Eye Fitting Straight Lines in the Modern Era

by Emily A. Robinson, Reka Howard, Susan VanderPlas

December 14, 2021

1 Introduction

We all use statistical graphics, but how do we know that the graphics we use are communicating properly? When creating a graphic, we must consider the design choices most effective for conveying the intended result. For instance, we must decide to highlight the relationship between two variables in a scatterplot by including a trend line, or adding color to highlight clustering (VanderPlas and Hofmann 2017). These design choices require that we understand the perception and visual biases that come into play when creating graphics, and as graphics are evaluated visually, we must use human testing to ground our understanding in empiricism.

Much of the research on the perception of visual features in charts has been conducted in psychophysics and test for accuracy and quantitative comparisons when understanding a plot. Cleveland and McGill conducted a series of cognitive tasks designed to establish a hierarchy of visual components for making comparisons (Cleveland and McGill 1984). For example, it is more effective to display information on an x or y axis rather than using color in order to reduce the visual effort necessary to make numerical comparisons. Other early studies evaluated human perception of lines. For instance, Cleveland and McGill (1985) found that assessing the position of points along an axis is easier than determining the slope of a line. The results of these cognitive tasks provide some consistent guidance for chart design; however, other methods of visual testing can further evaluate design choices and help us understand cognitive biases related to the evaluation of statistical charts.

1.1 Testing Statistical Graphics

Graphical tests are useful for studying the perception of statistical graphs. Studies might ask participants to identify differences in graphs, read information off of a chart accurately, use data to make correct real-world decisions, or predict the next few observations. All of these types of tests require different levels of use and manipulation of the information being presented in the chart. Early researchers studied graphs from a psychological perspective (Spence 1990; Lewandowsky and Spence 1989). These studies generally tested participants ability to detect a stimulus or a difference between two stimuli. In statistical graphics research, VanderPlas and Hofmann (2015a) uses methods of psychophysics to estimate the effect of sign illusions by adjusting the sine illusion up and down.

A major development in statistical graphics research is Wilkinson's Grammar of Graphics (Wilkinson 2013). The grammar of graphics serves as the fundamental framework for data visualization with the notion that graphics are built from the ground up by specifying exactly how to create a particular graph from a given data set. Graphics are viewed as a mapping from variables in a data set (or statistics computed from the data) to visual attributes such as the axes, colors, shapes, or facets on the canvas in which the chart is displayed. Software, such as Hadley Wickham's ggplot2 (Wickham 2011), aims to implement the framework of creating charts and graphics as the grammar of graphics recommends.

Efforts in the field of statistical graphics have developed graphical testing tools and methods such as the lineup protocol (Buja et al. 2009) to provide a framework for inferential testing. Through experimentation, methods such as the lineup protocol allow researchers to conduct studies geared at understanding human ability to conduct tasks related to the perception of statistical charts such as differentiation, prediction, estimation, and extrapolation (VanderPlas and Hofmann 2017, 2015b; Hofmann et al. 2012). The advancement of graphing software provides the tools necessary to develop new methods of testing statistical graphics.

1.2 Fitting Trends by Eye

Initial studies in the 20th century explored the use of fitting lines by eye through a set of points (Finney 1951; Mosteller et al. 1981). Common methods of fitting trends by eye involved maneuvering a string, black thread, or ruler until the fit is suitable, then drawing the line through the set of points. Recently, Ciccione and Dehaene (2021) conducted a comprehensive set of studies investigating human ability to detect trends in graphical representations from a psychophysical approach.

Mosteller et al. (1981) sought to understand the properties of least squares and other computed lines by establishing one systematic method of fitting lines by eye. Participants were asked to fit lines by eye to four scatterplots using an 8.5 x 11 inch transparency with a straight line etched completely across the middle. A latin square design with packets of the set of points stapled together in four different sequences was used to determine if there is an effect of order of presentation. It was found that order of presentation had no effect and that participants tended to fit the slope of the principal axis (error minimized orthogonally, both horizontal and vertical, to the regression line) over the slope of the least squares regression line (error minimized vertically to the regression line). These results support previous research on "ensemble perception" indicating the visual system can compute averages of various features in parallel across the items in a set (Chong and Treisman 2003, 2005; Van Opstal, Lange, and Dehaene 2011).

In 2015, the New York Times introduced an interactive feature, called You Draw It (Aisch, Cox, and Quealy 2015; Buchanan, Park, and Pearce 2017; Katz 2017). Readers are asked to input their own assumptions about various metrics and compare how these assumptions relate to reality. The New York Times team utilizes Data Driven Documents (D3) that allows readers to predict these metrics through the use of drawing a line on their computer screen with their computer mouse. After the reader has completed drawing the line, the actual observed values are revealed and the reader may check their estimated knowledge against the actual reported data.

1.3 Research objectives

In this paper, we establish 'You Draw It,' adapted from the New York Times feature, as a tool for graphical testing. The 'You Draw It' method is validated by replicating the study conducted by Mosteller et al. (1981). Based on previous research surrounding "ensemble perception," we hypothesize that visual regression tends to mimic principle component regression rather than a ordinary least squares regression. In order to assess this hypothesis, we introduce a method for statistically modeling the participant drawn lines using generalized additive mixed models.

2 Methods

2.1 Participants

Participants were recruited through through Twitter, Reddit, and direct email in May 2021. A total of 35 individuals completed 119 unique 'You Draw It' task plots; all completed you draw it task plots were included in the analysis. All participants had normal or corrected to normal vision and signed an informed consent form. The experimental tasks took approximately 15 minutes to complete. While convenience sampling took place, this is primarily a perceptual task and previous results have found few differences between expert and non-expert participants in this context (VanderPlas and Hofmann 2015b). Data collected in this study serves the purpose of a pilot study intended to lay the foundation for future data collection and provides promising results for understanding the perception of regression. Participants completed the experiment on their own computers in an environment of their choosing. The experiment was conducted and distributed through an RShiny application found here.

2.2 'You Draw It' Task

Data Driven Documents (D3), a JavaScript-based graphing framework that facilitates user interaction, is used to create the 'You Draw It' task plots. Integrating this into RShiny using the `r2d3` package, participants are asked to draw a trend-line using their computer mouse through a scatter-plot shown on their screen. In the study, participants are shown an interactive scatter-plot (Fig. 1) along with the prompt, "Use your mouse to fill in the trend in the yellow box region." The yellow box region moves along as the participant draws their trend-line until the yellow region disappears, indicating the participant has filled in the entire domain. Details of the development of the 'You Draw It' task plots will be addressed in future work.

2.3 Data Generation

All data processing was conducted in R statistical software. Data were simulated based on a linear model with additive errors with model equation parameters selected to reflect the four data sets (F, N, S, and V) used in Mosteller et al. (1981). See Appendix A for details.

2.4 Study Design

This experiment was conducted as part of a larger study; for simplicity, we focus on the study design and methods related to the you-draw-it paradigm. Each scatter-plot was the graphical representation of a data set that was generated randomly, independently for each participant at the start of the experiment. Participants in the study are shown two 'You Draw It' practice plots in order to train participants in the skills associated with executing the task. During the practice session, participants are provided with instruction prompts accompanied by a .gif and a practice plot. Instructions guide participants to start at the edge of the yellow box, to make sure the yellow region is moving along with their mouse as they draw, and that they can draw over their already drawn line. Practice plots are then followed by four 'You Draw It'

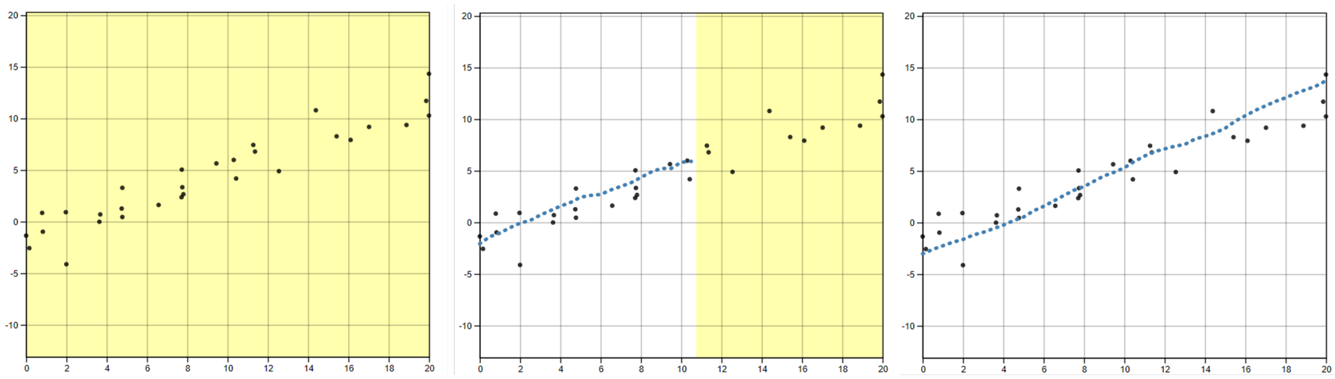

Figure 1: 'You Draw It' task plot as shown to particpants during the study. The first frame (left) illustrates what particpants first see with the prompt "Use your mouse to fill in the trend in the yellow box region." The second frame (middle), illustrates what the particpant sees while completing the task; the yellow region provides a visual cue for participants indicating where the participant still needs to complete a trend-line. The last frame (right) illustrates the participants finished trend-line before submission.

task plots associated with the current study. The order of the task plots was randomly assigned for each individual in a completely randomized design.

3 Results

3.1 Fitted Regression Lines

We compare the participant drawn line to two regression lines determined by ordinary least squares regression and regression based on the principal axis (i.e. Deming Regression). Appendix B illustrates the difference between an OLS regression line which minimizes the vertical distance of points from the line and a regression line based on the principal axis which minimizes the Euclidean distance of points (orthogonal) from the line.

3.2 Residual Trends

For each participant, the final data set used for analysis contains corresponding drawn and fitted values (OLS and PCA) for each participant and parameter combination. Using both a linear mixed model and a generalized additive mixed model, comparisons of vertical residuals in relation to the OLS fitted values and PCA fitted values were made across the domain. See Appendix C for details.

3.3 Linear Trend Constraint

Using the `lmer` function in the lme4 package (Bates et al. 2015), a linear mixed model (LMM) is fit separately to the OLS residuals and PCA residuals, constraining the fit to a linear trend. Parameter choice, x, and the interaction between x and parameter choice were treated as fixed effects with a random participant effect accounting for variation due to participant. Details on the LMM can be found in Appendix C.1.

Constraining the residual trend to a linear fit, Fig. 2 shows the estimated trend line of the residuals between the participant drawn points and fitted values for both the OLS regression line and PCA regression line. Estimated residual trend lines are overlaid on the observed individual participant residuals. Results indicate the estimated trends of PCA residuals (orange) appear to align closer to the $y = 0$ horizontal (dashed) line than the OLS residuals (blue). In particular, this trend is more prominent in parameter choices with large variances (F and N). These results are consistent to those found in Mosteller et al. (1981) indicating participants fit a trend-line closer to the estimated regression line with the slope of based on the first principal axis than the estimated OLS regression line thus, providing support for "ensemble perception."

3.4 Smoothing Spline Trend

Eliminating the linear trend constraint, the `bam` function in the mgcv package (Wood 2011) is used to fit a generalized additive mixed model (GAMM) separately to the OLS residuals and PCA residuals to allow for estimation of smoothing splines. Parameter choice was treated as a fixed effect with no estimated intercept and a separate smoothing spline for x was estimated for each parameter choice. A random participant effect accounting for variation due to participant and a random spline for each participant accounted for variation in spline for each participant. Details on the LMM can be found in Appendix C.2.

Allowing for flexibility in the residual trend, Fig. 3 shows the estimated trend line of the residuals between

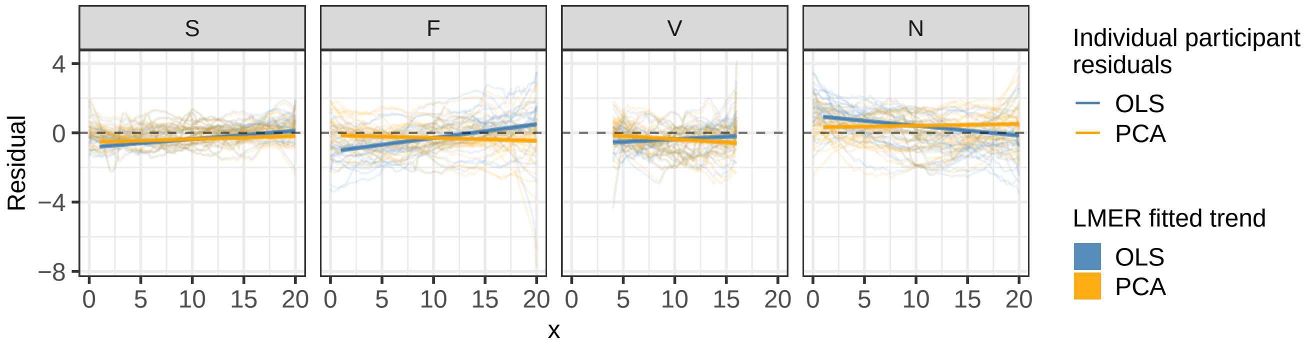

Figure 2: Estimated trend line of the residuals between the participant drawn points and fitted values for both the OLS (blue) regression line and PCA (orange) regression line constrained to a linear fit modeled by a linear mixed model. Estimated residual trends with 95% confidence bands are overlaid on the observed individual participant residuals.

the participant drawn points and fitted values for both the OLS regression line and PCA regression line. Estimated residual trends are overlaid on the observed individual participant residuals. The results of the GAMM align with those shown in Fig. 2 providing support that for scatter-plots with more noise (F and N), estimated trends of PCA residuals (orange) appear to align closer to the $y = 0$ horizontal (dashed) line than the OLS residuals (blue). By fitting smoothing splines, we can determine whether participants naturally fit a straight trend-line to the set of points or whether they deviate throughout the domain. In particular, in scatter-plots with smaller variance (S and V), we can see that participants began at approximately the correct starting point then deviated away from the fitted regression lines and correcting for their fit toward the end of their trend-line. In scatter-plots with larger variance (F and N), participants estimated their starting value in the extreme direction of the OLS regression line based on the increasing or decreasing trend but more accurately represented the starting value of the PCA regression line. As participants continued their trend-line, they crossed through the OLS regression line indicating they estimated the slope in the extreme direction. These results provide further insight into the curvature humans perceive in a set of points.

4 Discussion and Conclusion

The intent of research was to adapt 'You Draw It' from the New York Times feature as a tool and method for testing graphics and introduce a method for statistically modeling the participant drawn lines. We provided support for the validity of the 'You Draw It' method by replicating the study found in Mosteller et al. (1981). Using generalized additive mixed models, we assessed the deviation of the participant drawn lines from the statistically fitted regression lines. Our results found that when shown points following a linear trend, participants visually fit a regression line that mimics the first principle component regression as opposed to ordinary least squares regression. Data simulated with a larger variance provided strong support for a participants tendency to visually fit the first principle component regression. Our results indicate that participants minimized the distance from the their regression line over both the x and y axis simultaneously. These results provide support that humans perform "ensemble perception" in a statistical graphic setting. We allowed participants to draw trend lines that deviated from a straight line and gained an insight into the curvature the human eye perceives in a set of points.

5 Future Work

This study provided a basis for the use of 'You Draw It' as a tool for testing statistical graphics and introduced a method for statistically modeling participant drawn lines using generalized additive mixed models. Further investigation is necessary to implement this method in nonlinear settings and with real data. This tool could also be used to evaluate human ability to extrapolate data from trends. In the future, an R package designed for easy implementation of 'You Draw It' task plots will help make this tool accessible to other researchers.

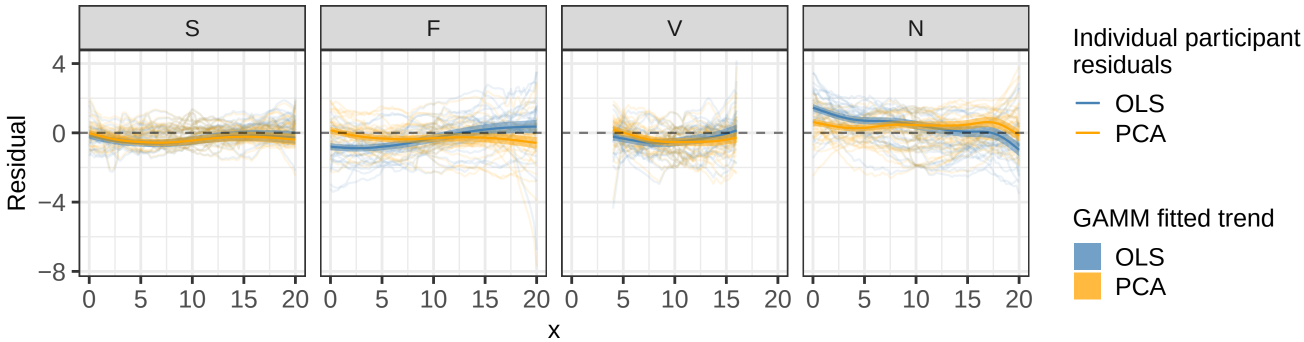

Figure 3: Estimated trend line of the residuals between the participant drawn points and fitted values for both the OLS (blue) regression line and PCA (orange) regression line determined by smoothing splines fit by a generalized additive mixed model. Estimated residual trends with 95% confidence bands are overlaid on the observed individual participant residuals.

References

Aisch, Gregor, Amanda Cox, and Kevin Quealy. 2015. "You Draw It: How Family Income Predicts Children's College Chances." *The New York Times*. The New York Times.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67 (1): 1–48. `https://doi.org/10.18637/jss.v067.i01`.

Buchanan, Larry, Haeyoun Park, and Adam Pearce. 2017. "You Draw It: What Got Better or Worse During Obama's Presidency." *The New York Times*. The New York Times.

Buja, Andreas, Dianne Cook, Heike Hofmann, Michael Lawrence, Eun-Kyung Lee, Deborah F Swayne, and Hadley Wickham. 2009. "Statistical Inference for Exploratory Data Analysis and Model Diagnostics." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367 (1906): 4361–83.

Chong, Sang Chul, and Anne Treisman. 2003. "Representation of Statistical Properties." *Vision Research* 43 (4): 393–404.

———. 2005. "Statistical Processing: Computing the Average Size in Perceptual Groups." *Vision Research* 45 (7): 891–900.

Ciccione, Lorenzo, and Stanislas Dehaene. 2021. "Can Humans Perform Mental Regression on a Graph? Accuracy and Bias in the Perception of Scatterplots." *Cognitive Psychology* 128: 101406.

Cleveland, William S, and Robert McGill. 1984. "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods." *Journal of the American Statistical Association* 79 (387): 531–54.

———. 1985. "Graphical Perception and Graphical Methods for Analyzing Scientific Data." *Science* 229 (4716): 828–33.

Finney, DJ. 1951. "Subjective Judgment in Statistical Analysis: An Experimental Study." *Journal of the Royal Statistical Society: Series B (Methodological)* 13 (2): 284–97.

Hofmann, Heike, Lendie Follett, Mahbubul Majumder, and Dianne Cook. 2012. "Graphical Tests for Power Comparison of Competing Designs." *IEEE Transactions on Visualization and Computer Graphics* 18 (12): 2441–48.

Katz, Josh. 2017. "You Draw It: Just How Bad Is the Drug Overdose Epidemic?" *The New York Times*. The New York Times.

Lewandowsky, Stephan, and Ian Spence. 1989. "The Perception of Statistical Graphs." *Sociological Methods & Research* 18 (2-3): 200–242.

Mosteller, Frederick, Andrew F Siegel, Edward Trapido, and Cleo Youtz. 1981. "Eye Fitting Straight Lines." *The American Statistician* 35 (3): 150–52.

Spence, Ian. 1990. "Visual Psychophysics of Simple Graphical Elements." *Journal of Experimental Psychology: Human Perception and Performance* 16 (4): 683.

Van Opstal, Filip, Floris P de Lange, and Stanislas Dehaene. 2011. "Rapid Parallel Semantic Processing of Numbers Without Awareness." *Cognition* 120 (1): 136–47.

VanderPlas, Susan, and Heike Hofmann. 2015a. "Signs of the Sine Illusion—Why We Need to Care." *Journal of Computational and Graphical Statistics* 24 (4): 1170–90.

———. 2015b. "Spatial Reasoning and Data Displays." *IEEE Transactions on Visualization and Computer Graphics* 22 (1): 459–68.

———. 2017. "Clusters Beat Trend!? Testing Feature Hierarchy in Statistical Graphics." *Journal of Computational and Graphical Statistics* 26 (2): 231–42.

Wickham, Hadley. 2011. "Ggplot2." *Wiley Interdisciplinary Reviews: Computational Statistics* 3 (2): 180–85.

Wilkinson, Leland. 2013. *The Grammar of Graphics*. Springer Science & Business Media.

Wood, S. N. 2011. "Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models." *Journal of the Royal Statistical Society (B)* 73 (1): 3–36.

Table 1: Designated model equation parameters for simulated data.

Parameter Choice	$y_{\bar{x}}$	β_1	σ
S	3.88	0.66	1.30
F	3.90	0.66	1.98
V	3.89	1.98	1.50
N	4.11	-0.70	2.50

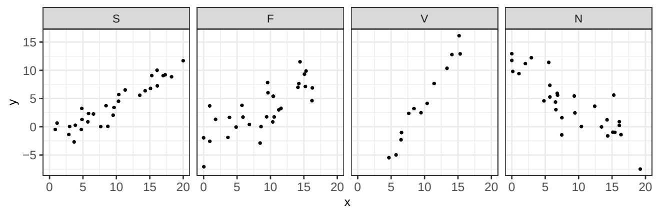

Figure 4: Example of simulated data points displayed in a scatter-plot illustrating the trends associated with the four selected parameter choices.

A Data Generation Procedure

All data processing was conducted in R statistical software. A total of $N = 30$ points $(x_i, y_i), i = 1, ...N$ were generated for $x_i \in [x_{min}, x_{max}]$ where x and y have a linear relationship. Data were simulated based on a linear model with additive errors:

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad (1)$$
$$\text{with } e_i \sim N(0, \sigma^2).$$

Model equation parameters, β_0 and β_1, were selected to reflect the four data sets (F, N, S, and V) used in Mosteller, Siegel, Trapido, & Youtz (1981) (Table 1). Parameter choices F, N, and S simulated data across a domain of 0 to 20. Parameter choice F produces a trend with a positive slope and a large variance while N has a negative slope and a large variance. In comparison, S shows a trend with a positive slope with a small variance and V yields a steep positive slope with a small variance over the domain of 4 to 16. Fig. 4 illustrates an example of simulated data for all four parameter choices intended to reflect the trends in Mosteller, Siegel, Trapido, & Youtz (1981). Aesthetic design choices were made consistent across each of the interactive 'You Draw It' task plots. The y-axis range extended 10% beyond (above and below) the range of the simulated data points to allow for users to draw outside the simulated data set range.

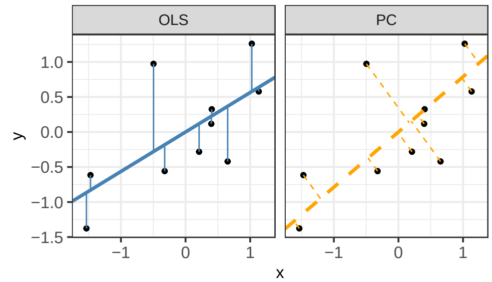

Figure 5: Comparison between an OLS regression line which minimizes the vertical distance of points from the line and a regression line based on the principal axis which minimizes the Euclidean distance of points (orthogonal) from the line.

B Fitted Regression Lines

We compare the participant drawn line to two regression lines determined by ordinary least squares regression and regression based on the principal axis (i.e. Deming Regression). Fig. 5 illustrates the difference between an OLS regression line which minimizes the vertical distance of points from the line and a regression line based on the principal axis which minimizes the Euclidean distance of points (orthogonal) from the line.

Due to the randomness in the data generation process, the actual slope of the linear regression line fit through the simulated points could differ from the predetermined slope. Therefore, we fit an ordinary least squares (OLS) regression to each scatter-plot to obtain estimated parameters $\hat{\beta}_{0,OLS}$ and $\hat{\beta}_{1,OLS}$. Fitted values, $\hat{y}_{k,OLS}$, are then obtained every 0.25 increments across the domain from the OLS regression equation, $\hat{y}_{k,OLS} = \hat{\beta}_{0,OLS} + \hat{\beta}_{1,OLS} x_k.$, for $k = 1, ..., 4x_{max} + 1$. The regression equation based on the principal axis was determined by using the `princomp` function in the stats package in base R to obtain the rotation of the coordinate axes from the first principal component (direction which captures the most variance). The estimated slope, $\hat{\beta}_{1,PCA}$, is determined by the ratio of the axis rotation in y and axis rotation in x of the first principal component with the y-intercept, $\hat{\beta}_{0,PCA}$ calculated by the point-slope equation of a line using the mean of of the simulated points, $(\bar{x}_i, \bar{y}_i)$. Fitted values, $\hat{y}_{k,PCA}$, are then obtained every 0.25 increment across the domain from the PCA regression equation, $\hat{y}_{k,PCA} = \hat{\beta}_{0,PCA} + \hat{\beta}_{1,PCA} x_k$.

C Residual Trends

For each participant, the final data set used for analysis contains $x_{ijk}, y_{ijk,drawn}, \hat{y}_{ijk,OLS}$, and $\hat{y}_{ijk,PCA}$ for parameter choice $i = 1, 2, 3, 4$, j $= 1, ...N_{participant}$, and x_{ijk}

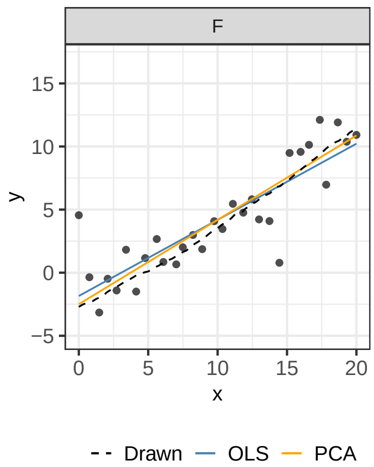

Figure 6: Illustrates the data associated with and collected for one 'You Draw It' task plot. Trend-lines include the participant drawn line (dashed black), the OLS regression line (solid steelblue) and the PCA regression line based on the principal axis (solid orange).

value $k = 1, ..., 4x_{max} + 1$. Using both a linear mixed model and a generalized additive mixed model, comparisons of vertical residuals in relation to the OLS fitted values ($e_{ijk,OLS} = y_{ijk,drawn} - \hat{y}_{ijk,OLS}$) and PCA fitted values ($e_{ijk,PCA} = y_{ijk,drawn} - \hat{y}_{ijk,PCA}$) were made across the domain. Fig. 6 displays an example of all three fitted trend lines for parameter choice F. Data used in the analyses are available to be downloaded from GitHub here.

C.1 Linear Mixed Model

Using the `lmer` function in the lme4 package (Bates, MÃ¤chler, Bolker, & Walker, 2015), a linear mixed model (LMM) is fit separately to the OLS residuals and PCA residuals, constraining the fit to a linear trend. Parameter choice, x, and the interaction between x and parameter choice were treated as fixed effects with a random participant effect accounting for variation due to participant. The LMM equation for each fit (OLS and PCA) is given by:

$$e_{ijk,fit} = [\gamma_0 + \alpha_i] + [\gamma_1 x_{ijk} + \gamma_{2i} x_{ijk}] + p_j + \epsilon_{ijk} \quad (2)$$

where

- $y_{ijk,drawn}$ is the drawn y-value for the i^{th} parameter choice, j^{th} participant, and k^{th} increment of x-value
- $\hat{y}_{ijk,fit}$ is the fitted y-value for the i^{th} parameter choice, j^{th} participant, and k^{th} increment of x-value corresponding to either the OLS or PCA fit
- $e_{ijk,fit}$ is the residual between the drawn and fitted y-values for the i^{th} parameter choice, j^{th} participant, and k^{th} increment of x-value corresponding to either the OLS or PCA fit
- γ_0 is the overall intercept
- α_i is the effect of the i^{th} parameter choice (F, S, V, N) on the intercept
- γ_1 is the overall slope for x
- γ_{2i} is the effect of the parameter choice on the slope
- x_{ijk} is the x-value for the i^{th} parameter choice, j^{th} participant, and k^{th} increment
- $p_j \sim N(0, \sigma^2_{participant})$ is the random error due to the j^{th} participant's characteristics
- $\epsilon_{ijk} \sim N(0, \sigma^2)$ is the residual error.

C.2 Generalized Additive Mixed Model

Eliminating the linear trend constraint, the `bam` function in the mgcv package (Wood, 2011) is used to fit a generalized additive mixed model (GAMM) separately to the OLS residuals and PCA residuals to allow for estimation of smoothing splines. Parameter choice was treated as a fixed effect with no estimated intercept and a separate smoothing spline for x was estimated for each parameter choice. A random participant effect accounting for variation due to participant and a random spline for each participant accounted for variation in spline for each participant. The GAMM equation for each fit (OLS and PCA) residuals is given by:

$$e_{ijk,fit} = \alpha_i + s_i(x_{ijk}) + p_j + s_j(x_{ijk}) \quad (3)$$

where

- $y_{ijk,drawn}$ is the drawn y-value for the i^{th} parameter choice, j^{th} participant, and k^{th} increment of x-value
- $\hat{y}_{ijk,fit}$ is the fitted y-value for the i^{th} parameter choice, j^{th} participant, and k^{th} increment of x-value corresponding to either the OLS or PCA fit
- $e_{ijk,fit}$ is the residual between the drawn and fitted y-values for the i^{th} parameter choice, j^{th} participant, and k^{th} increment of x-value corresponding to either the OLS or PCA fit
- α_i is the intercept for the parameter choice i
- s_i is the smoothing spline for the i^{th} parameter choice
- x_{ijk} is the x-value for the i^{th} parameter choice, j^{th} participant, and k^{th} increment
- $p_j \sim N(0, \sigma^2_{participant})$ is the error due to participant variation
- s_j is the random smoothing spline for each participant.