

Further Investigating the Effects of Point Size and Contrast on Correlation Perception in Scatterplots

ANONYMOUS AUTHOR(S)

It is paramount that the toolkit of the data visualization designer must continue to grow based on robust, reproducible experimental work. To this effect, we present a single-experiment study in which we combine previously explored data point size and contrast manipulations in scatterplots to influence viewer's estimates of correlation. We follow this with an extended exploration into how the combination of two established scatterplot manipulations can explain the relative contributions and mechanisms of each, in ways that the manipulations in isolation cannot. We make the case for more extensive testing of these manipulations such that they can be tuned to fully address a long-standing correlation underestimation bias by viewers of scatterplots in the literature. Simultaneously we present a rigorous experimental framework for fast, inexpensive, and empirical visualization design.

CCS Concepts: • **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

Additional Key Words and Phrases: correlation, scatterplot, perception, crowdsourced

ACM Reference Format:

Anonymous Author(s). 2018. Further Investigating the Effects of Point Size and Contrast on Correlation Perception in Scatterplots. In . ACM, New York, NY, USA, 11 pages. <https://doi.org/XXXXXXX.XXXXXXX>

1 INTRODUCTION

- why study scatterplots?
- correlation perception

We have previously demonstrated the potential for changes in point size and point contrast to bias peoples' estimates of correlation in scatterplots. Lowering the size and contrast of scatterplot points as they move further from the regression line partially corrects for the underestimation bias described above. We have hypothesized in our previous work that what is driving the changes in correlation estimation that we have seen is an increase in spatial uncertainty at the edges of the scatterplot relative to the center, which itself drives a reduction in the perceived width of the probability distribution said scatterplot represents. As part of our ongoing effort to tune scatterplots for more accurate correlation perception, our next step is to combine our point size and contrast manipulations.

2 RELATED WORK

2.1 Testing Correlation Perception

The testing of correlation perception has a long and rich history. Very early work had participants make discriminative judgements between scatterplots with different correlations [?], finding that performance on such a task became better as the objective r value increased. Work

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

© 2018 Association for Computing Machinery.

Manuscript submitted to ACM

2.2 Drivers of Correlation Perception

- width of perceived prob distribution
- shape of prediction ellipse
- probably some others

2.3 Transparency and Contrast

Changing the contrast of scatterplot points is standard practice to deal with issues of overplotting or clutter [2, 9]; scatterplots with very large numbers of points, especially with high degrees of overlap, suffer from low individual-point visibility caused by high point density. Lowering the contrast of all points addresses this, and makes data trends and distributions easier to see and interpret (see Figure 1).

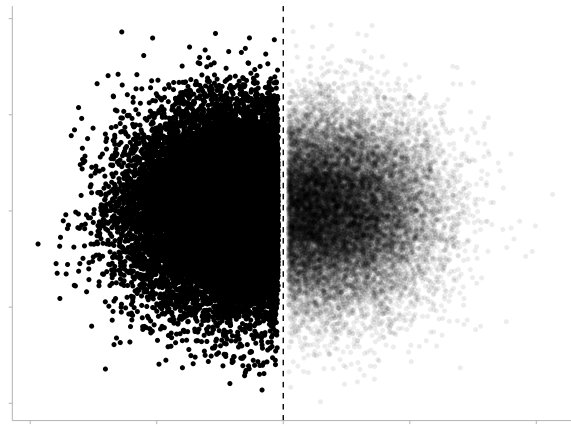


Fig. 1. Adjusting point contrast to address overplotting. Contrast between the points and the background is full ($\alpha = 1$, L) or low ($\alpha = .075$, R). The dataset used has 40,000 points.

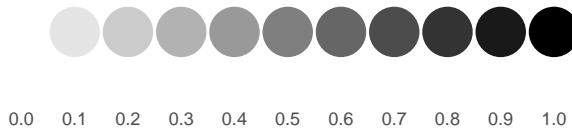


Fig. 2. Demonstrating the effects of different alpha values on point contrast.

In the present study we use the **ggplot2** package [19] to create our stimuli. This package uses an α parameter, or the level of linear interpolation [14] between foreground and background pixel values, to set the contrast of points. As demonstrated in Figure 2, an α value of 0 or 1 results in no interpolation and renders either the background or foreground pixel values respectively. Psychophysical definitions of perceived contrast are often based on what is being presented (i.e gratings) or are modelled to take into account phenomena of human vision (i.e visibility limits) [20]. Our crowdsourced methodology gives us no control the exact luminances of our stimuli, only over the *relative* differences in luminance between scatterplot points and backgrounds. For this reason we do not report absolute luminance nor make any attempt to adopt a formal definition of contrast; instead we simply report the α value used. Given that

our work aims to improve correlation perception without removing data from the scatterplot, we incorporate point visibility testing (see Section 3.5). Informal point visibility piloting suggested that our smallest, lowest contrast points had very low visibility. We therefore implemented an $\alpha = 0.2$ floor for these points, which we felt conferred a sufficient level of point visibility.

Previous work has found that lowering the contrast of *all* scatterplot points relative to the background can increase the level of underestimation error relative to full contrast, and that lowering point contrast *as a function of distance from the regression line* is able to bias correlation estimates upwards to partially correct for the underestimation bias [16]. Evidence suggests point salience/perceptual weighting and spatial uncertainty as drivers for this effect. Lower stimulus contrast is associated with lower salience [7], can bias judgements of mean point position [8] and increase error in positional judgements [18], and can result in greater uncertainty in speed perception [4]. Due to effects of both salience/perceptual weighting and spatial uncertainty operating in the same direction, previous work [16] has been unable to determine the extent to which each is responsible for the observed reduction in correlation estimation error.

2.4 Point Size

For discriminability reasons, scatterplots visualizing large datasets tend also to have smaller points. Bubble charts are a subclass of scatterplot which uses point size to describe a third variable, but what little experimental work there is on the impact of point size on correlation perception is inconclusive. Some work has found bias and variability in correlation perception to be invariant to changes in point size [?]; rensink_2014], while elsewhere a strong effect of changing point size as a function of distance from the regression line has been reported [?]. Evidence points towards a salience-dominant mechanism in the latter case, albeit with a small effect of spatial uncertainty. There is evidence that stimulus size is associated with lower levels of spatial certainty [?] but higher levels of salience [?], and so the opposing directionality of predicted effects of salience/perceptual weighting and spatial uncertainty on correlation estimation has allowed researchers to provide evidence for the mechanistic dominance of salience when point size is used.

2.5 Hypotheses

Based on previously established effects of adjusting point size and point contrast using identical decay functions in isolation, we presently make the following hypotheses about the combination of these functions. We hypothesize that; (H1) an increased reduction in correlation estimation error will be observed when standard orientation non-linear decay functions are used; (H2) the use of congruent inverted orientation decay functions will produce the least accurate estimates of correlation; and (H3) that owing to the greater strength of the size channel observed in previous work [?], there will be a significant difference in correlation estimates between the two incongruent orientation conditions.

3 METHODOLOGY

3.1 Crowdsourcing

Much prior work into correlation perception in scatterplots has taken place in-person, most often with graduate students with experience in statistics. While this work is valuable, especially to perception audiences, it can struggle to provide data that is resilient to different lay viewers and viewing contexts. In addition the ease and low-cost afforded to us by online, crowdsourced experimental work is unmatched. Given our intended HCI and design audience, we therefore choose to crowdsource all participants. We acknowledge however, that the technique has been affected by

low quality data and skewed demographics in the past [11? ?]. In light of these issues we follow published guidelines [11] to ensure the collection of high quality data. Namely, we use the Prolific.co platform [1] with stringent pre-screen restrictions; participants were required to have completed at least 100 studies using Prolific, and were required to have a prolific score of at least 100, representing a 99% approval rate. This is more strict than the 95% suggested in previous work [11], but

Discussions about the transparency, contrast, and luminance is inherently difficult within the context of online, crowdsourced experimental work. While the ease, low-cost, and resilience to different viewing contexts afforded to us by such work is advantageous, we must also consider the impact on our ability to precisely describe the experimental stimuli.

3.2 Open Research

This study was conducted according to the principles of open and reproducible research. All data and analysis code are available at (repository link removed for anon). This repository contains instructions for building a docker image to fully reproduce the computational environment used. This allows for full replications of stimuli, analysis, and the paper itself. Ethical approval for both experiments was granted by (removed for anon). Hypotheses and analysis plans were pre-registered with the OSF (links removed for anon).

3.3 Scatterplot Generation

The data used to generate the scatterplots were identical to that used in previous work [16?]. 45 scatterplot datasets were generated corresponding to 45 r values uniformly distributed between 0.2 and 0.99, as there is evidence that very little correlation is perceived below $r = 0.2$ [3, 5, 15]. Using so many values for r allows us to paint a broader picture of people’s perceptions than work using fewer values. Scatterplot points were generated based on bivariate normal distributions with standard deviations of 1 in each direction. Each scatterplot had a 1:1 aspect ratio, was generated as a 1000*1000 pixel .png image, and was scaled up or down according to a participant’s monitor such that they always occupied the same visual proportion. We used equation 1 to map residuals to size and contrast values. When adjusting point size, we further transform values using a scaling factor of 4 and a constant of 0.2 to ensure that the minimum point size in the present study is consistent with that of previous work [Strain et al. [16]; strain_2023b].

$$point_{size/contrast} = 1 - b^{residual} \quad (1)$$

0.25 was chosen as the value of b . This is both due to its previous usage in studies that the present work builds upon, and its production of a curve approximating the inverse around the identity line of the underestimation curve reported in previous work [Rensink [13]; strain_2023; strain_2023b]. We also acknowledge that there may be other, more suitable values of b , however extensively testing these is outside the scope of the present work. We used 2x2 combinations of this equation applied to point size and contrast in standard and inverted orientation forms. Scatterplot examples can be seen in Figure 4.

3.4 Modelling

We use linear mixed effects models to model the relationships between the combination of size and contrast decay conditions and participants’ errors in correlation estimates. Models such as these allow us to compare differences in our IVs across the full range of participant responses, as opposed to relying purely on aggregate data, as in ANOVA. These

models also afford us the ability to include random effects for participants and items. As per our pre-registrations we preferred maximal models, including random intercepts and slopes for participants and items. The structures of these models was identified using the **buildmer** package in R (version xx, [17]). This package takes a maximal random effects structure and then identifies the most complex model that converges, dropping terms that fail to explain a significant amount of variance.

3.5 Point Visibility Testing

Discussions about the size and contrast of particular scatterplot points are inherently difficult in the context of on-line, crowdsourced experiments. In addition to measures implemented to ensure high quality participant data (see Section 3.1), it is also key that we do not inadvertently remove data from scatterplots by including points whose size or contrast renders them invisible. We therefore included point visibility testing to ensure this. Participants viewed six scatterplots that were made up of a certain number of points. These points were of the same size and contrast as the smallest and lowest contrast points used in the experimental items. Participants were asked to enter in a textbox how many points were present. Participants scored an average of 74.89% ($SD = 32.25$). Despite our use of the contrast floor detailed in Section 2.3, it is clear that some of our small, low contrast points were not reliably visible, most likely due to low contrast between the point and background, as previous work (strain_2023b) found point visibility invariant to size. We suggest this is due to differences in monitors between participants. In reality this contrast floor would need to be calibrated on a per-monitor basis. Figure 3 shows distributions of participants' performances on the visual threshold tests. We also include performance on the point visibility test as a fixed effect in Section 4.1.

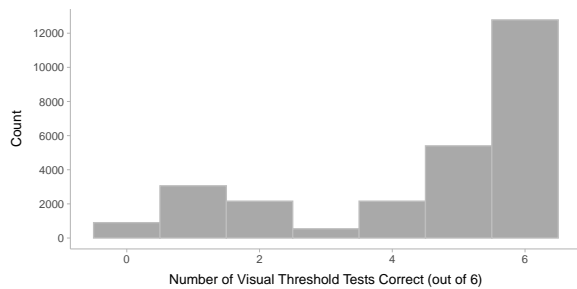


Fig. 3. Histograms of visual threshold testing performance.

3.6 Dot Pitch

We employed a method for obtaining the dot pitch of participants' monitors [10]. Combining this with monitor resolution information allows us to calculate the physical on-screen size of scatterplot points. Participants were asked to hold a standard size credit/ debit/ID card (ISO/IEC 7810 ID-1) up to their screen and resize an on-screen card until the two matched. We assumed a widescreen 16:9 aspect ratio and calculated dot pitch based on these measurements. Mean dot pitch was 0.34 ($SD = 0.05$). We include analyses with dot pitch as a fixed effect in Section 4.1.

3.7 Procedure

Both experiments were built using PsychoPy [12] and hosted on Pavlovia.org. Participants were only permitted to complete the experiments on a desktop or laptop computer. Each participant was first shown the participant information

sheet and provided consent through key presses in response to consent statements. They were asked to provide their age in a free text box, followed by their gender identity. Participants completed the 5-item Subjective Graph Literacy test [6], followed by the visual threshold task described in Section 3.5 and the screen scale task described in Section 3.6. Participants were given instructions, and were then shown examples of scatterplots with correlations of $r = 0.2$, 0.5 , 0.8 , and 0.95 , as piloting of a previous experiment indicated some of the lay population may be unfamiliar with the visual character of scatterplots. Section 4 contains further discussion of the potential training effects of this. Two practice trials were given before the experiment began. Participants worked through a randomly presented series of 180 experimental trials (90 in experiment 2?) and were asked to use a slider to estimate correlation to 2 decimal places. Visual masks preceded each scatterplot. Interspersed were 6 attention check trials which explicitly asked participants to ignore the scatterplot and set the slider to 0 or 1.

3.8 Participants

150 participants were recruited using the Prolific.co platform. Normal to corrected-to-normal vision and English fluency were required for participation. In addition, participants who had completed any of our previous studies into correlation estimation in scatterplots (references removed for anon) were prevented from participating.

Data were collected from 158 participants. 8 failed more than 2 out of 6 attention check questions, and, as per pre-registration stipulations, were rejected from the study. Data from the remaining 150 participants were included in the full analysis (50.7% male, 48.7% female, and 0.7% non-binary). Participants mean age was 30.6 ($SD = 8.6$). Participants' mean graph literacy score was 22.5 ($SD = 3.5$). The average time taken to complete the experiment was 37 minutes ($SD = 12.3$).

3.9 Design

We used a fully repeated-measures 2×2 factorial design in experiment 1. Each participant saw each combination of size and contrast decay condition plots for a total of 180 experimental items. Participants viewed these experimental items, along with 6 attention check items, in a fully randomized order. Examples of experimental items can be seen in Figure 4.

4 RESULTS

Our first two hypotheses were fully supported in this experiment. The combination of non-linear size and contrast decay functions produced the most accurate estimates of correlation, although this also resulted in a large correlation overestimation for many values of r (see Figure 7). Our second hypothesis was also supported; the combination of inverted size and inverted contrast decay conditions produced the least accurate estimates of correlation. We found no support for our third hypothesis; there was no significant difference in correlation estimates for non-linear size/inverted contrast decay plots and inverted size/non-linear contrast decay plots ($Z = -2.26$, $p = .11$), however we did find a significant interaction effect that provides evidence that the size decay function was stronger with regards to biasing people's estimates of correlation.

All analyses were conducted using R (version 4.3.1). Deviation coding was used for each of the experimental factors. We used the **buildmer** and **lme4** packages to build a linear mixed effects model where the difference between objective and rated r value was predicted by the size and contrast decay conditions used. A likelihood ratio test revealed that the model including point size and contrast decay conditions as fixed effects explained significantly more variance than the null ($\chi^2(3) = 5,286.81$, $p < .001$). There were significant fixed effects of size decay and contrast decay conditions, as

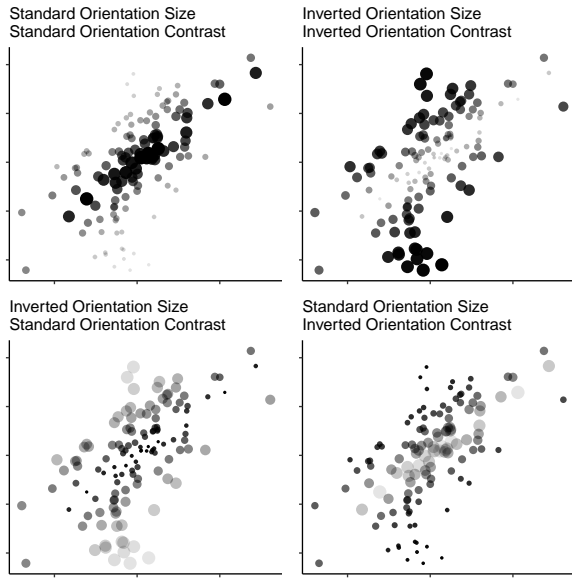


Fig. 4. ?(caption)

Table 1. Significances of fixed effects and interaction for experiment 1.

	Estimate	Standard Error	df	t-value	p
(Intercept)	0.08	0.013	103.32	6.27	<0.001
Size Decay	-0.14	0.005	148.39	-25.77	<0.001
Contrast Decay	0.12	0.002	26327.21	63.71	<0.001
Size Decay x Contrast Decay	0.15	0.004	26327.13	38.47	<0.001

Table 2. Pairwise comparisons for experiment 1. Our interaction is driven by the greater strength of the size channel, whether non-linear or inverted decay functions were used. Note the lack of significance for NL size/Inv contrast against Inv size/NL contrast comparison.

Contrast	Z.ratio	p.value
Non-linear Size x Inverted Contrast <-> Inverted Size x Inverted Contrast	-10.95	<0.001
Non-linear Size x Inverted Contrast <-> Non-Linear Size x Non-linear Contrast	72.29	<0.001
Non-linear Size x Inverted Contrast <-> Inverted Size x Non-linear Contrast	-2.26	0.108
Inverted Size x Inverted Contrast <-> Non-linear Size x Non-linear Contrast	46.13	<0.001
Inverted Size x Inverted Contrast <-> Inverted Size x Non-linear Contrast	17.84	<0.001
Non-linear Size x Non-linear Contrast <-> Inverted Size x Non-linear Contrast	-37.44	<0.001

well as a significant interaction between the two. Table 1 shows a summary of the model statistics. The experimental model has random intercepts for items and participants, and a random slope for the size decay factor with regards to participants.

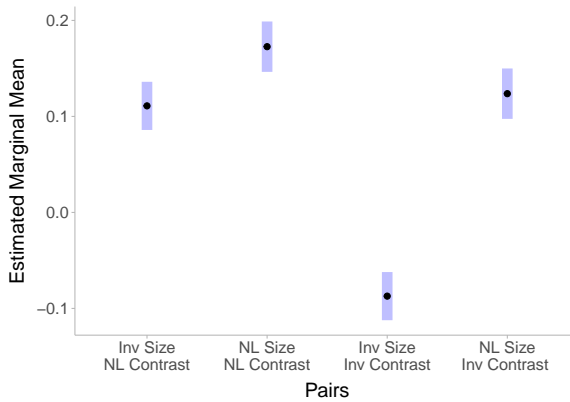


Fig. 5. ?(caption)

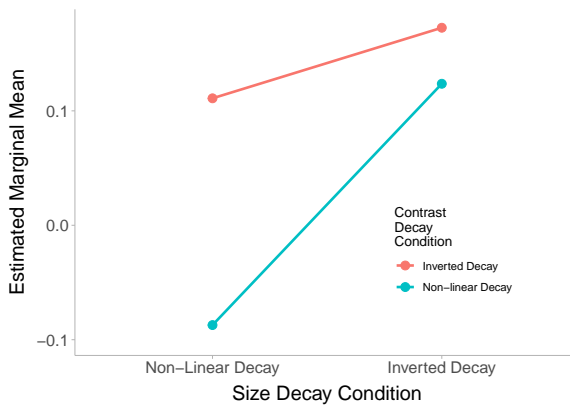


Fig. 6. Interaction plot showing the moderating effect of encoding channel on decay orientation.

The `emmeans` (cite) package was used to run pairwise comparisons. Figure 5 shows estimated marginal means for each combination of factors, and Figure 6 shows the form the interaction takes. The interaction is driven by there being a greater difference in estimates between NL and inverted size decay conditions when the contrast decay condition is held as non-linear, as opposed to when it is inverted. This is parsimonious with previous work demonstrating the greater capacity of the size encoding channel to bias participants' estimates of r [?], and findings that the inverted channel is generally weaker at biasing correlation in the opposite direction [16?]. Table 2 shows statistics for pairwise comparisons.

4.1 Additional Analyses

We find no effects of graph literacy ($\chi^2(1) = 3.50, p = .061$) or performance on the visual threshold task ($\chi^2(1) = 1.29, p = .257$), or dot pitch ($\chi^2(1) = 1.52, p = .218$) on participants' errors in correlation estimation.

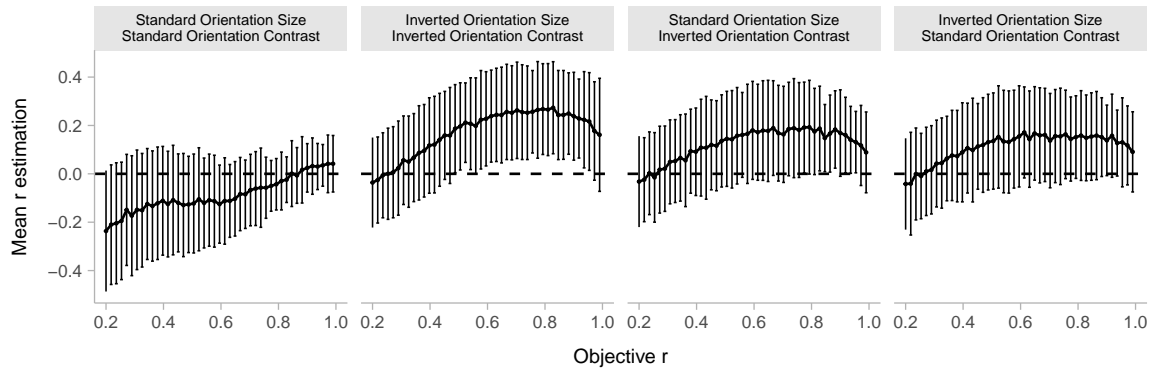


Fig. 7. Plots showing how participants' correlation estimation errors change as a function of the r value for each combination of size and contrast decay factors.

5 DISCUSSION

Our findings here provide further confirmatory evidence of what has been found previously with regards to the effects of point size and contrast manipulations on correlation estimation in scatterplots. Namely, that while both manipulations have a significant effect, the effect of changing point sizes is stronger, and that while we can influence correlation estimates in either direction, standard orientation manipulations are more powerful than inverted ones. The experiment described above additionally contributes evidence of a congruency effect; effects are most strong when there is orientation congruency between size and contrast manipulations.

The lack of support for our third hypothesis, that there would be a difference in correlation estimates between incongruent conditions, was surprising given the greater strength of the size channel relative to contrast. We take this as evidence that the combination of size and contrast manipulations is not additive, as we had initially suspected, which explains the interaction we see here. Integrating the data collected in the present study with previously collected data investigating the effects of the size and contrast manipulations in isolation allows us demonstrate this and to partial out the effects each manipulation.

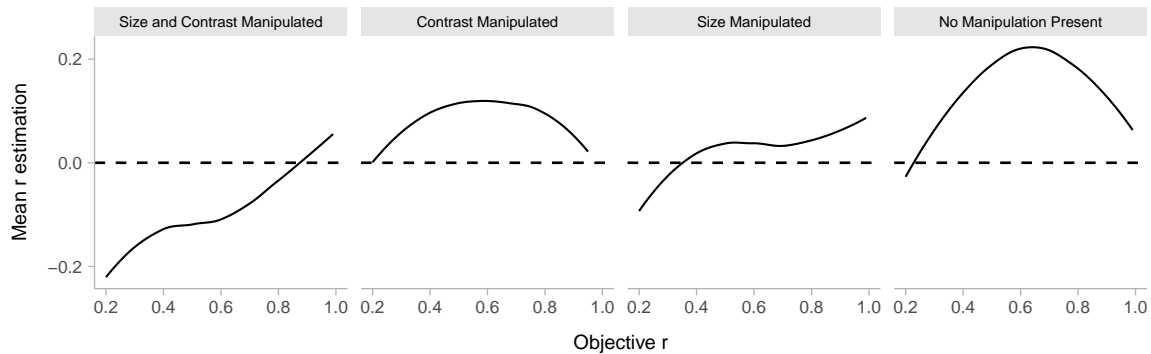


Fig. 8. From left to right, plotting r estimation error against the objective r value for the standard orientation condition in the present study, for standard orientation size and contrast manipulations in previous work, and for normal scatterplots averaged over identical conditions in previous work. Error bars have been left off this plot to make interpretation more simple.

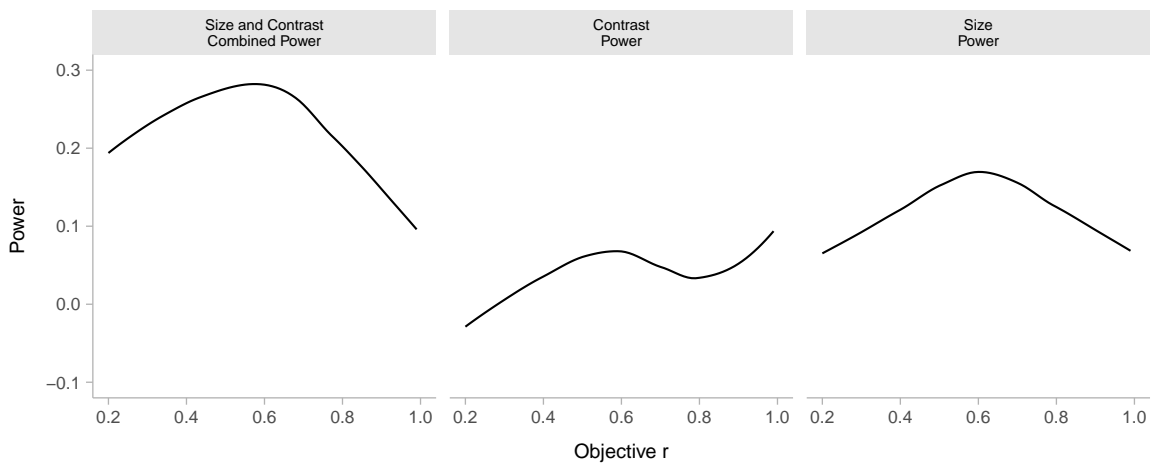


Fig. 9. additive_raw_pl = observed values for present study. standard_curve = no manipulation averaged across all experiments

As can be seen in Figure 8, combining the size and contrast manipulations results in an error curve of a mostly similar shape to that of the size manipulation. The difference that comes from adding the contrast manipulation lies in the severity of the effect and the orientation of the curve itself. Transforming the size underestimation curve from Strain et al. [16] into the size and contrast underestimation curve we have found in the present study gives us some insight into the effect of adding the two manipulations together.

REFERENCES

- [1] 2023. Prolific. <https://www.prolific.co>
- [2] Enrico Bertini and Giuseppe Santucci. 2004. Quality Metrics for 2D Scatterplot Graphics: Automatically Reducing Visual Clutter. In *Smart Graphics (Lecture Notes in Computer Science)*, Andreas Butz, Antonio Krüger, and Patrick Olivier (Eds.). Springer, Berlin, Heidelberg, 77–89. https://doi.org/10.1007/978-3-540-24678-7_8
- [3] Philip Bobko and Ronald Karren. 1979. The Perception of Pearson Product Moment Correlations from Bivariate Scatterplots. *Personnel Psychology* 32, 2 (1979), 313–325. <https://doi.org/10.1111/j.1744-6570.1979.tb02137.x>
- [4] Rebecca A. Champion and Paul A. Warren. 2017. Contrast Effects on Speed Perception for Linear and Radial Motion. *Vision Research* 140 (Nov. 2017), 66–72. <https://doi.org/10.1016/j.visres.2017.07.013>
- [5] W. S. Cleveland, P. Diaconis, and R. McGill. 1982. Variables on Scatterplots Look More Highly Correlated When the Scales Are Increased. *Science* 216, 4550 (June 1982), 1138–1141. <https://doi.org/10.1126/science.216.4550.1138>
- [6] Rocio Garcia-Retamero, Edward T. Cokely, Saima Ghazal, and Alexander Joeris. 2016. Measuring Graph Literacy without a Test: A Brief Subjective Assessment. *Medical Decision Making* 36, 7 (2016), 854–867. <https://doi.org/10.1177/0272989X16655334>
- [7] Christopher Healey and James Enns. 2012. Attention and Visual Memory in Visualization and Computer Graphics. *IEEE Transactions on Visualization and Computer Graphics* 18, 7 (July 2012), 1170–1188. <https://doi.org/10.1109/TVCG.2011.127>
- [8] Matt-Heun Hong, Jessica K. Witt, and Danielle Albers Szafir. 2021. The Weighted Average Illusion: Biases in Perceived Mean Position in Scatterplots. *IEEE Transactions on Visualization and Computer Graphics* (2021), 1–1. <https://doi.org/10.1109/TVCG.2021.3114783>
- [9] Justin Matejka, Fraser Anderson, and George Fitzmaurice. 2015. Dynamic Opacity Optimization for Scatter Plots. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, Seoul Republic of Korea, 2707–2710. <https://doi.org/10.1145/2702123.2702585>
- [10] Wakefield L Morys-Carter. 2023. ScreenScale. <https://doi.org/10.17605/OSF.IO/8FHQK>
- [11] Eyal Peer, David Rothschild, Andrew Gordon, Zak Evernden, and Ekaterina Damer. 2021. Data Quality of Platforms and Panels for Online Behavioral Research. *Behavior Research Methods* (Sept. 2021). <https://doi.org/10.3758/s13428-021-01694-3>
- [12] Jonathan Peirce, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindeløv. 2019. PsychoPy2: Experiments in Behavior Made Easy. *Behavior Research Methods* 51, 1 (Feb. 2019), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>

- [13] Ronald A. Rensink. 2017. The Nature of Correlation Perception in Scatterplots. *Psychonomic Bulletin & Review* 24, 3 (2017), 776–797. <https://doi.org/10.3758/s13423-016-1174-7>
- [14] Maureen Stone. 2008. Alpha, Contrast and the Perception of Visual Metadata. (2008), 5.
- [15] Robert F. Strahan and Chris J. Hansen. 1978. Underestimating Correlation from Scatterplots. *Applied Psychological Measurement* 2, 4 (Oct. 1978), 543–550. <https://doi.org/10.1177/014662167800200409>
- [16] Gabriel Strain, Andrew J. Stewart, Paul Warren, and Caroline Jay. 2023. The Effects of Contrast on Correlation Perception in Scatterplots. *International Journal of Human-Computer Studies* 176 (Aug. 2023), 103040. <https://doi.org/10.1016/j.ijhcs.2023.103040>
- [17] Cesko C. Voeten. 2023. *buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression*. <https://CRAN.R-project.org/package=buildmer> R package version 2.9.
- [18] C. Wehrhahn and G. Westheimer. 1990. How Vernier Acuity Depends on Contrast. *Experimental Brain Research* 80, 3 (May 1990). <https://doi.org/10.1007/BF00228001>
- [19] Hadley Wickham. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- [20] Silvia Zuffi, Carla Brambilla, Giordano Beretta, and Paolo Scala. 2007. Human Computer Interaction: Legibility and Contrast. In *14th International Conference on Image Analysis and Processing (ICIAP 2007)*. 241–246. <https://doi.org/10.1109/ICIAP.2007.4362786>