

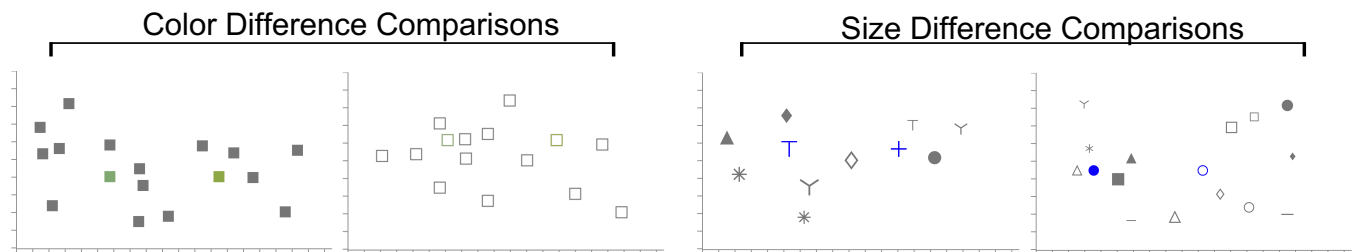
# Measuring the Separability of Shape, Size, and Color in Scatterplots

Stephen Smart

University of Colorado Boulder  
Boulder, Colorado  
stephen.smart@colorado.edu

Danielle Albers Szafr

University of Colorado Boulder  
Boulder, Colorado  
danielle.szafr@colorado.edu



**Figure 1:** We examine how mark shape, size and color influence data interpretation in multiclass scatterplots. We replicate prior findings showing an interdependence between size and color and find evidence that shapes also bias size and color perceptions between different shape categories.

## ABSTRACT

Scatterplots commonly use multiple visual channels to encode multivariate datasets. Such visualizations often use size, shape, and color as these dimensions are considered *separable*—dimensions represented by one channel do not significantly interfere with viewers’ abilities to perceive data in another. However, recent work shows the size of marks significantly impacts color difference perceptions, leading to broader questions about the separability of these channels. In this paper, we present a series of crowdsourced experiments measuring how mark shape, size, and color influence data interpretation in multiclass scatterplots. Our results indicate that mark shape significantly influences color and size perception, and that separability among these channels functions asymmetrically: shape more strongly influences size and color perceptions in scatterplots than size and color influence shape. Models constructed from the resulting data can

help designers anticipate viewer perceptions to build more effective visualizations.

## CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in visualization**;

## KEYWORDS

Visualization, Graphical Perception, Visual Channels, Separability, Crowdsourcing

## ACM Reference Format:

Stephen Smart and Danielle Albers Szafr. 2019. Measuring the Separability of Shape, Size, and Color in Scatterplots. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019)*, May 4–9, 2019, Glasgow, Scotland Uk. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3290605.3300899>

## 1 INTRODUCTION

Visualizations leverage a variety of visual channels to encode data and often combine channels to encode multiple data attributes at once. However, data represented using one channel may interfere with interpreting data along another. For example, visualizing one attribute using chroma and a second using lightness may degrade analysts’ abilities to accurately interpret either attribute independently: shifting lightness makes it more difficult to estimate chroma and varying chroma may complicate lightness perceptions [30]. Two visual channels are considered to be *integral* if encoding a data attribute using one channel interferes with selectively attending to the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

CHI 2019, May 4–9, 2019, Glasgow, Scotland UK

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-5970-2/19/05...\$15.00

<https://doi.org/10.1145/3290605.3300899>

other. Conversely, two visual channels are considered to be *separable* if one channel can be attended to without interference from the other [48]. Multivariate visualizations often leverage separable channels to design visualizations that allow analysts to accurately disentangle data along different dimensions. Channels such as size and color have conventionally been considered largely separable [70]. However, recent research [19, 64, 65] challenges the separability of color and size, suggesting a need for a deeper understanding of the interplay between different visual channels to inform effective visualization.

In this paper, we quantify the amount of perceptual interference between shape, size, and color. All three channels commonly encode values in many visualizations including maps and scatterplots. We measure how the interplay of shape, size, and color encodings influence our ability to distinguish data values along each channel and the symmetry of these effects. Understanding and quantifying the perceptual interference between these channels is vital to crafting visualizations that support accurate interpretation across multiple data dimensions. Our goal is to inform visualization designers and improve visualization effectiveness in the real world. We aim to achieve this by constructing actionable probabilistic models based on empirical evidence that can help predict how these visual channels interact with one another to recognize and account for potential biases in data analysis.

Our experiments look at three major categories of shapes used in existing commercial tools: filled geometries, unfilled geometries, and open shapes inspired by quantitative texton sequences (QTonS) [71] modified based on designs employed in existing tools. Our experiments leverage empirically validated methods to construct actionable probabilistic models of value difference perceptions [64, 66]. We apply these methods to model both color and size differences as a function of mark shape.

Our findings suggest that mark shape significantly affects perceptions of both mark size and color. For example, people more readily discriminate colors mapped to filled shapes compared to unfilled shapes, and shapes such as  $\top$ ,  $\blacktriangle$ , and  $\blacksquare$  appear larger than equally sized  $+$ ,  $\triangle$ , and  $\times$  shapes. Our results also indicate that these relationships are asymmetric: shape has a pronounced impact on size and color perception, while size and color have a smaller impact on shape perception for our tested ranges. We implemented the models constructed from our results as a D3 [6] extension that scales these channels based on known parameters of a visualization.

**Contributions:** The primary contributions of this work are empirical measures and models that predict how perceived differences in color and size vary depending on mark shape. Our models replicate prior work showing the interdependence of size and color [65] and provide an in-depth quantitative

analysis of the interplay between shape, size, and color in multivariate scatterplots. Our results allow for a more principled understanding of the separability of size, shape, and color for visualization design, providing empirically-grounded metrics for selecting between different encodings. These findings challenge current guidelines on separable and integral dimensions which we hope will spark future research around these principles rooted in empirical evidence from visualizations.

## 2 BACKGROUND

Shapes frequently encode categorical data in scatterplots. However, little empirical understanding exists to inform visualization designers about the effects of shape on interpreting data encoded using additional channels, such as color and size. We survey current empirical studies of separability from visualization and vision science to inform our work.

### Graphical Perception

Graphical perception measures the effectiveness of different visual channels for encoding data. These studies have explored how different visual channels might support a broad variety of tasks [57]. For example, they rank how well various channels support value comparison [2, 18, 21, 27], estimating quantities from specific visualization types [33, 62], investigating biases in particular channels [7, 23], and exploring certain statistical quantities like correlation [31, 55].

This work focuses on three visual channels—shape, color, and size—that prior studies have explored in isolation. For example, Skau & Kosara [62] measured how shapes factors in pie and donut charts affect interpretation accuracy. Ware [71] demonstrated how shape could encode ordinal data through Quantitative Texton Sequences (QTonS). We focus on shape encodings in scatterplots, where prior studies have primarily evaluated shape encodings for category discrimination tasks [39, 40, 68]. For example, Burlinson et al. [12] found that closed shapes support faster comparison than open shapes and that combining open and closed shapes may introduce perceptual interference in selection tasks. These studies provide preliminary insight into how the visual system processes shape information in visualizations in order to effectively leverage shape to support multiple analysis tasks.

Like shape, color is commonly used in a wide variety of visualizations. However, color perception is sensitive to viewing conditions [8, 36, 44, 46, 56, 58, 63] and visualization context [14, 52, 64, 65]. In visualizations, color can encode both qualitative and quantitative patterns. Quantitative color encodings rely on mapping data differences to sufficiently visible perceptual differences. As a result, developers often use either designer-crafted tools such as ColorBrewer [9] to select robust encodings or interpolate between differences using color spaces like CIELAB [24]. Additional features of

a color encoding affect interpretation accuracy, such as semantic alignment with the data [41, 60] and colormap design [4, 5, 42, 47] (see Zhou & Hansen [72] and Bujack et. al. [11] for surveys). However, most studies of color encodings focus on color as an isolated cue. Color’s sensitivity to mark geometries suggests that such an isolation assumption may provide only limited insight into how to effectively leverage color in visualizations [52, 64, 65].

Size is conventionally considered a more robust channel for encoding data values. For example, Cleveland & McGill [20] demonstrated that size is more accurate than color for value comparison tasks. Heer et al. [33] studied the effect of chart size on comparing values in time series visualizations. In text visualizations, size may also introduce biases due to the mark shapes created by word structures [3, 23]. We further explore how size-induced biases might influence separability in scatterplots.

Many graphical perception studies have explicitly explored graphical perception in scatterplots. For example, several studies measured people’s abilities to estimate correlation [31, 54, 61] and to detect (and even be primed to) value clusters [26, 59, 69]. Several recent techniques leverage findings from graphical perception studies of scatterplots to automate scatterplot design for legibility [15, 43] and to predict perceptual attributes associated with scatterplot analysis such as similarity [49] or class separability [59]. However, these investigations focus on analyses over either an entire design or a single channel. Multiclass and other multivariate scatterplots require designers to understand how different channels might interfere with data analysis across different subsets of dimensions. For example, mean position estimation is robust to multiple channels, but task performance varies with the salience of the channel employed [29]. In this work, we extend recent efforts in graphical perception for scatterplots to explore perceptual interactions between task-relevant and task-irrelevant channels to quantify potential biases in multivariate scatterplots.

### Separability and Conjunctive Search

Vision science traditionally describes separability as whether or not a dimension can be selectively attended to without interference from another dimension [28, 48]. Such studies related to integral and separable channels generally center on conjunctive search [67]—the process of identifying an object defined by multiple features, such as searching for a red X among a collection of blue X’s and red O’s. These studies aim to understand how the visual system processes an object’s feature information. For example, Jenkins et al. [34] found that attentional selection in searching for targets of a specified shape and size operates in parallel, allowing people to search for shape and size simultaneously. Alternatively, color tends

to attract attention more strongly than shape when searching for color-shape conjunctions [37].




Research in psychology has shown an interdependence between visual channels such as hue and brightness [8, 13, 30]. Pomerantz & Sager [51] found an interdependence between spatial configurations and symbol identification; however, the effect was asymmetric—symbol shape affected perceptions of spatial arrangements significantly more than arrangement affected symbol recognition. Cheng & Pachella [17] challenged the definitions of integral and separable dimensions and instead argued that channels corresponding to psychological attributes (i.e., an attribute that can be selectively attended to) are separable and channels that do not are integral.

Understanding of integral and separable channels in visualization has largely centered around expert heuristics [45, 70], with channels defined using terms like “fully separable” and “no significant interference.” Few studies have empirically explored interactions between separable channels in data visualizations. Acevedo & Laidlaw [1] found that people more precisely detect brightness differences than differences in other visual elements such as size or spacing in multivariate icon-based visualizations. Other studies have shown that quantity estimation is affected by both size and color [19, 23] with size perceptions being biased by specific hues. Spatial frequency and other factors of data complexity may influence the effectiveness of different color schemes depending on the target analysis task [52]. Kim & Heer found that size may interfere with positional data encodings [35].

More recent work has attempted to quantify such interactions between channels. For example, Stone et al. [64] and Szafir [65] modeled how mark size influences color perception, finding that larger and elongated marks increase our abilities to discriminate encoding colors. Demiralp et al. [25] studied the perceptual interactions between shape, size, and color similar to our study. They developed distance matrices based on Likert ratings, triplet comparison, and spatial arrangement tasks to help optimize visual encodings of color, shape, and size for discriminability, quantifying categorical and ordinal differences along these channels. We instead leverage psychophysical methods to capture smaller shifts in numeric perception at the scale of just noticeable differences and also evaluate how specific facets of a mark’s shape may systematically influence separability.

### 3 CONDITIONS AND HYPOTHESES

Our understanding of the separability of color, shape, and size encodings stems primarily from ordering tasks where participants compare large differences along each dimension [25, 70]. However, many visualization tasks require analysts to analyze small variations, such as value differences in continuous data. Prior work [64, 65] quantified interactions between size and color using probabilistic models that predict

Filled	
Unfilled	
mQTons	







**Figure 2:** We generated 16 candidate shapes drawing on encodings from popular commercial tools. Our shapes fall into three categories: filled shapes, unfilled shapes, and open shapes modified from QTons [71].

viewer perception based on visualization parameters. In this work, we construct similar models quantifying interactions between size, shape, and color through four crowdsourced experiments: two primary experiments to study the effect of mark shape on color and size perception and two measuring the symmetry of these effects.

We contextualize our investigation using scatterplots, which often use shape to distinguish categories of data points and color and size to encode continuous attributes. We tailor our experimental designs to reflect this scenario using the methodology from Szafir [65]—a binary forced-choice comparison of two target marks analogous to comparing two data points in a visualization. We constructed a candidate set of 16 mark shapes sampled from common visualization tools, including D3, Tableau, Excel, and Matlab (Figure 2). These shapes form three categories: filled shapes, unfilled shapes, and open shapes inspired by quantitative texton sequences (QTons) [71] modified based on designs used in existing commercial tools. For simplicity, we refer to these open shapes as modified QTons or *mQTons*. Our experiments tested shape, size, and color difference perceptions over six different mark sizes (6, 12, 18, 25, 37, and 50 pixels). All shapes were sized by filling a square box in either height or width as appropriate.

We used these conditions to evaluate three hypotheses related to the separability of mark shape, size, and color:

**H1—Colors will be more discriminable on denser shapes.**

We anticipate that colors will be more discriminable on filled shapes than equivalent unfilled shapes (e.g.,  vs. ) and that mQTons having angles less than 90° (e.g.,  and ) will perform comparable to filled shapes, while other mQTons will perform comparable to unfilled shapes. This hypothesis stems from prior work that shows colors become more discriminable as mark size increases. Filled shapes and mQTons such as  and  generally use a larger number of pixels compared to unfilled shapes and the other mQTons due to their density. We anticipate that density will exhibit similar gains.

**H2—Denser shapes will appear larger.**

When two shapes are close in diameter, the viewer might integrate other attributes such as area or visual density, privileging filled shapes and dense QTons. Additionally, unfilled shapes indirectly produce a slightly smaller negative shape by enclosing an area within its border [16]. This negative shape could interfere with size comparisons by directing viewer focus to the enclosed area.

**H3—Interference between channels will be asymmetric: shape will more significantly affect color and size than either will affect shape.**

Based on our observations and results from prior studies [25], unless the shapes are too small or too close to the background color to be readily identified, we anticipate that people will reliably compare shapes regardless of color or size.

## 4 GENERAL METHODS

Our goal is to understand separability in ways that inform visualization designers working in the context of real-world displays. Display conditions such as ambient illumination, display gamma, and resolution mean that a single design may look different on different displays. As the number of potential conditions is too large to feasibly account for each independently, prior studies have used crowdsourcing to attain large numbers of real world samples that help account for this variance in aggregate [32, 53, 65, 66]. These models, while limited in their abilities to account for the precise visual mechanisms at play, allow designers to predict perceptions in the real world with significantly higher precision than traditional laboratory models.

### Stimuli

Our experiments use a binary forced-choice design, asking participants to compare either the color, size, or shape of the two target marks. Participants saw a series of scatterplots rendered using D3 on a 375 x 250 pixel white background with 1 pixel gray axes without labels or numbers as shown in Figure 1. Each scatterplot contained two vertically aligned colored shapes embedded in a set of randomly selected gray distractor shapes. As the distance between marks can affect color perception [10], the two colored shapes were always rendered 125 pixels apart, approximating the width of foveal vision at arm’s length. Marks in visualizations are not always evenly spaced, but fixing the distance between the compared shapes allows us to control for potential confounding effects. Marks differed on one tested dimension (shape, size, or color) by a fixed amount.

We selected a set of six fixed sizes (6, 12, 18, 25, 37, and 50 pixels) based on those used in prior experiments [64, 65]. As designers have no methods to estimate aspects of

display resolution or viewing distance necessary for precise psychophysical calculations of visual angle, we present our results in terms of pixel sizes. We convert to visual angle for modeling, assuming a default pixel resolution of 96 PPI and a D65 whitepoint.

We inserted  $\sqrt{\frac{\text{visualization area}}{\text{mark diameter}^2}}$  distractor marks positioned according to a random sampling of a normal distribution ( $\mu = 0.5$ ,  $\sigma = 1.0$ ). Distractor shapes intersecting any other shape in the scatterplot were removed. We adjusted each distractor's size by a random value between  $[-20\%, 20\%]$  of the tested shape size to prevent distractor shapes from providing an anchoring size that might inadvertently support size comparisons. Distractor shapes add visual complexity to the scatterplots being rendered which helps us address the Isolation Assumption made in prior studies [65], increasing the ecological validity of our results.

## Procedure

Each experiment consisted of four phases (1) informed consent and screening, (2) tutorial, (3) formal study, and (4) demographics questionnaire. The specific question phrasings will be given in each experiment's relevant sections. Participants first provided informed consent for their participation and, for studies involving color perception, completed an Ishihara test to screen for color vision deficiencies [38]. While an online Ishihara test is limited due to variations in color representation across displays, it successfully caught color vision deficiencies in piloting. We additionally asked participants to self-report color vision deficiencies at the end of the study and excluded any participants self-reporting CVD. Upon successfully completing the screening, participants completed a series of tutorial questions to clarify any possible ambiguities in the experimental instructions. Participants had to answer all tutorial questions correctly to begin the formal study.

During the formal study, each participant completed a fixed number of trials sequentially in random order to mitigate transfer (e.g., learning or fatigue) effects. Participants recorded their responses for each trial using the 'f' and 'j' keys, as in Stone et al. [64]. Each scatterplot rested in the center of the page with the instructions, a reminder of keypress mappings, and a request to complete each trial as quickly and accurately as possible above the scatterplot. After each trial, a gray box covered the scatterplot for 0.5 seconds to mitigate potential contrast effects between subsequent trials. To ensure honest participation, each experiment contained several control trials with large differences between target marks. After completing all trials, participants reported basic demographic information and were compensated for their participation.

## Participant Recruitment

We recruited 1,930 participants across four experiments using Amazon's Mechanical Turk. All participants were from the United States, between the age of 18 and 65 years old, and had an approval rating of 95% or greater. 181 of these participants (9.4%) were excluded from our analyses for either failing to correctly answer a majority of the engagement checks (trials that were considerably easier to get correct than regular trials), self reporting a color vision deficiency or abnormal vision, or having a mean response time less than 0.5 seconds (mean overall response time: 2.05 seconds).

## Analysis

Across all our experiments, we collected equal numbers of samples for all combinations of our independent variables. Several of our independent variables were mixed-participants factors: they are counterbalanced between-participants with each participant seeing multiple settings of each variable. This allows us to balance statistical power with potential transfer effects. We analyzed effects from our independent variables using an ANCOVA treating interparticipant variation as a random covariate. All post-hoc analyses used Tukey's Honest Significant Difference Test (HSD,  $\alpha = .05$ ).

## 5 SHAPE AND COLOR

Our first pair of experiments measured the separability of shape and color by modeling color perception as a function of size for different shapes. The first experiment focuses on participants' abilities to accurately discern color differences for various mark shapes while the second leverages the same experimental paradigm to instead quantify the effects of color on shape perceptions, allowing us insight into the perceptual symmetry of this relationship.

### Experiment One: Effects of Shape on Color

We conducted a  $16$  (mark shape, mixed)  $\times 5$  (color difference, within)  $\times 3$  (CIELAB axis, between)  $\times 6$  (mark size, within) mixed-factors study to measure the effects of mark shape on color difference perceptions in scatterplots. The general procedure for this experiment is outlined in the General Methods section.

*Stimuli.* Our stimuli consisted of a series of fixed size scatterplots with two colorful marks (one mapped to a fixed color and the second to an adjusted color) embedded in a set of mid-grey ( $L^* = 50$ ) distractor marks. We generated a set of fixed colors by uniformly sampling the CIELAB gamut using 12.5 step increments along the  $L^*$  axis and 12 step increments along both the  $a^*$  and  $b^*$  axes. We removed any colors from our sample that, when adjusted, fell outside of the CIELAB gamut or that fell too close to grey to avoid confusion with our distractor marks. This sampling process yielded 79 fixed

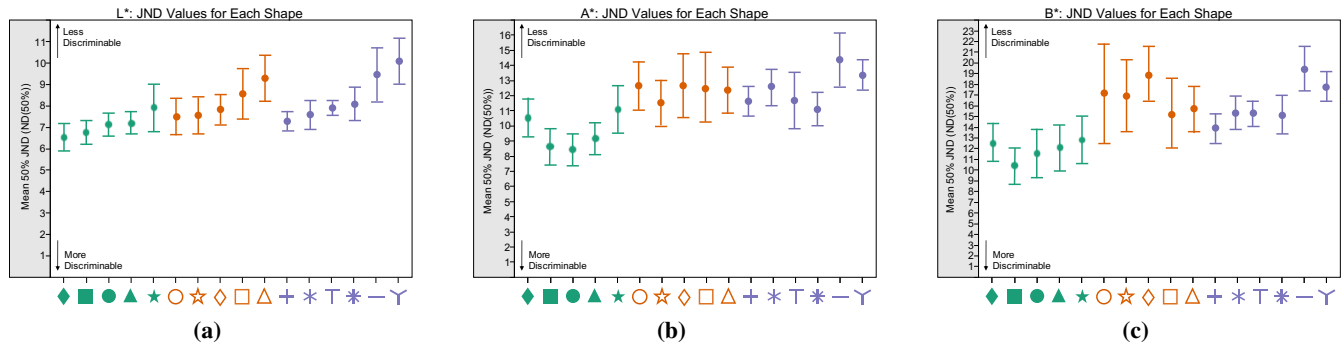


Figure 3: Mean just noticeable differences (JNDs) based on our data-driven models for all shapes across each color axis in CIELAB with error bars showing standard error. Unfilled shapes require larger color differences for people to accurately perceive a difference when compared to filled shapes across all three color axes. mQTons performed comparable to unfilled shapes; however, colors were more discriminable on  $\top$  and  $+$  compared to  $-$  and  $\vee$  contrary to *H1*.

colors where  $L^*$  ranged from 30 to 65,  $a^*$  ranged from -36 to 48, and  $b^*$  ranged from -48 to 48.

We mapped one of these fixed colors to one mark and adjusted the color of the second mark by a fixed color difference step along one axis of CIELAB. Steps corresponded to the size-adjusted JNDs tested in Szafir [65]: no difference,  $0.5ND(50)$ ,  $0.75ND(50)$ ,  $1ND(50)$ ,  $1.25ND(50)$ ,  $1.5ND(50)$ , and  $2ND(50)$ . Each participant saw target marks corresponding to two shapes rendered at six different sizes (6, 12, 18, 25, 37, and 50 pixels).

*Experimental Design.* Participants were asked to compare two colorful marks of the same shape and report whether they were the same color or different. We used discriminability rate (the proportion of correctly identified color differences) as our primary dependent measure. Adjusted color axis was a between-participants factor, while mark size (measured as the shape diameter) and color difference were within-participants factors. Mark shape was a mixed-participants factor. We chose to vary size and color difference within-subjects to help account for variance from display configurations. Each participant saw two different tested shapes and each combination of shape  $\times$  size  $\times$  color difference (along either  $L^*$ ,  $a^*$ , or  $b^*$ ) once. Shapes were pseudorandomly assigned to participants such that combinations of shape, size, and color difference were counterbalanced across all participants. Fixed color was randomly mapped to each trial, with each color tested once per participant. Each participant completed 79 trials, including 7 engagement checks to ensure honest participation.

**Results.** We recruited 683 total participants for this experiment. We excluded 77 participants from our analysis for either failing to correctly answer a majority of the engagement checks, self reporting a color vision deficiency or abnormal

vision, or having a mean response time less than 0.5 seconds. We analyzed data from the remaining 606 participants ( $\mu_{age} = 35.5, \sigma_{age} = 9.5$ , 296 male, 304 female, 6 DNR), resulting in 43,632 trials.

We analyzed the effect of shape on color difference perception across experimental factors (shape, color axis, color difference, and size) using a four-factor mixed-factors ANCOVA with interparticipant variation treated as a random covariate. We report significant effects relative to our primary research questions and include full data tables and analysis scripts at <https://bit.ly/2prjuRu>.

We found a significant effect of shape on color difference perception across all three axes in CIELAB ( $F_L(15, 590) = 7.5, p < .001$ ,  $F_a(15, 590) = 14.3, p < .001$ ,  $F_b(15, 590) = 19.6, p < .001$ ). We found partial support for **H1**: in general, colors were significantly more discriminable for filled shapes than for unfilled shapes. For example,  $\blacksquare$  ( $\mu_L = 56.5\% \pm 3.0\%$ ,  $\mu_a = 53.8\% \pm 3.3\%$ ,  $\mu_b = 58.4\% \pm 3.0\%$ ) and  $\blacktriangle$  ( $\mu_L = 51.1\% \pm 3.3\%$ ,  $\mu_a = 48.7\% \pm 3.3\%$ ,  $\mu_b = 49.2\% \pm 3.2\%$ ) significantly outperformed their unfilled counterparts,  $\square$  ( $\mu_L = 46.1\% \pm 3.3\%$ ,  $\mu_a = 40.0\% \pm 3.2\%$ ,  $\mu_b = 41.8\% \pm 3.3\%$ ) and  $\triangle$ . ( $\mu_L = 42.1\% \pm 3.2\%$ ,  $\mu_a = 34.4\% \pm 3.2\%$ ,  $\mu_b = 37.7\% \pm 3.2\%$ ).

mQTons, however, performed contrary to our expectations. Comparisons showed  $\top$  and  $+$  enabled significantly greater discriminability than unfilled shapes. For example, we found that color differences were perceived either significantly or marginally better across all three color axes when  $\top$  ( $\mu_L = 48.5\% \pm 3.3\%$ ,  $\mu_a = 41.1\% \pm 3.1\%$ ,  $\mu_b = 37.8\% \pm 3.2\%$ ) and  $+$  ( $\mu_L = 53.0\% \pm 3.3\%$ ,  $\mu_a = 39.4\% \pm 3.2\%$ ,  $\mu_b = 41.7\% \pm 3.3\%$ ) were used compared to  $\triangle$  ( $\mu_L = 42.1\% \pm 3.2\%$ ,  $\mu_a = 34.4\% \pm 3.2\%$ ,  $\mu_b = 37.7\% \pm 3.2\%$ ) despite using fewer pixels. These mQTons provided comparable discriminability to



filled shapes despite the lower density of colored pixels associated with the same size mark.

Additionally, we found a significant effect of size on color difference perception across all three color axes ( $F_L(5, 600) = 34.3$ ,  $p < .003$ ,  $F_a(5, 600) = 18.7$ ,  $p < .001$ ,  $F_b(5, 600) = 21.9$ ,  $p < .001$ ). Colors were generally more discriminable when larger sizes were used (e.g., 50 pixels) compared to smaller sizes (e.g., 6 pixels) which replicates previous experimental results [65].

**Modeling Results.** We used our data to construct a set of models using the methodology established in Stone et al. [64]. Our models predict color discriminability rates based on a mark's size and shape by computing 50% JNDs for each combination of shape  $\times$  size. We first used linear regression to fit a linear model through the origin to the mean discriminability rate as a function of color difference for each combination of shape, size, and tested axis, resulting in 96 models for each color axis. All of these models provided significant fits to our experimental data ( $p < .05$ ). We then computed 50% just noticeable color differences (JNDs) for each linear model and fit a second linear regression to these JNDs to model JND variation for each shape and axis as a function of inverse mark size, resulting in 48 total models ( $p < .05$  for all models). We can normalize each axis in the CIELAB  $\Delta E$  metric according to the outputs of these models in order to generate a specific difference model for each tested shape. All of our models can be found at <https://bit.ly/2prjuRu>.

Figure 3 summarizes the mean JNDs derived from our data for each combination of shape and axis aggregated over size, with error bars corresponding to the standard error of that mean to indicate each shape's sensitivity to size variation. We found that both the means and variance varied across different shapes. For  $L^*$ , the mean JND in the least discriminable shape,  $\gamma$  ( $\mu_{JND} = 10.07L^* \pm 3.1\%$ ) was 54.7% larger than the most discriminable shape,  $\blacklozenge$  ( $\mu_{JND} = 6.51L^* \pm 3.2\%$ ). These results suggest that mark shape and color are not readily separable: a mark's shape can significantly shift our abilities to discriminate fine-scale color differences, providing preliminary evidence of significant perceptual interference of shape on color encoding perception. As in prior studies, color perceptions were sensitive to size variation. For example, the mean JND value for  $L^*$  across all 16 shapes at our smallest fixed size (6 pixels) is  $11.30\Delta L^*$  but only  $6.48\Delta L^*$  for our largest fixed size (50 pixels).

## Experiment Two: Effects of Color on Shape

Our first experiment compared identical shapes to gauge the effect of shape on color perception and found evidence that color may not be strongly separable from mark shape. To evaluate the symmetry of our results, we repeated our experiment

using an inverted experimental paradigm: we mapped two differing target shapes to a fixed color to measure the influence of mark color on shape perception. We conducted a  $16$  (shape, within)  $\times 3$  (CIELAB axis, between)  $\times 6$  (mark size, within) mixed-factors study following the general procedure outlined in the General Methods section.

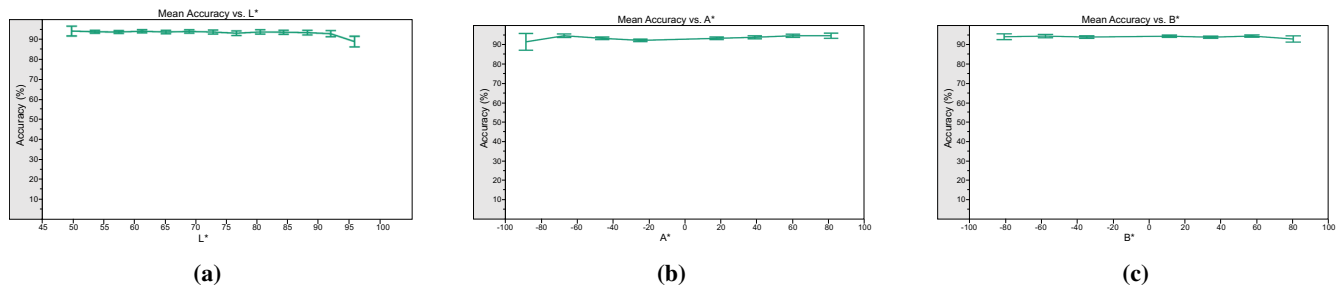
**Stimuli.** Shapes and sizes matched those from the first experiment. Each pair of target marks were mapped to a pair of shapes sampled with replacement from our candidate set, leading to 128 possible combinations of shape (112 differing pairs and 16 same-shape pairs).

We sampled colors uniformly from the CIELAB gamut. We mapped each pair of marks to a single fixed color and one of our 128 shape pairs. However, because we did not vary colors within each trial, we were able to sample a larger proportion of the gamut. This sampling resulted in 168 total colors where  $L^*$  ranged from 50 to 96 using 3.8 step increments,  $a^*$  ranged from -80 to 80 using 20 step increments, and  $b^*$  ranged from -60 to 80 using 20 step increments. We chose lighter colors ( $L^*$  from 50 to 96) because in piloting, error rates were highest when colors were close to the background luminance.

**Experimental Design.** Our experiment again used a binary forced-choice task measuring perceived shape difference across different mark colors. Participants were asked to compare the two colorful marks and report whether they were the same shape or different. We used the discriminability rate (the proportion of correctly identified shape differences) as our primary dependent measure. Fixed color was randomly mapped to each trial, with each color tested once per participant. Each participant completed 168 trials: one trial for each size and shape pair, four engagement checks, and 44 additional same-shape pairs to help balance the distribution of correct responses. Combinations of size, color, and shape were counterbalanced between participants.

**Results.** We recruited 219 total participants for this experiment. 8 were excluded from analysis for either failing to correctly answer a majority of the engagement checks, self-reporting a color vision deficiency or abnormal vision, or having a mean response time less than 0.5 seconds. We analyzed data from the remaining 211 participants ( $\mu_{age} = 36.4$ ,  $\sigma_{age} = 11.5$ , 63 male, 147 female, 1 DNR) resulting in 35,448 total trials.

We analyzed the effect of color on shape difference perception across experimental factors (shape pair, color axis, and size) using a three-factor mixed-factors ANCOVA with interparticipant variation treated as a random covariate. We found a significant effect of  $L^*$  on accuracy of perceived shape differences ( $F(1, 209) = 1.7$ ,  $p < .03$ ). However, Figure 4a shows that the magnitude of this effect is small, and significant variation in accuracies were only seen for colors with extremely high  $L^*$  values: the accuracy difference between



**Figure 4: Shape perception accuracy across all three color axes in CIELAB. Significance was only found in  $L^*$  for extremely light colors ( $L^* > 92$ ). Compared to Figure 3, we see little to no evidence of interference from color on shape perceptions.**

the highest and lowest  $L^*$  values was 4.4%. However, this behavior was asymptotic: we saw only observed degradation in shape perception beyond  $L^* = 92$ . As we measured these perceptions against a white background, this lightness value is near the threshold where the shapes will become difficult to distinguish from the background. We also found a significant interaction between shape and  $L^*$  on shape perception ( $F(15, 195) = 3.5, p < .01$ ), with pairs integrating  $*$  generating slightly lower overall accuracy ( $\mu_L = 93.3\% \pm 2.2\%$ ). We anticipate that the fine features on this shape are especially sensitive to the observed lightness threshold.

While the two dimensions (color and shape) are not immediately comparable, the large difference in the magnitude of the observed effects coupled with the asymptotic behavior we find in shape discrimination suggests that the effect of shape on color perceptions is significantly stronger than that of color on shape perceptions, indicating an asymmetry in the separability of these two channels (H3).

## 6 SHAPE AND SIZE

Our first study showed that mark shape, color, and size may not be separable: changing shapes can dramatically influence both perceived color difference and its sensitivity to changes in size. This interaction may be a result of mark shape biasing perceived size. To further explore the relationship between mark shape and size, we conducted a second pair of experiments measuring the separability of shape and size directly. Experiment Three measures how accurately participants can discern size differences for our 16 mark shapes to measure size perception biases between different shapes, while Experiment Four inverts Experiment Three to explore the effects of size on shape perception.

### Experiment Three: Effects of Shape on Size

We conducted a 16 (shape, within)  $\times$  6 (fixed size, within)  $\times$  4 (size difference, within) mixed-factors study to measure

the effects of shape on size difference perceptions in scatterplots. While our independent variables were tested within-participants, combinations of these variables were counter-balanced mixed-participants factors to avoid fatigue effects given the large number of possible combinations.

**Stimuli.** For this experiment, participants were shown a series of scatterplots as described in the General Methods section with two bright blue shapes embedded within a field of gray distractor shapes. We chose a bright blue ( $L^* = 32, a^* = 79, b^* = -107$ ) to highlight the two tested shapes and ease visual search. Shape pairs were drawn from the combinations of our 16 tested mark shapes resulting in 120 shape pairs.

Fixed sizes mirrored those in Experiment One, with one shape rendered at the fixed size and the second adjusted by adding 5%, 10%, or 15% to its diameter. These thresholds ranged from approximately one to two 50% JNDs for same-shape pairs in piloting. We elected to vary diameter rather than area as our size differences are small, and the geometries of tested shapes complicate normalizing areas between shape pairs. Diameter variation for continuous values in scatterplots is currently employed in tools like Tableau. To mitigate potential effects from aliasing, the actual adjusted differences in pixels were rounded to the nearest whole pixel. We reduced the set of adjusted sizes for our smallest marks as to remove any duplicate sizes caused by rounding errors.

**Experimental Design.** The general procedure for this experiment is outlined in the General Methods section. Participants reported which of the two blue shapes appeared to be greater in size. Our primary dependent measure was the accuracy rate (how often the larger of two shapes was correctly identified as largest). For equal sized pairs, an unbiased percept should lead to a 50% response rate. The tutorial instructed participants to consider size the greater of a mark's height or width to clarify potentially ambiguous conditions (e.g., those where size may disagree on any single dimension, such as — and ♦). While this clarification does add complexity to the task, participants had to correctly answer five tutorial questions



to ensure they understood how size was being defined in the experiment before proceeding to the main study.

In the formal study, each participant saw each shape combination once and each combination of fixed size and size difference five times, with combinations of shape pair, fixed size, and size difference mapped pseudorandomly and counterbalanced between participants. Each participant completed 123 trials including three engagement checks (trials where the shapes differed in length by 50%) to ensure honest participation.

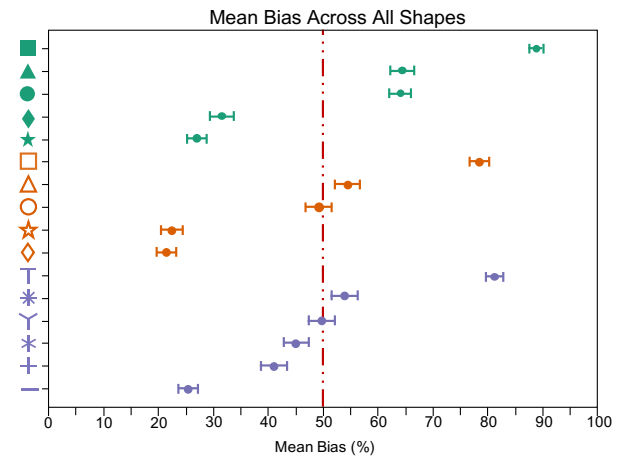
**Results.** We recruited 548 total participants for this experiment. 59 participants were excluded from analysis for either failing to correctly answer a majority of the engagement checks, self reporting a color vision deficiency or abnormal vision, or having a mean response time less than 0.5 seconds. We analyzed data from the remaining 489 participants ( $\mu_{age} = 36.5, \sigma_{age} = 11.6$ , 192 male, 293 female, 4 DNR) for a total of 60,147 trials.

We analyzed the effect of shape on size difference perception across our experimental factors (shape, fixed size, and size difference) using a four-factor ANCOVA. Shape was broken up into two factors: bigger shape (bs) and smaller shape (ss) based on the size differences discussed above.

We found a significant effect of shape on size perception ( $F_{bs}(15, 473) = 943.9, p < .001, F_{ss}(15, 473) = 940.2, p < .001$ ). Post-hoc comparisons showed that shapes having greater visual density along the top or bottom of the shape were generally perceived as having greater diameter (e.g.,  $\top$ ,  $\blacksquare$ ,  $\square$ ,  $\mu = 78.68\% \pm 9.90\%$ ). Additionally, filled shapes were generally perceived as larger than their unfilled counterparts, partially supporting **H2**. Contrary to **H2**, higher-order mQ-Tons (e.g.,  $\ast$ ,  $\times$ ,  $+$ ,  $\mu = 46.93\% \pm 7.16\%$ ) in general were consistently perceived as smaller than less dense shapes.

We also found a significant effect of fixed size on difference perception ( $F(5, 483) = 19.8, p < .001$ ). Participants compared sizes more accurately for larger fixed sizes (e.g., 50 pixels:  $\mu = 69.7\% \pm 1.0\%$ ) than for smaller fixed sizes (e.g., 6 pixels:  $\mu = 64.9\% \pm 1.1\%$ ). We also found a significant interaction effect between bigger shape and smaller shape ( $F(209, 279) = 11.9, p < .001$ ). This effect shows that certain pairs of shapes can also skew size perception.

**Modeling Results.** We fit our data to a set of psychometric functions that predict size perception based on the two shapes being compared, the fixed size, and size difference to model relative size perceptions as a function of the signed size difference between the two shapes. We elected not to include data involving our smallest fixed size (6 pixels) in our modeling because only three total size differences (compared to five and seven for our other fixed sizes) could be rendered within the desired range (-16.7%, 0%, and 16.7% corresponding to 5, 6, and 7 pixels). After removing our smallest fixed size, we



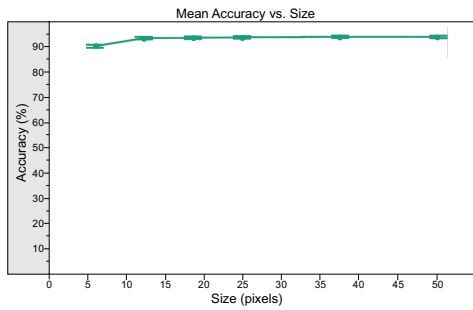
**Figure 5: Predicted mean bias in perceived size across all tested shapes based on our models. Error bars show standard error and the red line indicates no significant bias. Shapes such as  $\blacksquare$ ,  $\square$ , and  $\top$  were generally perceived larger compared to other shapes of the same size whereas  $\star$  and  $\diamond$  were perceived as smaller.**

constructed 600 models corresponding to each shape pair  $\times$  fixed size combination, excluding same shape pairs. Of these 600 models, 558 (93%) of them provided significant fits to the data ( $p < .05$ ).

We can use the y-intercept of these models to estimate the likelihood that one shape will be perceived as larger than another for any pair of shapes. Figure 5 shows the mean bias for each tested shape computed using our models. This figure shows biases consistent with our inferential observations: shapes with visual mass near the top or bottom of the shape tend to be seen as larger overall (e.g.,  $\blacksquare$  compared to  $+$ ). By seeding our functions according to target visualization parameters, we can use the well-fit models generated from this data to normalize our size computations between encoding shapes to help account for potential biases in size perceptions due to shape. Our results suggest that this bias can be quite substantial: for example,  $\top$  shapes were perceived as larger than any other tested shape in 82% of trials whereas  $\star$  shapes were only reported as larger in 27% of trials. This bias suggests that separability between data attributes encoded using shape and size is limited as certain shapes are likely to be perceived as significantly larger even if rendered at comparable sizes. A full list of models is available at <https://bit.ly/2prjuRu>.

#### Experiment Four: Effects of Size on Shape

Due to the substantial biases in perceived size differences between tested shapes, we conducted a 16 (shape, within)  $\times$  6 (fixed size, within)  $\times$  4 (size difference, within) mixed-factors study to measure the symmetry of these effects of



**Figure 6: Participants accurately perceived shape differences robust to size. Significant variation in accuracy was only found in our smallest tested fixed size (6 pixels), and the magnitude of this variation is quite small (accuracy only 4.5% lower than with 50 pixel shapes).**

size on shape perception in scatterplots. While our independent variables were tested within-participants, combinations of these variables were counterbalanced mixed-participants factors to avoid fatigue affects given the large number of possible combinations. Our primary dependent measure was the discriminability rate (how frequently participants correctly reported shape differences).

*Experimental Design.* The general procedure for this experiment is outlined in the General Methods section. Participants again saw a series of scatterplots containing two bright blue shapes within a field of gray distractor shapes. Shape sizes matched those tested in Experiment Three, with each scatterplot mapping shapes to a single target size. Similar to Experiment Two, participants were asked if the two blue shapes in the scatterplot were the same shape or different. Each participant saw 168 trials corresponding to 120 shape pairs plus an additional 48 same-shape pairs. Trials that rendered shapes at the largest size (50 pixels) were used as engagement checks. Shape, fixed size, and size difference were distributed across trials identically to Experiment Three.

*Results.* We recruited 480 total participants for this experiment. 37 were excluded from analysis for either failing to correctly answer a majority of the engagement checks, self-reporting a color vision deficiency or abnormal vision, or having a mean response time less than 0.5 seconds. We analyzed data from the remaining 443 participants ( $\mu_{age} = 37.2$ ,  $\sigma_{age} = 12.5$ , 167 male, 275 female, 1 DNR), resulting in 74,424 total trials.

We analyzed the effect of size on shape perception across our experimental factors (shape, fixed size, and size difference) using a four-factor mixed effects ANCOVA. Shape was again broken up into two factors: bigger shape (bs) and smaller shape (ss) based on the size differences discussed above. We found a significant effect of size ( $F(5, 437) = 47.3$ ,  $p < .001$ ) as well as size difference ( $F(3, 437) = 8.5$ ,  $p <$

.001) on the rate of perceived shape differences. Similar to Experiment Two however, the magnitude of this effect is small, and significant variation in accuracies are only seen for our smallest fixed size (6 pixels,  $F_{bs}(75, 367) = 1.5$ ,  $p < .01$ ,  $F_{ss}(75, 367) = 1.9$ ,  $p < .001$ ). The difference in shape discriminability between 50 pixel shapes and 6 pixel shapes was 4.5%. As with color, the observed behavior again appears to be asymptotic for small shapes. One possible explanation for this effect is that the shapes have become sufficiently small that fine resolution details can no longer be efficiently rendered or detected. However, further testing is needed to confirm this hypothesis.

Our results again suggest an asymmetry between effects of shape on size and size on shape for our tested mark parameters: accuracy differences for size perceptions across shapes were much larger than those for shape perceptions at varying sizes. We are limited in our abilities to directly compare shape and size as differences along these two encoding vectors as our sampled sizes and shapes are not necessarily aligned in their difficulty; however, as with color, we see that these asymmetries hold across even large size differences. The two experiments also use the same set of parameters and stimuli, but generate significantly different results, further indicating an asymmetry in the separability of size and shape.

## 7 DISCUSSION

We measured the separability of shape, size, and color in multiclass scatterplots to understand the perceptual interplay of these encodings. Our results show:

- *Shape significantly affects color difference perception:* Colors were more discriminable when filled shapes were used compared to unfilled shapes. Shapes such as T and + performed comparably to denser shapes.
- *Color has some effect on shape perception:* Extremely light colors complicated shape perceptions but the magnitude of the effect was small.
- *Shape significantly affects size perception:* Sizes were perceived to be greater for shapes that contain more visual density on the top or bottom of the shape (e.g., T, ■, □). Filled shapes were perceived as larger than their unfilled counterparts.
- *Size has some effect on shape perception:* Extremely small sizes complicated shape perceptions but the magnitude of the effect was small.

Color difference perception accuracy was generally highest when filled shapes were used, partially supporting H1. Our results also replicated results from previous experiments (Szafr [65] and Stone et al. [64]), showing that color difference perception varies proportionally to mark size. However, contrary to our expectations, color difference perceptions in open shapes were not well explained by shape density. We

expected the less dense mQTons ( $-$ ,  $\top$ ,  $\gamma$ , and  $+$ ) to perform similar to unfilled shapes, and the denser mQTons ( $*$  and  $\blacksquare$ ) to perform similar to filled shapes. However, we found no clear patterns in mQTON shape and color perceptions. For example, our fourth mQTON ( $+$ ) outperformed other mQTons for the  $L^*$  and  $b^*$  axes in CIELAB and had similarly high discriminability for  $a^*$ . This shape uses fewer pixels than  $*$  and  $\blacksquare$ , and uses a comparable number of pixels to  $\top$  and  $\gamma$ , yet consistently supported higher discriminability.

This preliminary evidence suggests that there may be alternative mechanisms that cause the biases observed in our data. One phenomenon that could be affecting color perceptions for mQTons is illusory contours [50]. For example, viewers could be perceiving an illusory unfilled cube when looking at  $\gamma$ , or an illusory circular border around  $*$  and  $\blacksquare$ . These illusory contours may cause our visual system to blend the shape color with the background color. However, we would need to further test these hypotheses in a laboratory environment to ascertain the precise factors involved in mQTON perception.

Our second experiment demonstrated an asymmetry in the perceptual interdependence of color and shape: only extremely light colors that are rarely used in practice significantly degraded shape perceptions. This effect is likely caused by participants either not being able to find target shapes or not being able to resolve the details of the target shapes because the background color of the scatterplot (white) and the light-colored target shapes are not sufficiently discriminable. This limited effect suggests an asymmetric relationship exists between shape and color: shape much more strongly affects color perception than vice versa, supporting **H3**.

We anticipated some of the variation we saw in our first pair of experiments may be explained by variations in size perceptions across shapes: if some shapes appear larger, they may also support greater discriminability. Our results showed that filled shapes are generally perceived as larger than unfilled shapes, partially supporting **H2**. Holes created by unfilled shapes may cause people to attend to the negative shape created by the hole and causing unfilled shapes to appear smaller. However, if we group shapes generally perceived as larger (e.g.,  $\top$ ,  $\blacksquare$ , and  $\square$ ), only one of these shapes is filled. Our results suggest that visual density along the top or bottom of a shape is a more significant determinant of perceived size biases. The two target shapes in our stimuli were always vertically aligned, which may partially explain these results; however, shapes like  $\blacklozenge$  also span the full vertical mark space. Our results also suggest that the variations seen in Experiment One cannot be explained by mark size biases. For example,  $\blacklozenge$  generally provided high discriminability; however, it was perceived as smaller than 12 of our 15 tested marks. Similarly,  $+$  provided the best color discriminability of all mQTons, but was perceived as smaller than 4 of the 5 tested mQTons. Future studies in laboratory environments would help us better

understand the perceptual mechanisms behind shape, color, and size interference to evaluate these hypotheses.

We found a limited effect of size on shape perception; like Experiment Two, the magnitude of this effect was substantially smaller than the effect of shape on size and only present for the smallest tested mark sizes, again supporting **H3**. When mark size is extremely small (smaller than any practical visualization would use), shapes may not render accurately due to the limited amount of pixel space. This would lead to a decrease in shape perception accuracy as the geometry of the shape itself would change.

As our goal is to improve visualization effectiveness in the real world, it is important to consider how these results might also affect higher level tasks such as correlation or average. While we evaluate binary comparisons, these shifts likely affect higher level tasks. Errors while performing higher level tasks are likely to be as high as those for comparison tasks as the perception of each individual mark is skewed, which skews the overall distribution. For example, scatterplots using  $\blacksquare$  and  $*$  to categorize datapoints and using a color encoding could potentially skew correlation or average interpretations because colors are more discriminable on  $\blacksquare$  compared to  $*$  which would skew the overall distribution of color differences perceived. Our modeling approach allows us to approximate these errors and shift encodings appropriately.

Separable versus integral channels have always been important guidelines for visualization designers because understanding how viewers will perceive the information being presented is critical for maximizing visualization effectiveness. Our results collectively suggest that shape, size, and color are significantly less separable than conventionally thought. These results show strong asymmetries among these visual channels. This work offers quantified empirical guidance for reasoning about and accounting for integrality in multivariate visualizations.

## Limitations and Future Work

While crowdsourcing allows us to model visualization perception in real-world contexts, it represents a trade-off of control for ecological validity. This trade-off is well-studied in graphical perception [22, 25, 32, 42, 52], but does limit our results in notable ways. For example, hardware settings could cause colors to appear brighter or more discriminable for some participants. We recruited a large number of participants to account for this limitation, and our results from Experiment One closely replicate results from prior studies that show an interdependence between color and size, including both laboratory studies [14] and our own work with web-based visualizations [64, 65]. Variation in screen resolution could cause marks to appear different in size for different participants. While we use absolute pixel sizes in our experiments, our models emphasize the relative difference in mark sizes

which are consistent despite variation in screen resolution and consistent with visualization encodings. Because of this, our results focus on the ordinal nature of our tested sizes rather than the absolute pixel size.

Additionally, anti-aliasing could cause some lines to be rendered slightly differently for different participants since anti-aliasing is based on browser configuration. To help mitigate the influence of this, we use D3 [6], one of the most popular visualization libraries, which renders SVGs that leave rasterization to the browser and beyond a visualization designer's control. Anti-aliasing would substantially impact shapes containing lines less than a pixel wide which would only affect our smallest tested size (6 pixels). We elected to remove samples involving 6-pixel marks from our models constructed from Experiment Three. The primary goal of this work is to inform visualization designers and improve visualization effectiveness in the real world. Designers rarely know the hardware settings and environment of the viewer, so the choice to use crowdsourcing for these experiments simulates what designers work with in the real world.

Our experiments used a limited sample of shapes, sizes, and colors but generally kept the samples within the range of practical visualizations. Future work could use different backgrounds and more carefully controlled mark shapes. Further, we used limited samples to control for the large number of conditions. Future work could increase the number of samples integrated into our models to reduce model variance and to allow for a more in-depth statistical evaluation of model variation across each condition.

Conventional guidelines around size encoding often recommend using area rather than diameter. We chose to use diameter to encode size to provide consistency in how we normalized and adjusted shapes with very different geometries and because our size manipulations were sufficiently small that we anticipate the differences in diameter and area encodings were negligible. This decision is well-grounded in common visualization tools: several visualization tools used in the real world, such as Tableau, encode size using diameter. However, future experiments should confirm these results hold for other size encodings such as area.

While our results show disadvantages to using unfilled shapes with respect to color and size biases, unfilled shapes can help disambiguate marks that are overlapping. For example, Tableau uses unfilled shapes by default to avoid overdraw in scatterplots. Extending these studies to consider scatterplots with overdraw would further illuminate trade-offs in using filled versus unfilled shapes in multivariate visualizations.

### Implementation

The lack of an easily discernable mapping between mark geometry, size perceptions, and color perceptions make it

difficult for people to account for potential biases created in multidimensional scatterplots and other visualizations combining these encodings. The measures and models constructed in this work can be used to make adjustments to the color and size ranges used in these visualizations as a function of mark shape in order to improve visualization effectiveness.

To support the use of these models in practice, we have implemented the models constructed from Experiment One as a D3 extension (which can be found here: <https://bit.ly/2prjuRu>) to normalize color encoding scales according to the size and shape parameters of a target visualization. This package takes parameters such as mark size and shape and returns the minimum distance ( $\Delta E$  in CIELAB) that mark colors need to be apart so that 50% of the population will perceive a difference to support perceptually corrected color interpolation. By integrating these models into a common visualization tool, we hope to allow developers to fluidly account for interference between these channels to generate more effective multidimensional visualizations.

## 8 CONCLUSION

In this paper, we measure how the interplay of shape, size, and color encodings influence our ability to distinguish data values along each channel and measure the symmetry of these effects. Our results suggest that shape, size, and color are less separable than conventionally thought and that some of this variation can be readily modeled. We found that colors are generally more discriminable when filled shapes are used compared to unfilled shapes and shapes such as + and T performed comparably to denser shapes despite using fewer pixels. We also found preliminary evidence that shapes containing more visual density along the top or bottom edges (e.g., T, ■, □) are perceived as larger than shapes that do not (e.g., −, +, ★). These effects appear asymmetric—shape more strongly affects color and size difference perception than color or size affect shape perception. From our experimental results, we constructed models that predict viewer perception based on what shapes are being used in a scatterplot and implemented these models as an extension in D3. We hope that this work can be used to spark future research to construct new guidelines for separable and integral visual channels grounded in empirical evidence.

## 9 ACKNOWLEDGMENTS

We thank Matt Whitlock for his support and thoughtful feedback. This research was funded by NSF CRII: CHS # 1657599.

## REFERENCES

- [1] Daniel Acevedo and David Laidlaw. 2006. Subjective quantification of perceptual interactions among some 2D scientific visualization methods. *IEEE Transactions on Visualization and Computer Graphics* 12, 5 (2006), 1133–1140.

- [2] Danielle Albers, Michael Correll, Steve Franconeri, and Michael Gleicher. 2014. A task driven framework for visualizing time series data. In *Proc. 2014 ACM Human Factors in Computing Systems*. ACM.
- [3] Eric Alexander, Chih-Ching Chang, Mariana Shimabukuro, Steven Franconeri, Christopher Collins, and Michael Gleicher. 2016. The biasing effect of word length in font size encodings. In *Poster Compendium of the IEEE Conference on Information Visualization*.
- [4] Lawrence D Bergman, Bernice E Rogowitz, and Lloyd A Treinish. 1995. A rule-based tool for assisting colormap selection. In *Proceedings of the 6th conference on Visualization '95*. IEEE Computer Society, 118.
- [5] David Borland and Russell M Taylor. 2007. Rainbow color map (still) considered harmful. *IEEE computer graphics and applications* 27, 2 (2007).
- [6] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D<sup>3</sup> data-driven documents. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2301–2309.
- [7] Nadia Boukhelifa, Anastasia Bezerianos, Tobias Isenberg, and Jean-Daniel Fekete. 2012. Evaluating sketchiness as a visual variable for the depiction of qualitative uncertainty. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2769–2778.
- [8] D.H. Brainard and B.A. Wandell. 1992. Asymmetric color matching: how color appearance depends on the illuminant. *Journal of the Optical Society of America A* 9, 9 (1992), 1433–1448.
- [9] Cynthia A Brewer, Geoffrey W Hatchard, and Mark A Harrower. 2003. ColorBrewer in print: a catalog of color schemes for maps. *Cartogr. Geogr. Inf. Sci.* 30, 1 (2003), 5–32.
- [10] Alžběta Brychtová and Arzu Çöltekin. 2017. The effect of spatial distance on the discriminability of colors in maps. *Cartography and Geographic Information Science* 44, 3 (2017), 229–245.
- [11] Roxana Bujack, Terece L Turton, Francesca Samsel, Colin Ware, David H Rogers, and James Ahrens. 2018. The good, the bad, and the ugly: A theoretical framework for the assessment of continuous colormaps. *IEEE transactions on visualization and computer graphics* 24, 1 (2018), 923–933.
- [12] David Burlinson, Kalpathi Subramanian, and Paula Goolkasian. 2018. Open vs. Closed Shapes: New Perceptual Categories? *IEEE transactions on visualization and computer graphics* 24, 1 (2018), 574–583.
- [13] Tara C Callaghan. 1984. Dimensional interaction of hue and brightness in preattentive field segregation. *Perception & psychophysics* 36, 1 (1984), 25–34.
- [14] Robert C Carter and Louis D Silverstein. 2010. Size matters: Improved color-difference estimation for small visual targets. *Journal of the Society for Information Display* 18, 1 (2010), 17–28.
- [15] Haidong Chen, Wei Chen, Honghui Mei, Zhiqi Liu, Kun Zhou, Weifeng Chen, Wentao Gu, and Kwan-Liu Ma. 2014. Visual abstraction and exploration of multi-class scatterplots. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1683–1692.
- [16] Lin Chen and Wu Zhou. 1997. Holes in illusory conjunctions. *Psychonomic Bulletin & Review* 4, 4 (1997), 507–511.
- [17] Patricia W Cheng and Robert G Pachella. 1984. A psychophysical approach to dimensional separability. *Cognitive Psychology* 16, 3 (1984), 279–304.
- [18] W.S. Cleveland and R. McGill. 1984. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *J. Amer. Statist. Assoc.* 79, 387 (1984), 531–554.
- [19] William S Cleveland and William S Cleveland. 1983. A color-caused optical illusion on a statistical graph. *The American Statistician* 37, 2 (1983), 101–105.
- [20] William S Cleveland and Robert McGill. 1984. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *J. Amer. Statist. Assoc.* 79, 387 (1984), 531–554.
- [21] Michael Correll, Danielle Albers, Steven Franconeri, and Michael Gleicher. 2012. Comparing averages in time series data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1095–1104.
- [22] Michael Correll, Dominik Moritz, and Jeffrey Heer. 2018. Value-Suppressing Uncertainty Palettes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 642.
- [23] Michael A Correll, Eric C Alexander, and Michael Gleicher. 2013. Quantity estimation in visualizations of tagged text. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2697–2706.
- [24] Commission Internationale de l'Éclairage. 1978. Recommendations on uniform color spaces, color-difference equations, psychometric color terms. *Paris: CIE* (1978).
- [25] Gagatay Demiralp, Michael S Bernstein, and Jeffrey Heer. 2014. Learning perceptual kernels for visualization design. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1933–1942.
- [26] Ronak Etemadpour, Robson Carlos da Motta, Jose Gustavo de Souza Paiva, Rosane Minghim, Maria Cristina Ferreira de Oliveira, and Lars Linsen. 2014. Role of human perception in cluster-based visual analysis of multidimensional data projections. In *Information Visualization Theory and Applications (IVAPP), 2014 International Conference on*. IEEE, 276–283.
- [27] Johannes Fuchs, Fabian Fischer, Florian Mansmann, Enrico Bertini, and Petra Isenberg. 2013. Evaluation of alternative glyph designs for time series data in a small multiple setting. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3237–3246.
- [28] Wendell R Garner. 1976. Interaction of stimulus dimensions in concept and choice processes. *Cognitive psychology* 8, 1 (1976), 98–123.
- [29] Michael Gleicher, Michael Correll, Christine Nothelfer, and Steven Franconeri. 2013. Perception of Average Value in Multiclass Scatterplots. *IEEE TVCG* 19, 12 (2013), 2316–2325.
- [30] Richard L Gottwald and WR Garner. 1975. Filtering and condensation tasks with integral and separable dimensions. *Perception & Psychophysics* 18, 1 (1975), 26–28.
- [31] Lane Harrison, Fumeng Yang, Steven Franconeri, and Remco Chang. 2014. Ranking Visualizations of Correlation Using Weber's Law. *IEEE Transactions on Visualization and Computer Graphics* (2014).
- [32] J. Heer and M. Bostock. 2010. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the 28th international conference on Human factors in computing systems*. ACM, 203–212.
- [33] Jeffrey Heer, Nicholas Kong, and Maneesh Agrawala. 2009. Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1303–1312.
- [34] Michael Jenkins, Anna Grubert, and Martin Eimer. 2017. Target objects defined by a conjunction of colour and shape can be selected independently and in parallel. *Attention, Perception, & Psychophysics* 79, 8 (2017), 2310–2326.
- [35] Younghoon Kim and Jeffrey Heer. 2018. Assessing effects of task and data distribution on the effectiveness of visual encodings. In *Computer Graphics Forum*, Vol. 37. Wiley Online Library, 157–167.
- [36] J.H. Krantz. 2001. Stimulus delivery on the Web: What can be presented when calibration isn't possible. *Dimensions of Internet Science* (2001), 113–130.
- [37] Jeongmi Lee, Carly J Leonard, Steven J Luck, and Joy J Geng. 2018. Dynamics of Feature-based Attentional Selection during Color-Shape Conjunction Search. *Journal of cognitive neuroscience* (2018), 1–15.
- [38] H Legrand, G Rand, and C Rittler. 1945. Tests for the detection and analysis of color-blindness I. The Ishihara test: An evaluation. *Journal*

- of the Optical Society of America 35 (1945), 268. Issue 4.
- [39] Stephan Lewandowsky and Ian Spence. 1989. Discriminating strata in scatterplots. *J. Amer. Statist. Assoc.* 84, 407 (1989), 682–688.
  - [40] Jing Li, Jarke J van Wijk, and Jean-Bernard Martens. 2009. Evaluation of symbol contrast in scatterplots. In *Visualization Symposium, 2009. PacificVis' 09. IEEE Pacific*. IEEE, 97–104.
  - [41] Sharon Lin, Julie Fortuna, Chinmay Kulkarni, Maureen Stone, and Jeffrey Heer. 2013. Selecting Semantically-Resonant Colors for Data Visualization. In *Computer Graphics Forum*, Vol. 32. Wiley Online Library, 401–410.
  - [42] Yang Liu and Jeffrey Heer. 2018. Somewhere Over the Rainbow: An Empirical Assessment of Quantitative Colormaps. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 598.
  - [43] Luana Micallef, Gregorio Palmas, Antti Oulasvirta, and Tino Weinkauff. 2017. Towards perceptual optimization of the visual design of scatterplots. *IEEE transactions on visualization and computer graphics* 23, 6 (2017), 1588–1599.
  - [44] K.T. Mullen. 1985. The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *The Journal of Physiology* 359, 1 (1985), 381–400.
  - [45] Tamara Munzner. 2014. *Visualization Analysis and Design*. CRC Press.
  - [46] B. Oicherman, M.R. Luo, B. Rigg, and A.R. Robertson. 2008. Effect of observer metamerism on colour matching of display and surface colours. *Color Research and Applications* 33, 5 (2008), 346–359.
  - [47] Lace Padilla, P Samuel Quinan, Miriah Meyer, and Sarah H Creem-Regehr. 2017. Evaluating the Impact of Binning 2D Scalar Fields. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 431–440.
  - [48] Stephen E Palmer. 1999. *Vision science: Photons to phenomenology*. MIT press.
  - [49] Anshul Vikram Pandey, Josua Krause, Cristian Felix, Jeremy Boy, and Enrico Bertini. 2016. Towards understanding human similarity perception in the analysis of large sets of scatter plots. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 3659–3669.
  - [50] Susan Petry and Glenn E Meyer. 2012. *The perception of illusory contours*. Springer Science & Business Media.
  - [51] James R Pomerantz and Lawrence C Sager. 1975. Asymmetric integrality with dimensions of visual pattern. *Perception & Psychophysics* 18, 6 (1975), 460–466.
  - [52] Khairi Reda, Pratik Nalawade, and Kate Ansah-Koi. 2018. Graphical Perception of Continuous Quantitative Maps: the Effects of Spatial Frequency and Colormap Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 272.
  - [53] Katharina Reinecke, David R Flatla, and Christopher Brooks. 2016. Enabling Designers to Foresee Which Colors Users Cannot See. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2693–2704.
  - [54] Ronald A Rensink. 2017. The nature of correlation perception in scatterplots. *Psychonomic bulletin & review* 24, 3 (2017), 776–797.
  - [55] Ronald A Rensink and Gideon Baldridge. 2010. The perception of correlation in scatterplots. In *Computer Graphics Forum*, Vol. 29. Wiley Online Library, 1203–1210.
  - [56] P. Rizzo, A. Bierman, and M.S. Rea. 2002. Color and brightness discrimination of white LEDs. In *International Symposium on Optical Science and Technology*. International Society for Optics and Photonics, 235–246.
  - [57] Bahador Saket, Alex Endert, and Cagatay Demiralp. 2018. Task-Based Effectiveness of Basic Visualizations. *IEEE Transactions on Visualization and Computer Graphics* (2018).
  - [58] A. Sarkar, L. Blondé, P. Le Callet, F. Atrousseau, P. Morvan, J. Stauder, et al. 2010. A color matching experiment using two displays: design considerations and pilot test results. In *Proceedings of the Fifth European Conference on Color in Graphics, Imaging and Vision*.
  - [59] Michael Sedlmair and Michaël Aupetit. 2015. Data-driven Evaluation of Visual Quality Measures. In *Computer Graphics Forum*, Vol. 34. Wiley Online Library, 201–210.
  - [60] Vidya Setlur and Maureen C Stone. 2016. A linguistic approach to categorical color assignment for data visualization. *IEEE transactions on visualization and computer graphics* 22, 1 (2016), 698–707.
  - [61] Varshita Sher, Karen G Bemis, Ilaria Liccardi, and Min Chen. 2017. An Empirical Study on the Reliability of Perceiving Correlation Indices using Scatterplots. In *Computer Graphics Forum*, Vol. 36. Wiley Online Library, 61–72.
  - [62] Drew Skau and Robert Kosara. 2016. Arcs, angles, or areas: individual data encodings in pie and donut charts. In *Computer Graphics Forum*, Vol. 35. Wiley Online Library, 121–130.
  - [63] M. Stokes, M.D. Fairchild, and R.S. Berns. 1992. Precision requirements for digital color reproduction. *ACM Transactions on Computer Graphics* 11, 4 (1992), 406–422.
  - [64] Maureen Stone, Danielle Albers Szafr, and Vidya Setlur. 2014. An engineering model for color difference as a function of size. In *Color and Imaging Conference*, Vol. 2014. Society for Imaging Science and Technology, 253–258.
  - [65] Danielle Albers Szafr. 2018. Modeling Color Difference for Visualization Design. *IEEE transactions on visualization and computer graphics* 24, 1 (2018), 392–401.
  - [66] Danielle Albers Szafr, Maureen Stone, and Michael Gleicher. 2014. Adapting color difference for design. In *Color and Imaging Conference*, Vol. 2014. Society for Imaging Science and Technology, 228–233.
  - [67] Anne Treisman and Sharon Sato. 1990. Conjunction search revisited. *Journal of experimental psychology: human perception and performance* 16, 3 (1990), 459.
  - [68] Lothar Tremmel. 1995. The visual separability of plotting symbols in scatterplots. *Journal of Computational and Graphical Statistics* 4, 2 (1995), 101–112.
  - [69] André Calero Valdez, Martina Ziefle, and Michael Sedlmair. 2018. Priming and anchoring effects in visualization. *IEEE transactions on visualization and computer graphics* 24, 1 (2018), 584–594.
  - [70] Colin Ware. 2000. *Information visualization* (3 ed.). Morgan Kaufmann.
  - [71] Colin Ware. 2009. Quantitative texton sequences for legible bivariate maps. *IEEE Transactions on Visualization and Computer Graphics* 15, 6 (2009).
  - [72] Liang Zhou and Charles Hansen. 2016. A survey of colormaps in visualization. *IEEE Transactions on Visualization and Computer Graphics* 22, 8 (2016).