# Evaluating Just Noticeable Differences in Layered Opacity Visualizations

Caterina Ponti*
University of San Francisco

Shano Liang†
Worcester Polytechnic Institute

Lane Harrison‡
Worcester Polytechnic Institute

Alark Joshi§
University of San Francisco

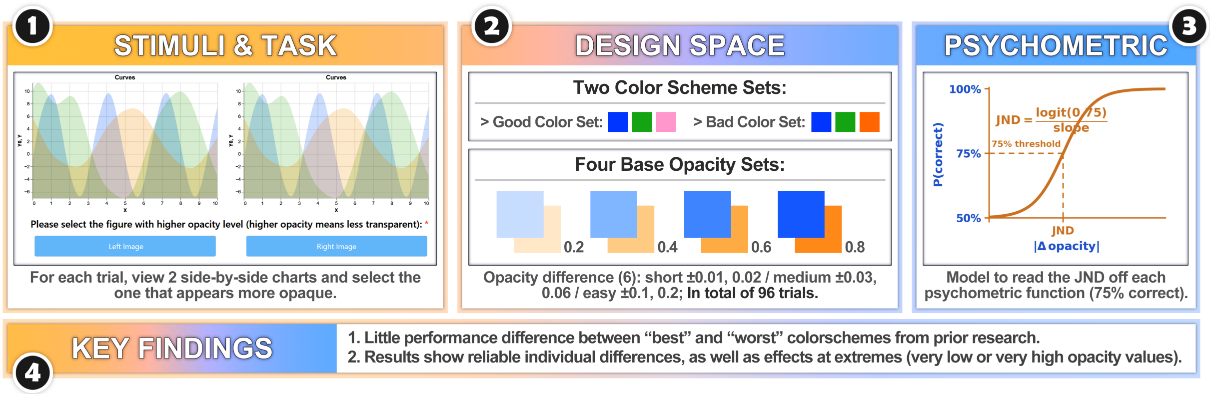


Figure 1: Overview of our study. We measure just-noticeable differences (JNDs) in layered opacity charts. (1) On each trial, participants view two side-by-side charts and select the one that appears more opaque; (2) Stimuli span a controlled design of four base opacity levels (0.2–0.8), six opacity differences, and the two color schemes drawn from Meng et al. [6]: a recommended ("*good*") scheme and a non-recommended ("*bad*") scheme, for 96 trials per participant; (3) We fit a Bayesian logistic mixed-effects model to the responses and read the JND off each psychometric function as the opacity difference giving 75% correct performance; (4) Results suggest little difference between colorschemes, but larger effects for individual differences and variance when comparisons approach extreme values (0.0 or 1.0).

## Abstract

Opacity is a widely used channel in data visualization, but it remains less well understood compared to channels such as color, length, size, etc. Recent work from Meng et al. [6], investigated the impact of opacity across competing color schemes, finding that certain color schemes were associated with better participant accuracy. We examine these effects further in a controlled two-alternative forced choice setup, to determine whether opacity differences are truly equal across possible opacity comparison ranges. In a within-subjects study with 96 trials, including two competing color schemes (*best* and *worst* from Meng et al. [6]) and 48 opacity pairs, we find little differences between color schemes, but larger individual differences in accuracy. Further, results show stable performance in middle opacity ranges, with more errors occurring when comparing extreme values. We discuss potential implications for design guidelines and further study, and make our study materials, analysis scripts, and data available at https://osf.io/zv9dx/overview?view_only=38d03cea1b3d42788593e3e6b1016cfd.

**Index Terms:** layered opacity, just noticeable differences, visual perception

## 1 Introduction

Understanding how people perceive visual differences is essential for designing effective visualizations. In this study, we investigate how accurately people can detect *differences in opacity when interpreting layered opacity charts*. Specifically, we use a two-alternative forced choice task and estimate JNDs across opacity levels and color schemes to estimate the change in opacity where observers can reliably perceive differences.

Building on existing literature on perception comparison in visualizations [3, 4, 8], we focused on layered opacity chart comparisons and examined participants' ability to detect opacity differences using a fixed set of comparison stimuli. We tested four baseline opacity levels (0.2, 0.4, 0.6, and 0.8) and compared each to stimuli with opacity differences of $\pm 0.01$ and $\pm 0.02$ (short), $\pm 0.03$ and $\pm 0.06$ (medium), and $\pm 0.1$ and $\pm 0.2$ (long). Trials were presented in randomized order, and every participant completed trials in both the "good" and "bad" color conditions, which we adapted from Meng et al. [6].

We analyzed how visual opacity discrimination accuracy varies as a function of the magnitude of the opacity difference between layered opacity plots, the baseline opacity level of stimuli, and the experimental condition (good vs. bad). Collected trial data were modeled using logistic regression with the base opacity level and color-scheme serving as the main factors, allowing for variation in parameters between participants to test for potential individual differences.

Results suggest that color-scheme had relatively little impact on performance in terms of estimated JNDs. However, results did show more variation between participants, with some detecting even small differences reliably. Comparisons at the extremes (e.g. near 0.0 or 1.0) also resulted in variance in performance. We discuss potential implications of these findings in terms of design and future studies.

## 2 Related Work

Many prior studies in visualization evaluate how well people perceive differences or estimate values based on carefully controlled visual stimuli, often reflecting how data is encoded in practice. Pre-

*e-mail: cponti@dons.usfca.edu
†e-mail: sliang1@wpi.edu
‡e-mail: ltharrison@wpi.edu
§e-mail: apjoshi@usfca.edu

vious studies have explored how changes in visual attributes (correlation [8, 9, 12], color [11, 12] and opacity [10, 6]) affect people's performance.

## 2.1 Scatterplots, Correlation, and JNDs

Rensink et al. [8] evaluated the visual perception of correlation in scatterplots by measuring the observers' sensitivity to differences in correlation. Precision was measured using the Just Noticeable Difference (JND) (in the context of correlation perception) that is defined as the minimum difference between two side-by-side correlations that can be discriminated 75 percent of the time. Rensink et al. showed that the perception of positive correlation in scatterplots can be modeled using Weber's Law [1]. Subsequently, Harrison et al. [3] used Weber's Law to evaluate the effectiveness of commonly used visualizations, demonstrating that nine commonly used visualizations follow Weber's law. The Just Noticeable Difference (JND) is a key component in this evaluation, as it defines the smallest detectable change in correlation perception, helping to establish perceptual thresholds. Kay et al. [4] reviewed the generality of Weber's Law and found that a log-linear model provides a better description of the relationship between JNDs and objective correlation strength. Their Bayesian linear mixed-effects model accounts for individual variability. It suggests that correlation perception may follow a more complex, non-linear pattern than previously assumed. Elliott and Rensink [2] further showed that overlapping visual populations can interfere with the perception of correlation. Their findings suggest that perceptual thresholds may change when multiple visual layers compete for attention.

Beyond correlation estimation, Micallef et al. [7] demonstrated that scatterplot design parameters such as point size, transparency, and sampling influence perceptual accuracy, highlighting the importance of accounting for visual encoding choices when modeling perceptual thresholds.

## 2.2 Color and Opacity Studies

Lu et al. [5] addressed challenges in transparency perception. When color overlaps with transparency, they blend, creating ambiguous color mappings and introducing false colors. To solve this issue, Lu et al. introduced a color-name-aware (CNA) optimization framework that produced consistent color choices and transparency levels while maintaining a clear visual distinction between all parts of the graph. They also showed that their method works for histograms and other visualizations like parallel coordinate plots and Venn diagrams. Meng et al. [6] investigated how color hue and opacity affect depth perception in semi-transparent visualizations. Their findings underscore that both color and opacity interact to shape perceptual accuracy in layered visual representations, further motivating the need to model perceptual thresholds such as JNDs in visualization design, especially when looking at the joint effect of opacity and color. Specifically, they identified 4 key predictors of depth order perception: hue front, hue back, opacity back, and opacity front. They also found that participants made *fewer errors when certain colors appeared in specific layer positions*: blue, green, and orange had fewer errors when placed in the front layer, and orange and pink did so when placed in the back layer. However, their study did not examine how participants perform when comparing side-by-side layered opacity charts that differ in opacity levels. Building on these findings, our experiment draws primarily from the methodologies and insights of Kay et al. [4] and Meng et al. [6] to investigate perceptual threshold in layered opacity visualizations.

## 3 Methods

The experiment investigated opacity discrimination under two color conditions derived from color recommendations reported in previous work [6]. The recommended color combination consisted of blue and green elements in the front layer and orange and pink

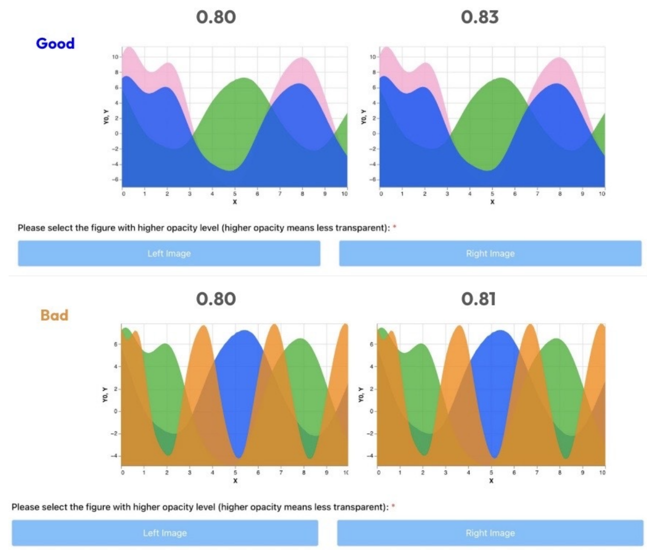


Figure 2: This figure shows two example stimuli with the interface seen by the participants. The top row showing the "good" palette showing the 0.8 stimuli with the 0.83 stimuli (JND distance of 0.3). The bottom row shows the "bad" palette with the 0.8 stimuli and the 0.81 stimuli. The stimuli in the bottom row are indistinguishable according to the results of our study. You can see a demo of the study at `https://opacityjnd.web.app`.

elements[1] in the back layer. Previous findings showed that blue, green, and orange ("good palette") produced lower error rates when displayed in the front layer, whereas orange and pink produced lower error rates when displayed in the back layer. The non-recommended color combination consisted of orange elements in the front layer and blue and green elements in the back layer ("bad palette"). This combination was associated with higher error rates in prior work and served as a comparison condition. The goal in including these was to determine whether error rates across "good" and "bad" palettes were associated with particular opacity levels and differences, as well as to examine individual performance. Stimuli were rendered using Vega with its default HTML5 Canvas renderer against a white background (#FFFFFF; RGB 255, 255, 255). Opacity was applied using the browser's default Canvas 2D source-over compositing in gamma-encoded sRGB.

Stimuli were presented at four base opacity levels (0.2, 0.4, 0.6, and 0.8) in both color conditions. On each trial, two images differing in opacity were displayed side by side with the participant asked to choose the figure with the higher opacity, as shown in Figure 2. Figure 2 shows the "good" palette combination in the top row with two stimuli (with opacity 0.80 and 0.83 (+0.03 = JND distance from the original stimuli)) and two stimuli (with opacity 0.80 and 0.81) with the "bad" palette in the bottom row. The opacity difference varied across trials. For each opacity, the comparison stimuli distances ranged from short ($\pm 0.01$) to long ($\pm 0.2$).

Participants completed both color conditions in a mixed design. Each participant performed 96 trials corresponding to all combinations of four base opacity levels and two color conditions (12 stimuli distances * 2 color conditions * 4 opacities). On every trial, participants viewed two images and indicated which image appeared to have *greater opacity* by selecting either the left or the right image. Response accuracy was recorded for each trial.

[1]The *HSL* (Hue, Saturation, and Lightness) values of these colors are $(219,100,50)$ for blue, $(119,58,44)$ for green, $(332,100,83)$ for pink, and $(31,100,51)$ for orange.

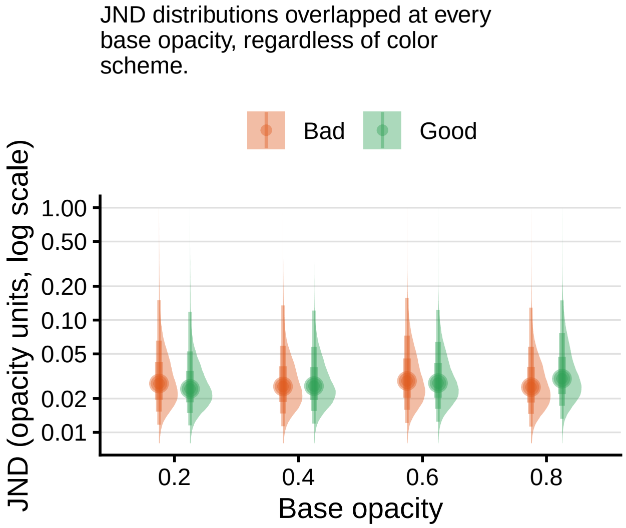


Figure 3: Full posterior JND distributions by color condition and base opacity (log scale). Distributions largely overlap between the recommended (good) and non-recommended (bad) conditions at every base opacity level.

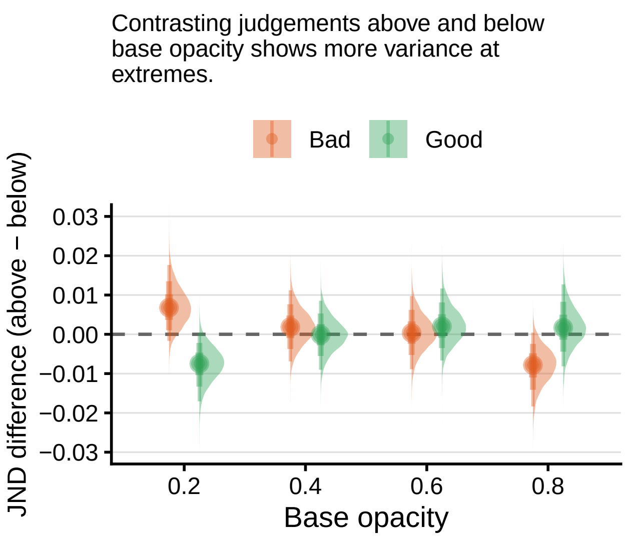


Figure 4: Within-condition direction contrast (above − below JND) by color condition and base opacity. Positive values indicate the "below" direction was easier to detect (lower JND); negative values indicate "above" was easier. Bad and good conditions sit on opposite sides of zero at base opacity 0.2 and flip at base opacity 0.8.

# 4 Results

## 4.1 Participants

A total of 30 participants were recruited through Prolific to complete the online opacity experiment. Initial analyses revealed potential confusion among participants who flipped responses. As these also fell more than two standard deviations below mean performance, they were excluded. Mean accuracy across participants was 75.3% (SD = 20.9%), resulting in an exclusion threshold of 33.5%. Three participants were excluded, with a final sample of 27 participants. Participant reported demographics including 18 men (66.7%), 8 women (29.6%), and 1 non-binary (3.7%). Participants ranged in age from 18 to 64 years. With respect to education, 13 participants (48.1%) reported holding a bachelor's degree or equivalent, 8 (29.6%) reported a high school diploma or equivalent as their highest level of education, and 6 (22.2%) reported holding a master's degree or equivalent.

## 4.2 Statistical Analysis

Accuracy was modeled using a Bayesian logistic mixed-effects model[2]. The probability of a correct response was predicted by a separate slope for each combination of color condition and base opacity (i.e., an interaction with no main-effect or intercept terms), so that predicted accuracy is exactly chance when the opacity difference is zero. By-participant random effects were included for the intercept, the opacity-difference slope, the color-condition effect, and their interaction, to account for individual differences in both overall sensitivity and in how sensitivity varied by color condition:

$$\begin{aligned} P(\text{correct}) \sim\ & |\Delta\text{opacity}| : \text{condition} : \text{base opacity} - 1 \\ & + (|\Delta\text{opacity}| \times \text{condition} \mid \text{participant}) \end{aligned}$$

Just Noticeable Difference (JND) was defined as the *opacity difference required to achieve 75% correct performance*. At the population level, where the fixed effects have no intercept, the JND is simply the opacity difference at which the fitted slope reaches the 75%-correct criterion:

$$\text{JND} = \frac{\text{logit}(0.75)}{\text{slope}}$$

Larger JND values indicate lower sensitivity, meaning a greater opacity difference is needed to achieve the same level of performance. Larger opacity differences were associated with higher probabilities of correct responses across all eight conditions × base-opacity combinations.

## 4.3 Just-Noticeable Difference Estimates

JND values were defined as the opacity difference corresponding to 75% predicted accuracy. Posterior median JND estimates and 95% credible intervals are shown, in full distributional form, in Figure 3. At base opacity 0.2, the bad condition yielded a somewhat larger median JND (0.0271, 95% CrI [0.0107, 0.166]) than the good condition (0.0243, 95% CrI [0.0107, 0.126]), though the credible intervals overlap substantially and the posterior probability that bad exceeded good was $P(\text{bad} > \text{good}) = 0.774$. This pattern was not consistent across the opacity range: contrasts at base opacities 0.4 and 0.6 were inconclusive, and at base opacity 0.8 the point estimate nominally reversed (good: 0.0300 vs. bad: 0.0251), though this contrast was also not clearly credible ($P(\text{bad} > \text{good}) = 0.096$). Overall, JND did not vary monotonically with base opacity, and pooling across the full opacity range showed little evidence of a consistent difference between color conditions.

[2] Data and analysis scripts available at: https://osf.io/zv9dx/overview?view_only=38d03cea1b3d42788593e3e6b1016cfd

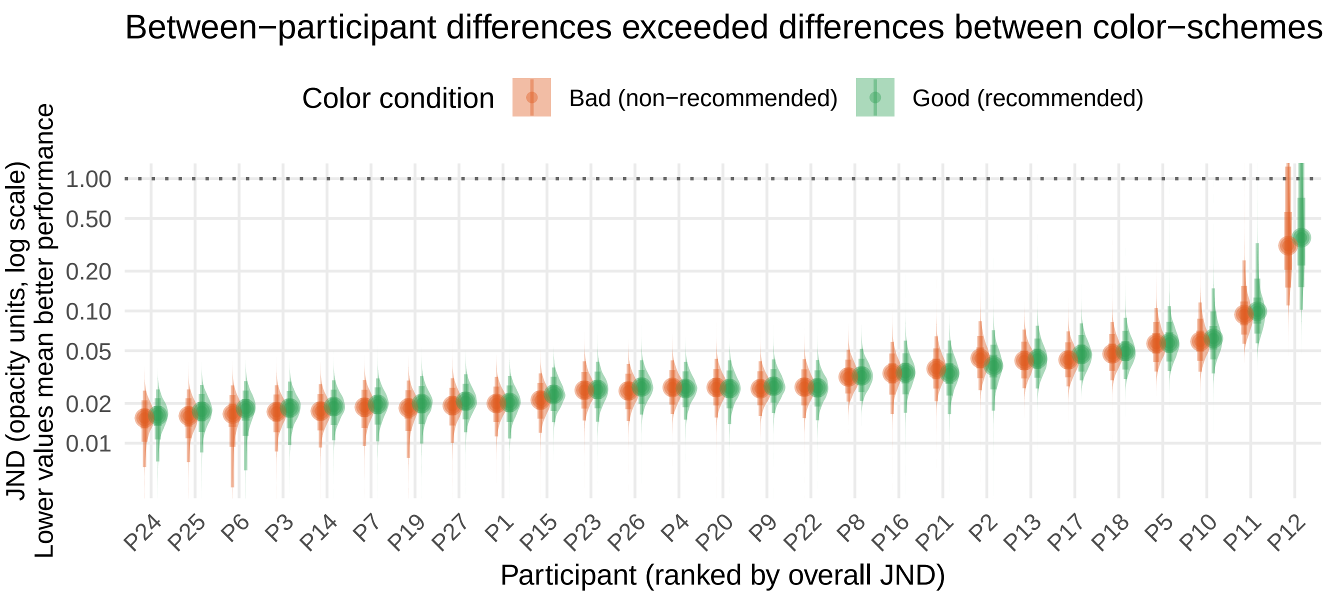


Figure 5: Between-participant differences in overall sensitivity. Each participant's JND (by color condition), averaged across base opacity, ranked from most to least sensitive. Between-participant spread is substantially larger than the difference between color conditions within any given participant.

### 4.4 Overall Color-Scheme Effects

Prior work on layered opacity visualizations has devoted substantial attention to how color-scheme choices interact with opacity to shape perceptual accuracy [6]. Given this emphasis, a natural first question is whether the recommended ("good") and non-recommended ("bad") color schemes produced systematically different JNDs once results are pooled across the full range of base opacities and comparison directions. As shown in Figure 3, the full posterior JND distributions for the two color schemes overlap substantially at every base opacity level, and none of the four base-opacity contrasts reached conventional credibility. We therefore find *no reliable overall effect of color scheme on opacity discrimination in this task.*

### 4.5 Individual Differences

In contrast to the modest overall effect of color scheme, participants varied more substantially from one another in their overall ability to discriminate opacity differences (Figure 5). Some participants performed extremely well, reliably discriminating opacity differences of roughly 1 to 1.5%, while others required a difference several times larger to perform at the same level, with a broad middle range of participants falling in between. One participant (P12) stands out with a JND far outside the range of the rest of the sample and a correspondingly wide credible interval, consistent with near-chance performance on a subset of trials rather than genuine low perceptual sensitivity.

For example, at a base opacity of 0.80 in the recommended ("good") color condition– the best-performing participant could reliably tell 0.80 apart from a comparison opacity of just 0.82. In comparison, the observed worst-performing participant (excluding P12, given the near-chance performance) would need the comparison pushed all the way to 0.93 before reliably seeing differences between stimuli.

### 4.6 Differentiating Above/Below Trials

To test whether the effect of color condition depended on whether the comparison stimulus was more or less opaque than the base ("above" vs. "below"), trials were split by this direction, and the model was refit with this additional factor. Figure 4 shows the resulting within-condition contrast (above − below JND) for each color condition and base opacity.

This split revealed a pattern that the pooled results in Figure 3 could not show: at the extremes of the tested opacity range (0.2 and 0.8), the two color schemes diverged in opposite directions specifically for comparison stimuli that were more opaque than the base. At base opacity 0.2, the "bad" scheme yielded a higher JND (worse performance) than the "good" scheme in this direction; at base opacity 0.8, this pattern reversed, with the good scheme instead yielding the higher JND. Contrasts in the opposite direction (comparison less opaque than the base) and at the two intermediate base opacities did not show this pattern. Rather than a uniform or monotonic difference between schemes, any true effect of color scheme, therefore, appears to be confined to specific combinations of comparison direction and base opacity, consistent with the absence of a simple overall effect reported above.

## 5 Discussion

Results suggest that opacity discrimination was consistent for individual participants, regardless of the "good" or "bad" color scheme. This invites further investigation on what actually drives error rates in visualization tasks involving opacity, such as the depth-order judgments studied by Meng et al. [6] or the scatterplot design parameters (point size, transparency, sampling) studied by Micallef et al. [7]. One possibility is that these tasks differ in their underlying mechanisms, judging depth order or estimating correlation, versus our task which required discrimination between opacity values. These tasks may recruit distinct perceptual, cognitive, or decision-making processes despite all involving judgements of opacity at some level. Future studies might explore the relationship between these tasks in a single study so that within-participant comparisons are possible across tasks and visualization types.

While color scheme did not have a main effect, results did reveal two other effects to consider. Most notable is the effect of individual differences. Some participants were remarkably adept at the task, correctly discriminating opacity when it differed by about

1-1.5%. However, some participants needed a 5% or more difference to reliably be able to distinguish between opacity differences. While further investigation is warranted (e.g., see limitations below on visualization types and tasks), this may point towards practical guidelines of designing opacity difference ranges in 5% or greater "steps", to help ensure participants can reliably distinguish differences.

Results also showed modest differences at the extremes of the ranges tested, with error rates of the "bad" scheme improving at higher ranges and error rates of the "good" scheme improving in lower ranges. Curiously, the middle opacity values seem unaffected. We do not believe this points towards a potential future recommendation as much as a need for further studies. We recommend testing more color schemes and increased base opacity ranges (e.g., getting close to the extremes) to determine whether there are any effects of color scheme and performance at extreme ranges.

Also worth noting are limitations. Primarily, we tested a single visualization type, layered area charts, due to their common use of opacity. However, many other visualizations often use the opacity channel to help deal with various forms of overlap and overplotting. For the present work, our goal was to evaluate the overall methodology and test for effects across opacity values and individuals. Another limitation is the task itself. Discrimination is not the same as estimating values and should not be taken as having a direct connection to higher-level visualization task performance. At the same time, discrimination has been used to investigate visualization stimuli in controlled settings for some time now, often revealing useful insights into the relationship between perceptual performance and data visualizations.

The present study also lays groundwork for several potential future investigations. Future work might expand on the modeling approach, for example by calculating differences not based on opacity differences but based on perceptual color differences for each stimulus pair. Future studies should also include more in-depth training trials, which may help with the formation of their strategies. Participants could even be asked to describe their comparison strategies, which might reveal cases where participants appeared to reverse judgments, as well as important insights into the underlying reasons behind the performance differences between participants.

## References


[1] G. Ekman. Weber's law and related functions. *The Journal of Psychology*, 47(2):343–352, 1959. 2

[2] M. A. Elliott and R. A. Rensink. Interference in the perception of two-population scatterplots. *Journal of vision*, 15 12:893, 2015. 2

[3] 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. 1, 2

[4] 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. 1, 2

[5] K. Lu, L. Zhu, Y. Wang, Q. Zeng, W. Song, and K. Reda. Colorname aware optimization to enhance the perception of transparent overlapped charts. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 2

[6] Z. Meng, Y. Yang, Q. Zeng, K. Lu, L. Lu, C. Tu, F. Yang, and Y. Wang. Seeing through the overlap: The impact of color and opacity on depth order perception in visualization. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2025. 1, 2, 4

[7] L. Micallef, G. Palmas, A. Oulasvirta, and T. Weinkauf. Towards perceptual optimization of the visual design of scatterplots. *IEEE Transactions on Visualization and Computer Graphics*, 23:1588–1599, 2017. 2, 4

[8] R. A. Rensink and G. Baldridge. The perception of correlation in scatterplots. In *Computer graphics forum*, vol. 29, pp. 1203–1210. Wiley Online Library, 2010. 1, 2

[9] G. Strain, A. J. Stewart, P. Warren, and C. Jay. The effects of contrast on correlation perception in scatterplots. *Int. J. Hum. Comput. Stud.*, 176:103040, 2023. 2

[10] G. Strain, A. J. Stewart, P. A. Warren, and C. Jay. Effects of point size and opacity adjustments in scatterplots. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 2024. 2

[11] C. Tseng, G. J. Quadri, Z. Wang, and D. Szafir. Measuring categorical perception in color-coded scatterplots. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023. 2

[12] Z. Zeng and L. Battle. A review and collation of graphical perception knowledge for visualization recommendation. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2021. 2