Response to reviewers of JCGS-22-045

Eye Fitting Straight Lines in the Modern Era

Emily A. Robinson, Reka Howard, and Susan VanderPlas

2022

Reviewer(s)' Comments to Author

All responses to reviewers' comments are written in blue.

Reviewer: 1

Summary: The authors present a study where human participants had to fit straight lines through points in a scatterplot. On average, the participants' fitted lines closely matched a principal axis line and not the least squares regression line. Overall, the article is well written.

I have a series of minor comments, though - see below. Specific minor comments:

- p.3, l.11 (& more): your spelling of "scatter-plot" with the hyphen is unconventional; simply use without hyphen (that will also match the 3 occurrences in your references)

  – Changed all spellings of "scatter-plot" to "scatterplot" within the manuscript and figure captions.

- p.3, l.32: use past tense when referring to published past work, i.e., "focused" here

  – Accepted change to past tense: "focus" to "focused".

- p.3, l.37: -> "Cleveland (1993) provided" (& more)

  – Accepted changes to past tense: Cleveland (1993) "provides" to "provided"; Kosslin (2006) "examines" to "examined"; cognitive tasks "provide" to "provided".

- p.3, l.32: "viewers ability" -> "viewers' ability"

  – Accepted grammatical change.

- p.3 or p.6, bottom: add some references from the late 1980s/early 1990s that are related to the visual assessment of structural changes in time series visualizations; this reference comes to mind first (although different sources cite it differently): Unwin, A.R., and Wills, G. (1988) "Eyeballing Time Series," in Proceedings of the 1988 ASA Statistical Computing Section, 263-268. A google scholar search for "Eyeballing Time Series" results in a few related references.

    – Added "Eyeballing Time Series" reference to reference list on p.4,l18.

- p.4, l.26 (& more): use consistent capitalization here: "Wilkinson's Grammar of Graphics (Wilkinson 2013). The grammar of graphics"; I would always capitalize it as "Grammar of Graphics" as a proper name

    – Capitalized all "grammar of graphics" to "Grammar of Graphics".

- p.6, l.42: use single quotes for `You Draw It` as in the abstract

    – Added single quotes to 'You Draw It'.

- p.7, top: be specific and indicate in which of the sections of your paper each of these points is being addressed, in other words, provide a specific overview what can be found in Sections 2 - 5 of your paper (and also mention the supplementary materials in 1 or 2 sentences)

    – Added information about specific parts contained in each of the paper sections.

- p.7, l.34: you had 35 participants and got 119 plots; later (p.10, l.38) you state that each participant had to evaluate 4 plots; provide this information here immediately; can you summarize why there are 21 plots missing, e.g., some plots were incomplete, some participants were dropping out after just 1 or 2 plots, etc.

    – After double checking the data, it was actually 131 plots - we are working with pilot data and thus had this number from another set. We clarified and made this number now reproducible by using r-code in the rmd template file rather than manually entering it. We also discussed the missing plots by adding: "Data were collected as a part of a pilot study meant to test the applet; therefore, either voluntary participant dropout or disconnection from a server not designed to accommodate large magnitudes of participants resulted in missing plots in our data set for analysis."

- p.7, l.51 & p.17, l.27 & l.32: "here" is meaningless for printed materials; list the full URLs (but still make them clickable)

    – All URLs have been changed from "here" to show URL text.

- p.8, l43: provide reference for the r2d3 R package

    – Added reference for r2d3 R package.

- p.8, l.53: "R statistical software" -> "R software environment for statistical computing and graphics" (based on https://www.r-project.org/); also add the typical reference for R

  – Changed wording for use of R software and included typical reference for R.

- p.8, l.55: "i = 1, …N" -> "i = 1, …, N,"

  – Added "," into indexing on p11, l34 and p7, l55

- p.9, top: $y_{\bar{x}}$ in Table 1 and β_0 in (1) are not properly defined and used; are they the same?

  – We aimed to replicate the trends from Mosteller et al. (1981). This original manuscript gives vague descriptions of the coefficients used to describe their trends. The way they defined their parameters was by using the y-value at the mean of the x-values ($y_{\bar{x}}$) and the slope. When simulating our data, we first generate and jitter x-values across the domain, therefore, x-bar may differ for each unique data set due to the random jitter. In turn, the actual y-intercept (β_0) will vary slightly from data set to data set, but each trend will go through the same $(\bar{x}, y_{\bar{x}})$ point. We changed the equation notation to follow the point-slope equation rather than changing notation.

- p.9, l.25-37: can you further describe these four models with some more details and extend the table, e.g., list the domains in the table as well (and not only in the text); also indicate how many points were used in each scatterplot (e.g., was this random or fixed?; plot V in Fig 2 seems to have much fewer points than the other three); do the 4 letters have a certain meaning?

  – We double checked the data simulation process based on the shiny applet code. There was an extra line of code in the manuscript file that incorrrectly filtered points out. There are were 30 points simulated for each parameter choice which is stated on p7 l53. A new "set" of data was simulated randomly for each participant with the same underlying parameters and a fixed number of points. We added the domain for each parameter choice to the table. The parameter choice names were selected to reflect the labels used in Mosteller et al. (1981). We clarified this on p8 l25 with "Model equation parameters, $beta_1$, $y_{\bar{x}}$, and parameter choice letter names (F, S, V, N) and V), were selected to reflect the four data sets used and labeled in Mosteller et al…" See analysis in supplementary material for code on the data simulation procedure.

- p.10, l.38-39: I suppose each participant got one plot of type S, F, V, N ? - say so

  – We clarified that each participant got one of each of the plot types in Section 2.4.

- p.10-p.11: use PCA or PC and then use these consistently; if "principal axis" is the full term, I would even use PA as abbreviation as PCA is a widely used abbreviation for

principal component analysis and could confuse readers here; if you switch to PA, also adjust the main text below

- Notation was changed to "PA" to indicate the principal axis.

- p.11, l.35 & l.48: be consistent, e.g., "increment" vs. "increments" and missing specification of k in the 2nd part

  - Changed to "increment" for both to be consistent and added specification of k in l.48.

- p.12, l.50: it is not clear to me what you mean by "constraining the fit to a linear trend", even after reading the following part; provide some additional explanations

  - We clarified what we meant by "constraining the fit to a linear trend" by adding explantion about participants being allowed to freely draw: "The 'You Draw It' method does not restrict participants to draw a straight line as other methods would, such as using a ruler. Instead, participants are allowed to freely draw a line with potential curvature. Using the 'lmer' function in the lme4 package, a linear mixed model (LMM) was fit separately to the OLS residuals and PA residuals, emulating the effect of constraining participants to draw a linear trend."

- p.13, l.26: keep the same order of F, S, V, N as before, i.e., S F V N from Table 1 to avoid confusing the reader

  - The order of the parameter choices has been changed to be consistent (F, S, V, N) for all tables, text, and figures.

- p.14, Fig 5: this uses even another order (F N S V); stay with the same order (S F V N from Table 1) in all consecutive models and figures to avoid confusing the reader (same for Fig. 6)

  - The order of the parameter choices has been changed to be consistent (F, S, V, N) for all tables, text, and figures.

- p.14, l.28: use one single Wood reference for the R package and the most important one that describes the method; leave out the three other Wood references here

  - Selected one reference (book) which we used for the analysis.

- p.14, bottom - p.15, top: this model is very similar to the previous one; simply state y_i,j,k,drawn etc. are the same as for model (2) and only specify what is new/different here, e.g., the s_i

  - Clarified similarities between the linear mixed model equation and the generalized additive mixed model equation by stating "Defining $e_{ijk,fit}$ the same as in equation (2) above, the GAMM equation...".

- p.16, Discussion: Add some references from cognitive/human movement sciences that human arm movement is a complex task. Drawing a straight line is not as easy as it sounds. Could this lead to the large residuals when x approaches 20 ?!? Here are some candidate articles (based on the abstracts). Do some further reading and find additional related references.

  - https://doi.org/10.1007/BF00241501
  - https://doi.org/10.1145/2820619.2820633
  - These are some really great suggestions that lead to some interesting and fun reads related to this work. We reviewed the suggested readings and followed some references to other related articles. A few sentences have been added to the discussion section to address some of the cognitive/human movement sciences and influences on our study.

- p.17, Future work:

  - I worked through the 12 plots at https://emily-robinson.shinyapps.io/you-draw-it-validation-applet/ and had considerable problems to draw a straight line with a mouse, even when I intended to do so. Similarly, I fail in the classroom when I try to add a regression line to points sketched out on a blackboard. My best solution is to visually estimate the point of means (x-bar, y-bar), use a long ruler, and rotate it so that it goes through the point of means and most of the points in the scatterplot.

  - This leads to a modified version of this study: Give the participants a horizontal line segment at the bottom of the graph and let them shift and rotate this line segment until they think they have found the best fit for the point cloud. This likely would remove some of the big errors we see in Figs 5 & 6 as x approaches 20.

  - A second option for future work that goes beyond Mosteller et al. (1981) is to create scatterplots with one (or multiple) extreme outliers. What would humans fit then? Still, the PCA line or rather a robust regression line that ignores the outliers (likely not the OLS line that may be totally off from a line that goes through most of the data). [Apparently, this will also depend on the instructions and the feedback that is provided in the training part of such a study.]

  - Neither of these has to be implemented in the current study, but listing both options as natural extensions for the straight line assessment seems to be worthwhile.

  - These would be some great extensions to validating 'You Draw It' as a method for graphical testing. We added these suggestions as well as suggested a study which compares the indirect interaction (mouse and pointer on the screen) and direct interaction (digital pen on a tablet) method of using 'You Draw It'.

- p.17-21, References:

  - capitalize journal names, e.g., Psychological bulletin & IEEE transactions on visualization and computer graphics

5

- capitalize proper names, e.g., in obama's presidency
- for books, include the city/state or city/country, e.g., in Cleveland (1993)
- Deming (1943) is incomplete (is this a book or a journal article?)
- in Kosslyn & Kosslyn, spell out OUP (and indicate the city/state and not USA)
- check capitalization of nouns; if I am not mistaken, JCGS uses upper case letters for all nouns [some of these apply to other references as well]
- Reviewed references and adjusted bibtex file and style.

Reviewer: 2

The authors provide empirical evidence that data from subjects performing a "You draw it" task replicate previous findings in graph perception, namely the finding that, when asked to fit a line on a scatterplot, participants draw a line that minimizes the Euclidean distance of the points to the fit instead of the vertical one (such as in OLS). The paper is written in a good English although the sections describing the statistical analyses are quite complex and a little verbose.

My main concern about the manuscript is about its novelty. In fact, authors state that they confirm previous findings by Mosteller et al. (1981). However, the study by Ciccione and Dehaene (2021) that they cite in the manuscript, has already provided the exact same results: people minimize the orthogonal distance of points to the fit when asked to provide their best estimation of the regression line. The authors are surely aware of this (since they account for it on page 5, line 24); however, throughout the abstract, the rest of the introduction and the discussion, the authors present their results as novel and simply extending Mosteller study. In fact, what they called "a principal axis component" is mathematically equivalent to the "Deming regression" (as it is called in Ciccione & Dehaene), that minimizes the orthogonal distance of the points.

In my opinion, the actual "novel" aspects of the presented research work are: 1) the fact that an online tool initially meant for the general public can replicate empirical findings from the literature on graph perception; 2) the use of generalized additive mixed models. Concerning this second aspect, I am seriously wondering whether it was necessary at all to use such complex models to prove that participants minimized the orthogonal distance of the points to the fit. Simpler statistical analyses could have been (or could be) provided. However, since I am not an expert of these models, I recognize that it might be of interest for a specialized readership.

In conclusion, I strongly believe that the manuscript should concentrate on the actual novel aspects mentioned above (the use of an online tool and the use of generalized additive mixed models) and not on an aspect that the literature has already empirically tested and statistically confirmed and formalized. If such a major revision is undertaken by the authors, the article might be worth publishing (especially within the scope of the journal).

Below I provide a series of other aspects that should also be considered.

- Authors should avoid any form of plagiarism: for example, lines 43 to 46 (page 2) and line 55 (page 8) and line 22 (page 9) are identical excerpts from Ciccione & Dehaene.

  – Thank you for catching this. Amongst literature review and edits, we did not realize we had directly quoted. We have rephrased the information in each of the indicated locations as they are important points to include.

- Page 3, line 20. The method of adjustment is introduced but not explained, leaving the reader wondering what it refers to.

  – We describe the method of adjustment and added a reference.

- Page 3, line 50. The line-up protocol is introduced but, again, no clear explanation is provided.

  – We added a couple sentences to describe the lineup protocol in addition to the Buja (2009) reference. "A statistical lineup is a plot consisting of smaller panels where the viewer is asked to identify the target panel containing the real data from among a set of decoy null plots which display data under the assumption there is no relationship. If the viewer can identify the target panel randomly embedded within the set of null panels, this suggests that the real data is visually distinct from data generated under the null model."

- Page 5, line 21. The concept of "ensemble perception" is not clearly related to what is described just before: how can the finding itself of human fitting a principal axis on scatterplot "support" (as the authors say) the work on ensemble perception? The link needs to be clarified or avoided.

  – We removed the sentence discussing "ensemble perception" in relation to the Mosteller (1981) paper.

- Page 8, table 1. I think it's a little bit confusing to see that the intercept parameter is called yx whereas in the equation in the text is called b0. The authors should choose one notation or the other and stay consistent through the text.

  – We aimed to replicate the trends from Mosteller et al. (1981). This original manuscript gives vague descriptions of the coefficients used to describe their trends. The way they defined their parameters was by using the y-value at the mean of the x-values ($y_{\bar{x}}$) and the slope. When simulating our data, we first generate and jitter x-values across the domain, therefore, x-bar may differ for each unique data set due to the random jitter. In turn, the actual y-intercept (β_0) will vary slightly from data set to data set, but each trend will go through the same ($\bar{x}, y_{\bar{x}}$) point. We changed the equation notation to follow the point-slope equation rather than changing notation.

- Page 10, line 39. The reader is invited to refer to an R function for the description of the regression equation based on the principal component axis. Authors should provide an actual formalization of the equation.

  – We reran the analysis using the mcreg function in the mcr package which implements Deming regression (equivalent to the principal axis regression). We clarified and provided a reference for this in the "Fitted Regressions" Results section.

- Page 12, line 48 and page 15, line 45. Again (as in page 5, line 21), the authors' work is directly presented as providing support for ensemble perception. Ensemble perception simply states that human perception tends to average, in parallel, the information coming from various items/locations in the visual field. Evidence for a principal axis component extraction does not provide a direct piece of evidence for ensemble perception. Authors should better clarify this link or avoid introducing the concept of ensemble perception.

  – We agree that ensemble perception is not necessarily the focus of this paper and have removed introducing the concept.

Editor's Comments to Author

Thank you for your submission and interesting work. Two reviewers have provided feedback and suggestions on your manuscript. There are a few points I would like to emphasize and request clarification:

1. In your manuscript, it would be helpful if you could clarify and emphasize the novelty of your experiment, data collection, and analysis. In particular, how does your work differ from Ciccione and Dehaene (2021)?(Ciccione, L. and Dehaene, S., 2021. Can humans perform mental regression on a graph? Accuracy and bias in the perception of scatterplots. Cognitive Psychology, 128, p.101406.)

- We clarifed the differences between our work and Ciccione (2021) as our study use a more flexible method of testing graphics and introduces a way to analyze the data using GAMM's. While this study was designed to replicate results found in previous studies as a way to validate our method, our intent is to set the foundation for using 'You Draw It' for testing statistical graphics as the tool extends nicely to the non-linear setting. We added this clarification into the end of Section 1.2.

2. In Section 3, it is helpful to note the R functions used as comments or footnotes, but you should also clearly describe the method of estimation. In other words, what are the R functions doing to estimate the model?

- We added the method of estimations based on the R documentation for the lmer and bam functions.

3. Could you clarify if the participants were known to be experts in statistics or related fields?

- We added a sentence describing that the data is from a pilot study and thus participants from the recruitment methods likely have an interest in advancing research in data visualization: "As this is a pilot study, participants from Twitter and Reddit pages related to data visualization voluntarily completed the study and likely have an interest in fields related to statistics and want to help advance research in graphics."

4. Throughout the manuscript, references to "principle components" should be "principal components."

- Based on a reviewer's suggestion, we have changed terminology to principal axis.