A Baseline Study of Emphasis Effects in Information Visualization

Abstract

Emphasis effects – visual changes that make certain elements more prominent – are commonly used in information visualization to draw the user's attention or to indicate importance. Although theoretical frameworks of emphasis exist (that link visually diverse emphasis effects through the idea of visual prominence compared to background elements), most metrics for predicting how emphasis effects will be perceived by users come from abstract models of human vision which may not apply to visualization design. In particular, it is difficult for designers to know, when designing a visualization, how different emphasis effects will compare and what level of one effect is equivalent to what level of another. To address this gap, we carried out two studies that provide empirical evidence about how users perceive different emphasis effects, using three visual variables (colour, size, and blur/focus) and eight strength levels. Results from gaze tracking, mouse clicks, and subjective responses show that there are significant differences between visual variables and between levels, and allow us to develop an initial understanding of perceptual equivalence. We developed a model from the data in our first study, and used it to predict the results in the second; the model was accurate, with high correlations between predictions and real values. Our studies and empirical models provide valuable new information for designers who want to understand and control how emphasis effects will be perceived by users.

Index Terms: Human-centered computing—Visualization—Visualization techniques—Perception; Human-centered computing—Visualization—Visualization design and evaluation methods

1 Introduction

Emphasis effects are visual changes that make certain elements more prominent, and are commonly used in information visualization to draw the user's attention or to indicate importance. Emphasizing important data points is a common method used by designers to support the user when gradually exploring the data – or in narrative visualization [25], when known aspects of the data are presented to the users. An effective emphasis effect will alter a data point's visual features [4, 22], such that a viewer's attention will be guided to the region of interest [61].

The goal of emphasis is to alter important data points to appear more visually prominent and can be achieved through the use of a variety of visual effects [17, 22, 23, 59]. For example, a visualization can use colour changes to emphasize some data points, and differences in the visual prominence of the selected data points will be achieved from variations in color, a visual variable known to guide a user's attention [24].

Although theoretical frameworks of emphasis exist that link visually diverse emphasis effects through the idea of visual prominence compared to background elements [22], we still know little about how emphasis effects will be perceived by users. In particular, we know little about what visual effects, and what magnitudes of those effects, will be most quickly recognized as emphasis by the viewer of a visualization; in addition, we know little about how different effects compare and what level of an effect is equivalent to what level of another.

Many metrics for predicting how emphasis effects will be perceived by users come from abstract models of human vision which may not apply to visualization design. These abstract models from human vision are generally constructed using large and visually isolated stimuli under optimal conditions. Models of human visual attention are effective at predicting perceptibility in isolation but not within a field of distractors, and do not work well with even minor changes to the visual field [3, 58].

Visualizations, in contrast, often consist of large numbers of a variety of marks viewed using a wide range of devices and environments - and designers may use a variety of techniques to emphasize data points. Current guidelines do not address how different emphasis effects are perceived by viewers in visualizations, or provide an equivalence metric for perceived emphasis so designers can choose effects correctly. Without effective models of visual prominence in visualizations, designers lack information on know how different visual effects compare, and don't know what magnitude of effect to use to appropriately guide a viewer's attention to an area of interest.

To address this gap, we carried out two studies that provide empirical evidence about how users perceive different emphasis effects, using three visual variables (colour, size, and blur/focus) and eight strength levels. It is important to note that the three emphasis effects are qualitatively different - for example, colour and size manipulate just the emphasized element, whereas blur/focus manipulates everything but the emphasized element - and so our goal is not simply to identify which effect is most perceivable, but rather to establish how the effects compare to one another at different magnitudes. To do this, our first study established a baseline of perceived visual prominence through eye-tracking data, interaction logs, and subjective ratings in simulated static scatterplot visualizations. We then built a model from the first study's data using logarithmic curves, that provides a prediction of equivalence between the three emphasis effects. Our second study then examined perceived emphasis in a more realistic context, by looking at visual prominence in complex visualizations that are taken from real-world applications (the MASSVIS dataset [7]). We evaluated our model by using it to predict the results of the second study for three different measures; the model was accurate, with R^2 values as high as 0.96.

Our two studies provide new findings about how people perceive three emphasis effects and their magnitudes in visualizations:

- There were significant differences in both studies for emphasis effect: blur/focus was most prominent, and colour least prominent, with size in between depending on magnitude.
- There were also significant differences between the magnitude levels for all effects, providing a graduated way to increase or decrease perceived prominence.
- A predictive model based on logarithmic curves fit the Study 1 data well, and was accurate at predicting perceived emphasis in Study 2 (particularly in terms of subjective ratings).

Our studies provide an initial empirical foundation for understanding how visual effects operate and are experienced by viewers when used for emphasis in visualizations – and although more work is needed to refine and broaden the models, our work provides useful new information for designers who want to control how emphasis effects will be perceived by users.

2 Related Work

Emphasis is essential to InfoVis and is used to highlight regions of interest in a visualization. While there is a large body of research in this domain, much of the work seeks to understand how the underlying perceptual system operates – limiting the possibility of extracting design lessons from low-level data and findings. We survey current empirical studies of perception from visualization and vision science to inform our work.

2.1 Visual Attention and Graphical Perception

There are various theories and computational models for selective visual attention, but in general, most theories agree that attention operates by alternately selecting "features" from a number of incoming subsets of sensory data for further processing [50]. Early work suggests a two-stage process: first, a bottom-up, pre-attentive stage, which is automatic and independent of a task [50], where attention is guided to the most salient items in a scene [69]; followed by a second, slower, top-down stage that is driven by current tasks and goals [34, 49, 61]. Within this model, the conjunction of basic features (such as colour and orientation) stems from "binding" features together (known as Feature Integration Theory [61]).

A second theory, the Guided Search Theory, extends the two-stage process by proposing that attention can be biased toward targets of interest (e.g., a user looking for a red circle) in the top-down phase by encoding particular visual characteristics [68]: for example, assigning a higher weight to the red colour. Recently-proposed attention theories, however, challenge the two-stage model suggesting there is also bias to prioritize items that have been previously selected, thus proposing a three-stage model: current goals, selection history, and physical salience (bottom-up attention) [1].

Attention is commonly examined through visual search experiments which usually ask participants to determine whether a target is present in a scene with distractors; reaction times (RT) and accuracy are used to model the relationship between the response and the number of distractors. Frequently, the term "popout" is used to describe a target item that is easily identified due to its unique visual properties in these searches. While the noticeability of specific visual characteristics such as colour and size cues have also been examined from an attention perspective [21, 43], a related area of research called *graphical perception* takes a more in-depth look at the suitability of different visual channels, and at how choices in visual variables for encoding data affect visualization effectiveness [15].

Graphical perception studies have explored how different visual channels might support a variety of tasks for visualization [51]. Bertin was among the first to study the ability of visual variables to encode information, suggesting that variations in individual visual variables is an effective tool for encoding information and achieving noticeability [4]. Particularly, Bertin suggests that selective visual variables, such as position, size, colour hue, or texture allows viewers to immediately detect variables.

Following Bertin, researchers in multiple disciplines such as cartography [41], statistics [15], and computer science [42] have conducted human-subjects experiments and have derived rankings of visual variables for nominal, ordinal, or quantitative data [15, 41, 42, 56]. In addition to comparing the effectiveness of alternative visual variables for visualization, researchers have investigated how other design factors such as aspect ratios [13], chart sizes [26], and animations [62] influence the effectiveness of charts.

Graphical perception studies have focused on measuring how the visual encoding of variables affect the accuracy of estimating and understanding values of the underlying data; insights from studies in graphical perception, however, can also be applied in manipulating data points in a visualization to guide a viewer's attention to an area of particular importance.

2.2 Emphasis in Visualization

Emphasis is essential to information visualization by offering support to a user when exploring data, for instance through highlighting areas of interest when brushing and linking across multiple views to emphasize relationships [35]. Emphasis is also important when presenting known aspects of data to a user through narrative visualization [53]. The goal of emphasis is to manipulate the visual features of an important data point to make it visually prominent, such that a viewer's bottom-up attention is attracted to the point [22].

While distortion and magnification techniques – which create emphasis effects by simultaneously manipulating a visual variable's size and positioning have been a focus of infovis researchers for creating emphasis [11, 32, 38], new techniques such as blur [36], motion [27], and flicker [64] have recently emerged. Given these new emphasis techniques, Hall *et al.* suggested a categorization of emphasis effects into two main groups based on time variation: time-invariant and time-variant effects [22].

Time-invariant emphasis effects such as highlighting (colouring a data point in a visualization), and blurring (where an data point is shown in focus while the other elements are blurred) do not change with time, and do not use features such as fly-in, fade-in or other transitions [22]. Time-variant emphasis effects, such as motion, flickering, or zooming, by contrast, involve time variations, commonly achieved through animations that alter the appearance of a data point [22].

While there are many ways in which a data point in a visualization can be emphasized, all visual techniques generate emphasis by making the focus mark (i.e., the target) visually more prominent by making it sufficiently dissimilar from the other elements (i.e., the non-target marks) in at least one visual channel [22]. For example, blur/focus, magnification, and highlighting create emphasis by making one data point more visually prominent than others (e.g., sharper, bigger, or a different colour).

There are three main properties of a visual channel that could influence the effectiveness of the visual prominence of an emphasized target mark against the set of non-emphasized marks: the similarity between targets and non-targets, the similarity of all non-targets, and the channel offset (i.e., the lowest value of the non-targets) [63]. Similarity theory shows that visual search efficiency decreases with increased target/non-target similarity and with decreased similarity between the non-targets [18]).

In a similar theory, the relational account of attention theory suggests that the perceived similarity between targets and non-targets can be modeled by the magnitude of a vector in feature space pointing from the target to the closest non-target [2]. If users are given a feature direction in a visual search task (e.g., find the brightest or largest), attention will be guided to the mark that differs in the given direction from the other marks. In this theory, however, the non-target similarity does not have an influence on the visual prominence of a target.

Findings from classic psychophysics and visual search experiments, however, cannot always be applied directly to data visualization. Simple changes such as adding links between dots to simulate a node-link diagram, or changes to contrast effects due to a background luminance have shown to have considerable effects on the results from experiments [3, 58]. These results reinforce the need for empirical evaluations of visualizations to validate theory and evaluate real-world visualization applications.

2.3 Evaluations of Perception in Visualization

Evaluations of perception in visualization have focused on understanding the details of integral and separable channels [57] and the interactions between separable channels. Smart and Szafir found that separability among shape, colour, and size perception functions asymmetrically, with shape found to have the strongest influence on size and colour perception over size's or colour's influence on shape

perception [57]. Other studies have shown that size perception is biased by specific hues, and quantity estimation in visualizations are affected by both size and colour [14, 16].

Scatterplots are one of the most effective visualizations for visual judgments due to data points being positioned along a common scale [25]. Several studies have explicitly explored graphical perception in scatterplots, with many recent techniques being developed to automate scatterplot design [12], and to predict perceptual attributes that may affect scatterplot analysis such as similarity or separability. However, these studies and techniques primarily focus on analyses over single-channel features for scatterplot design to improve legibility or its suitability for data comparison [19].

Eye-tracking evaluations are a popular and effective tool for understanding how users view and visually explore visualizations [5,6]. For example, eye-tracking has been used to understand how different tasks and visual search strategies affect cognitive processes through fixation patterns [45, 46], and has also been used to evaluate specific visualization types [10, 29, 30], for comparing multiple types of visualizations [20], and for evaluating decision making and interaction in visualization [6, 33].

Free-viewing is a common technique for evaluating human perception of visual stimuli. Participants are not given a task and are instructed to freely look around the image, which avoids task-dependent effects. As some attention theories suggest that attention can be guided by a high-level task [1, 69], free-viewing allows attention to be guided by image elements in a bottom-up manner. This assumption has guided researchers to the use of free-viewing for collecting ground truth data for evaluating saliency and attention in visualizations.

However, despite the extensive body of research from vision science on graphical perception, prior research has been focused on evaluating factors that may affect the visual prominence of a specific emphasis effect [63], or in empirically ranking visual variables for encoding data [15, 26]; few guidelines discuss the issue of how different emphasis effects are perceived by viewers in visualizations, or consider issues of equivalence for perceived emphasis. Therefore, in the evaluations described next, we set out to determine the viewer's perception of visual prominence, and the effectiveness of a variety of emphasis effects at a wide range of intensity levels.

3 GENERAL METHODS

Data visualizations are used both to reveal patterns in data through exploration, and to communicate specific information to a viewer. When building visualizations for communication, a designer may need to draw a user's attention to a specific data point in order to better reveal the narrative focus of the visualization, and this can effectively be done by increasing the perceptual difference in the visual variables of the underlying data.

In the following two studies, we experimentally evaluate how specific emphasis effects are experienced by a viewer. Our first study was designed to determine the baseline visual prominence of eight levels of three emphasis effects using different visual variables (blur/focus, colour, and size). In simple scatterplot visualizations, we visually emphasized one data element, and gathered eye-tracking data, mouse clicks, and subjective ratings of visual prominence. Our second study built on the first; it used a similar paradigm but increased the complexity of the visualizations by using subset of the MASSVIS dataset - a repository of static data visualizations obtained from a variety of publicly-available online sources intended for a wide audience [7].

Our theoretical starting point for these studies was the mathematical framework of emphasis effects in data visualizations developed by Hall *et al.*, where visually diverse emphasis effects can be linked through the idea of visual prominence compared to background elements [22]. Our first study extends this previous work to determine the visual prominence of emphasis effects through eye-tracking metrics, click data, and user's subjective ratings. Using eye movement data makes it possible to examine which areas of a visualization viewers attend to and how their attention can be guided by applying emphasis effects. Combining eye tracking, interaction logs and subjective methods allowed us to collect a more diverse set of data, allowing us to analyze how participants' actions were guided by their perception of the different effects. This rich data allows us to better understand how users perceive commonly used effects that designers can use to emphasize a particular element in a visualization.

3.1 Emphasis Effects and Levels

Many modern visualization software and libraries utilize a wide range of emphasis techniques. For example, Chart.js, a commonly-used visualization library for the web, increases a mark's size when a user clicks on it to generate emphasis, while Tableau uses a combination of blur/focus and size to emphasize a mark. Based on an informal survey of visualization tools, we chose three visual variables for our study – colour, blur/focus, and size – that are commonly used to provide emphasis in many different contexts. (It is important to note that our goal of measuring the relative perceptibility of different effects is not strongly tied to any particular visual variable).

- *Colour.* Emphasizing an element using colour means changing the hue of the data element to be different from the standard element colour; colour is well known to "pop out" when there is adequate difference between the highlighted item and the other elements, and colour change is widely used to indicate importance.

- *Size.* Emphasizing an element using size means increasing the area of the data element such that it is bigger than other elements. Size also pops out, and is used in several visualization tools for interactive highlighting.

- *Blur/Focus.* Emphasizing an element using blur/focus means applying a blur filter (e.g., Gaussian blur) to all of the elements in the visualization except for the emphasized element (which remains sharp). This effect is therefore qualitatively different from colour and size because it affects a much larger fraction of the overall view.

For each type of emphasis effect, we chose several levels of the visual variable so that we could test the effect at different levels of magnitude (eight levels for Study 1, and three levels for Study 2). We sampled mark sizes, colour differences, and blur strength along increasing levels of difference between the target and the distractor, allowing us to compare our results for each effect and level - we term these levels of difference as 'magnitude of difference'. For some of the visual variables, the magnitude of difference range was constrained at both ends (e.g., there is a fixed range of hues between red and blue); for other variables, such as blur or size, the range was constrained only at one end (e.g., blur/focus and size start from the sharpness and size of the distractors and range up to an arbitrary upper end).

For colour, we chose eight magnitude levels using a colour difference metric that normalizes the colour space to provide a closer fit between perceptual and geometric differences between colours [48]. ΔE is a metric devised to understand and measure how the human eye perceives colour difference, where a difference of 2.3 is roughly equal to one Just Noticeable Difference (JND) [54]. By utilizing ΔE, we can more accurately compare a wider range of colours, utilizing all the colours of a colour space to compare differences and comparing their change in visual perception. We use the current ΔE standard, CIEDE2000 [55], as our primary colour difference metric, which has added corrections to account for lightness, chroma, and hue. For the colour levels used in the first study, we chose eight fixed colour differences (i.e., the difference between emphasized

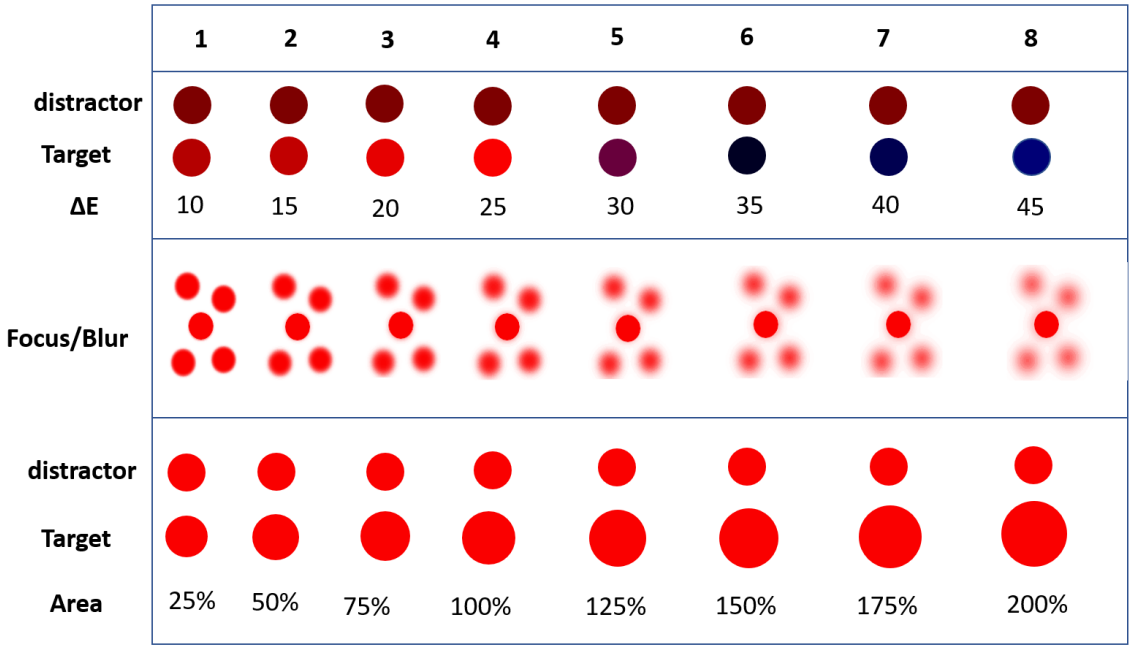


Figure 1: Visual variables and magnitude of difference levels. Rows 1-3: Colour, Blur/Focus, Size (distractors upper, targets lower).

and non-emphasized elements) ranging from ΔE 10 to ΔE 45 (see Figure 1) The empirical results we describe below confirm that the increasing ΔE values did result in increasing perceptibility of the emphasized data element (e.g., see Fig 4).

For size, we chose eight fixed size differences (difference in mark area between emphasized and non-emphasized content) from 25% to 200%. As shown in Figure 1, the size differences indicate area rather than diameter (since area is more perceptually noticeable).

For blur/focus, we chose eight different blur intensities (applied to the non-emphasized areas of a visualization) implemented using GIMP's Gaussian Blur function – with blur radius ranging from 1 to 8. Note that size and blur/focus do not have difference metrics similar to colour's ΔE; therefore, for these effects we chose levels that cover a wide range of perceived prominence for all targets. A subset of the emphasized visual targets, and their corresponding distractors are shown in Figure 1.

Our first study measured perceptibility of the three emphasis effects and the eight magnitude levels using artificial static scatterplot visualizations rendered using Chart.js[1]. Scatterplots were rendered on a white background using one-pixel gray axes. The second study used the same three effects, but only three of the eight levels; we used the visual variable to manipulate elements in visualizations taken from the MASSVIS dataset.

3.2 Apparatus

To record eye movement and interaction data we used an SMI Red-m eye tracker running at 60 Hz on a Dell 24-inch monitor (screen resolution of 1980x1080) connected to a Windows 10 PC. The viewing distance was approximately 60 cm (Figure 2). Gaze data was recorded using SMI Experiment Center and analyzed with SMI BEGAZE software. Users' heads were not fixed, but they were instructed to avoid unnecessary head movements. The experiment was conducted in an indoor laboratory with normal lighting conditions. All questionnaire data was collected through web-based forms.

[1] https://www.chartjs.org/

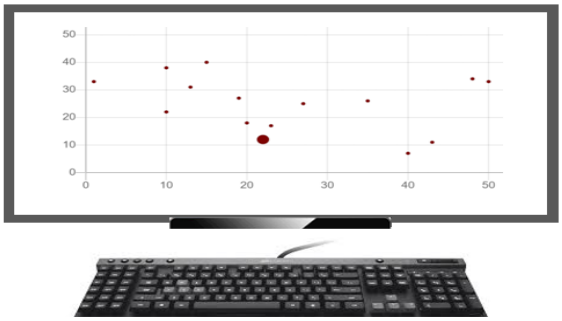


Figure 2: Setup and visualization graphic presented to participant for one trial

4 STUDY 1: ESTABLISHING A BASELINE FOR PERCEIVED EMPHASIS

4.1 Participants

Twenty-one participants were recruited from the local university pool. We excluded three participants from our analysis either for self-reporting a colour vision deficiency, or for high eye-tracking deviation; this left eighteen people (7 male, 11 female) who were given a $10 honorarium for their participation. The average age of the participants was 26 (SD 4.5). All participants continuing to the study reported normal or corrected-to-normal vision and no colour-vision deficiencies, and all were experienced with mouse-and-windows applications (10 hrs/wk). Six participants reported previous experience with information visualizations from previous university courses.

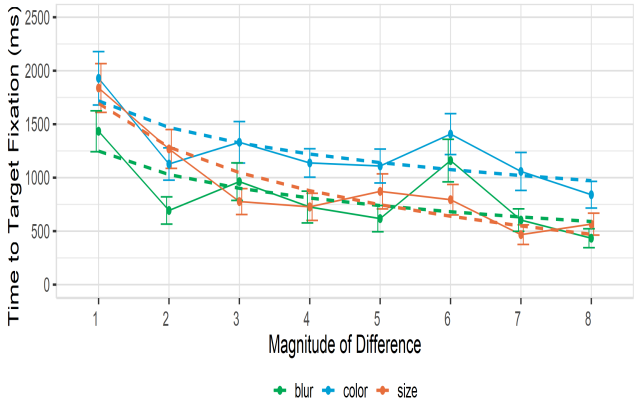


Figure 3: Time to Target Fixation. Empirical means (solid lines) and log curve (dashed lines). R^2 values for logarithmic curves: blur = 0.45, colour =0.60, and size = 0.87.

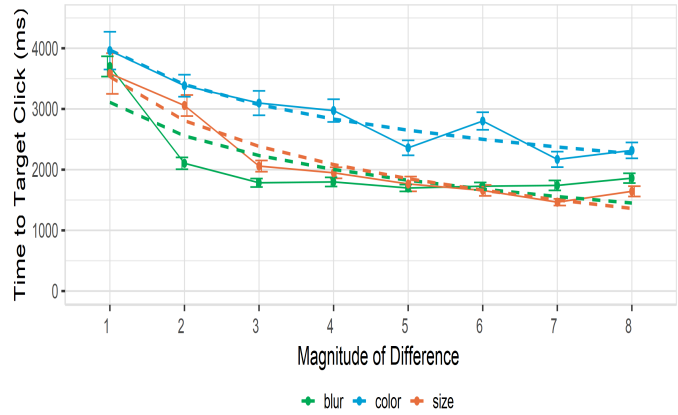


Figure 4: Time to Target Click. Empirical means (solid lines) and log curve (dashed lines). R^2 values for logarithmic curves: blur = 0.68, colour =0.90, and size = 0.96.

4.2 Study Procedure

Participants completed informed consent forms and demographic questionnaires. Participants then completed a colour vision test: we checked for colour blindness using ten of the Ishihara test plates [31]. Next, we used the five-point calibration procedure from the SMI experimental suite to calibrate the eye tracker. Once the eyetracker calibration step was completed, participants carried out a series of trials with our scatterplot visualizations. The instructions given to participants were to visually explore each visualization and click on the element they felt was most emphasized.

Participants were presented with an order-balanced presentation of the visual stimuli. Each visualization contained one target mark (an emphasized stimuli) and twenty randomly-placed distractor marks, avoiding overlaps. While spatial distance between marks can influence colour difference perceptions [9], we elected to construct our scatterplots with variable element spacing to increase the visual complexity of the stimuli for increased ecological validity. The three emphasis effect types were presented at their 8 magnitude-of-difference levels, and each emphasis level was presented 5 times. Each target maintained the same appearance for each of the 5 trials of the level. The monitor was blanked after each trial (after the participants clicked on an element) and the study software then asked the participant to rate the perceived visual prominence of the target mark, on a 1-7 scale.

4.3 Study Design and Analysis

To compare the relative perceptibility of the three emphasis effects, we needed to determine the differences between the visual variables and the magnitudes of the effects. We used an analysis-of-variance approach to explore these differences: in particular, a repeated-measures within-participants design, with factors Emphasis Effect (blur/focus, colour, size) and Magnitude of Difference (levels 1-8). Dependent measures were: time to eye fixation on the target, time to the user's mouse click on the target, the user's total fixation time on the target, and the user's subjective rating of the target's emphasis.

We then use our analysis results to explore relative differences between the emphasis effects, and fit curves to our empirical data in order to develop an initial equivalence model.

5 STUDY 1 RESULTS

5.1 Time to Target Fixation, Time to Target Click, and Fixation Times

We analyzed differences between emphasis effect and magnitude of difference on participant's time to target fixation, target click, and fixation time in an Area of Interest (AOI) surrounding the emphasized visual target. We report effect sizes for significant RM-ANOVA results as general eta-squared η^2 (considering .01 small, .06 medium, and >.14 large [40]). For all follow up tests involving multiple comparisons, the Holm correction was used.

Time to Target Fixation. RM-ANOVA showed significant main effects of *Emphasis Effect* ($F_{2,34} = 17.73, p < 0.001, \eta^2 = 0.07$), and *Magnitude of Difference* ($F_{7,119} = 8.23, p < 0.001, \eta^2 = 0.19$) on time to target fixation. RM-ANOVA found no interaction between *Emphasis Effect* $\times$ *Magnitude of Difference*. These data are shown in Fig 3. Overall, across all Magnitude of Differences, participants fixated on targets fastest in the Blur/Focus condition (828 ms), followed by Size (913 ms) and Colour (1242 ms). Post-hoc t-tests showed significant ($p < 0.01$) differences between each emphasis pair except for Blur $\rightarrow$ Size. Across all emphasis effects, time to target fixation was the fastest at a magnitude of difference of 8 (613 ms), and the slowest at a difference of 1 (1733 ms). A similar post-hoc t-test was applied for pairs of magnitude of differences and showed a significant difference for 1 $\rightarrow$ 2-8 (p<0.001), and 3 $\rightarrow$ 7 (p<0.05).

Time to Target Click. RM-ANOVA showed significant main effects of *Emphasis Effect* ($F_{2,34} = 40.99, p < 0.001, \eta^2 = 0.24$), and *Magnitude of Difference* ($F_{7,119} = 56.45, p < 0.001, \eta^2 = 0.45$) on target click. These data are illustrated in Fig 4. Participants clicked on focused targets fastest in the Blur condition (2051 ms), followed by size (2141 ms) and colour (2882 ms). Holm-corrected post-hoc t-tests showed significant ($p < 0.01$) differences between each emphasis pair except for Blur $\rightarrow$ Size. Averaged across all emphasis effects, time to target click was fastest at Magnitude 7 (1791 ms), and the slowest at a Magnitude of 1 (3748 ms).

Target Fixation Time. RM-ANOVA showed a significant main effect of *Magnitude of Difference* ($F_{7,119} = 6.65, p < 0.001, \eta^2 = 0.12$) on fixation time, but no difference between *Emphasis Effects* ($F_{2,34} = 2.76, p = 0.08$). Averaged across magnitude, total fixation time was similar among the emphasis effects (1990ms for size; 2260 ms for both size and colour). Averaged across all effects, a Magnitude of 7 had fixation time of 2470ms, while a Magnitude of 1 had the least time at 1913 ms. Post-hoc t-tests showed significant

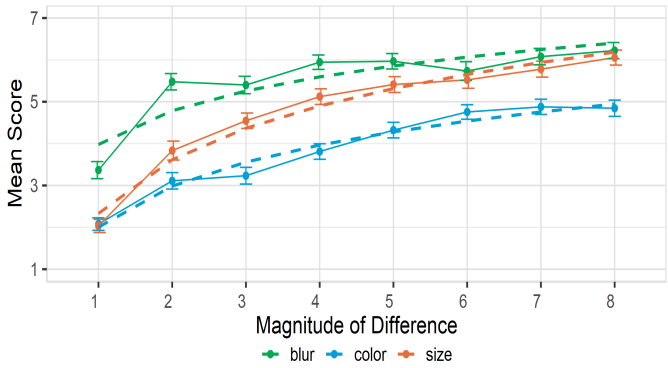


Figure 5: Perceived Prominence of Emphasis Effects. Empirical means (solid lines) and log curve (dashed lines). R^2 values for logarithmic curves: blur = 0.80, colour =0.97, and size = 0.98.

(all $p < 0.05$) differences on difference pairs $1 \rightarrow 5$, $1 \rightarrow 7$, $3 \rightarrow 7$, and $7 \rightarrow 8$.

5.2 Subjective Perception of Visual Prominence

After the presentation of each visualization, participants were asked to rate how visually prominent the emphasized data point appeared to them. Mean response scores are shown in Figure 5. We used the Aligned Rank Transform [67] with the ARTool package in R to enable analysis of the subjective prominence responses using RM-ANOVA. For subjective ratings of perceived emphasis there were main effects of *Emphasis Effect* ($F_{2,408} = 56.38, p < 0.001$), *Magnitude of Difference* ($F_{7,408} = 24.98, p < 0.001$), but no interaction between *Effect x Magnitude* ($F_{14,408} = 1.43, p < 0.13$). Results from these analyses follows those from Time to Mouse Click, in which sharp objects in the Focus/Blur emphasis condition were, on average, perceived as most visually prominent, followed by Size and Colour - with an increasing perceived visual prominence as the difference between emphasized and non-emphasized data points increase.

5.3 Participant Preferences and Comments

At the end of the study session we asked participants to state which emphasis effect they felt was the most visually prominent, least prominent, and to provide further comments on their responses. Overall, focus/blur was found to be perceived as most prominent, with seven participants overall rating blur/focus as most prominent, six rated size as most prominent, while four stated colour as most prominent. One participant stated that none seemed to stand out as most prominent. Participant comments for the three emphasis effects reflect the empirical findings, favouring blur/focus. On preferring focus/blur, one participant reported, "[In focus/blur] other data points were very blurry and hard to distinguish so the clear one stood out more that if the colour were different or the size were different (i.e. could only focus on the emphasized one, compared to the other types where you could still view the non-emphasized points)". Another commented on blur/focus being preferred as "[blur/focus] clearly hid the other circles". Participants that favoured size emphasis reported that size may be easier for quick comparisons; one participant remarked "It is easier for the eye to visualize a bigger/smaller size in comparison to other dots vs trying to see a colour difference of a similar size dot".

5.4 Building an Initial Equivalence Model

We used the raw data from Study 1 to build initial predictive models of time to target fixation, time to click, and subjective rating of emphasis – and although more data will be needed to refine the predictions, we are able to capture some of the main differences between the three emphasis effects that we examined. Our models are simple functions fit to the raw empirical data; we use logarithmic functions they are commonly used to describe human performance in signal-detection and perceptual studies [66]. We fit the functions to the data using R (lm(mean $\sim$ log(magnitude of difference)); we could then use R's 'predict' function to get predicted values. The fitted logarithmic curves for time to target fixation, time to click, and subjective ratings are shown in Figures 3, 4, and 5. Captions for these figures also state the R^2 values for the accuracy of the fitted functions to the data: for time to fixation the curve was only moderately accurate, but for time to click and subjective ratings, the accuracy was much higher.

The logarithmic curves provide a simple model that allows investigation of equivalence between the three effects. For all three measures, the models allow us to observe some main features of the relationships: first, colour is consistently less perceptible than the other two effects, both in terms of performance data and subjective ratings; second, size and blur/focus are very similar at level 3 and above of both performance measures, but at levels 1 and 2, size is somewhat weaker; third, size and blur/focus are more clearly separated in subjective ratings, with clear differences up to level 5.

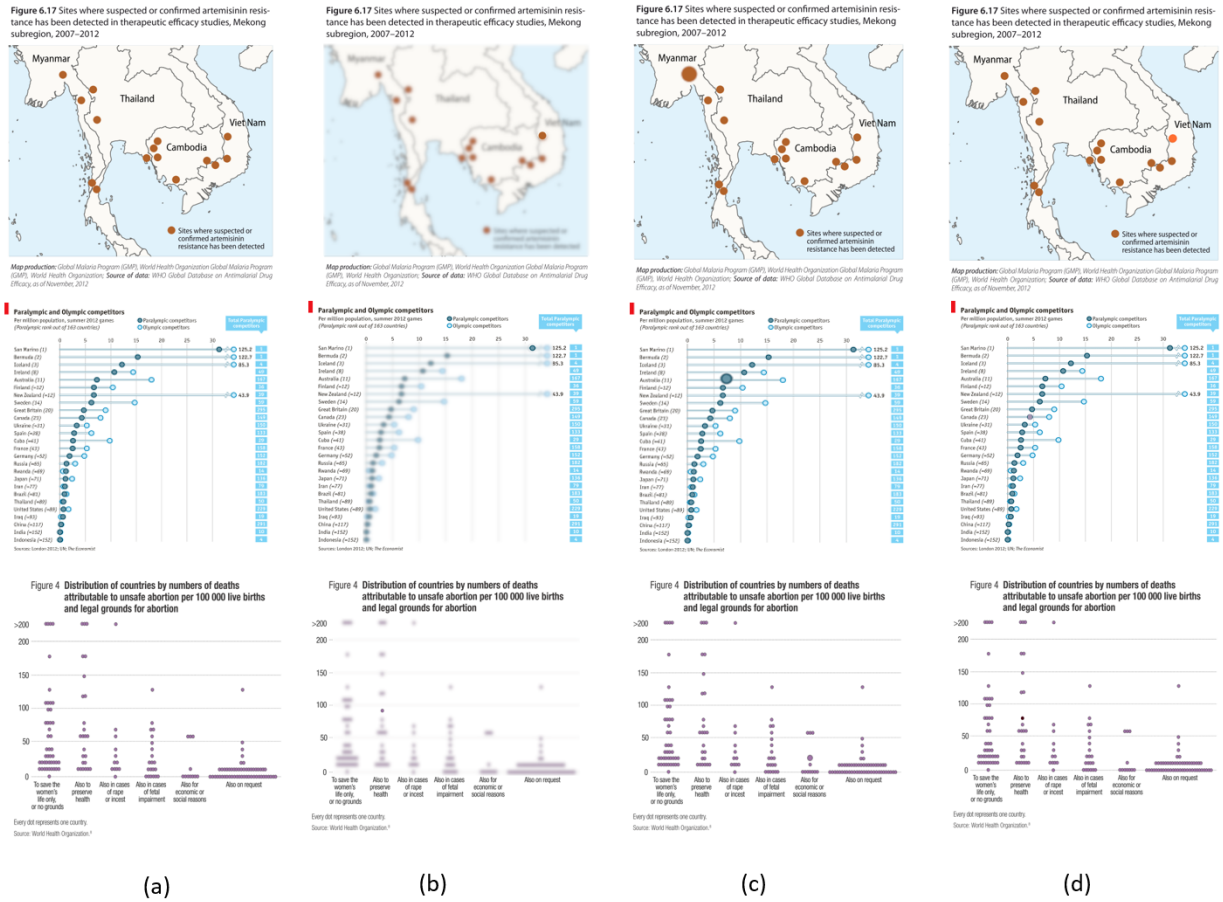
These models, once validated, can allow simple calculation of equivalence between effects. As an example of how the calculation works, consider a scenario where a designer needs to change from a blur/focus emphasis effect to one that uses colour; interpolation of the curves of Figure 4 indicate that to translate the perceived emphasis of level 1 of blur/focus, a designer would need to use a colour effect of approximately level 7. However, before we can consider using the models for equivalence, we need to verify that they are robust enough to work with other visualizations. We do this by predicting data from Study 2 with the models developed from Study 1 data, as described below.

6 STUDY 2: PERCEPTION OF EMPHASIS IN COMPLEX VISUALIZATIONS

In contrast to the scatterplots used in Study 1, many visualizations include other visual factors such as background graphics, labels, titles, annotations and other embellishments that may affect how a user's attention is guided and ultimately how an emphasis effect is perceived. Therefore, we need to understand how users perceive emphasis effects in more complex visualizations. We designed our study following a similar method to Study 1, but evaluated emphasis effects in complex, real-world visualization graphics from the MASSVIS database [7].

6.1 Image Data

As the emphasis effects we are studying are not particularly targeted towards a specific visualization type, we chose the MASSVIS database [7] as the source for image data. The dataset contains 5000 static data visualizations obtained from a variety of online sources, and real-world applications and are targeted to a broad audience - as such, making it a popular choice to understand how users in general understand data visualizations. We selected a subset of 16 visualizations from the dataset covering a variety of visualization types, including maps, and scatter plots. Each of the 16 visualizations had one emphasis effect applied at a time (Fig 6), which were then used to evaluate how users perceive the different emphasis effects in our experiment. We included baseline (no emphasis effect applied) graphics to investigate and compare whether users already perceived an area of the graphic as emphasized.



(a) (b) (c) (d)

Figure 6: Example stimulus display for study 2 (a) baseline, (b) focus/blur, (c) size/area, (d) colour; Rows 1-2 at level magnitude of difference 4. Row 3 shows magnitude of difference 7.

6.2 Emphasized Stimuli

Following study 1 baseline results, we decided to sample mark sizes, colour differences, and blur strength along three uniform steps (1, 4, and 7), giving us the performance range we saw in the baseline results, and allowing us to compare our results for each effect. Example graphics with an emphasized data point are illustrated in Fig 6.

6.3 Experimental Design and Procedure

The experiment followed a similar procedure to that of Study 1; After providing informed consent and going through the eye-tracker calibration, participants were instructed to explore each visualization and to click on the area they felt was most emphasized. Participants were presented an order-balanced presentation of the visual stimuli. Baseline graphics had no emphasized marks, test graphics contained one randomly-placed test mark in the graphic. After each stimulus presentation, participants were asked to rate the perceived visual prominence of the emphasized point they selected.

6.4 Participants

Twenty four participants were recruited from the local university pool. We excluded four participants from our analysis for high eye-tracking deviation, or failure to follow experiment instructions. The remaining twenty participants (9 male, 9 female, 2 non-binary) were given a $10 honorarium for their participation. The average age of the participants was 26 (SD 6.02) and all reported normal or corrected-to-normal vision and no colour-vision deficiencies; all were experienced with mouse-and-windows applications (10 hrs/wk), and 6 had previous visualization experience. We used the same experimental setup described in Study 1.

6.5 Study Design and Analysis

The study used a repeated-measures within-participants design, with factors Emphasis Effect (blur/focus, colour, and size) and Magnitude of Difference (Levels 1, 4, and 7 from Study 1). We used the same four dependent variables: time to fixate on target, time to target click, total fixation time, and subjective rating of perceived emphasis.

7 STUDY 2 RESULTS:

We again analyzed emphasis effect and magnitude of difference on participant's time to target fixation, target click, and target fixation time. We again report effect sizes as general eta-squared η^2, and use Holm correction for followup tests.

Time to Target Fixation. RM-ANOVA found no main effect of *Emphasis Effect* ($F_{2,38} = 1.78, p = 0.18$) on time to target fixation, but did find an effect of *Magnitude of Difference* ($F_{2,38} = 3.80, p < 0.01, \eta^2 = 0.31$). There was an *Emphasis Effect* $\times$ *Magnitude of Difference* interaction ($F_{4,76} = 3.09, p < 0.01, \eta^2 = 0.05$). These data are shown in Fig 7. Post-hoc t-tests showed significant ($p < 0.01$) differences between each magnitude-of-difference pair. Averaged

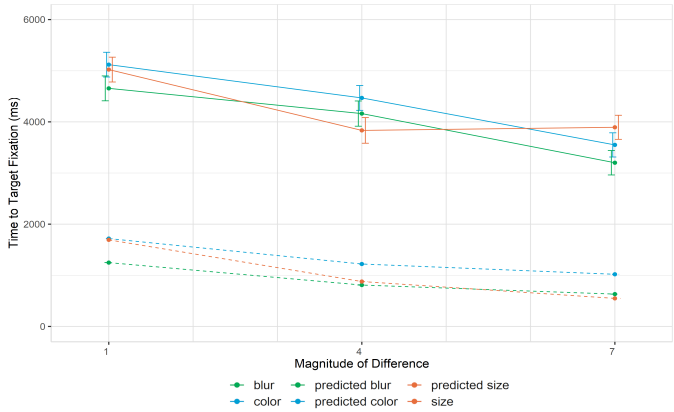


Figure 7: Time to Target Fixation in Complex Visualizations. Empirical means (solid lines) and predicted means (dashed lines).

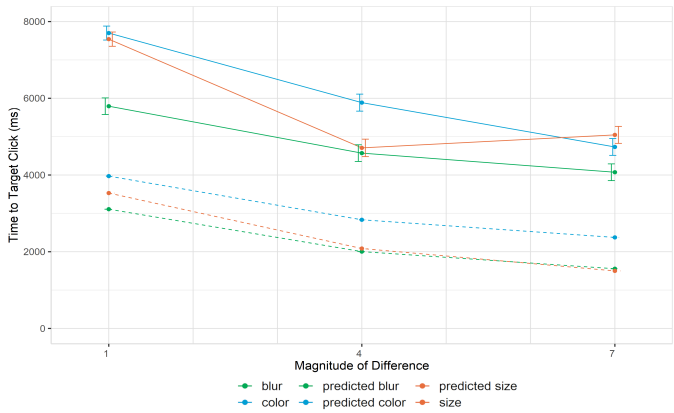


Figure 8: Time to Target Click in Complex Visualizations. Empirical means (solid lines) and predicted means (dashed lines).

across all emphasis effects, time to click on an emphasized data point was fastest at an Magnitude of Difference of 7 (3548 ms), and the slowest at a Magnitude of 1 (4932 ms).

Time to Target Click. RM-ANOVA showed significant main effects of *Emphasis Effect* ($F_{2,38} = 41.18, p < 0.01, \eta^2 = 0.32$) and *Magnitude of Difference* ($F_{2,38} = 58.04, p < 0.01, \eta^2 = 0.62$) on target click time, and an *Emphasis Effect* $\times$ *Magnitude of Difference* interaction ($F_{4,76} = 14.64, p < 0.01, \eta^2 = 0.15$). These data are illustrated in Fig 8. Similar to Study 1, focused targets in the blur condition were clicked on fastest (4812 ms), followed by Size (5787 ms) and Colour (3106 ms). Holm-corrected post-hoc t-tests showed significant ($p < 0.01$) differences between each emphasis pair. Averaged across all emphasis effects, time to click on an emphasized data point was fastest at an magnitude of difference of 7 (4616 ms), and the slowest at a Difference of 1 (7012 ms).

Target Fixation Time. RM-ANOVA showed a significant main effect of *Emphasis Effect* ($F_{2,38} = 3.41, p < 0.01, \eta^2 = 0.04$) and *Magnitude of Difference* ($F_{2,38} = 15.08, p < 0.01, \eta^2 = 0.22$) on total fixation time, and a *Emphasis Effect* $\times$ *Magnitude of Difference* ($F_{4,76} = 4.30, p < 0.01, \eta^2 = 0.08$) interaction. Averaged across magnitude of differences, fixation time for blur/focus was 1369ms and 1240 ms for both Size and Colour. Averaged across all effects, a magnitude of difference of 7 gathered the most attention with a

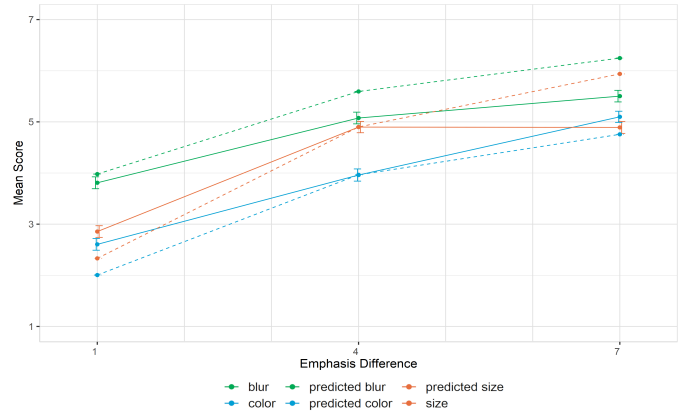


Figure 9: Perceived Prominence of Emphasis Effects in Complex Visualizations. Empirical means (solid lines) and predicted means (dashed lines).

fixation time of 1494ms, while a difference of 1 had the least fixation time at 1080 ms. Post-hoc t-tests showed significant (all $p < 0.01$) differences for *Magnitude of Difference* but no difference among *Emphasis Effects*.

7.1 Subjective Perception of Visual Prominence in Complex Visualizations

After the presentation of each visualization, participants were asked to rate the visual prominence of the emphasized data point. Mean response scores are shown in Fig 9. We used the Aligned Rank Transform [67] with the ARTool package in R to enable analysis of the subjective responses using RM-ANOVA. RM-ANOVA showed there were main effects of *Emphasis Effect* ($F_{2,171} = 16.05, p < 0.001$), *Magnitude of Difference* ($F_{2,171} = 60.00, p < 0.001$), and an interaction between *Emphasis Effect x Magnitude of Difference* ($F_{4,171} = 2.57, p = 0.03$). Results from these analyses are shown in Figure 9 and follow those from Time first to Mouse Click, in which sharp objects in the Focus/Blur effect were perceived as most visually prominent, followed by Size and Colour - with an increasing perceived visual prominence as the difference increased.

7.2 Participant Preferences and Comments

After completing the study, participants provided their preferences and general comments on the emphasis effects they identified. Participant comments echoed our other findings. Participants made several comments on how the focus/blur emphasis effect helped them to rapidly identify content. On preferring blur/focus, one participant stated "It [emphasized point] just popped out more than the rest, provided more contrast". Another participant reflected "Because it [blur/focus] didn't allow me to see the others, I focused all my attention to the point that was not blurry". One participant favoured size, stating "[size] always drew my eye immediately".

When asked whether there were other areas of a visualization that got their attention, one participant remarked "The titles and information, I was trying to read them and see if that would have helped somehow to identify what was emphasized", while another participant stated "I occasionally looked at the titles to see what the information was representing".

7.3 Consistency Across Studies and Validation of the Model

We used the models built from Study 1 data to predict the data gathered for each effect and magnitude used in Study 2, and then

compared the empirical data points to the predicted values (predictions are shown in Figures 7, 8 and 9 as dotted lines). Although the absolute values of the predictions are lower than the true values, the predictions do capture many of the characteristics of the Study 2 results, as discussed below. We tested the correlation between the predicted and empirical values: for time to target fixation, the correlation was 0.82 (R^2=0.87); for time to target click, correlation was 0.92 (R^2 = 0.94); for subjective ratings, correlation was 0.96 (R^2 = 0.96).

If equivalence models are to be useful, the perceptibility of emphasis must be reasonably reliable across different visualization situations. Our two studies involve two visual settings: plain scatterplots in Study 1, and more complex visualizations in Study 2 (with background graphics and colours, text, and multiple visual styles). Nevertheless, there are several similarities between the two sets of results (as indicated by the very strong correlation scores). In both studies, the colour effect was less perceivable (higher time to target fixation and target click time, and lower subjective ratings); however, the earlier difference between colour and size at the highest magnitude is now gone. As in study 1, the blur/focus effect is again consistently more perceivable (and is rated as more prominent). Also as in Study 1, there was a similar improvement in performance as the magnitude of the effect increases; there was less of a clear logarithmic curve for some of the emphasis effects (although this would be less apparent with the three magnitude levels used in Study 2).

The most obvious difference between the predicted and real values is that times for both fixation and clicking were substantially higher with the MASSVIS visualizations. However, this was an expected difference because of the additional visual information available in each image – and because all emphasis effects were affected similarly, any equivalence calculations using the model will be unaffected.

The subjective responses were particularly well predicted by the Study 1 model (see Figure 5), with the predicted points being accurate both in terms of absolute score and the relationship between the effects. This is a particularly valuable finding, because as discussed below, it may be that the user's perception of emphasis is a more important measure for designers than the user's gaze patterns or click behaviour.

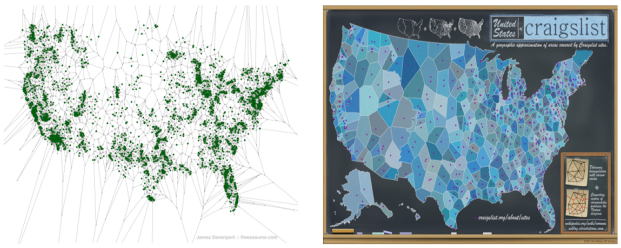
The main point where the predictions were inaccurate – both for performance data and for subjective ratings – was the perceptibility of the size effect at level 7. After reviewing the stimuli for this condition, there are two possible reasons for the empirical results being different from predicted values. First, two of the visualizations (see Figure 10) contained a large number of data points and many visual elements overall, and previous research has shown that it is more difficult to recognize objects in a cluttered environment due to *visual crowding*, which can create a visual-perception bottleneck [39]. Second, when data points in these visualizations are dense, composite blobs with several overlapping points create marks that are larger than the default size. Although none of our target elements were in or beside these blobs, the presence of varying-size elements in the visualization may have forced participants to do a more careful visual search.

This anomaly with the size effect points to another useful aspect of having a predictive model, however: that is, the identification of empirical results that are not as expected and that may need to be investigated further.

8 Discussion

8.1 Summary of Findings

We investigated how users perceive colour, size, and blur/focus when used as emphasis effects in both basic scatterplots and more complex visualizations. Our evaluations provide several findings:



(a) (b)

Figure 10: Study 2 graphics. Graphic (a) contained a larger number of data points with composite blobs, leading to visual crowding. Graphic (b) has multiple visual elements (shapes, colours, and text), reducing the effect of size emphasis on a data point.

- Across both studies, blur/focus led to fastest target fixation and target click, and was rated highest in terms of visual prominence by participants; size also led to fast performance and high ratings of visual prominence at higher magnitude levels (with one exception); colour led to the slowest performance and lowest ratings for prominence.
- Across both studies, increasing the magnitude of the effect consistently increased visual prominence (again, with the same one exception).
- A predictive model based on logarithmic curves fit the Study 1 data well, and was reasonably accurate at predicting emphasis in Study 2 (particularly the subjective ratings).

In the following sections, we consider possible explanations for these results, look at how our findings and models can be used to assist designers in building visualizations with emphasis, and discuss limitations and directions for future research in this area.

8.2 Explanation of Results

Differences Between the Emphasis Effects

We saw consistent differences in fixation time, click time, and subjective ratings for our three emphasis effects, and the reasons for these differences arise from each technique's fundamental properties (as introduced earlier). First, blur/focus is an effect that manipulates the entire visualization except for the emphasized data element, and so has advantages over single-element techniques like colour and size. In particular, the blur effect guarantees that there will be no inadvertent competing visual stimuli that could slow the user's visual search (as happened with size at level 6 in Study 2), because all other elements are blurred. Second, the relative advantage for size over colour in our studies can be explained by the inherent limit on colour difference (i.e., there is a maximum difference between any two colours) whereas size difference has an unlimited upper end. The study results show that our range of magnitudes for size was larger than our range for colours – which points to the need for a better understanding of equivalences between effects.

Effectiveness of the Predictive Model

The model built from Study 1 data provided accurate predictions of the results in Study 2 (R^2 values of 0.87, 0.94, and 0.96 for fixation time, click time, and subjective rating), and the model correctly represented the overall relationships between the emphasis effects and the changes expected with increasing magnitude level. The success of the predictive model shows that perception of emphasis is

consistent between our two experimental settings – plain scatterplots in Study 1, and realistic scatterplots with other visual features in Study 2. In addition, the model was a useful tool for identifying results that need further exploration, including the greater overall response times for the MASSVIS dataset, and the anomalous performance of size at level 7. Of course, these anomalies can only be spotted when there are empirical results to compare to the model, but it is likely that the model will be used in concert with empirical testing until it matures with the addition of more data in different settings.

The Size-at-Level-Seven Anomaly

As described above, the size effect at level seven was less prominent than expected, with two possible reasons: visual crowding from other elements in the visualizations, and inadvertent size variance from overlapping data points (see Figure 10). This result clearly indicates that there can be emergent properties in real-world visualizations that interfere with the user's perception of emphasis, and thus a planned emphasis effect must be considered in light of other visual elements. These real-world interactions are another motivation to have equivalence metrics, so that designers can switch from one emphasis effect to another (and preserve the prominence of the emphasized element) when interference is discovered.

While our setup of the presentation of visual stimuli ensured that distractor marks and stimuli would not overlap, changes in the distance between distractors and the emphasized elements may affect their noticeability. Because effects of visual crowding occur with a wide range of objects, colours and shapes [65], this phenomena may have affected other individual data points as well; but our explicit decision to not control the distance between points in Study 1 means that our results provide a more valid representation of the challenges faced by designers when emphasizing elements in a crowded visualization. As noted above, global effects such as blur/focus are less affected by visual crowding, as blurring non-targets partially eliminates them from a user's view, leaving only the focused element available. We note that it is possible to quantify the overall degree of visual complexity in an image, and in future work this could be added to our models as a factor (i.e., further studies could examine perceived emphasis at different levels of crowding).

8.3 Implications for Design

Our findings are applicable in a number of different visualization contexts. Visualization designers often need to draw a user's attention to important data points; our studies improve understanding of how visual cues are detected as emphasis effects and offer insights to their perceived visual prominence. While the current set of visual stimuli examined was relatively small, we intend to explore further visual variables in future studies.

A first design implication is that global visual effects such as blur/focus can achieve a high perceived visual prominence, and is relatively unaffected by a visualization's background. Perceived differences for other variables such as colour and size can be affected by the non-target elements, but by blurring the non-target objects in a visualization, the focused item is less likely to be affected by visual crowding. In visualizations with a large number of objects (such as different colours and shapes), blurring non-targets may achieve the highest noticeability – however, blur/focus cannot be used in visualizations where the user needs to inspect elements that are not emphasized.

Second, predictive models of perceived visual prominence can be valuable tools for designers. Although our model is still only a first step, it was already able to predict the results of Study 2, and can already be used to consider the equivalence between perception of the three effects that we tested. (We note that the model should not be used to calculate exact conversion factors between the effects, but rather to understand general relationships and approximate relative magnitudes). As further studies are carried out and more data is added, models like ours can become resources for designers that can accelerate the design of a narrative visualization. It is interesting that the model was most accurate at predicting people's subjective ratings of prominence, which raises the question of which metric is most important. It may be that subjective perception is a better measure for a model, because when a designer adds emphasis to a visualization, they typically want the viewer to know that the item is being emphasized – that is, what the viewer thinks is being emphasized is possibly more important than what their eye is drawn to first.

A third design consideration is for designers utilizing colour as a way to emphasize certain data points. It should be noted that a subset of users suffer from various genetic conditions which cause atypical forms of colour perception - in such cases, a different emphasis effect may be more appropriate. Designers may wish to use our metrics and results to evaluate the effectiveness of a different visual effect to achieve the same perceived importance. Our future work intends to evaluate the use of various visual cues for emphasis effects and compare the sets for individuals with normal vision and users with a vision deficiency.

Beyond visualization, our findings can also be applicable in other domains. For example, interface designers may wish to use our results as a way of devising methods of providing visual feedback. For instance, visual feedback during "find" tasks in different software software such as web browsers and pdf readers varies - with some software opting for colour highlighting an item when found, while others increase its size or use a combination of both. To effectively guide a user's attention to an item, designers can use perceived visual prominence as a method to evaluate and compare different visual effects.

8.4 Generalizability, Limitations, and Future Work

Our studies tested a limited range of visualizations (i.e., scatterplot presentations), so the application of our results should be limited to that type; in Study 2, however, we did test a wide variety of different visual styles taken from real-world examples, and so we believe that our findings will be robust across a range of real scatterplots. In future work, we plan to extend our work to other types of visualizations and other real-world scenarios. We also tested only a single emphasized data point, and an opportunity to extend to our work is to investigate visualizations that emphasize multiple points. Multiple points of emphasis also provides us with another opportunity to test the predictions of the model – that is, if two data elements are emphasized with different effects that our model predicts should be equally prominent, which will the user fixate on first? (We note that this kind of comparison is only possible with single-element effects such as size and colour).

The difference levels for the visual variables tested in our experiments are intended to be generalizable for the design of emphasized elements in typical visualizations. However, although we tested a wide range of magnitude of differences, it is possible that our findings are influenced by the magnitude of differences we tested (as noted above in terms of the range of difference that is possible with each visual variable). We also plan to carry out studies that look at how magnitude of emphasis is affected by clutter and by other mappings of visual variables to data variables.

Other factors in generalization should be considered as well. Colour perception models rely on a simplified model of the world that assume perfect viewing conditions. While this assumption is necessary for understanding the visual system, complexities of the real world such as the viewing environment [37], lighting conditions [8, 47], and display device [52] may affect visual perception. Our experimental viewing conditions were controlled and remained stable throughout the studies, however, future work could extend these results to larger user samples and different viewing conditions, using crowd-sourcing methods [25].

There are several additional opportunities for extending our findings. We explored emphasis effects with static visual variables (time-invariant in terms of Hall et al.'s framework [22]) but there are many other effects that could be tested, including depth, transparency, or shape. Additionally, future research should investigate time-variant emphasis effects with dynamic visual variables such as flicker or motion and extend our results to interactive visualizations.

We evaluated our emphasis effects based on empirical metrics such as time to target fixation, and time to mouse click. There are other ways emphasis effects can be evaluated. For instance, the MASSVIS dataset contains a comprehensive set of user attention maps on the visualizations [7]. We intend to analyze viewer's attention maps on the visualizations, comparing the visualization's attention maps with and without an emphasis effect applied.

Finally, we elected to use the CIE2000 as it is commonly used in visualization and has been methodologically validated in past studies [28, 60]. Future work may consider the use of other colour difference models or colour spaces, such as CIECAM02 [44]. We anticipate investigating a number of different colour spaces will result in more accurate models of colour difference perceptions for visualization design.

9 Conclusion

Emphasis is an essential component of InfoVis, and is used by designers to draw a user's attention or to indicate importance. However, it is difficult for designers to know how different emphasis effects will compare and what level of one effect is equivalent to what level of another when designing visualizations. We carried out two user studies to evaluate the visual prominence of three emphasis effects (blur/focus, colour, and size) at various strength levels, and developed a predictive model that can indicate equivalence between effects. Results from our two studies provide the beginnings of an empirical foundation for understanding how visual effects operate and are experienced by viewers when used for emphasis in visualizations, and provide new information for designers who want to control how emphasis effects will be perceived by users.

Acknowledgments

Removed for anonymous review.

References

[1] E. Awh, A. V. Belopolsky, and J. Theeuwes. Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in cognitive sciences*, 16(8):437–443, 2012.

[2] S. I. Becker. The role of target–distractor relationships in guiding attention and the eyes in visual search. *Journal of Experimental Psychology: General*, 139(2):247, 2010.

[3] R. V. d. Berg, F. W. Cornelissen, and J. B. T. M. Roerdink. Perceptual dependencies in information visualization assessed by complex visual search. *ACM Trans. Appl. Percept.*, 4(4):3:1–3:21, Feb. 2008. doi: 10.1145/1278760.1278763

[4] J. Bertin, W. J. Berg, and H. Wainer. *Semiology of graphics: diagrams, networks, maps*, vol. 1. University of Wisconsin press Madison, 1983.

[5] T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf, and T. Ertl. State-of-the-art of visualization for eye tracking data. In *EuroVis (STARs)*, 2014.

[6] T. Blascheck, L. MacDonald Vermeulen, J. Vermeulen, C. Perin, W. Willett, T. Ertl, and S. Carpendale. Exploration Strategies for Discovery of Interactivity in Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 2626(c):1–1, 2018. doi: 10.1109/TVCG.2018.2802520

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

[8] D. H. Brainard and B. A. Wandell. Asymmetric color matching: how color appearance depends on the illuminant. *JOSA A*, 9(9):1433–1448, 1992.

[9] A. Brychtová and A. Çöltekin. The effect of spatial distance on the discriminability of colors in maps. *Cartography and Geographic Information Science*, 44(3):229–245, 2017.

[10] M. Burch, N. Konevtsova, J. Heinrich, M. Hoeferlin, and D. Weiskopf. Evaluation of traditional, orthogonal, and radial tree diagrams by an eye tracking study. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2440–2448, 2011.

[11] M. S. T. Carpendale and C. Montagnese. A framework for unifying presentation space. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pp. 61–70. ACM, 2001.

[12] H. Chen, W. Chen, H. Mei, Z. Liu, K. Zhou, W. Chen, W. Gu, and K.-L. Ma. Visual abstraction and exploration of multi-class scatterplots. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1683–1692, 2014.

[13] W. S. Cleveland. *Visualizing Data*. Hobart Press, 1993.

[14] W. S. Cleveland and W. S. Cleveland. A color-caused optical illusion on a statistical graph. *The American Statistician*, 37(2):101–105, 1983.

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

[16] M. A. Correll, E. C. Alexander, and M. Gleicher. Quantity estimation in visualizations of tagged text. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pp. 2697–2706. ACM, New York, NY, USA, 2013. doi: 10.1145/2470654.2481373

[17] H. Doleisch, M. Gasser, and H. Hauser. Interactive feature specification for focus+ context visualization of complex simulation data. In *VisSym*, vol. 3, pp. 239–248, 2003.

[18] J. Duncan and G. W. Humphreys. Visual search and stimulus similarity. *Psychological review*, 96(3):433, 1989.

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

[20] J. H. Goldberg and J. I. Helfman. Comparing information graphics: a critical look at eye tracking. In *Proceedings of the 3rd BELIV'10 Workshop: BEyond time and errors: novel evaLuation methods for Information Visualization*, pp. 71–78. ACM, 2010.

[21] C. Gutwin, A. Cockburn, and A. Coveney. Peripheral popout: The influence of visual angle and stimulus intensity on popout effects. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 208–219. ACM, 2017.

[22] K. W. Hall, C. Perin, P. G. Kusalik, C. Gutwin, and S. Carpendale. Formalizing emphasis in information visualization. *Comput. Graph. Forum*, 35(3):717–737, June 2016. doi: 10.1111/cgf.12936

[23] H. Hauser. Generalizing focus+ context visualization. In *Scientific visualization: The visual extraction of knowledge from data*, pp. 305–327. Springer, 2006.

[24] C. Healey and J. Enns. Attention and visual memory in visualization and computer graphics. *IEEE transactions on visualization and computer graphics*, 18(7):1170–1188, 2011.

[25] J. Heer and M. Bostock. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. *Proceedings of the 28th Annual Chi Conference on Human Factors in Computing Systems*, pp. 203–212, 2010. doi: 10.1145/1753326.1753357

[26] J. Heer, N. Kong, and M. Agrawala. 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*, pp. 1303–1312. ACM, 2009.

[27] J. Heer and G. Robertson. Animated transitions in statistical data graphics. *IEEE transactions on visualization and computer graphics*, 13(6):1240–1247, 2007.

[28] J. Heer and M. Stone. Color naming models for color selection, image editing and palette design. p. 1007, 2012. doi: 10.1145/2207676.2208547

[29] W. Huang. Using eye tracking to investigate graph layout effects. In *2007 6th International Asia-Pacific Symposium on Visualization*, pp. 97–100. IEEE, 2007.

[30] W. Huang, P. Eades, and S.-H. Hong. A graph reading behavior: Geodesic-path tendency. In *2009 IEEE Pacific Visualization Symposium*, pp. 137–144. IEEE, 2009.

[31] S. Ishihara. *Test for colour-blindness*. Kanehara Tokyo, Japan, 1987.

[32] A. Keahey. The generalized detail-in-context problem. In *Proceedings of the 1998 IEEE Symposium on Information Visualization*, INFOVIS '98, pp. 44–51. IEEE Computer Society, Washington, DC, USA, 1998.

[33] S.-H. Kim, Z. Dong, H. Xian, B. Upatising, and J. S. Yi. Does an eye tracker tell the truth about visualizations?: findings while investigating visualizations for decision making. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2421–2430, 2012.

[34] C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. In *Matters of intelligence*, pp. 115–141. Springer, 1987.

[35] R. Kosara, H. Hauser, and D. L. Gresh. An interaction view on information visualization. *State-of-the-Art Report. Proceedings of EUROGRAPHICS*, pp. 123–137, 2003.

[36] R. Kosara, S. Miksch, and H. Hauser. Focus+ context taken literally. *IEEE Computer Graphics and Applications*, 22(1):22–29, 2002.

[37] J. H. Krantz. Stimulus delivery on the web: What can be presented when calibration isn't possible. *Dimensions of Internet science*, pp. 113–130, 2001.

[38] Y. K. Leung and M. D. Apperley. A review and taxonomy of distortion-oriented presentation techniques. *ACM Trans. Comput.-Hum. Interact.*, 1(2):126–160, June 1994. doi: 10.1145/180171.180173

[39] D. M. Levi. Crowding—an essential bottleneck for object recognition: A mini-review. *Vision research*, 48(5):635–654, 2008.

[40] T. R. Levine and C. R. Hullett. Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research*, 28(4):612–625, 2002.

[41] A. MacEachren. *How Maps Work: Representation, Visualization, and Design*. Guilford Publications, 2004.

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

[43] A. Mairena, C. Gutwin, and A. Cockburn. Peripheral notifications in large displays: Effects of feature combination and task interference. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 640. ACM, 2019.

[44] N. Moroney, M. D. Fairchild, R. W. Hunt, C. Li, M. R. Luo, and T. Newman. The ciecam02 color appearance model. In *Color and Imaging Conference*, vol. 2002, pp. 23–27. Society for Imaging Science and Technology, 2002.

[45] M. Pohl, M. Schmitt, and S. Diehl. Comparing the readability of graph layouts using eyetracking and task-oriented analysis. In *Computational Aesthetics*, pp. 49–56, 2009.

[46] M. Raschke, T. Blascheck, M. Richter, T. Agapkin, and T. Ertl. Visual analysis of perceptual and cognitive processes. In *2014 International Conference on Information Visualization Theory and Applications (IVAPP)*, pp. 284–291. IEEE, 2014.

[47] P. Rizzo, A. Bierman, and M. S. Rea. Color and brightness discrimination of white leds. In *Solid State Lighting II*, vol. 4776, pp. 235–246. International Society for Optics and Photonics, 2002.

[48] A. R. Robertson. Historical development of cie recommended color difference equations. *Color Research & Application*, 15(3):167–170, 1990.

[49] R. Rosenholtz. A simple saliency model predicts a number of motion popout phenomena. *Vision research*, 39(19):3157–3163, 1999.

[50] R. Rosenholtz, J. Huang, and K. A. Ehinger. Rethinking the role of top-down attention in vision: Effects attributable to a lossy representation in peripheral vision. *Frontiers in psychology*, 3:13, 2012.

[51] B. Saket, A. Endert, and Ç. Demiralp. Task-based effectiveness of basic visualizations. *IEEE transactions on visualization and computer graphics*, 25(7):2505–2512, 2018.

[52] A. Sarkar, L. Blondé, P. L. Callet, F. Autrusseau, P. Morvan, and J. Stauder. A color matching experiment using two displays: design considerations and pilot test results. In *Conference on Colour in Graphics, Imaging, and Vision*, vol. 2010, pp. 414–422. Society for Imaging Science and Technology, 2010.

[53] E. Segel and J. Heer. Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics*, 16(6):1139–1148, 2010.

[54] G. Sharma and R. Bala. *Digital color imaging handbook*. CRC press, 2002.

[55] G. Sharma, W. Wu, and E. N. Dalal. The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, 30(1):21–30, 2005.

[56] B. G. Shortridge. Stimulus processing models from psychology: Can we use them in cartography? *The American Cartographer*, 9(2):155–167, 1982.

[57] S. Smart and D. A. Szafir. Measuring the Separability of Shape, Size, and Color in Scatterplots. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, pp. 1–14, 2019. doi: 10.1145/3290605.3300899

[58] M. Stone and L. Bartram. Alpha, contrast and the perception of visual metadata. 01 2009.

[59] H. Strobelt, D. Oelke, B. C. Kwon, T. Schreck, and H. Pfister. Guidelines for effective usage of text highlighting techniques. *IEEE transactions on visualization and computer graphics*, 22(1):489–498, 2015.

[60] D. A. Szafir, M. Stone, and M. Gleicher. Adapting color difference for design. In *Color and Imaging Conference*, vol. 2014, pp. 228–233. Society for Imaging Science and Technology, 2014.

[61] A. M. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive psychology*, 12(1):97–136, 1980.

[62] R. Veras and C. Collins. Saliency Deficit and Motion Outlier Detection in Animated Scatterplots. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, (Chi):1–12, 2019. doi: 10.1145/3290605.3300771

[63] M. Waldner, A. Karimov, and E. Gröller. Exploring visual prominence of multi-channel highlighting in visualizations. In *Proceedings of the 33rd Spring Conference on Computer Graphics*, SCCG '17, pp. 8:1–8:10. ACM, New York, NY, USA, 2017. doi: 10.1145/3154353.3154369

[64] M. Waldner, M. Le Muzic, M. Bernhard, W. Purgathofer, and I. Viola. Attractive flicker—guiding attention in dynamic narrative visualizations. *IEEE transactions on visualization and computer graphics*, 20(12):2456–2465, 2014.

[65] D. Whitney and D. M. Levi. Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in cognitive sciences*, 15(4):160–168, 2011.

[66] F. A. Wichmann and N. J. Hill. The psychometric function: I. fitting, sampling, and goodness of fit. *Perception & psychophysics*, 63(8):1293–1313, 2001.

[67] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pp. 143–146. ACM, New York, NY, USA, 2011. doi: 10.1145/1978942.1978963

[68] J. M. Wolfe. Guided search 4.0: A guided search model that does not require memory for rejected distractors. *Journal of Vision*, 1(3):349–349, 2001.

[69] J. M. Wolfe and T. S. Horowitz. What attributes guide the deployment of visual attention and how do they do it? *Nature reviews neuroscience*, 5(6):495, 2004.