5:601–617, 1889.
[36] D. McCready. On size, distance, and visual angle perception. *Percept. Psychophys.*, 37(4):323–334, Apr. 1985.
[37] T. Moher. Embedded phenomena: supporting science learning with classroom-sized distributed simulations. In *Proc. CHI'06*, pages 691–700. ACM.
[38] B. Nissen and J. Bowers. Data-things: Digital fabrication situated within participatory data translation activities. In *Proc. CHI'15*, pages 2467–2476. ACM.
[39] J. F. Norman, J. T. Todd, V. J. Perotti, and J. S. Tittle. The visual perception of three-dimensional length. *J. Exp. Psychol. Hum. Percept. Perform.*, 22(1):173–186, Feb. 1996.
[40] A. J. Parkes. *Phrases of the kinetic: dynamic physicality as a dimension of the design process*. PhD thesis, Massachusetts Institute of Technology, 2009.
[41] L. V. Peterson and W. Schramm. How accurately are different kinds of graphs read? *Audiovisual communication review*, 2(3):178–189, 1954.
[42] F. Phillips and E. J. L. Egan. Crossmodal information for visual and haptic discrimination. In *IS&T/SPIE Electronic Imaging*, pages 72400H–72400H–15. International Society for Optics and Photonics, 5 Feb. 2009.
[43] J. Plateau. Sur la mesure des sensations physiques, et sur la loi qui lie lintensité de ces sensations à lintensité de la cause excitante. *Bulletins de l'Académie Royale des Sciences, des Lettres et des Beaux-Arts de Belgique*, 23:376–388, 1872.
[44] E. C. Poulton. *Bias in quantifying judgements*. Taylor & Francis, 1989.
[45] Z. Pousman, J. T. Stasko, and M. Mateas. Casual information visualization: Depictions of data in everyday life. *IEEE TVCG*, 13(6):1145–1152, 2007.
[46] M. K. Rasmussen, E. W. Pedersen, M. G. Petersen, and K. Hornbæk. Shape-changing interfaces: A review of the design space and open research questions. In *Proc. CHI'12*, pages 735–744.
[47] B. E. Rogowitz, L. A. Treinish, S. Bryson, and Others. How not to lie with visualization. *Computers in Physics*, 10(3):268–273, 1996.
[48] B. Schneider and R. Bissett. "ratio" and "difference" judgments for length, area, and volume: are there two classes of sensory continua? *J. Exp. Psychol. Hum. Percept. Perform.*, 14(3):503–512, Aug. 1988.
[49] O. Shaer and E. Hornecker. Tangible user interfaces: Past, present, and future directions. *Found. Trends Hum.-Comput. Interact.*, 3(1–2):1–137, 2010.
[50] I. Spence. Visual psychophysics of simple graphical elements. *J. Exp. Psychol. Hum. Percept. Perform.*, 16(4):683–692, Nov. 1990.
[51] R. J. Stanek. A parametric study of volume and surface judgments. *Percept. Psychophys.*, 6(1):16–18, 1969.
[52] S. S. Stevens. On the psychophysical law. *Psychol. Rev.*, 64(3):153–181, May 1957.
[53] S. S. Stevens. Issues in psychophysical measurement. *Psychol. Rev.*, 78(5):426, Sept. 1971.
[54] S. Stusak, J. Schwarz, and A. Butz. Evaluating the memorability of physical visualizations. In *Proc. CHI'15*, pages 3247–3250. ACM, to appear, 2015.
[55] S. Stusak, A. Tabard, F. Sauka, R. A. the, and A. Butz. Activity sculptures: Exploring the impact of physical visualizations on running activity. *IEEE Trans. IVis. Comput. Graph.*, 20(12):2201–2210, Dec. 2014.
[56] S. Swaminathan, C. Shi, Y. Jansen, P. Dragicevic, L. A. Oehlberg, and J.-D. Fekete. Supporting the design and fabrication of physical visualizations. In *Proc. CHI'14*, pages 3845–3854. ACM.
[57] F. Taher, J. Hardy, A. Karnik, C. Weichel, Y. Jansen, K. Hornbaek, and J. Alexander. Exploring interactions with physically dynamic bar charts. In *Proc. CHI'15*, pages 3237–3246. ACM, 2015.
[58] J. Talbot, V. Setlur, and A. Anand. Four experiments on the perception of bar charts. *IEEE TVCG*, 20(12):2152–2160, Dec. 2014.
[59] A. S. Taylor, S. E. Lindley, T. Regan, D. Sweeney, V. Vlachokyriakos, L. Grainger, and J. Lingel. Data-in-place: Thinking through the relations between data and community. In *Proc. CHI'15*, pages 2863–2872. ACM.
[60] M. Teghtsoonian. The judgment of size. *Am. J. Psychol.*, 78:392–402, Sept. 1965.
[61] R. Teghtsoonian and M. Teghtsoonian. The effects of size and distance on magnitude estimations of apparent size. *Am. J. Psychol.*, 83(4):601–612, Dec. 1970.
[62] A. Vande Moere. Beyond the tyranny of the pixel: Exploring the physicality of information visualization. In *IV'08*, 2008.
[63] A. Vande Moere and S. Patel. Analyzing the design approaches of physical data sculptures in a design education context. In *Visual Information Communications International (VINCI'09)*, 2009.
[64] J. Vandekerckhove, D. Matzke, and E.-J. Wagenmakers. Model comparison and the principle of parsimony. *Oxford Handbook of Computational and Mathematical Psychology. Oxford University Press, Oxford*, 2014.
[65] J. Zacks, E. Levy, B. Tversky, and D. J. Schiano. Reading bar graphs: Effects of extraneous depth cues and graphical context. *J. Exp. Psychol. Appl.*, 4(2):119, 1998.