Modeling Just Noticeable Differences in Charts

Min Lu, Joel Lanir, Chufeng Wang, Yucong Yao, Wen Zhang, Oliver Deussen, and Hui Huang

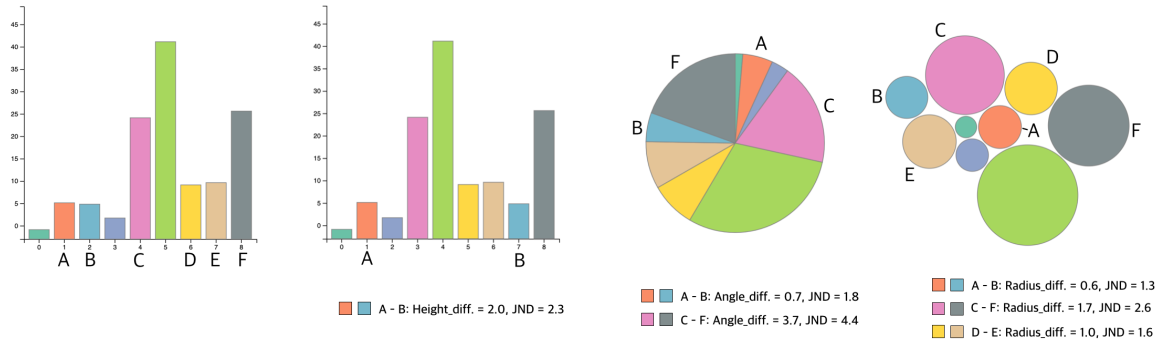

Fig. 1. Four charts of the same dataset, but with different pairs below the Just Noticeable Difference (JND) threshold. From left to right: a bar chart with no pairs below JND, the difference between all pairs of bars in the graph is noticeable; with a different order, there is a pair of bars below JND (A-B); two pairs of indistinguishable fans are detected in the pie chart (A-B and C-F); three pairs of circles are detected as not distinguishable in the bubble chart (A-B, C-F and D-E).

Abstract—One of the fundamental tasks in visualization is to compare two or more visual elements. However, it is often difficult to visually differentiate graphical elements encoding a small difference in value, such as the heights of similar bars in bar chart or angles of similar sections in pie chart. Perceptual laws can be used in order to model when and how we perceive this difference. In this work, we model the perception of Just Noticeable Differences (JNDs), the minimum difference in visual attributes that allow faithfully comparing similar elements, in charts. Specifically, we explore the relation between JNDs and two major visual variables: the intensity of visual elements and the distance between them, and study it in three charts: bar chart, pie chart and bubble chart. Through an empirical study, we identify main effects on JND for distance in bar charts, intensity in pie charts, and both distance and intensity in bubble charts. By fitting a linear mixed effects model, we model JND and find that JND grows as the exponential function of variables. We highlight several usage scenarios that make use of the JND modeling in which elements below the fitted JND are detected and enhanced with secondary visual cues for better discrimination.

Index Terms—Visual perception, Charts, Just noticeable difference, Modeling.

1 Introduction

Charts have been used for the presentation of quantitative information for decades. Quantitative values are mapped to visual attributes such as position, length, or angle in a common scale, by which viewers can intuitively compare and understand the data elements. One of the basic tasks in the reading of charts is to compare two or more visual elements. In a basic bar chart, for example, the height of bars represents the quantitative measure of the element. Positioning bars aligned, side by side, allows for a quick comparison, where one can easily tell the relative relations among the bars. Still, a basic problem in bar chart is that it remains difficult to tell apart close quantities. It is difficult to distinguish between bars with similar heights, or to know when bars represent exactly the same quantity (e.g., the bars ‘A’ and ‘B’ in the second bar chart of Figure 1). This pitfall is even more apparent in other common charts, such as bubble or pie charts in Figure 1. This weakness in perceiving the noticeable difference in chart elements, motivates this work.

Comparison of visual elements that have a small difference between them, relates to the concept of *Just Noticeable Difference (JND)* [15]. JND is a pyschophysical concept defined as the minimum level a stimulus that needs to be changed in order for people to be able to perceive it. The well-known Weber’s law claims that the noticeable difference in a stimulus is proportional to the intensity of the stimulus. For visual stimuli, this usually translates into the length or size of the object. However, research in visualization shows that the perception and discrimination of objects in charts relates not only to the targeted objects themselves, but also to the whole configuration of the visualization, which may include contextual factors such as the spatial setting of the charts (e.g., aligned or stacked bar chart) [7], neighbourhood elements [43], scale of the chart [36] and more.

Inspired by these findings, this work examines the JND within the context of visualization charts, and models it as a function of both the *intensity* of the objects (e.g., their height or size), following Weber’s law, and the *distance* between them. We examine three conventional charts: bar charts, bubble charts, and pie charts, measuring JNDs for these charts through a series of discrimination tests performed by 28 participants. As a result, for each chart, we first identify the main visual variables that have a significant effect on the JND, showing that distance affects the bar chart, intensity affects the pie chart and both distance and intensity affect the bubble chart. Then, we apply mixed effects models to quantify JNDs depends on these variables for the three chart types. Finally, we show how using such a fitted model, we are able to identify and enhance groups of elements with values below the JND threshold, that otherwise for most users, would be difficult to perceptually discriminate.

Thus, the key contributions of this paper are as follows: First, we identify the effect that intensity and distance have on the JND in three common charts: bar chart, pie chart and bubble chart. Second, we

- *Min Lu, Chufeng Wang, Yucong Yao, Wen Zhang and Hui Huang are with Shenzhen University. E-mail: {lumin.vis, chufengwang96, yyc173966019, zhangwen.thu, hhzhiyan}@gmail.com*
- *Joal Lanir is with The University of Haifa. E-mail: ylanir@is.haifa.il*
- *Oliver Deussen is with University of Konstanz. E-mail: oliver.deussen@uni-konstanz.de*
- *Hui Huang is the corresponding author of this paper*

model this effect by a mixed effects model that can be used to predict the JND in these three charts. Third, we demonstrate how this model can be applied for visually enhancing the perception of barely visible differences. Finally, the dataset and code from our experiments is released as open source for further JND related research.

2 Just Noticeable Difference

The *Just Noticeable Difference (JND)*, sometimes also called the difference threshold, is the minimum amount by which a stimulus intensity must be changed to be noticeable. For example, for sound, the JND would be the smallest *change* in volume that a person could sense.

How well humans distinguish properties of different stimuli has been formally studied since the middle of the 19th century. JNDs were first investigated by the German psycho-physiologist Ernst Heinrich Weber in 1834 in a series of experiments. Gustav Fechner, Weber's student, formulated the results into Weber's Law which states that the size of the Just Noticeable Difference is a constant proportion of the original stimulus [3]. In its classic version, it is modeled as $\delta(a) = ca$, where a is the intensity of elements, $\delta(a)$ is the JND, and c is the constant Weber fraction. Gustav Fechner further formulated *Fechner's Law* [28], in which it is stated that the stimulus's perceived strength is proportional to the logarithm of the objective stimulus strength. Subsequently, researchers in psychophysics elaborated *Weber-Fechner Laws* for various circumstances, such as visual discrimination of brightness [15], orientation [39] and more.

Weber and Fechner's work focused mostly on differences in weight and light intensity. However subsequent studies found mixed support for either Weber's law or Fechner's law with different types of stimuli [17]. A deviation of Fechner's Law was proposed by Stevens, who stated that the sensation is better modeled by a power law such as, $S = cI^{\alpha}$, where S is the sensation magnitude, I is the stimulus intensity and α depends on the sensory modality. Stevens showed that different stimuli such as brightness and loudness are better modelled using a power law, and proceeded in calculating exponent values for various stimuli [35]. Still, critics argue about the validity of Stevens's law, also saying that the power law can be deduced mathematically from Fechner's logarithm function [26].

Looking at how Weber's law applies to different visual elements, the lengths of lines was reported to follow Weber's law already in Weber's initial report and was further discussed by Fechner who conducted an experiment dealing with the discriminability of distances between two points [9]. This was also confirmed by further studies [29]. Other works claimed that a power law, similar to what Stevens has modelled was found to best fit judgments of perceiving the differences in the lengths of lines [13]. Area perception was also found to follow Stevens's law, by Augustin and Roscher, who looked at perceived differences in area of squares [1]. However, only a few studies have examined JND in the context of visualizations. We aim to fill this gap, and examine JND within three conventional visualization charts.

3 Related Work

Next, we survey works related to the understanding of charts and specifically, comparison tasks in charts, followed by a discussion of works that deal with perception modeling in information visualization.

3.1 Visual Comparison

Visual comparison is a fundamental task in visual analysis [4, 11, 38]. Several studies looked at how the performance of visual comparisons is affected by various contextual factors. Talbot et al. [37] investigate the performance of estimating relative sizes (i.e., the relative height of a shorter bar compared to a taller bar) as a function of the distance between them and surrounding distractors. Their results show that the larger the separation is between bars, the more difficult it is to compare between their heights. Regarding distractors (i.e., height of bars separating the targets), the results of the study were ambiguous, however other studies have managed to find and show that the perception of a bar is affected by its neighbours (neighbourhood effect) [43].

Other studies looked at different factors that may affect comparison tasks. Srinivasan et al. [34] evaluated the usefulness of overlays in such tasks for four variants of bar charts and report that charts with difference overlays facilitate a wider range of comparison tasks. Kim et al. [21] study how positive and negative visual framing within charts affects the feedback from users (framing effect). Going beyond 2D charts, Zacks et al. [42] especially looked at relative comparisons among 3D bars. Our work is in line with the above findings, but focuses on perceptual modelling of human performance in a comparison task, distinguishing similar visual elements that encode nearby values.

3.2 Perception Modeling in Visualization

A great deal of research has been conducted studying the visual perception of different types of visual channels and their impact on visualizations. The seminal work of Cleveland and McGill [6] provides an initial framework for the ordering of visual variables, comparing elementary perceptual tasks such as position on a common and non-aligned scale, length, direction, angle and more. Later, Mackinlay [27] extended the study to non-quantitative perception tasks, evaluating the performance of visual variables in encoding ordinal and nominal information. More recently, Heer and Bostock [16] performed an approximate replication of Cleveland and McGill's study using crowdsourcing. They confirmed the relative rankings among the different types of visual variables.

Looking to model perception, researchers measure and model how viewers perceive visual objects within the context of an entire chart or the context of different tasks. Many such studies focused on user perception of scatterplots. Looking at high-level perception tasks such as the detection of correlation, Rensink and Baldridge [30] modelled human perception of correlations in scatterplots finding that it is congruous with Weber's and Fechner's law. Gleicher et al. [12] explored mean value judgements in multi-class scatterplots and found that viewers can efficiently make comparative mean judgements across a variety of conditions and encodings. Harrison et al. [14] extended the study of scatterplots to examine the overall perception of correlations between two variables. Following this work, Kay and Heer [20] provide a secondary analysis of the data by Harrison et al., but consider individual differences in the precision of estimations. Later on, Yang et al. [41] modulate not only correlation values, but also candidate visual features that best align with participants' judgments. Looking at the apparent order in the data, Chung et al. [5] also employed a crowdsourcing study to investigate the perception of order in a sequence of elements with different visual channels. Finally, Hughes [19] estimates Weber's constant for 2D and 3D bar charts and draws the conclusion that the JND for 3D charts would be larger than for 2D charts. However, in his experiment, bar charts only consisted of two targeted bars.

Our work is based on the above in-depth studies for visual perception, but aims at modelling JND for comparison tasks of visual items, especially studying when the difference between items cannot be perceived. We examine this with three different chart types, and study not only the effect that the stimuli size (e.g., the height of the bar in a bar chart) might have on the JND, but also the distance between the comparable items in the chart.

4 Experiment

We aim to analyze the perception of JND in chart elements in order to understand (and possibly aid) how people compare visual elements that have close quantities. In particular, we study three popular charts: bar charts, pie charts and bubble charts. In standard psychophysical experiments, targets are presented alone, without any context. In our case, the JND of targeted visual elements are analyzed in the context of element comparison tasks within a chart. We focus on two variables that may affect the perception of comparison:

Object Intensity. Previous works in psychophysics demonstrate that the *object intensity* has a significant effect on the JND and can be quantified by Weber's Law: the increment noticeable intensity threshold is proportional to the object's intensity. For example, in a noisy environment, people might shout to be heard while a whisper works in a quiet room. In our experiment, object intensity specifically refers to visual cue that encodes the data, i.e., height of bar in bar charts, radius of circle in bubble charts, and angle of fan in pie charts.

Separation Distance. In a visualization, the JND for items of a chart may be affected by other contextual factors as well. We focus here on distance between elements because distance was shown to be one of the most dominant factors affecting element comparison in charts [37]. *Separation distance*, is measured by the Euclidean distance between two bars in a bar chart, or two circles in a bubble chart. For pie charts, the separation distance is modulated by the angular difference between two targeted fans. While other contextual factors such as color or neighboring items may also affect JND (i.e., a small, yet existent neighborhood effect of close-by bars was found in bar charts [43])), we focus on distance as it is strongly supported by existing literature and seems to have the highest potential to affect JNDs.

When performing our experiments, we follow a custom methodology of JND examination by using the method of *Constant Stimuli*. It studies when and how similar elements with equally increased or decreased differences can be distinguished.

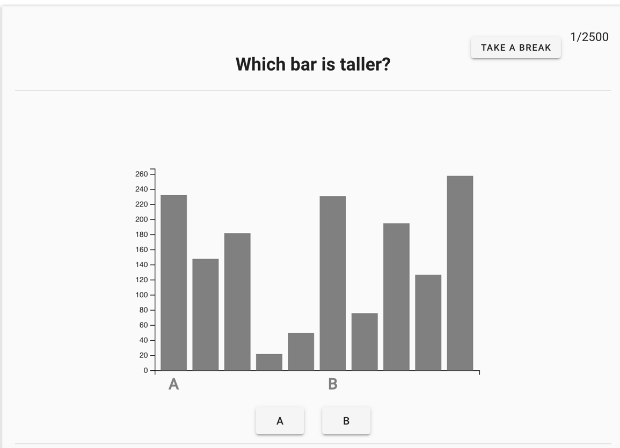

Fig. 2. Experiment Interface for the Bar Chart: ‘Which bar is taller, A or B?”

4.1 Conditions

We examine three types of charts in our experiments: bar charts, pie charts and bubble charts (see Fig. 1). For each of them, we vary two independent variables: object intensity and separation distance, each having 5 levels. Each experimental *condition* refers to a unique combination of separation distance and object intensity under which JNDs are measured. Next, we describe the configuration of the stimuli in each of the three chart types. Note that all stimuli are measured in pixels in a controlled display environment (see Section 4.3).

Bar Chart All bar charts were generated with the same configuration: 10 bars horizontally aligned, each bar is 33 pixel in width, the horizontal gap between bars is fixed to 9 pixels. In each trial, two bars were targeted to be compared, while the rest of eight bars were generated with random height.

The separation distance of the two targeted bars was set to five levels, according to the number of bars between them, *N*, where *N* is either 0, 2, 4, 6, or 8. That is, the distance between the target bars varied from 9 pixels to 345 pixels on 5 different samples. Object intensity (the height of standard stimuli bar in each trial) was sampled at 5 levels, between 50 pixels to 250 pixels, 50 pixels per level. Thus, 25 conditions in total were generated.

Pie Chart Pie charts were also generated with a common configuration: 8 fans, composing a pie with a radius of 135 pixel. In each trial, the two targeted fans were set at a certain angle and angular distance, then the remaining six fans were generated at a random angle, but in a way that a whole pie chart was generated.

For the two targeted fans, the angle of the standard stimuli was sampled at 5 levels, from 10 to 130 degree, with steps of 30 degrees. The separation distance (smaller angle between the two fans) was set to 5 levels from 0 to 100 degrees. Due to the constraint of not exceeding 360 degrees, three conditions were sorted out as being invalid (one condition with fans of 100 degree but 100 degree separation, and two conditions with fans of 130 degree, but 80 or 100 degree separated), resulting in a total of 22 conditions.

Bubble Chart Bubble charts were generated with a common configuration: 10 circles, placed on a canvas of 500 pixel width and 400 pixel height. In each trial, the two targeted circles were set first with having a certain radius and distance between them, then the rest of eight circles were generated at random sizes and positions without overlap.

The radius of the standard stimuli circle (one of the two target circles) was sampled at 5 levels, from 10 pixel to 50 pixel. The distance between the two circles was sampled at 5 levels, from 0 pixels to 200 pixel, with steps of 50 pixels. Note that with zero pixel distance the two circles are tangent. In total, 25 conditions were rendered.

4.2 JND Measurement

For each condition, we followed the classic JND experimental methodology using the method of *Constant Stimuli*. In this method, a number of stimulus values that are likely to encompass the JND case are chosen and presented in a quasi-random order to observers. For each stimulus presentation, the observer reports whether its intensity is ‘stronger’ (i.e., ‘taller’ in the case of bar chart, ‘larger’ in bubble chart and pie chart) than the standard stimulus. Then the proportion of ‘stronger’ responses is calculated for each stimulus level. A so-called *Psychometric Function* is constructed by taking the stimulus intensity as the x-axis and proportion of ‘stronger’ as y-axis (see Figure 3(left)) [8]. JND is defined as the difference between two x-axis values at which the function crosses 0.5 and 0.75 [24].

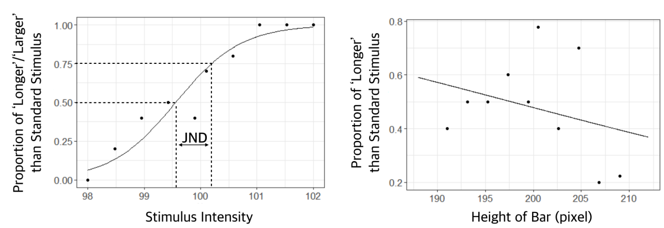

Fig. 3. JND Calculation: (left) a normal fitted Psychometric Function example with a sigmoidal shape; (right) an ill-fitted example in bar chart (explained in Section 5) , which is caused by a severe disorder in the data (as the height of comparison stimuli increased, the participant didn’t consistently report more answers that they were taller than the standard stimulus).

For example, if we look at a bar chart in which there are two separating bars in a distance (e.g., in a distance of 83 pixels) and an object intensity of 100 pixel, we generate a sequence of trials in which the height of the standard stimulus bar (denoted by ’A’) is fixed at 100 pixels, while the height of the comparison stimulus (denoted by ’B’) varies within a certain range $Height_B = [Height_A - \varepsilon_{lower}, Height_A + \varepsilon_{upper}]$. The range is bounded by two values, ε_{lower} and ε_{upper}. The lower bound is smaller than the standard stimuli which can easily be distinguished from it, the upper bound is larger than the standard stimuli and can also easily be distinguished from it. We now evenly sample 10 levels within this range and repeat 10 trials at each level. For example, in Fig. 3(left), $(\varepsilon_{lower}, \varepsilon_{upper})$ is set to $(-2, 2)$ pixels, with ten comparison stimuli evenly sampled in-between. For each comparison, the proportion of responses that stimulus is identified ‘taller’ or ‘larger’ than the standard stimulus is computed and then Psychometric Function is constructed. The JND is computed as the intensity difference at the proportion of 0.5 and 0.75.

Overall, each participant performed 2500 trials with bar charts (25 conditions x 10 comparisons x 10 responses), 2200 trials with pie charts (22 conditions x 10 comparisons x 10 responses), and 2500 trials with bubble charts (25 conditions x 10 comparisons x 10 responses).

In order to set the bounds for all conditions of all three charts, for each condition, one author of this work proposed a pair of

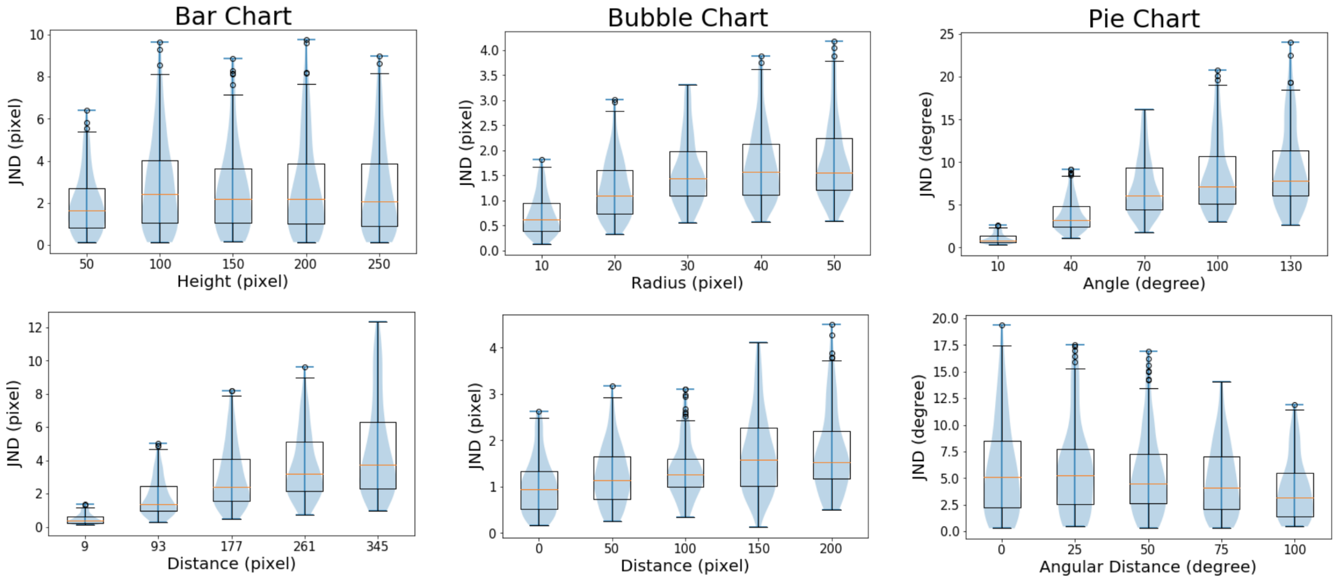

Fig. 4. Distribution of JND according to the two independent variables: (left) in bar chart, a strong positive correlation can be observed between JND and the distance between bars, while a weak to non-existent correlation between JND and the height of bars; (middle) both distance and radius of circles show positive correlation with JND in bubble chart; (right) in pie chart, JND shows a strong correlation to the angles of fans to be compared, while weak correlation with angular distance.

($\varepsilon_{lower}, \varepsilon_{upper}$), e.g., (-5, 5) pixels. This value was calibrated by a second author, until the differences by the lower bound and upper bound were easily perceived.

4.3 Procedure

We recruited 28 participants from a local university, 15 females and 13 males, with an average age of 21 years (range 19-25), 22 out of the 28 participants had a major in science or engineering, six participants had a major in humanities and social science. Five of them were post graduate students, the rest were undergraduates. Twenty-two participants reported themselves not having previous experiences in visual discrimination tasks or any professional knowledge in visualization, six participants self-reported with a medium or pretty good background in visualization. All participants reported having normal or corrected-to-normal visual acuity as well as not having any type of color-blindness.

We divided the experiment into three sessions of different chart types. The 28 participants got tested in all of the three sessions. Each participant performed the tests independently in each session. The participant was asked to sit in front of a personal computer with a 14-inch display screen (12.2″ x 6.86″ display size, 157 pixels per inch, 0.161mm dot pitch), at a distance of about 45cm from the screen. The testing interface was presented on a full-screen display (Figure 2). At the beginning of the first session, we collected demographic information, including age, gender, academic field, academic level, and experience in data visualization. Then some basic instructions were given. Participants were asked to provide the best answer possible according to their immediate perception. Participants were encouraged to take a break whenever they got tired.

Each participant first went through a short practice block with several trials to get acquainted with the task. Then the formal trials began. In each trial, a chart was presented to the participant (bar, pie or bubble) on which two stimuli were presented. Fig. 2 shows an example of the user interface of one trial. The experiment had only one type of task: *Which object is taller/larger?* with two-alternative choices *A*, and *B*. Participants were not returned feedback about how they did well in each trial. The participants were notified about their progress on the top-right corner of the interface. Anytime during the experiment, the participants could take a break and later resume the test. For each trial, the response was recorded. Each session took around one hour to complete, averagely 1.5 second every trial. All trials were presented at random order, within each trial left and right order of standard stimulus and comparison stimulus was also randomized.

5 Modeling

Responses from all 28 participants were collected. From the responses to each of the conditions and each participant, a psychometric function was fitted by using Generalized Linear Model with logit link function, from which the JND value can be estimated. All raw data from the experiment, the calculated JNDs, and analysis codes are provided online `https://github.com/deardeer/JND-in-Charts`.

Within the resulting values, several ill-fitted JNDs were found. As exemplified in Fig. 3(right), ill-fitted JNDs are cases in which the proportion of the comparison stimulus taller/larger than the standard stimulus changes dramatically instead of consistently increasing as the intensity of the comparison stimulus increase. Therefore we removed JNDs with extremely large or extremely small values. This way, we excluded six outliers out of 700 JNDs (28 participants x 25 conditions) in bar charts which are either smaller than 0.1 or larger than 50 pixel, five outliers out of 616 JNDs (28 participants x 22 conditions) for the pie charts which are either smaller than 0.1 or larger than 60 degree, and one outlier out of 700 JNDs (28 participants x 25 conditions) in bubble charts which are either smaller than 0.1 or larger than 25 pixel.

After pre-processing, we plotted the distribution of the JNDs of the three charts over the two independent variables. Fig. 4 shows the effects of the two variables on the JND in each chart. A strong positive correlation can be observed between the JND and the distance between bars in the bar chart, while we see only a weak to non-existent correlation between the JND to the height of the bars. In the pie chart, JND shows a strong correlation to the angles of fans to be compared, while it is not correlated to angular distance. In the bubble chart, both the distance and radius of circles show positive correlation with JND. To examine these correlations, we first applied analysis of variance (ANOVA) to identify whether there is a main effect for each of the variables in each of the charts. Then we applied our model settings to fit the JND to the corresponding changes of these variables.

5.1 Variables with Main Effects

We performed a two-way within subject ANOVA with a linear mixed model to evaluate the two independent variables, *separation distance*

Table 1. Analysis of Variance.

	Distance	Intensity	Distance:Intensity
Bar	F(4,642)=6.9 p<.0001	F(4,642)=0.03 p=0.999	F(16,642)=1.6 p=0.055
Pie	F(4,562)=0.1 p=0.978	F(4,562)=34.7 p<.0001	F(13,562)=1.9 p=0.026
Bubble	F(4,647)=10.6 p<.0001	F(4,647)=24.5 p<.0001	F(16,647)=0.7 p=0.788

and *intensity*, for each of the three charts.

As shown in Table 1, for bar charts only the main effect of the separation distance on JND is significant, while the main effect of the intensity (height of bars) and the interaction effect are not significant. For pie charts only the intensity (angle of fan) has a significant main effect on the JND, while the main effect of separation distance and the interaction effect are not found to be significant at a significance level of 0.01. For bubble charts the ANOVA result shows that both, the main effects of intensity (radius of circle) and separation distance are significant. The interaction effect is not significant .

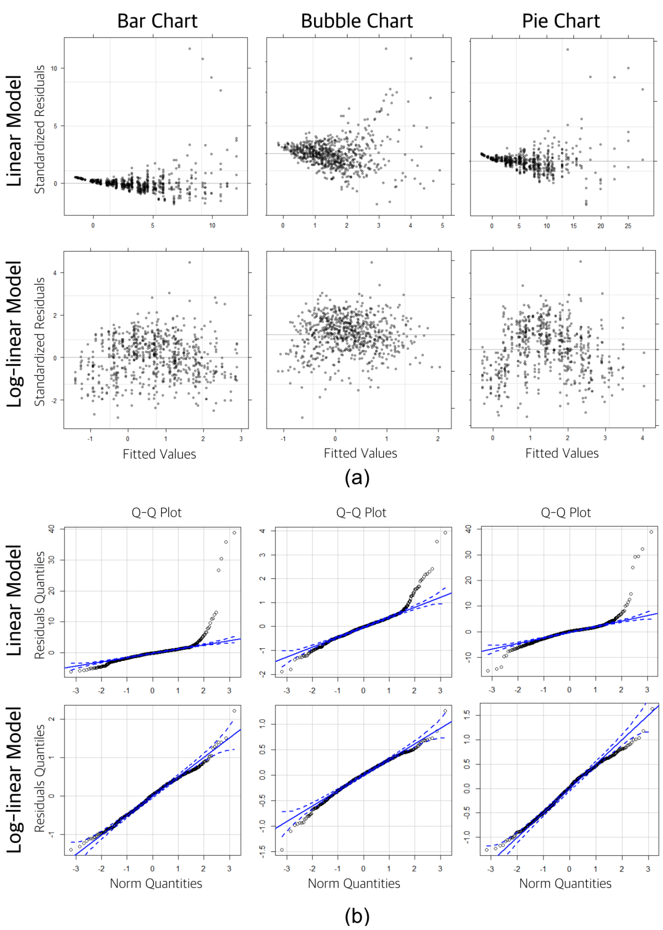

Fig. 5. Comparison of fits of the linear mixed effect model and the log-linear mixed effect model: (a) the residual plot of the linear model shows *non-constant variance*, while the residual plot of the log-linear model shows *constant variance*. (b) the Q-Q plot of the linear model shows *non-normality* of residuals, while the Q-Q plot of the log-linear model shows improvement of normality

5.2 JND Model Setting

Linear Mixed Effects Model We begin our modeling by recalling our experimental design. The collected JND data is non-independent and imbalanced: (1) non-independent: each participant was repetitively measured over all conditions during the experimental procedure. We cannot apply a simple linear regression due to the clear violation of the independence assumption in the linear regression model [10]. (2) imbalanced: as described earlier, a very few mistaken JNDs and extremely large/small JNDs were excluded. This results in an imbalanced dataset for repeated measures. Therefore we start with a linear mixed effect model [2, 40], which takes participant effects into account by adding a participant varying-intercept random effect u_k to a simple linear model.

For each of the three charts c, we have one response variable, the JND, and two independent variables, the Separation Distance $Distance$ and the Object Intensity $Intensity$. Since interaction effects were not significant for all charts according to the result of ANOVA, the linear mixed effect model can be described in a general form as follows:

$$JND_{i,c} = \beta_{c,0} + \beta_{c,1}Distance_i + \beta_{c,2}Intensity_i + u_k + \varepsilon_i \qquad (1)$$

Here, u_k is the random effect which describes the performance difference of participant k, it is randomly drawn from normal distribution $N(0,\tau^2)$; ε_i is the random error which follows a normal distribution $N(0,\sigma^2)$. Note that for each chart situation, the $JND_{i,c}$ is equal to a linear function of $Distance_i$ and $Intensity_i$ with the offset $\beta_{c,0}$ and the corresponding slopes $\beta_{c,1}$ and $\beta_{c,2}$ plus a random offset u_k specified per participant and a normally-distributed random error ε_i. According to the ANOVA result described in Sect. 5.1, we set $\beta_{bar,2} = 0$ for bar charts and $\beta_{pie,1} = 0$ for pie charts.

Log-linear Mixed Effects Model Some violations of the model assumptions for the linear mixed effect model are inherent to our data. The most crucial violation of model assumptions is the non-constant variance (heteroscedasticity), illustrated in Fig. 5(a)[1]. As shown in our model definition in Equation 1, the variance of the random error, σ^2, should be a constant. The upper row of Fig. 5(a) shows residuals plots of the linear mixed effects models. The fitted JND_c increases with increasing scale of the residual; this is inconsistent with the assumption of a constant variance in the linear mixed effects model.

To meet assumption of a constant variance (homoscedasticity) for the random error, a common and efficient approach is to apply a log transformation to the data [25]. We performed a log transformation to JND_c, then the log-linear mixed effects model would now look like:

$$\log(JND_{i,c}) = \beta_{c,0} + \beta_{c,1}Distance_i + \beta_{c,2}Intensity_i + u_k + \varepsilon_i \qquad (2)$$

Here $u_k \sim N(0,\tau^2)$ and $\varepsilon_i \sim N(0,\sigma^2)$. The log transformation has additional advantage by avoiding meaningless predictions such as negative predictions for JNDs. Since the inverse transformation of the log transformation would map values in $(-\infty,\infty)$ to $(0,\infty)$, the log linear model ensures that the predicted values of JNDs are always greater than 0, while this is not the case for the linear mixed effect model.

Comparison of Two Models The results of linear mixed effects model and log-linear mixed effects model are presented in Figure 5 and 6. Log-linear mixed effects model outperforms linear mixed effects model. Specifically, (1) *Homoscedasticity*. It shows that there is no longer a violation of the constant variance assumption in model fitting by comparing the residual diagnosis of the log-linear model with the linear model (Fig. 5(a)). We performed homoscedasticity test of residuals by using Breusch-Pagan test. The linear models do not hold the assumption of homoscedasticity for residuals ($p < .0001$) for all charts, while the log-linear models hold the assumption of homoscedasticity for residuals. (2) *Normality of residuals*. It displays that log-linear models have less skewness and kurtosis (Fig. 6) in the residuals than the linear model. The Shapiro-Wilk tests for normality of residuals and Q-Q plots (Fig. 5(b)) also show an improvement of normality.

5.3 Discussion

Our results provide a model that quantitatively approximates the JND for bar, pie and bubble charts, taking into consideration of two variables,

[1]Models in this paper were fitted using nlme package in R.

	Coefficients									
Linear model	β_0	p	β_1	p	β_2	p	Normality of residuals	Skewness	Excess kurtosis	Homosce-dasticity
Bar	0.6211	p=0.2189	0.0156	p<.0001	--	--	p<.0001	6.6566	66.8330	p=0.0068
Pie	1.0621	p=0.2175	--	--	0.0860	p<.0001	p<.0001	3.7003	30.3479	p=0.0005
Bubble	0.2820	p= 0.0609	0.0041	p<.0001	0.0290	p<.0001	p<.0001	1.4643	7.6624	p=0.0024
Log-linear Model										
Bar	-0.4653	p=0.0006	0.0065	p<.0001	--	--	p=0.0012	0.0124	0.2919	p=0.3942
Pie	0.2405	p=0.0247	--	--	0.0187	p<.0001	p=0.0019	-0.1634	-0.3535	p=0.1030
Bubble	-0.7697	p<.0001	0.0031	p<.0001	0.0235	p<.0001	p<.0001	-0.4606	0.9276	p=0.1405

Fig. 6. The coefficients and regression diagnostics of the linear models and log-linear models

intensity and distance. Inspired by Weber's law, we started the model setting from a linear mixed regression model, with two variables and their interaction effect considered. Later the linear mixed regression is found to fit the experiment data better with a log transformation. The model shows that JND grows as the exponential function of affected variables in charts. In this work, we performed a log-linear model by which the error structure was changed. However there are alternative models, such as generalized linear mixed model which keep the error structure unchanged. Future work is encouraged to investigate the benefits of different models.

As our results indicate, JND perception for bar charts was not proportional to the bar's height. At a first glance, this might be surprising, since previous work has found that the perception of line lengths can be modelled by Weber's law [29]. However, this can be explained by the horizontal alignment of the bars in a bar chart, which echoes the conjectures raised by Cleveland et al. [6] that the primary elementary task in bar chart is judging positions along a common scale. Also, our finding that distance between bars significantly correlating with JND is inline with the results found by Talbot et al. [37], that comparisons between adjacent bars are more accurate than between widely separated bars. Thus, when comparing bars, viewers do not compare the length of the entire bars but rather focus on relative differences of their top positions. We believe height would matter much more in other forms of bar charts such as stacked bars, where the bars are not necessarily aligned. In our experiment, the heights of the bars located between the two bars to be compared were generated randomly for each trial. Future work should examine and possibly add a neighboring effect [43] of close-by bars to the JND model.

Results of the pie chart show that the intensity (i.e., the angle of the pie segment) affected the perception of JND following Weber's law. However, the angular distance between the segments does not have a significant effect. When reading a pie chart, several perceptual cues may be used [18]. Angle, area and arc length may all have some effect on estimation perception of pie segments [31]. However, several perceptual studies have shown that angle was perceived as somewhat less important, and that area emerged as the most dominant perceptual factor [23,31]. It seems that also for a comparison task of two segments of a single pie chart, area perception is dominant. In this work, we build up the JND model according to angle, which could be extended to area or arc length of pie segments in the future.

6 Chart Enhancement with the JND Model

The JND model quantifies the perception of similar objects in charts. In this section, we demonstrate how this quantification can be used to enhance charts by alleviating perceptual ambiguities.

6.1 Visual Quality Measurement

The JND measurement can serve as a metric to evaluate the representation quality of a chart. A chart is considered with high visual interpretability when all data elements can be discriminated from each other (i.e., their difference is above JND). On the other hand, if there are pairs of elements with differences below the JND, the chart has a lower visual interpretability.

For example in Figure 1, four different charts visualize the same dataset, but are detected with different pairs of elements below the JND threshold. In the left-most bar chart of Figure 1, no below-JND pair is detected, i.e., all of the bars can be perceived differently from each other. When bars are placed in a different order, the second bar chart has a lower visual interpretability since a pair of bars become perceptually indistinguishable. In the pie chart there are two pairs of below-JND elements, and in the bubble chart, the visual interpretability gets worse where three pairs of fans are non-distinguishable.

Thus, the JND model can be useful when choosing or optimizing a visual representation by first examining the JNDs between elements. As the example shows in Figure 1, we can use JNDs to detect which visual encoding works best for a given data set, or take it as an optimization factor when trying different visual configurations (e.g., changing the layout of bubble chart, reordering fans in pie charts, etc.). Of course, there are cases where the data has similar or very close values which will result in below-JND items in any representation. However, to come back to our example: although not as accurate in depicting values and differences as the bar chart, a designer might wish to use a bubble chart for aesthetic or other reasons. Using JND, she can evaluate whether the elements in the chart are distinguishable given her data.

6.2 Enhancement of Below-JND Objects

The JND measurement model is able to predict when and where there might be difficulty in comparing or assessing differences between items in a chart. Using this prediction, a secondary visual enhancement can be selectively added to groups of elements who are below the JND (i.e., below-JND objects). Such a visual enhancement can help chart readers to be aware of small differences between elements, and enable ordering of items to help differentiate visual elements representing similar quantities.

Figure 7 demonstrates the idea of detecting and visually enhancing elements that are below JND. The three charts show data at three time intervals. As more data blocks appear in-between bar 'A' and bar 'B', the distance between 'A' and 'B' changes, such that the 'A' and 'B' are detected to be below the JND threshold in 2018 (the right most chart). Visual enhancements, in this example text annotations, can be adaptively added if discrimination is important.

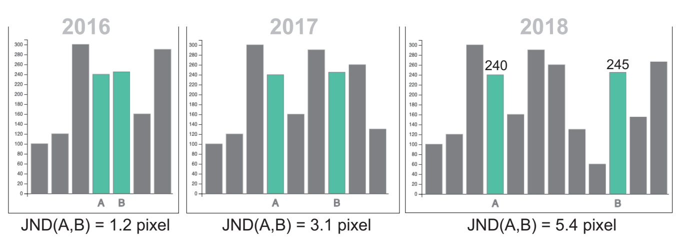

Fig. 7. Adaptive visual enhancement of bars with below-JND values for better discrimination.

In Figure 7, we show a simple case where textual annotations are adaptively added to a pair of below-JND objects. A more general case would deal with the discrimination of multiple below-JND objects, i.e., groups of below-JND objects. To detect below-JND groups, a practical approach is to model the chart as a graph where each node is a data object of the chart, and edges connect pairs of objects which are below the JND threshold. Then, connected components can be detected in the graph, each of which is a group of below-JND objects.

One conventional way to enhance a chart to enable better detection of JND objects is to add common grid-lines for finer comparison of close values [22]. Another plausible way is to overlay secondary visual cues to facilitate the ordering among below-JND objects. According to the seminal work by Mackinlay [27], who ranked the effectiveness of graphical presentations by perceptual tasks, ordinal excel visual cues, e.g., density, texture, can be considered. Figure 8 shows a bubble chart example. The top-right side of Figure 8 shows a simple enhancement design with small marks added to the below-JND bubbles. The number of marks indicates the quantity order in the group of similar bubbles. In the bottom of Figure 8, we show some possible designs of a five-scale ordinal comparison. Compared to text annotations and grid-lines, these visual markers are integrated into the visual objects, therefore they require less displaying space and are not as cluttered. This is especially so for bubble charts where bubbles are placed in a compact layout. However, these markers may create an overload on the perception of the main visual channel. Text annotations might be easier to discern as it is easier to separate texts and visual objects. In this work, we focus on the JND modeling and leave enhancements methods of below-JND objects as an open question. Future works can examine various enhancing designs and examine which enchantment type is better under which circumstance.

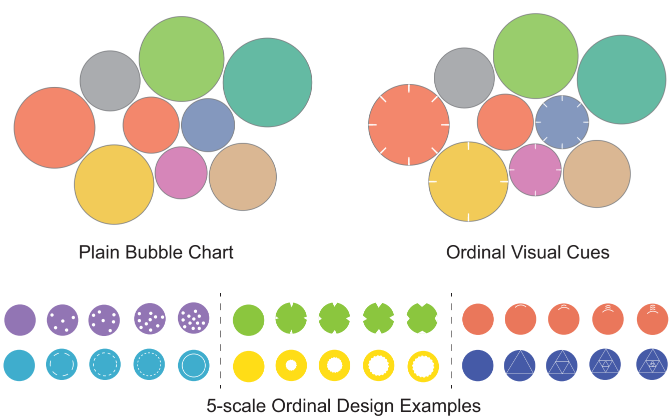

Fig. 8. Bubble chart with below-JND enhancement: (top-left) original bubble chart; (top-middle) common grid-line style enhancement; (top-right) ordinal-icon style enhancement; (bottom) some other ordinal designs.

6.3 An Experiment to Examine Below-JND Enhancement

For further validation, we conducted a second experiment to examine the performance of an object comparison task for objects below and above our fitted JND. We compared error value and time cost in a discrimination task of different JND-values to a baseline that includes annotations, to see how effective are our predictions and when visual enhancement of below-JND objects are needed.

We sampled nine conditions for each chart: 3 different distances and 3 different object intensities ([90, 170, 230] height x [100, 220, 330] distance in bar chart, [23, 35, 47] radius x [80, 140, 200] distance in bubble chart, [34, 76, 118] angle x [25, 55, 85] angular distance in pie chart). For each condition, six comparisons were generated, three of which included objects whose difference was below the fitted JND value (below-JND objects whose difference $Diff = JND_{2.5\%} * (1+\delta)$, where $\delta \in \{-0.15, -0.3, -0.45\}$) and three comparisons with objects whose differences were above the fitted JND (above-JND objects with difference $Diff = JND_{97.5\%} * (1+\delta)$, where $\delta \in \{0.15, 0.3, 0.45\}$).

For each comparison we prepared two versions: *plain* and *annotated*. The plain chart was a regular chart that did not include any aids, while the annotated chart included numerical values on top of the objects that had to be compared. Figure 9 shows the annotated chart. The plain chart was exactly the same, but without the numerical annotations.

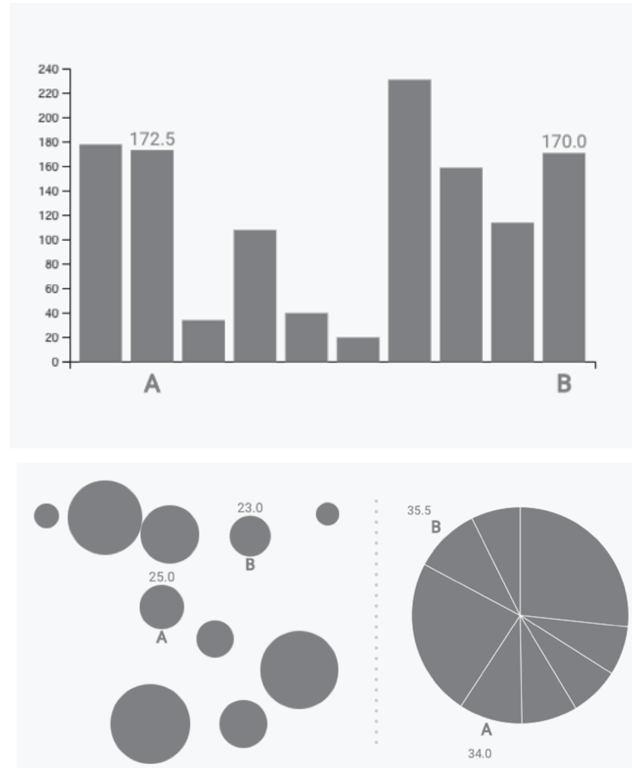

Fig. 9. Annotated charts used in the below and above JND experiment. (Note that charts were scaled here, therefore the annotations look smaller than what they actually look like in the study.)

Twenty four (24) participants were recruited. Each participant saw a total of 324 trials (3 charts * 9 conditions * 6 comparisons * 2 versions), with trial order being randomized. 16 out of the 24 self-reported with little expertise in visualization. In each trial, participants were asked to pick the taller/larger object from the two marked ones. Overall, the experiment set-up and procedure was the same as the experiment introduced in Section 4.3. Time cost and accuracy were recorded for each test. Figure 10 shows the experimental results with the left column showing the accuracy and the right column showing the time cost.

Looking at error rate, for the three types of charts, the annotated conditions have unsurprisingly a very low error rate. In the plain chart condition, for all three chart types the error rate of below-JND objects is much higher than that of the above-JND objects. When comparing the plain and the annotated condition, the error rate of the above-JND of the plain condition is pretty close to that of the annotated condition, but it soars for the below-JND values. This clear rise in error rate can be noticed in all charts between the -0.15 (below JND) and the +0.15 (above JND) delta values. These observations suggest the necessity of chart enhancements when comparing below-JND objects.

For the time cost of correct answers, we see that for bar charts the participants answered more quickly in the annotated condition compared to the plain condition. This was not the case for bubble or pie charts. It is interesting to note that for the plain condition, in both the bubble and the pie charts, there is a slight drop off in the above-JND time cost compared to the below-JND values, sometimes falling even below the time cost of the annotated chart. With more annotations (e.g., if all values in the chart will be annotated), it is likely that the time cost of the annotated chart would increase. Thus, adaptive annotations of only below-JND values as well as design enhancements that utilize less visual clutter are important. Further work should be done to examine time-costs when there are more annotations.

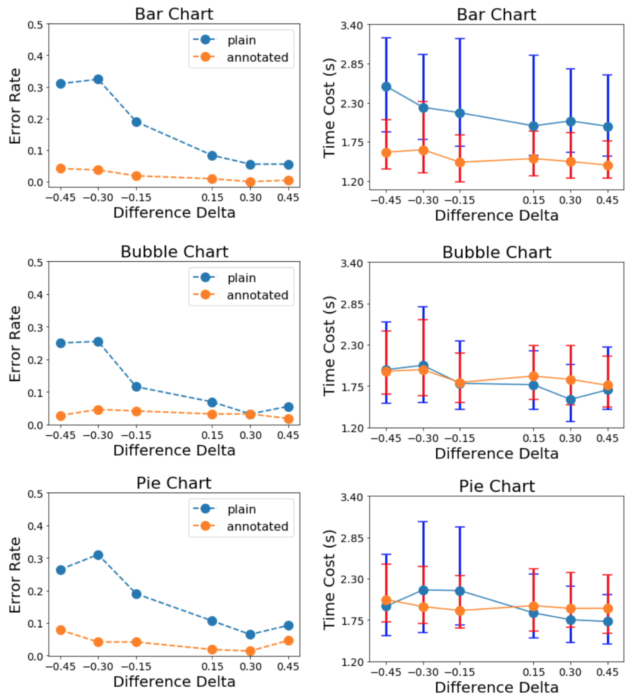

Fig. 10. Error rates and time costs for plain and annotated charts: (left) Error rates, i.e., the proportion of wrong answers in relation to the differences; (right) average time cost per trial with correct answer.

7 Conclusion

Discriminating quantities of chart elements is a fundamental chart reading activity [18, 33]. This work continues the research line of perception modelling and takes one of the first steps in modeling JND in visualizations, to facilitate with the quantity discrimination in charts. Specifically, JND in charts is modelled to two visual variables: object intensity and separation distance. Using experimental data collected from 28 participants in an elementary discrimination task, JND is examined in three conventional charts in relation to object intensity and distance. In the bar chart, distance between bars on JND is found statically significant, while the height of bars is not. As the distance becomes larger, JND becomes larger. For the pie chart, JND depends on the angle of fan sections to be compared, while no significance is found for the fans' angular distance. For the bubble chart, both intensity and distance are found significantly related to JND. A log-linear mixed effects model is found and implemented that fits JND as an exponential function of the detected variables. We show the application of the model by suggesting to use chart enhancements, adding adaptive gridlines, annotations, or other enhancements when below-JND objects are detected.

This work situates the JND study in the field of psychometric in the context of visualization. Since chart perception is often affected by the entire chart environment, it may be possible to refine our model considering other factors such as chart size, scale, or bar width (in the bar chart). These examinations are left for future works.

In this work, we performed a laboratory experiment, which had a limited group of participants with a similar background. This might inject some bias into the JND models. One future work would thus be to extend this study to crowd-sourcing experiments (e.g., [16]) to involve participants of larger diversity. Also, we studied JNDs in a controlled environment, in which charts were displayed on a 600x600 canvas on the screen of a personal laptop The exact JND model might be dependent on the display size and quality. While we believe our main findings of the effects of intensity/distance on JNDs in charts are applicable to all displays, the exact model parameters might need to be set differently on different displays. Also, it would be interesting to study JND in other visual environments, such as hand-held devices, tiled or larger displays or virtual reality scenarios, and see whether the constructed models remain valid in these situations. Finally, there might be other contextual factors which could potentially influence the JND such as color (e.g., color perception of chart elements was found to be affected by other visual channels of these elements such as size and shape [32, 36]), neighboring element sizes (neighborhood effect) or other factors. We encourage following work to study these potential factors and refine the JND models.

Acknowledgments

We sincerely thank Prof. Daniel Cohen-Or for his exceptional enthusiasm and knowledge that inspired this work, Prof. Ye Fei from Capital University of Economics and Business for his help in JND modeling. We also would like to celebrate the first year of Min's baby Rixin with this work. This work is supported in parts by NSFC (61802265, U2001206), GD Talent Program (2019JC05X328), DEGP Key Project (2018KZDXM058), Shenzhen Science and Technology Program (RCJC20200714114435012), Shenzhen Support Program (20200812122821001), DFG Center of Excellence 2117 "Centre for the advanced Study of Collective Behaviour" (ID: 422037984), National Engineering Laboratory for Big Data System Computing Technology, and Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ).

References

[1] T. Augustin and T. Roscher. Empirical evaluation of the near-miss-to-weber's law: a visual discrimination experiment. *Psychology Science Quarterly*, 50(4):469, 2008.

[2] D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, Articles*, 67(1):1–48, 2015.

[3] E. G. Boring. Sensation and perception in the history of experimental psychology. 1942.

[4] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE transactions on visualization and computer graphics*, 19(12):2376–2385, 2013.

[5] D. H. Chung, D. Archambault, R. Borgo, D. J. Edwards, R. S. Laramee, and M. Chen. How ordered is it? on the perceptual orderability of visual channels. In *Computer Graphics Forum*, vol. 35, pp. 131–140. Wiley Online Library, 2016.

[6] 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.

[7] W. S. Cleveland and R. McGill. Graphical perception and graphical methods for analyzing scientific data. *Science*, 229(4716):828–833, 1985.

[8] W. H. Ehrenstein and A. Ehrenstein. Psychophysical methods. In *Modern techniques in neuroscience research*, pp. 1211–1241. Springer, 1999.

[9] G. T. Fechner. *Elemente der psychophysik*. Breitkopf u. Härtel, 1860.

[10] J. Fox. *Applied regression analysis and generalized linear models*. Sage Publications, 2015.

[11] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, 2011.

[12] M. Gleicher, M. Correll, C. Nothelfer, and S. Franconeri. Perception of average value in multiclass scatterplots. *IEEE transactions on visualization and computer graphics*, 19(12):2316–2325, 2013.

[13] J. Guilford. A generalized psychophysical law. *Psychological Review*, 39(1):73, 1932.

[14] L. Harrison, F. Yang, S. Franconeri, and R. Chang. Ranking visualizations of correlation using weber's law. *IEEE transactions on visualization and computer graphics*, 20(12):1943–1952, 2014.

[15] S. Hecht. The visual discrimination of intensity and the weber-fechner law. *The Journal of general physiology*, 7(2):235–267, 1924.

[16] 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*, pp. 203–212. ACM, 2010.

[17] M. Heidelberger. *Nature from within: Gustav Theodor Fechner and his psychophysical worldview*. University of Pittsburgh Pre, 2004.

[18] J. G. Hollands and I. Spence. The discrimination of graphical elements. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 15(4):413–431, 2001.

[19] B. M. Hughes. Just noticeable differences in 2d and 3d bar charts: A psychophysical analysis of chart readability. *Perceptual and Motor Skills*, 92(2):495–503, 2001.

[20] M. Kay and J. Heer. Beyond weber's law: A second look at ranking visualizations of correlation. *IEEE transactions on visualization and computer graphics*, 22(1):469–478, 2015.

[21] Y.-H. Kim, J. H. Jeon, E. K. Choe, B. Lee, K. Kim, and J. Seo. Timeaware: Leveraging framing effects to enhance personal productivity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 272–283. ACM, 2016.

[22] 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.

[23] R. Kosara. Evidence for area as the primary visual cue in pie charts. In *2019 IEEE Visualization Conference (VIS)*, pp. 101–105. IEEE, 2019.

[24] T. Kuroda and E. Hasuo. The very first step to start psychophysical experiments. *Acoustical Science and Technology*, 35(1):1–9, 2014.

[25] E. Limpert, W. A. Stahel, and M. Abbt. Log-normal distributions across the sciences: Keys and clues. *BioScience*, 51(5), 2001.

[26] D. M. MacKay. Psychophysics of perceived intensity: A theoretical basis for fechner's and stevens' laws. *Science*, 139(3560):1213–1216, 1963.

[27] J. Mackinlay. Automating the design of graphical presentations of relational information. *Acm Transactions On Graphics (Tog)*, 5(2):110–141, 1986.

[28] M. W. Matlin and H. J. Foley. *Sensation and perception*. Allyn & Bacon, 1992.

[29] H. Ono. Difference threshold for stimulus length under simultaneous and nonsimultaneous viewing conditions. *Perception & Psychophysics*, 2(5):201–207, 1967.

[30] R. A. Rensink and G. Baldridge. The perception of correlation in scatterplots. In *Computer Graphics Forum*, vol. 29, pp. 1203–1210, 2010.

[31] D. Skau and R. Kosara. Arcs, angles, or areas: Individual data encodings in pie and donut charts. In *Computer Graphics Forum*, vol. 35, pp. 121–130. Wiley Online Library, 2016.

[32] S. Smart and D. A. Szafir. Measuring the separability of shape, size, and color in scatterplots. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2019.

[33] I. Spence and S. Lewandowsky. Displaying proportions and percentages. *Applied Cognitive Psychology*, 5(1):61–77, 1991.

[34] A. Srinivasan, M. Brehmer, B. Lee, and S. M. Drucker. What's the difference?: Evaluating variations of multi-series bar charts for visual comparison tasks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 304. ACM, 2018.

[35] S. S. Stevens. On the psychophysical law. *Psychological review*, 64(3):153, 1957.

[36] D. A. Szafir. Modeling color difference for visualization design. *IEEE transactions on visualization and computer graphics*, 24(1):392–401, 2017.

[37] J. Talbot, V. Setlur, and A. Anand. Four experiments on the perception of bar charts. *IEEE transactions on visualization and computer graphics*, 20(12):2152–2160, 2014.

[38] C. Tominski, C. Forsell, and J. Johansson. Interaction support for visual comparison inspired by natural behavior. *IEEE Transactions on visualization and computer graphics*, 18(12):2719–2728, 2012.

[39] E. Vandenbussche, R. Vogels, and G. A. Orban. Human orientation discrimination: changes with eccentricity in normal and amblyopic vision. *Investigative Ophthalmology & Visual Science*, 27(2):237–245, 1986.

[40] B. Winter. Linear models and linear mixed effects models in R with linguistic applications. *arXiv preprint*, arXiv:1308.5499, 2013.

[41] F. Yang, L. T. Harrison, R. A. Rensink, S. L. Franconeri, and R. Chang. Correlation judgment and visualization features: A comparative study. *IEEE Transactions on Visualization and Computer Graphics*, 25(3):1474–1488, 2019.

[42] J. Zacks, E. Levy, B. Tversky, and D. J. Schiano. Reading bar graphs: Effects of extraneous depth cues and graphical context. *Journal of experimental psychology: Applied*, 4(2):119, 1998.

[43] M. Zhao, H. Qu, and M. Sedlmair. Neighborhood perception in bar charts. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 232. ACM, 2019.