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

ANONYMOUS AUTHOR(S)

Abstract goes here

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, 9 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. In the present study we hypothesize that; an increased reduction in correlation estimation error will be observed when congruent non-linear functions are used; the use of congruent inverted conditions will produce the least accurate estimates of correlation; and that owing to the greater strength of the size channel observed in **?**], there will be a significant difference in correlation estimates between the two incongruent conditions.

2 RELATED WORK

2.1 Testing Correlation Perception

2.2 Drivers of Correlation Perception

2.3 Transparency and Contrast

Changing the contrast of scatterplot points is standard practice to deal with issues of overplotting or clutter (**?**]; **?**]); 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. 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 (**?**]). The balance of evidence points towards an uncertainty-based mechanism for these effects. Lower contrast can increase error in positional judgements (**?**]), can result in greater uncertainty in speed perception (**?**])

- formalising contrast
- ggplot etc
- use of contrast floor
- rehash contrast paper main points about contrast/alpha
- explain luminance, alpha, etc
- clarify use of alpha = 0.2 floor

2.4 Point Size

- interplay between size and spatial uncertainty
- hype up how strong the size effect was last time

3 METHODOLOGY

3.1 Crowdsourcing

- issues with crowdsourcing
- how those issues were solved
- why did we choose to crowdsource

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

Both experiments were 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 our previous work [**? ?**]. 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 $r = 0.2$ [**? ? ?**]. 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.

$$point_{size/contrast} = 1 - 0.25^{residual} \tag{1}$$

In experiment 1, we used 2x2 combinations of this equation in standard and inverted forms. Scatterplot examples can be seen in Figure 2. In experiment 2 we fine tune this equation, adding the objective r value and the previously observed biases into the equation in order to flatten the underestimation curve.

3.4 Modelling

In both experiments we use linear mixed effects models to model the relationships between our independent variables (the combination of size and contrast decay conditions in experiment 1, and the use of the tuned manipulation in experiment 2) 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, Voeten [2]). This package takes a maximal random effects structure and then identifies the most complex model that converges, dropping random effects terms that fail to explain a significant amount of variance.

3.5 Point Visibility Testing

It is key that our manipulations do not remove data from the scatterplot. In order to ensure that our points were visible, we included visual threshold testing. 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 to be used in that experiment. Participants were asked to enter in a textbox how many points were present. In experiment 1, 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 our 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 calibrated on a per-monitor basis. Figure 1 shows distributions of participants' performances on the visual threshold tests. We also include performance as fixed effects in experimental models in Section 4 and

3.6 Dot Pitch

We employed a method for obtaining the dot pitch of participants' monitors (Morys-Carter [1]). 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 in experiment 1 was 0.34 (*SD* = 0.05) and xx in experiment 2. We include analyses with dot pitch as a fixed effect below.

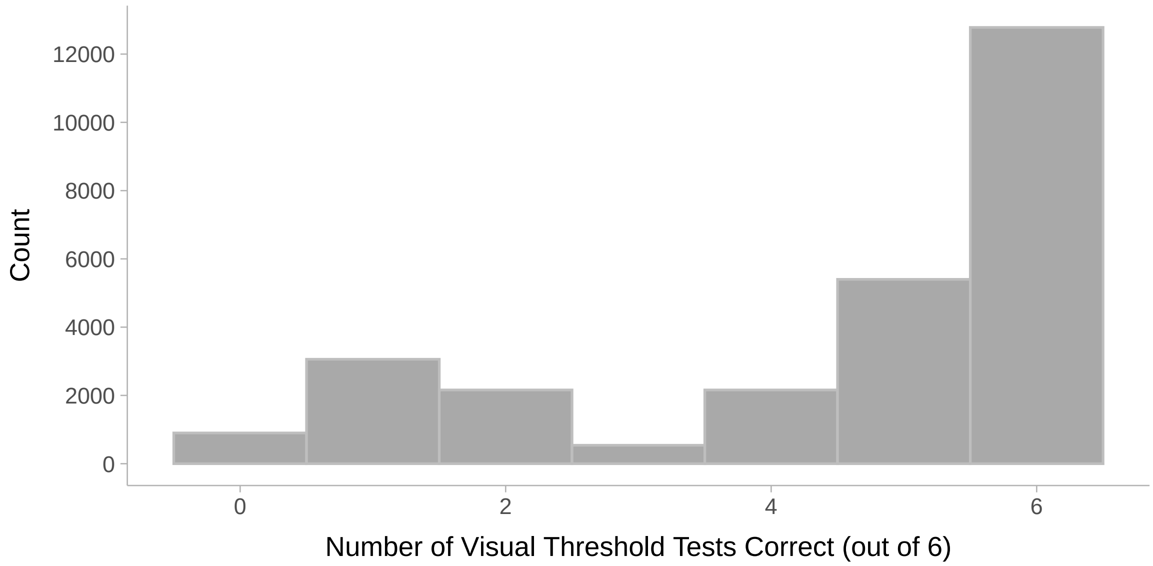

Fig. 1. Histograms of visual threshold testing performance for experiment 1 (L) and experiment 2 (R).

3.7 Procedure

Both experiments were built using PsychoPy (**?**]) 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 (**?**]), 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. To ensure high quality data, and in accordance with guidelines published in **?**], participants were required to have successfully completed a minimum of 100 studies on Prolific. In addition, participants who had completed any of our previous studies into correlation estimation in scatterplots [**? ?** , and a previous pre-study] were prevented from participating.

Data were collected from 158 participants. 8 failed more than 2 our 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 2.

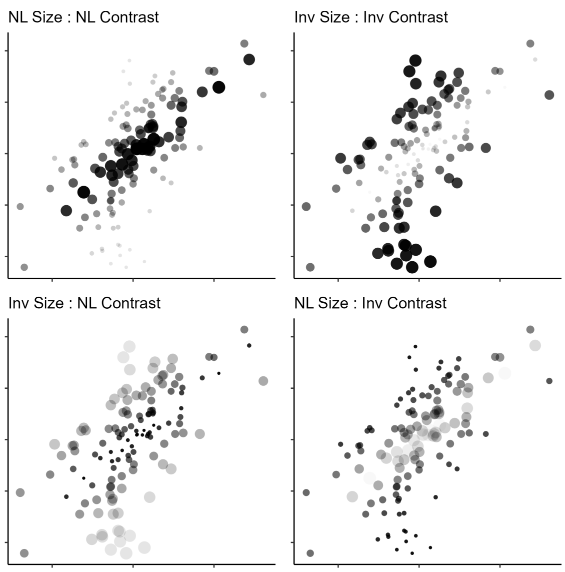

Fig. 2. Examples of the experimental items used. NL = non-linear decay, Inv = inverted decay.

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 5). 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

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

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 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.

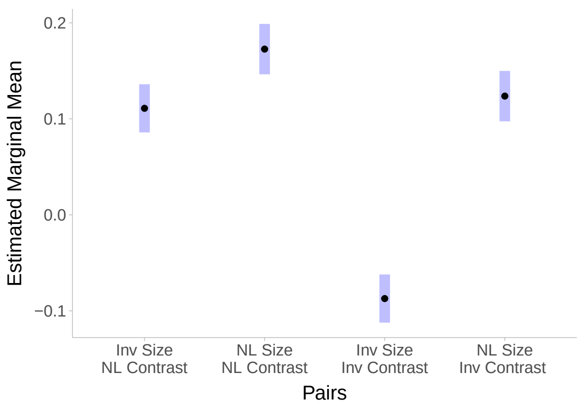

Fig. 3. **?(caption)**

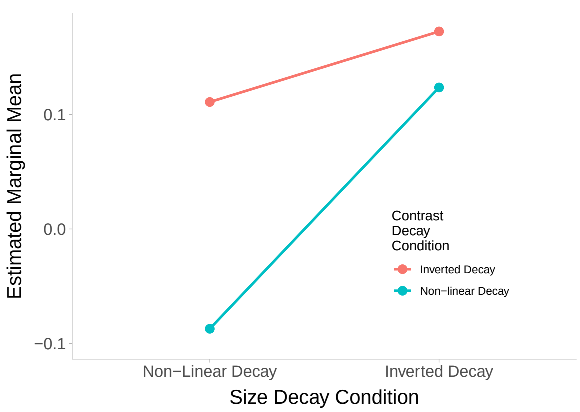

Fig. 4. Interaction plot showing the moderating effect of encoding channel on decay orientation.

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

The **emmeans** (cite) package was used to run pairwise comparisons. Figure 3 shows estimated marginal means for each combination of factors, and Figure 4 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 our previous work demonstrating the greater capacity of the size encoding channel to bias participants' estimates of r (**?**]), and our findings that the inverted channel is generally weaker at biasing correlation in the opposite direction (**?**]; **?**]). Table 2 shows statistics for pairwise comparisons.

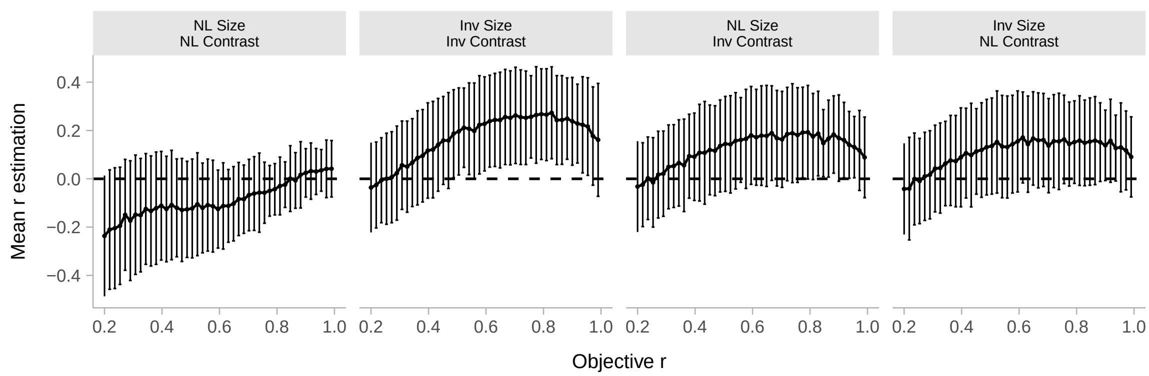

Fig. 5. Plots showing how participants' correlation estimation errors change as a function of the r value for each combination of size and contrast decay factors.

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.

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.

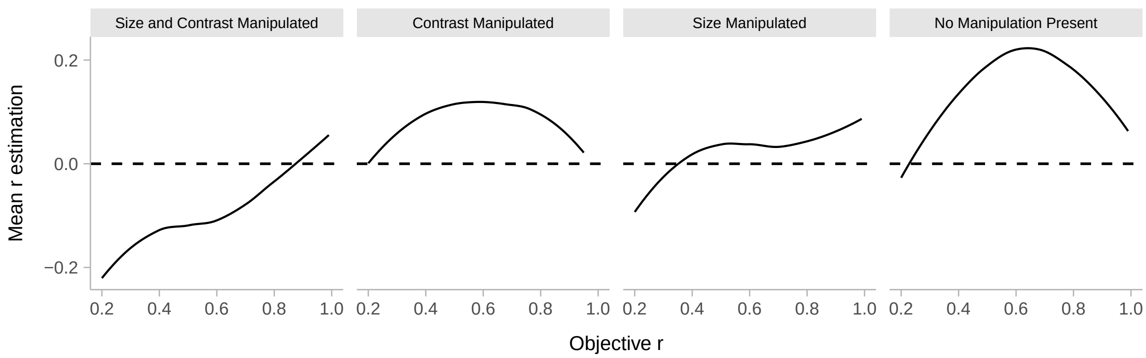

Fig. 6. From left to right, plotting r estimation error against the objective r value for the standard orientation condition in the present study, for standard orientation size and contrast manipulations in previous work, and for normal scatterplots averaged over identical conditions in previous work. Error bars have been left off this plot to make interpretation more simple.

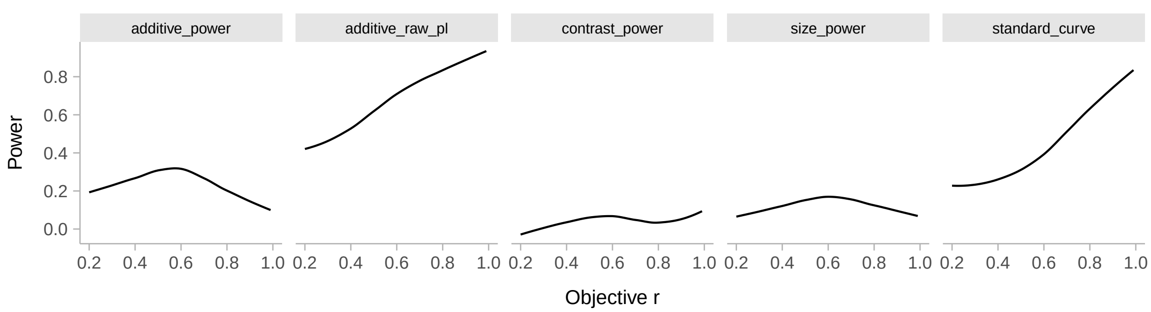

Fig. 7. additive_raw_pl = observed values for present study. standard_curve = no manipulation averaged across all experiments

As can be seen in Figure 6, 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 ?] 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.

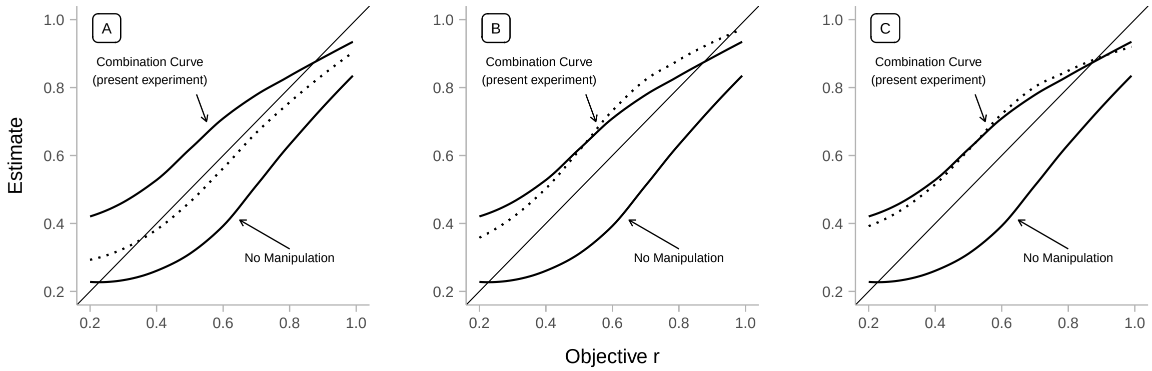

Fig. 8. Transformations required to derive the Combination Curve observed in the present experiment using the size and contrast curves from previous work. The dotted line represents the size curve (A), two times the size curve (B), and two times the size curve rotated by 6 degrees clockwise.

Figure 8 illustrates the transformations we must do to replicate the results we have found in the present study using results from previous work (**?**]; **?**]). Given the similarity between the shape of the combination curve found in the present study and that of the size only manipulation found in **?**], we use this as a starting point. Adding the contrast manipulation has the effect of doubling the power of the size curve, followed by a change in orientation that can be approximated by a 6°clockwise rotation. With the knowledge that the addition of the contrast manipulation can be thought of as doubling and rotating the size factor, and having worked out t, a translation function to approximate the $y = x$ optimum estimation line, we are able to generate the following set of equations. S = size, C = contrast, O = observed (present study), r = rotation function, t = translation function.

$$S + C = Or(S + C) = OC = Sr(2S) = Otr(2S) = tO \tag{2}$$

REFERENCES

[1] Wakefield L Morys-Carter. 2023. ScreenScale. https://doi.org/10.17605/OSF.IO/8FHQK

[2] 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.