Gridlines Mitigate Sine Illusion in Line Charts

Clayton Knittel
Google

Jane Awuah
Georgia Institute of Technology

Steven Franconeri
Northwestern University

Cindy Xiong Bearfield*
Georgia Institute of Technology

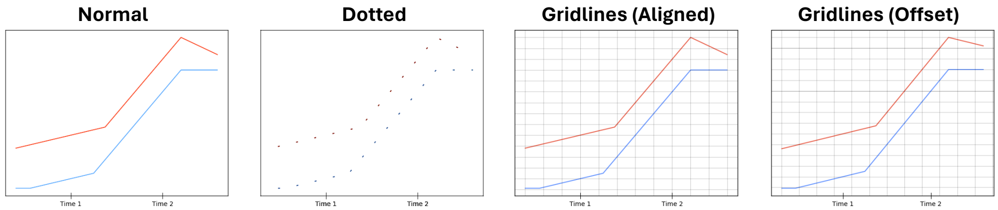

Figure 1: Four conditions: default, dotted, gridlines (aligned and offset) tested in this paper.

Abstract

The sine illusion is an underestimation of the difference between two lines when both lines have increasing slopes. We evaluate three visual manipulations on mitigating sine illusions: dotted lines, aligned gridlines, and offset gridlines via a user study. We asked participants to compare the deltas between two lines at two time points and found aligned gridlines to be the most effective in mitigating sine illusions. Using data from the user study, we produced a model that predicts the impact of the sine illusion in line charts by accounting for the ratio of the vertical distance between the two points of comparison. When the ratio is less than 50%, participants begin to be influenced by the sine illusion. This effect can be significantly exacerbated when the difference between the two deltas falls under 30%. We compared two explanations for the sine illusion based on our data: either participants were mistakenly using the perpendicular distance between the two lines to make their comparison (the perpendicular explanation), or they incorrectly relied on the length of the line segment perpendicular to the angle bisector of the bottom and top lines (the equal triangle explanation). We found the equal triangle explanation to be the more predictive model explaining participant behaviors.

Index Terms: sine illusion, gridlines, perception, bias, thresholds

1 Introduction

First formally introduced by Cleveland and McGill in 1984 [7], the sine illusion describes a perceptual error where more quickly changing pairs of lines can lead to bigger underestimates of the delta between them. This illusion is common in the real world. Multi-line line charts, stream graphs, and area charts all harbor opportunities for this bias, whenever viewers compare vertical distances between lines [28].

When comparing the deltas, the sine illusion happens as viewers rely on one of many potentially irrelevant visual cues as proxies for the actual distance between two lines For example, viewers might compare the areas of the regions near the two points of comparisons between the two lines rather than just the vertical distances, similar to the Müller-Lyer line illusion [9, 15], where viewers include the arrow tips in the length comparison to overestimate the line length, and the 'hull area' proxy presented by Jardine et al. [16], where the viewer perceives an implied hull bounded by the bars when comparing the means of bar sets. Others suggest that, instead of relying on the vertical distance between two points with a common X-axis value, viewers might rely on the orthogonal distance (or the minimal distance) which leads to non-vertical comparisons between points on the two lines that do not share a common x-axis value [28, 4], similar to doing a Deming regression [23]. We contrast these two explanations by pitting two models in competition to best model the sine illusion. We refer to the two models as the 'perpendicular model' and the 'equal triangle' model, which we describe more closely in Section 3.5.

Existing work has investigated visualization solutions to mitigate sine illusions. For example, Bu et al. developed Sinestream to reduce the effect of sine illusions in stream graphs [4]. It manipulates the geometry of a stream graph by the bottom-most curve such that the orthogonal and vertical orientations of the lines align. In this paper, we test two more alternative designs to mitigate sine illusion: dotted lines and gridlines. In the dotted lines design, we break the area surrounding the points of comparison by separating the lines into spaced dots. This increases the perceptual difficulty of viewers relying on overall dimensions (i.e., the 'hull area' proxy) and areas to make their judgment. In the gridlines design, we add gridlines to the line chart such that the vertical lines can anchor and nudge viewers to compare the vertical distance rather than orthogonal distances between the two lines. We also manipulate the ratio of vertical distance between the two lines at the two points of comparison. This allows us to identify the threshold for when sine illusion begins to significantly interfere with a viewer's ability to correctly compare the deltas between two lines.

Contribution: We contribute an experiment demonstrating the sine illusion in line charts and model the severity of the illusion as a function of the ratio of the vertical distance between the two points of comparison. We provide a perceptual foundation to inform visualization design that mitigates sine illusions, and a quantitative model describing the influence of sine illusions as a function of the deltas between the two lines at the points of comparison.

2 Background and Related Work

People can be cognitively and perceptually biased when interpreting data visualizations [27, 29, 10, 31, 11, 22]. For example, visualization readers can overly focus on salient features when making sense of data [11]. They can gravitate toward specific trends [3], colors and highlights [1], larger font sizes [13], and specific annotations [26] that are aligned with their beliefs and agendas [33, 32].

People are biased by their mental prototypes when making sense of data. For example, when participants are estimating the height of bars, they underestimate bars that are taller than they are wide and overestimate bars that are wider than they are tall, which suggests that they see a bar mark more as a prototypical square [5]. This

*e-mail: cxiong@gatech.edu

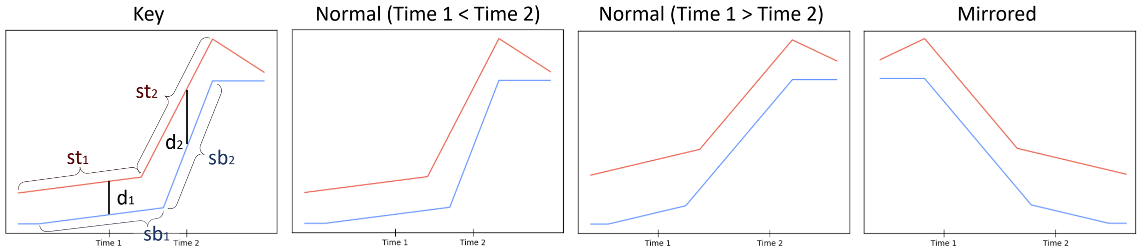

Figure 2: Left: Key explaining the line aspects we manipulated in this paper. Middle: Two examples showing conditions where the delta at Time 1 and Time 2 is greater respectively. Right: The mirrored condition of the Time 1 ¿ Time 2 example.

effect generalizes to stacked bars and dot plots. When participants were asked to visually redraw the y-position of stacked bars and dot plots, smaller values were overestimated and larger values were underestimated, with an average error of approximately 10% [17]. When they were asked to reproduce the values verbally instead of visually, participants can be anchored by round numbers such as five's and ten's [24]. When reading line charts or regressions, people also tend to see the lines as closer to 45 degrees (often referred to as 'bank to 45 degrees' [6]), as they see the lines as closer to a prototypical angle bisector [12]. These works on human perception of visualizations uncover the biasing effects of cognitive prototypes and provide insights for crafting more effective visualizations that mitigate biases. For example, Heer et al. [12] developed optimization techniques to reduce the banking to 45-degree bias by automatically identifying trends in data and generating a tailored chart scale. They migrated such a bias by aligning what people intuitively see in data with the objectively correct interpretation.

We take a similar approach to studying sine illusions. The sine illusion is a perceptual distortion when viewers misjudge the alignment and spacing between a pair of data values, both of which follow a sine-wave pattern [28]. This illusion could appear in non-sinusoidal curves as well, making it more omnipresent [14, 8, 25]. As shown in Figure 2, when comparing the distance at Time 1 (d1) and Time 2 (d2), the sine illusion can happen when the slopes of the lines increase over time. This increasing slope distorts the perceived distance between the two lines (commonly referred to as the 'delta'). The delta at an earlier point is perceived to be larger than the delta at a later point, despite the reverse being true. There has been some research aimed at mitigating its distortive effects. For example, Reimann et al. demonstrated that visual aids like "lollipops" in scatterplots could align visual and statistical fit estimates, potentially reducing the impact of the sine illusion. [23] We join existing efforts to study the sine illusion, offering concrete guidelines to inform visualization design to mitigate this bias.

3 Experiment

We compare three visualization designs in mitigating the sine illusion in line charts with two lines. We manipulated the ratio of the deltas between the lines at two points to obtain a threshold above which the sine illusion begins to significantly bias viewer perception. We further model and compare two potential explanations of participants' behavior.

We recruited 62 participants, with an average age of 20.65 (*SD* = 4.98), for this study, using the online crowdsourcing platform Amazon's Mechanical Turk [19]. Participants were compensated at a rate of 12 USD per hour. Thirty-one participants completed the study and contributed 50,777 trials of valid responses.

3.1 Experimental Design and Procedure

We generated charts depicting two increasing lines (blue and red). To establish the illusion task, we identified two time points on the chart (Time 1 and Time 2), and manipulated the vertical distance between the blue and the red line at these two points, as shown in Figure 2. Participants were tasked with comparing the distance between the red and blue lines at both time points to determine whether the differences between them were larger at Time 1 or Time 2. We refer to the distance between the two lines at Time 1 to be $d1$, and the distance between the two lines at Time 2 to be $d2$.

To determine a threshold past which sine illusion begins to have a strong effect based on the ratio of $d1$ and $d2$, we manipulate their ratio between 0.5 to 1.0, based on pilot studies that suggest that a ratio below 0.5 between $d1$ and $d2$ leads to generally highly accurate performance. To avoid a combinational explosion of values to display for $d1$ and $d2$, we fixed the value of $d1$ to be 0.1 while varying the value of $d2$ at 0.025 increments, ranging from 0.075 to 0.2. This manipulation required us to change the slope of the second and third segments of the blue line, which we refer to as $sb1$ and $sb2$, as well as the first and second segments of the red line, which we refer to as $st1$ and $st2$, as shown in Figure 2. To account for the potential effects of the slopes on participants' ability to compare the distances at Time 1 and Time 2 without the loss of generalizability, we randomly generated values of $sb1$ and $sb2$ to satisfy the following constraints: (1) the red line always remains above the blue line during Time 1 and Time 2, and (2) for each combination of $sb1$, $sb2$, $st1$, $st2$, three values remained constant while the fourth value is varied to allow us to control the specific effect of the varied component. We created a charting space of 1.75 (x) by 2.25 (y) with 0.25 intervals. These two constraints produced 1638 charts.

With this current design, because both lines are positively slanted, their slopes at Time 1 are always smaller than their slopes at Time 2. To account for potential left-right position-driven bias in participant response, we mirrored all of our charts so that both lines are negatively slanted in the mirrored conditions, as shown in Figure 2. This creates $1638 * 2 = 3276$ charts. This counterbalancing also accounts for potential response biases to prevent participants from achieving a higher accuracy rate by only selecting Time 1.

For each chart, we varied the design of the lines to be either default or dotted, or we added gridlines to the default version of the chart, as shown in Figure 2, creating $3276 * 3 = 9828$ charts. For the 3276 gridline charts, we recognize that whether the gridlines aligned with the marker at Time 1 and Time 2 also might affect comparison accuracy. Therefore, as a counterbalancing, for half of the gridline charts (randomly selected from the stimuli pool), we aligned the vertical gridlines with Time 1 and Time 2, and for the other half, we offset them, as shown in Figure 1.

Procedure: Participants were instructed to set their browser window to 100% before the study. We asked them to compare the revenue of two companies: A and B, over two time intervals, Time 1 and 2. They selected the time at which the difference in revenue between A and B appeared bigger. Figure 2 shows a condition where the difference at Time 1 is bigger and a condition where the difference at Time 2 is bigger. All participants went through a practice trial with feedback. We instructed participants to respond as accurately and fast as they could. They could also take a break at any time from the experiment. In total, everyone saw all 1638 charts, and the experiment took approximately one hour. After the experiment, they were given an MTruk completion code and were redirected to a Qualtrics survey to enter their demographic information.

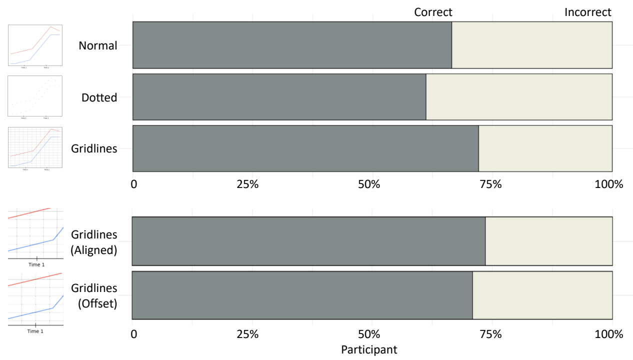

Figure 3: Overall accuracy for the three conditions and gridlines two counterbalancing conditions (aligned and offset).

3.2 Results: Overall

In analyzing our results, we filtered for trials that were completed in under 21791 milliseconds and more than 200 milliseconds as a quality control. We picked the lower bound because this task of comparing Time 1 and Time 2 requires an eye movement, and existing work in human perception suggests that it takes about 200 milliseconds for eye movement to begin [21]. We picked the upper bound based on one standard deviation above the mean response time. After filtering, we obtained 50,777 trials of valid responses, with an average response time of 2062.39 milliseconds.

In our following analysis, we also filtered out the trials where the difference between the two lines at Time 1 equals that at Time 2. These equality trials were added to prevent participants from consistently selecting one time over the other. Since there are no valid correct answers in these equality trials, we excluded them from our analysis. In general, for these 7457 equality trials (14.7% of the total trials), participants were 46.6% likely to choose Time 1. As shown in the top section of Figure 3, participants performed this task above chance (50%) across all conditions. In 66.5% of the trials, participants drew correct conclusions from the normal condition, 61.2% from the dotted lines condition, and 72.1% from the gridlines condition. A logistic linear regression predicting accuracy with condition shows that, compared to the normal condition, participants are about 1.30 times more likely to obtain the correct answer with the gridlines condition ($p < 0.001$). For the dotted lines condition, they were only 0.79 times as likely to obtain the correct answer ($p < 0.001$).

Takeaway: Making lines dotted does not mitigate the sin illusion bias, but adding gridlines can help.

3.3 Mirroring

We also mirrored all our displays to counterbalance our experiment accounting for the effect of the x-axis order (relative horizontal position of the sine illusion bias) for Time 1 and Time 2, as shown in Figure 2. On average, in 68.8% of the trials, participants obtained the correct answer from the default normal condition, and in 63.9% of the trials, participants obtained the correct answer from the mirrored condition. A logistic regression revealed that participants were only 0.78 times as likely to get the answer correct when the figure is mirrored, compared to the default (Est = -0.244, SE = 0.022, p ¡ 0.001). This suggests that the x-axis ordering and relative line positions may have a biasing effect to exaggerate sine illusions. We further discuss potential research avenues to better understand such bias in Section 5.

3.4 Effect of Gridlines On/Offset

We took a closer look at the gridline condition to understand how it might have helped increase accuracy. As shown in Figure 1, we manipulated the gridlines to either align the gridlines with Time 1 and Time 2 or offset them such that no vertical line goes through Time 1 or Time 2. We constructed a logistic regression predicting accuracy with whether the gridlines were offset or aligned. We found that aligning the gridlines increased accuracy to be 1.14 times that of the offset condition ($p = 0.00122$). Overall, as shown in the bottom section of Figure 3, participants performed with 73.53% accuracy when the gridlines are aligned and with 70.86% accuracy when the gridlines are offset.

3.5 Relationship between Times 1 and 2 on Accuracy

We further investigate the driving factor behind the varying accuracy levels for the comparison task between Time 1 and Time 2. Specifically, we examine how accuracy changes depending on the difference between the deltas at Time 1 and Time 2. We created the stimuli by varying the ratio between the Time 1 delta and Time 2 delta to range from 0.5 to 0.8. For example, when the ratio between Time 1 delta and Time 2 delta is 0.5, it means the difference between the red line and the blue line at Time 1 is half that at Time 2. The closer this ratio is to 1, the more difficult this comparison task is.

As shown in the left-most panel in Figure 4, as the ratio between Time 1 and Time 2 increases, the overall accuracy decreases for all three conditions. However, the decrease is less steep for the gridlines condition compared to the normal and dotted conditions. One unit increase in the ratio between Time 1 delta and Time 2 (delta i.e., the difference between Time 1 and Time 2 becomes smaller) leads to the accuracy in task performance to be 3.37% as accurate as the previous tier. For the dotted condition, task performance decreases to be only 1.40% as accurate. For the normal condition, task performance decreases to be 0.99% as accurate.

We computed participants' task accuracy based on whether the delta is bigger in Time 1 or Time 2. However, because we observed a low accuracy, we suspect that participants were not making their decision by comparing the vertical distances. We propose two alternative heuristics participants relied on when responding to the comparison task via two models:

(1) Perpendicular: this heuristic is about taking the perpendicular distance between the two lines, anchored on the bottom line at Time 1 and at Time 2, as shown in the middle panel in Figure 4. This idea is inspired by existing work that suggests that sine illusion happens because people rely on the orthogonal instead of the vertical distance between the two lines [4].

(2) Equal Triangle: this heuristic takes the length of the line segment perpendicular to the angle bisector of the bottom and top lines, see the right panel in Figure 4. This is inspired by the 'hull area' proxy from Jardine et al. [16] and the theory proposed by Day et al. [9] which suggest that participants could have considered the overall dimensional area (i.e., similar to drawing a circle with radius equal to the delta between the two lines at the point of comparison) surrounding the points of comparison when making the decision.

We re-compute the task accuracy by assuming that participants were relying on these two heuristics when comparing Time 1 and Time 2. For example, for the perpendicular heuristic, we compute the perpendicular distance at Time 1 and Time 2. If the perpendicular distance at Time 1 is greater than Time 2, and participants selected Time 1 to be greater, even if the vertical distance at Time 1 is smaller, we would consider their response correct. Under this setup, if the task accuracy increases, then we can infer that participants were more likely to rely on the perpendicular heuristic when

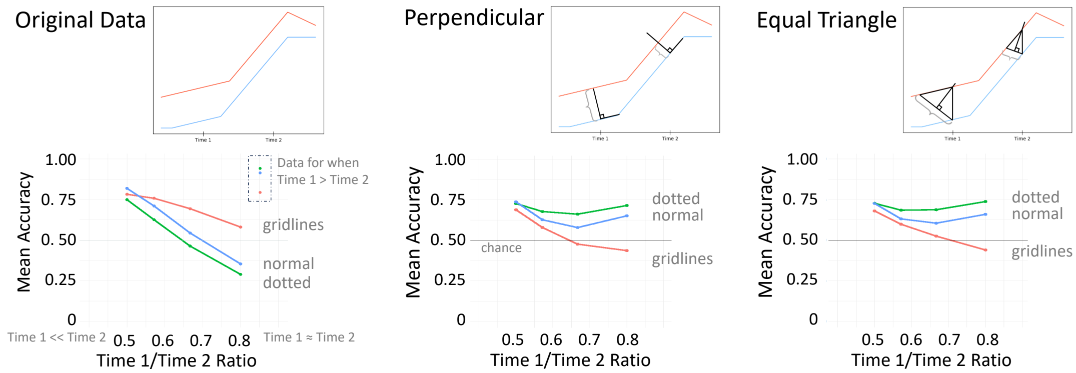

Figure 4: Accuracy by the ratio of deltas at Time 1 and Time 2, for the original data, in comparison to the two models: perpendicular and equal triangle. The higher accuracy under both models suggests that participants relied on the orthogonal distance to make their comparisons.

completing the task rather than the vertical distance. Comparing the three line charts in Figure 4, we see that participants' task accuracy increases under the assumptions of the perpendicular and equal triangle models. We conducted a t-test comparing the overall accuracy between these two models and found the equal triangle model to have higher accuracy (t = 2.75, p = 0.0060).

Takeaway: When participants compare the deltas between two lines, they are less likely to rely on the vertical distance between the lines. Rather, they compute the orthogonal distance, best modeled by looking at the length of the line segment perpendicular to the angle bisector of the bottom and top lines.

3.6 When Time 1 Delta is Actually Larger

When the delta at Time 1 is *larger* than that at Time 2, relying on the sine illusion would help the participants. In these cases, participants performed *worst* with the gridlines condition and *best* with the dotted lines condition. When they viewed the dotted line charts, their accuracy was 4.18 times higher than the gridline condition. When they viewed the default normal condition, their accuracy was 2.75 times higher than the gridlines condition. However, despite this reversal in condition effectiveness, participants overall performed significantly better (t = 61.84, p ¡ 0.001) in the scenario where Time 1 delta is actually larger (ratio of Time 1 over Time 2 is greater 1, $M_{accuracy}$ = 0.89), compared to the scenario where Time 1 delta is smaller (ratio of Time 1 over Time 2 is smaller than 1, $M_{a}ccuracy$ = 0.61). See annotation on the left-most panel in Figure 4.

While further investigation is needed to draw causal conclusions, we suspect that the gridlines made participants double-guess their response instead of relying on the sine illusion, which decreased accuracy. When the gridlines are present, participants can compare the length of the gridlines between the two lines at Time 1 and Time 2 to make their decisions. But length comparison is subjected to perceptual bias, as length encoding is not an extremely precise encoding channel [18, 5, 6]. As a result, they make mistakes and perform with lower accuracy. However, when the gridlines are turned off, participants tend to rely on the sine illusion and compare the equal triangle distance for the task, which resulted in Time 1 being perceived as larger. Thus when the Time 1 delta is actually larger, relying on the sine illusion, rather than making a length comparison, results in a higher overall accuracy across all conditions, and relatively lower accuracy for gridlines.

4 Design Guidelines and Summary of Findings

We produced a model that predicts the likelihood and severity of the sine illusion in line charts based on the ratio of the deltas between the two points of comparison. We found that, in general, accuracy drops below chance when the ratio falls above 0.7 for default line charts, and below 75% when the ratio falls above 0.5. This means that when participants are comparing the difference between two lines at two different time points, if that difference is less than 50%, then people will struggle. And if that difference is less than 30%, people will struggle significantly. This threshold can be mitigated by adding gridlines, especially when the gridlines are aligned with the points of comparison. Adding gridlines can shift the threshold, such that people will only start to struggle when the ratio between the two vertical distances when the difference is less than 20%. While we have not tested to identify optimal design options for the gridlines, existing work by Bartram et al. identified adopting an alpha value between 0.1 and 0.45 for gridlines might be the most preferred and effective [2].

In general, people make mistakes in this task because they are not comparing the vertical distance between the two lines. They are comparing the length of the line segments perpendicular to the angle bisector of the bottom and top lines, as modeled by the equal triangle heuristics.

5 Limitation and Future Directions

We modeled the effect of the ratio between the vertical distance at the two comparison points. As shown in the left panel in Figure 2, factors such as the line slopes at the two points of comparison, could also influence the severity of sine illusions. A preliminary logistic regression model predicting comparison accuracy with the line slopes at the two points of comparison suggests these factors to be significant (details can be found in the supplementary materials at https://osf.io/kq87n/). Most notably, the slope of the top line at Time 1 seems to have the strongest effect on comparison accuracy (OR = 2.26, p ¡ 0.001). Future work could examine these effects to create a predictive model that takes data values of the lines as input, computes the relative slopes and deltas between two lines, and outputs a likelihood of the viewer seeing sine illusions.

Further, we only tested one design of the line charts to avoid a combinatorial explosion of conditions. Future work can consider alternative colors, line thickness, and line types with different spacing to validate the generalizability of our findings. Moreover, considering that the amount of data can impact perception [17, 18, 30], future work can explore the effect of data set size on the illusion.

Finally, we only modeled two heuristics. Participants could engage in other strategies when completing the task. For example, people might perceive the pairs of lines as shapes, so that changing perspectives produced by eye movements would not distort their percept [20]. Instead of reading the values following rules of graphical interpretation, participants might be comparing the width and height (major/minor axes) of that shape. Future experiments should consider think-aloud protocols or offline studies to elicit those strategies and make even better models of perception.

Acknowledgments

The authors wish to thank Judith Uchidiuno, as well as NSF awards IIS-2237585 and IIS-2311575 for support.

References

[1] K. Ajani, E. Lee, C. Xiong, C. N. Knaflic, W. Kemper, and S. Franconeri. Declutter and focus: Empirically evaluating design guidelines for effective data communication. *IEEE Transactions on Visualization and Computer Graphics*, 2021. doi: 10.1109/TVCG.2021.3068337 1

[2] L. Bartram and M. C. Stone. Whisper, don't scream: Grids and transparency. *IEEE Transactions on Visualization and Computer Graphics*, 17(10):1444–1458, 2010. doi: 10.1109/TVCG.2010.237 4

[3] C. X. Bearfield, L. van Weelden, A. Waytz, and S. Franconeri. Same data, diverging perspectives: the power of visualizations to elicit competing interpretations. *IEEE Transactions on Visualization and Computer Graphics*, 30(6):2995–3007, 2024. doi: 10.1109/TVCG.2024.3388515 1

[4] C. Bu, Q. Zhang, Q. Wang, J. Zhang, M. Sedlmair, O. Deussen, and Y. Wang. Sinestream: Improving the readability of streamgraphs by minimizing sine illusion effects. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1634–1643, 2020. doi: 10.1109/TVCG.2020.3030404 1, 3

[5] C. R. Ceja, C. M. McColeman, C. Xiong, and S. L. Franconeri. Truth or square: Aspect ratio biases recall of position encodings. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1054–1062, 2020. doi: 10.1109/TVCG.2020.3030422 1, 4

[6] W. S. Cleveland. A model for studying display methods of statistical graphics. *Journal of Computational and Graphical Statistics*, 2(4):323–343, 1993. doi: DOI: 10.2307/1390686 2, 4

[7] 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. doi: 10.2307/2288400 1

[8] 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. 2

[9] R. H. Day and E. J. Stecher. Sine of an illusion. *Perception*, 20(1):49–55, 1991. doi: 10.1068/p200049 1, 3

[10] E. Dimara, S. Franconeri, C. Plaisant, A. Bezerianos, and P. Dragicevic. A task-based taxonomy of cognitive biases for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 2018. doi: 10.1109/TVCG.2018.2872577 1

[11] 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. doi: 10.1109/TVCG.2011.127 1

[12] J. Heer and M. Agrawala. Multi-scale banking to 45 degrees. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):701–708, 2006. doi: 10.1109/TVCG.2006.163 2

[13] M. Hegarty, M. S. Canham, and S. I. Fabrikant. Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(1):37, 2010. doi: 10.1037/a0017683 1

[14] H. Hofmann and M. Vendettuoli. Common angle plots as perception-true visualizations of categorical associations. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2297–2305, 2013. 2

[15] C. Q. Howe and D. Purves. The müller-lyer illusion explained by the statistics of image–source relationships. *Proceedings of the National Academy of Sciences*, 102(4):1234–1239, 2005. doi: 10.1073/pnas.0409314102 1

[16] N. Jardine, B. D. Ondov, N. Elmqvist, and S. Franconeri. The perceptual proxies of visual comparison. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):1012–1021, 2019. doi: 10.1109/TVCG.2019.2934786 1, 3

[17] C. M. McColeman, L. Harrison, M. Feng, and S. Franconeri. No mark is an island: Precision and category repulsion biases in data reproductions. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1063–1072, 2020. doi: 10.1109/TVCG.2020.3030345 2, 4

[18] C. M. McColeman, F. Yang, T. F. Brady, and S. Franconeri. Rethinking the ranks of visual channels. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):707–717, 2021. doi: 10.1109/TVCG.2021.3114684 4

[19] Amazon's mechanical turk. `www.MTurk.com`, 2020. 2

[20] J. Ninio. Geometrical illusions are not always where you think they are: a review of some classical and less classical illusions, and ways to describe them. *Frontiers in human neuroscience*, 8:856, 2014. 4

[21] D. Purves, G. Augustine, D. Fitzpatrick, L. Katz, A. LaMantia, J. McNamara, and S. Williams. Neuroscience 2nd edition. sunderland (ma) sinauer associates. *Types of Eye Movements and Their Functions*, 2001. 3

[22] G. J. Quadri and P. Rosen. A survey of perception-based visualization studies by task. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):5026–5048, 2021. doi: 10.1109/TVCG.2021.3098240 1

[23] D. Reimann, N. Ram, and R. Gaschler. Lollipops help align visual and statistical fit estimates in scatterplots with nonlinear models. *IEEE Transactions on Visualization and Computer Graphics*, 2022. doi: Lollipops help align visual and statistical fit estimates in scatterplots with nonlinear models 1, 2

[24] T. Savalia, C. Ceja, R. Cowell, and C. Xiong. Data shape and response modalities can bias estimations of average data location in visualizations. *Journal of Vision*, 22(14):3413–3413, 2022. doi: 10.1167/jov.22.14.3413 2

[25] M. Schonlau. Hammock plots: visualizing categorical and numerical variables. *Journal of Computational and Graphical Statistics*, pp. 1–16, 2024. 2

[26] C. Stokes, C. X. Bearfield, and M. A. Hearst. The role of text in visualizations: How annotations shape perceptions of bias and influence predictions. *IEEE Transactions on Visualization and Computer Graphics*, 2023. doi: 10.1109/TVCG.2023.3338451 1

[27] A. C. Valdez, M. Ziefle, and M. Sedlmair. A framework for studying biases in visualization research. In *DECISIVe 2017: Dealing with Cognitive Biases in Visualisations*, 2017. 1

[28] S. VanderPlas and H. Hofmann. Signs of the sine illusion—why we need to care. *Journal of Computational and Graphical Statistics*, 24(4):1170–1190, 2015. doi: 10.1080/10618600.2014.951547 1, 2

[29] E. Wall, L. M. Blaha, L. Franklin, and A. Endert. Warning, bias may occur: A proposed approach to detecting cognitive bias in interactive visual analytics. In *2017 ieee conference on visual analytics science and technology (vast)*, pp. 104–115. IEEE, 2017. doi: 10.1109/VAST.2017.8585669 1

[30] C. Xiong, L. Padilla, K. Grayson, and S. Franconeri. Examining the components of trust in map-based visualizations. In *1st EuroVis Workshop on Trustworthy Visualization, TrustVis 2019*, pp. 19–23. The Eurographics Association, 2019. doi: 10.2312/trvis.20191186 4

[31] C. Xiong, A. Sarvghad, Ç. Demiralp, J. M. Hofman, and D. G. Goldstein. Investigating perceptual biases in icon arrays. *Proceedings of the 2022 ACM Conference on Human Factors in Computing Systems (CHI '22)*, 2022. doi: 10.1145/3491102.3501874 1

[32] C. Xiong, C. Stokes, Y.-S. Kim, and S. Franconeri. Seeing what you believe or believing what you see? belief biases correlation estimation. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):493–503, 2022. doi: 10.1109/TVCG.2022.3209405 1

[33] C. Xiong, L. Van Weelden, and S. Franconeri. The curse of knowledge in visual data communication. *IEEE Transactions on Visualization and Computer Graphics*, 26(10):3051–3062, 2019. doi: 10.1109/TVCG.2019.2917689 1