Title
Information sampling explains Bayesian learners' biases in correlation judgment

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors
Zhang, Xiumei
He, Lisheng
Bhatia, Sudeep

Publication Date
2022

Peer reviewed

Information sampling explains Bayesian learners' biases in correlation judgment

Xiumei Zhang (0203100537@shisu.edu.cn) [1]
School of Business and Management, Shanghai International Studies University
Shanghai, China

Lisheng He (hlisheng@shu.edu.cn) [1]
SILC Business School, Shanghai University
Shanghai, China

Sudeep Bhatia (bhatiasu@sas.upenn.edu)
Department of Psychology, University of Pennsylvania
Philadelphia, PA, USA

Abstract

Correlation judgments are at the core of belief formation. In previous studies of correlation judgment from 2D scatterplots, observers underestimate correlations, and display stronger underestimation biases when the scatterplot is shown in a landscape view than in a portrait view. Yet, it is unclear how these biases arise. Here, we propose that observers are Bayesian learners who perform "mental regression" using the observed data points in graph. Accordingly, judgment errors can arise from biased visual information sampling. We test our model's predictions with two eye-tracking experiments and find that the Bayesian learning model, applied to information obtained from visual fixation data, replicates classic behavioral findings. The model also predicts trial-level estimation biases at a high accuracy level. Our study shows how computational models trained on process-level data can shed light on the cognitive mechanisms underlying belief formation, and yield theory-driven practical implications for data visualization and statistical communication.

Keywords: Correlation judgment; Bayesian learning; intuitive statistician; information sampling; computational modeling

Introduction

Correlation judgments are at the core of everyday belief formation. Individuals may use the correlation between two health-related variables to judge risk and make medical decisions (Dawes et al., 1989); investors may rely on a stock's past performance to predict future performance, guiding financial decisions (Shefrin & Statman, 1985); business leaders may use historical cases to judge the effectiveness of a strategy and make business decisions (Lifchits et al. 2021); and policy makers may use perceived correlations to infer causal mechanisms and make important social, economic, and political decisions. Correlation perception also underlies various social psychological phenomena including stereotyping, conditioning and attribution (Trolier & Hamilton, 1986; Denrell, 2005; Mahrholz et al., 2018). More generally, the perceived correlation between signals allows individuals and groups to explain data, predict future events, and make evidence-based decisions.

Despite their importance, correlation judgments are often inaccurate, leading to biases in beliefs and decisions. Biases happen either when the information is given sequentially (Erlick, 1966) or presented simultaneously in a scatterplot (Bobko & Karren, 1979; Cleveland et al., 1982). In this paper, we focus on scatterplots. In such judgments, observers repeatedly underestimate the actual correlation shown in the graph (Rensink & Baldridge, 2010).

Various visual features in the scatterplot play a role in the subjective perception of correlations. In a recent review, Yang et al. (2018) identify 29 features that can be categorized into eight concepts, including the length of prediction ellipse, side length of the bounding box and the standard deviation of pairwise distance. They also find that some visual features alone may outperform the actual magnitude of correlation in predicting the subjective perception of correlation from the graph. As many of the features can be simply manipulated by the stimulus display format (e.g. by changing the axis range) with no change to the data points in graph, this result has had important implications for statistical communication.

Why does the perceived correlation follow the psychophysical laws that mostly describe low-level stimuli? Why does display format influence perceived correlation? Recently, scientists have begun to investigate the cognitive mechanisms underlying biased correlation judgments. For example, Rensink (2017) finds that observers extract image entropy in a way that is analogous to ensemble coding (Alvarez, 2011; Haberman & Whitney, 2012) and make inferences about correlation based on this entropy. In related trend judgment tasks, Ciccione et al. (2021) suggest that the observers are able to performs "mental regression" on the scatter graphs, even within a short time window. These theoretical claims are in line with the "intuitive statistician" hypothesis, an influential theory that claims that our cognitive

[1] X.Z. and L.H. share the first authorship.

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

system makes statistical inferences from sensory input in a way akin to probabilistic models (Peterson & Beach, 1967; Oaksford & Chater, 2007; Griffiths et al. 2010).

Despite the key role of visual input, to our best knowledge there is no existing research that has directly measured observers' eye fixation data while they make correlation judgment from scatterplots. At a given moment, the visual input to the observer corresponds to the *attended* points in the foveal or perifovea region, rather than all the points in the graph. Therefore, the visual information samples using techniques such as eye tracking should be essential to the study of correlation judgment from scatterplots. Eye-tracking also allows us to investigate trial-level variations due to differences in visual information sampling.

In this paper, we provide a computational account for correlation judgments from scatterplots. Our approach is motivated by the "intuitive statistician" hypothesis in which our cognitive system is equipped with an ideal learning machinery (Peterson & Beach, 1967). Here, we propose that decision makers are Bayesian learners who perform "mental regressions" that take as input the attended points (i.e. pairs of x and y coordinates) and update the slope of the regression line as output. As, the sampling of evidence may be subject to biases due to complex attention and memory processes (Wei & Stocker, 2015, Yang et al. 2016; Sanborn & Chater, 2016; Chater et al. 2020), a Bayesian learner applied to visual fixation data can generate biases in the eventual judgment. We test our theory in two eye-tracking experiments that gathered participants' gaze patterns while viewing the scatterplots. In so doing, we make a new formal connection between visual information sampling and correlation judgment from scatterplots.

Psychophysical Laws

The scatterplot has been one of the greatest inventions in the history of statistical graphics (Friendly & Denis, 2005). It displays bivariate relationships in a way that readers can obtain with ease. Not only can people discriminate which graph shows stronger relatedness (Doherty et al. 2007; Rensink & Baldridge, 2010), but they can also estimate the relatedness on an absolute scale (Cleveland et al. 1982). Scatterplots have been widely used in academia, business and even legal court cases (Bobko & Karren, 1979).

The most widely studied bivariate relatedness is the Pearson product-moment correlation, R. Experimental work has repeatedly found that observers underestimate the correlation from scatterplots, especially at intermediate levels with $0.2 \leq R \leq 0.6$ (Bobko & Karren, 1979). This underestimation bias is robust among both novices and statistically sophisticated observers (Strahan & Hansen 1978; Cleveland et al. 1982). It has been argued that this underestimation bias follows classic psychophysical laws and a number of mathematical functions have been proposed to describe the subjective perception of correlation from scatterplots (see Rensink, 2017 for a review). For example, through a series of psychophysics experiments, Rensink and Baldridge (2010) claim that the just noticeable difference (JND) in discriminating between scatterplots follows the Weber's law and the estimation of correlation magnitude follows a modified Fechner's law that the perceived correlation is a *convex*, rather than concave, function of the presented correlation.

Visual Features

The perception of correlation from scatterplot is not simply a function of the actual correlation, but is also influenced by a number of visual properties of the display (see Doherty & Anderson, 2009 for a review). For example, the presence of outliers in the scatterplot influences the precision of correlation judgment. Observers typically underestimate the effect of outliers on the actual strength of correlation (Bobko & Karren, 1979; Meyer et al., 1997). The perceived strength of covariation also increases with the number of points in the graph (Lauer & Post, 1989; Ciccione et al., 2021).

Display features that have nothing to do with actual data points in graph also influence the perceived strength of correlation. For example, the perceived strength increases with the slope of the regression line, which can be simply manipulated by axis stretching or shrinkage (Meyer & Shinar, 1992; Meyer et al., 1997). The perceived strength also increases with the density of the point cloud, the latter of which can be altered by simply manipulating the range of the axes (Cleveland et al., 1982; Boynton, 2000). Yang et al. (2018) compare 49 different visual features and evaluate how well each of them predicts the observers' perceived correlation. In a comprehensive model comparison, they find that features that correspond to dispersion measures along the regression line (e.g. the standard deviation of distance between points and the line) predict the perceived strength at an accuracy higher than the objective correlation.

Statistical Learning with Visual Sampling

Although the above papers document important biases at play in correlation judgment, they do not explain the cognitive mechanisms responsible for the biases. Pearson correlation is an abstract mathematical concept, rather than a type of low-level stimuli (e.g. length, weight, luminance). It is unclear how such an abstract concept is directly linked to any of the visual features shown in the graph, and why the evaluation of such an abstract concept should be sensitive to psychophysical properties of low-level stimuli.

Here we propose that our cognitive system makes statistical inferences based on the visual input from the point cloud in a scatterplot. We assume that the observer is an ideal Bayesian learner who fits a linear trend line using the observed sets of points (i.e., their x and y coordinates), as Bayesian linear regression does. The perceived strength of covariation can be decoded from the Bayesian linear regression. Further, the visual input, which we capture using eye-tracking, may not be an unbiased sample of the point cloud. Some points can be attended to more often than others. Even if the observer detects all the points, they may have differential weights in the "mental regression", depending on their distances to the central vision. The points in foveal

vision have the highest weights, followed by those in the perifovea and peripheral vision. If people have biases in visual information sampling, then they may also develop biases in correlation judgment (even if they are able to extract statistical inferences from visual observations accurately).

Our statistical learning theory is built upon the influential *intuitive statistician* hypothesis (Peterson & Beach, 1967), which posits that humans perceive variables and probabilistic relations in the environment as in the probability theory and statistics framework. This approach has had enormous success in various cognitive domains such as categorization, language comprehension, learning and decision making (Oaksford & Chater, 2007; Griffiths et al., 2010). In a similar vein, a few recent studies have claimed that observers are able to extract statistical information from scatterplots to form subjective perception of correlation (Rensink, 2017; Ciccione et al., 2021). We also draw upon contemporary advances in information sampling in visual and cognitive sciences (Wei & Stocker, 2015, Yang et al. 2016; Sanborn & Chater, 2016; Chater et al. 2020). We integrate those diverse strands of research by combining computational modeling with eye-tracking.

Experiments

We ran two eye-tracking experiments to test our Bayesian learning model of correlation judgments. In Experiment 1, participants made correlation judgments from scatterplots in a square display. Experiment 2 involved two within-participant conditions, Landscape condition and Portrait condition, that displayed the scatterplots in landscape or portrait views respectively. In both experiments, participants' eye fixations while viewing the scatterplots were recorded by the eye-tracker.

Methods and Materials

Participants. A total of 109 undergraduate or graduate students (58 female; aged 22.22 ± 2.57) participated in Experiment 1 and 103 (62 female; aged 20.57 ± 1.76) participated in Experiment 2. All participants had normal or corrected to normal vision and 51% of them had attended at least one statistics-related course and 82% knew what *correlation* entailed. We excluded six additional participants from the experiments either because they failed the 9-point calibration in the eye-tracker or that their key-press data were missing in at least 25% trials.

Stimuli. The scatterplots for both experiments were generated using *mvrnorm* in R (Venables & Ripley, 2002). Each scatterplot contained 100 bivariate normally distributed data points, with different levels of correlation between *x*- and *y*-coordinates. The distributions of the *x*- and *y*-coordinates were both set at the standard normal distribution. Therefore, the Pearson correlation is equal to the slope of the regression line. No numeric information was on display (see Figure 1a). Therefore, from the participants' perspective, the variance of the generated data points did not matter.

We set 10 different Pearson's correlation coefficients, ranging from 0 to 0.9 (in an increment of 0.1). The same code was repeated four times to generate four distinct scatterplots with the same correlation coefficients, resulting in a total of 40 scatterplots. In Experiment 1, all 40 scatterplots were presented in the same 1000×1000px format. In Experiment 2, half of the scatterplots were displayed in a Landscape format (1200×800px) while the other half were displayed in a Portrait format (800×1200px). Therefore, the *x*-axis was longer than the *y*-axis in the Landscape condition, whereas the *y*-axis was longer than the *x*-axis in the Portrait condition (see Figure 3 for examples).

Procedures. Both experiments were run on a Tobii Pro Spectrum Eye-tracker, equipped with a 1920×1080px monitor (capturing gaze data at the speed of 1200 Hz). Upon arrival, participants were seated in front of the eye-tracking monitor, with approximately 55~65cm between their eyes and the screen. They were asked to adjust the seat height to make themselves conformable while looking at the screen, though we asked them to keep their bodies and heads as still as possible throughout the experiment.

In both experiments, we showed our participants 40 scatterplots and asked them to judge the correlation coefficient from the scatterplots (Figure 1a). Each trial began with a fixation + at the screen center (2 seconds), followed by a scatterplot that was displayed for 3 seconds. After the scatterplot disappeared, participants indicated their correlation judgment on a 10-point scale, ranging from 0 to 9, using a number-keyboard. Participants were allowed a maximum of 20 seconds to make the response. Before the formal experiment, participants were present three practice trials to familiarize themselves with the task. To ensure eye-tracking data quality participants performed a 9-point calibration and validation procedure twice, before the exercise session and before the formal experiment. Written consent from all participants was obtained.

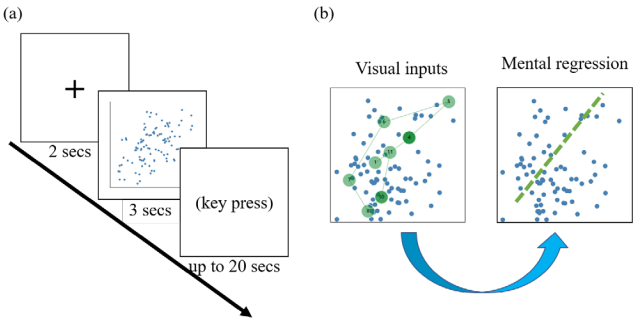

Figure 1: Experimental and modeling setups. (a) Trial procedures; (b) Bayesian learning model takes the observed data points from the graph (fixations on the left) and updates the slope of the regression line (the dotted line on the right) like a Bayesian linear regression.

Eye gaze data were collected during the 3-second scatterplot viewing phase. Each eye fixation was projected to the coordinates in the scatterplot. Therefore, we were able to identify, at each moment, which data points in the scatterplot were focused on and conversely which data points were out of the participants' central vision. Accordingly, the points at

or closer to central vision should receive more weight in the mental regression, while the points far from the central vision received less weight. To automate this process, we used a squared exponential kernel to compute a point's weight based on its onscreen distance to the fixation point: $w = \alpha \cdot e^{-\beta \cdot d^2}$, where α ($0 < \alpha \leq 1$) and β ($0 < \beta \leq 10$) are two free parameters and d represents the data points' onscreen distance to the fixation point (Ting et al., 2016). This weight, along with the data points, was applied to train the Bayesian learning model for correlation judgment predictions.

Bayesian learning model. We assume that the subjects are ideal Bayesian learners who fit a linear regression line using the observed sets of data points (i.e., their x and y coordinates), as well as their weights, as in a (Bayesian) linear regression (Figure 1b). Although it is possible that the observer can detect all the points in some cases, the points in foveal and peripheral vision should have differential weights in the mental regression. Thus, we operationalized learning in this setting as Bayesian regression with weighted inputs. The posterior belief of $\boldsymbol{\theta} = [\theta_0, \theta_1]$ (θ_0: intercept; θ_1: slope) can be written as $p_t(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}) = \frac{p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{w},\boldsymbol{\theta})}{\int p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{w},\boldsymbol{\theta})}$, where $\boldsymbol{x}$ and $\boldsymbol{y}$ are the set of observed points and $\boldsymbol{w}$ is the vector of weights corresponding to the set of points. The model' predicted Pearson's correlation is $R = \theta_1 \cdot \frac{sd(\boldsymbol{x},\boldsymbol{w})}{sd(\boldsymbol{y},\boldsymbol{w})}$, where θ_1 is the posterior mean estimation of the slope and $sd(\boldsymbol{x}, \boldsymbol{w})$ and $sd(\boldsymbol{y}, \boldsymbol{w})$ denote the weighted standard deviations of the observed data points on the x- and y-axis respectively.

By assuming a Gaussian prior distribution for $\boldsymbol{\theta}$ ($\theta_0 \propto \mathrm{N}(0,\ 1), \theta_1 \propto \mathrm{N}(0.45,\ 1)$), in conjugate to the likelihood function $p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{\theta})$ for linear regression, the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w})$ is also a multivariate Gaussian distribution. This allows us to compute the posterior estimation efficiently, with no need to run time-consuming simulations. Note that we set the prior of $\theta_1 \propto \mathrm{N}(0.45,\ 1)$, because participants were told that the presented correlations ranged from 0 to 0.9 and the mean correlation presented in all stimuli was 0.45.

Results

Summary of Behavioral Data. The behavioral results were highly consistent with previous findings. Overall, participants significantly underestimated the correlation from scatterplots (Figure 2a). This underestimation bias was particularly strong at the intermediate levels, $0.2 \leq R \leq 0.7$, in both experiments. On the individual level, 62% of the participants in Experiment 1 and 66% in Experiment 2 underestimated the correlations. On the trial level, the strength of correlation in 60% and 75% of the stimuli was underestimated for Experiments 1 and 2 respectively. When the actual correlation was low ($R = 0, 0.1$ or 0.2), participants hardly discriminated the difference between them in estimation, such that 0 and 0.1 correlations were always overestimated while 0.2 correlations were underestimated. Note that in our experimental design, 0 correlations could only be overestimated, since participants were not allowed to report negative correlations.

In Experiment 2, we observed that the underestimation bias was stronger in the Landscape condition than in the Portrait condition (Figure 2b, $p < .001$ in the mixed-effect linear regression). On the individual level, this asymmetry emerged in 65% of our participants in Experiment 2 while only 32% showed the reverse. On the trial level, 70% of the stimuli in the Portrait condition were underestimated while 80% of the stimuli in the Landscape condition were underestimated.

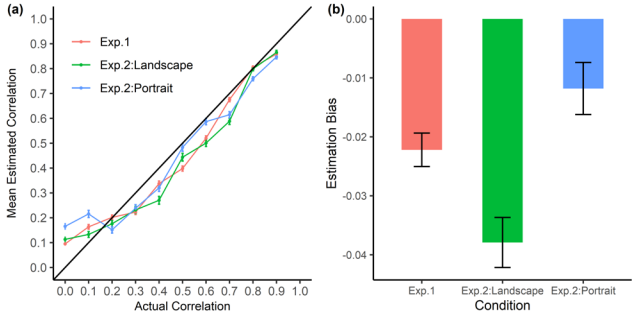

Figure 2: Behavioral data. (a) Estimated correlations at different levels of actual correlations; (b) Mean estimation biases (i.e., estimated correlation – actual correlation) across all trials. Error bars denote standard errors.

Summary of Eye-movement Data. The eye tracker gathered our participants' eye fixations during the three-second free viewing phase. In Experiment 1, participants made an average of 6.52 fixations per trial (sd = 0.27) . In Experiment 2, participants made 7.79 fixations (sd = 0.34) per trial in Landscape condition, and 7.92 fixations per trial (sd = 0.21) in Portrait condition respectively. We mapped the fixation coordinates to the that of the scatterplots and used the transformed coordinates to determine the weight of the data points for the mental regression (see the following section for details).

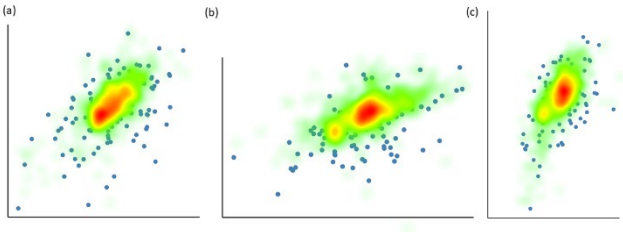

Figure 3: Heatmap of eye fixations on the scatterplots using the data of all participants. We used three scatterplots with actual correlation = 0.5 for illustration. (a) Experiment 1; (b) The Landscape condition of Experiment 2; (c) The Portrait condition of Experiment 2.

Figure 3 displays the density of eye fixations spanning the scatterplots, using the eye-movement data from all participants. Overall, participants' fixations were mostly focused on the central regions of the scatterplots. Peripheral regions of the point cloud attracted much less attention. This

tendency was robust across all experimental conditions. In other words, the visual sampling of the data points in graph was susceptible to a bias toward central regions. In contrast, the data points on periphery were underrepresented.

Aggregate Model Predictions. In this paper, we assumed that the participants are Bayesian learners who took as input the attended data points in the graph and updated the regression line as output. The eye-tracker allowed us to monitor the data points in graph that were attended to in each fixation. Thus, we trained the Bayesian learning model (i.e. Bayesian linear regression) on the attended (weighted) data points.

As mentioned earlier, we used a squared exponential kernel to determine the weights for different data points: $w = \alpha \cdot \mathrm{e}^{-\beta \cdot d^2}$, where α $(0 < \alpha \leq 1)$ and β $(0 < \beta \leq 10)$. We ran a grid search to show to how model predictions varied with the two free parameters. Figure 3 shows that our model's key predictions were very robust with different values for α and β. In Figures 3a-c, we show that our model consistently predicted underestimation of correlations from scatterplots. The model also predicted a stronger underestimation bias in the Landscape condition (Figure 3b) than in the Portrait condition (Figure 3c) (all p's < .001 with the 100 different parameter values).

Unsurprisingly, the magnitude of underestimation varied with the parameters, especially the decay parameter β. Overall, the predicted underestimation bias became stronger as β increased. Consider $\beta = 0$. In this setting, all data points in the graph are given equal weights in the Bayesian learning model, regardless of their distances to the fixation point. The Bayesian learning model would thus predict no overall underestimation or overestimation. On the other hand, when $\beta = 10$, the data points far from the fixation are given small weights.

Overall, our Bayesian learning model precisely predicted the key behavioral findings in correlation judgment from scatterplots by simply using the participants' eye-fixation data while viewing the graph.

Trial-level Prediction Accuracy. The Bayesian learning model also predicted estimation biases accurately on the trial level. We show such trial-level predictions in Figure 5, setting $\alpha = 0.1$ and $\beta = 10$. Figure 5a shows that there was little underestimation when the presented correlation was below 0.2, but the underestimation was rather strong when the presented correlation was between 0.2 and 0.8. The predicted underestimation attenuated when the presented correlation was 0.9. This pattern was again consistent with previous research and our behavioral findings that underestimation mostly occurred at intermediate correlation levels. Unlike human participants, the simulated Bayesian learners were able to discriminate between small correlations from scatterplots and predicted rather accurately at low correlation levels (with presented R≤0.2).

Finally, we evaluated how accurately the Bayesian learning model predicted the actual estimation biases. Consider $\alpha = 0.1$ and $\beta = 10$ for example. Figures 5b-c show that the model can predict trial-level variation in actual estimation biases accurately in both Experiments 1 and 2. Notably, the model not only successfully predicted underestimation at the intermediate levels (with presented R between 0.2 and 0.7), but also predicted overestimation at the low levels (with presented R below 0.2). That was interesting because the simulated Bayesian learners' estimation was not bounded between 0 and 0.9, though our human participants' estimation was. Other parameter values for α and β produced similar patterns. Figures 3d-f shows similar correlation coefficients using different α and β values.

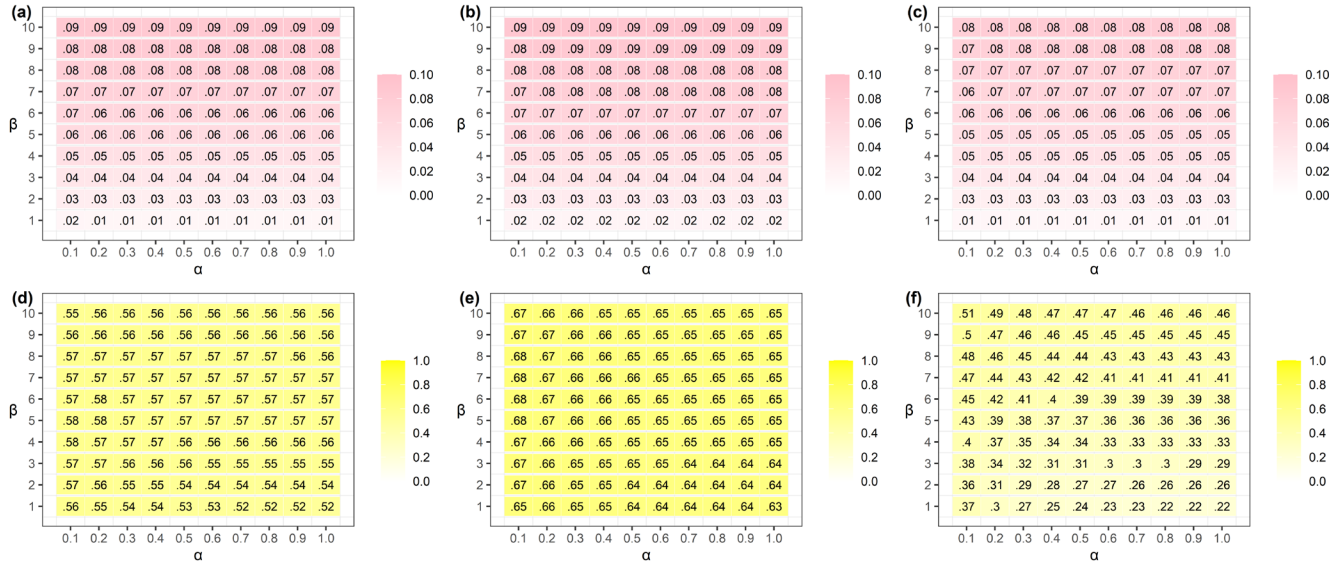

Figure 4: Model predictions using different squared exponential kernel parameters. (a)-(c) show the predicted underestimation biases in Experiment 1, Experiment 2 (Landscape), Experiment 2 (Portrait) respectively. All cells predicted an underestimation bias and the cell values denote the magnitude of underestimation. (d)-(f) show the Pearson's correlation

coefficients between the predicted biases and actual biases on the trial level in Experiment 1, Experiment 2 (Landscape), Experiment 2 (Portrait) respectively.

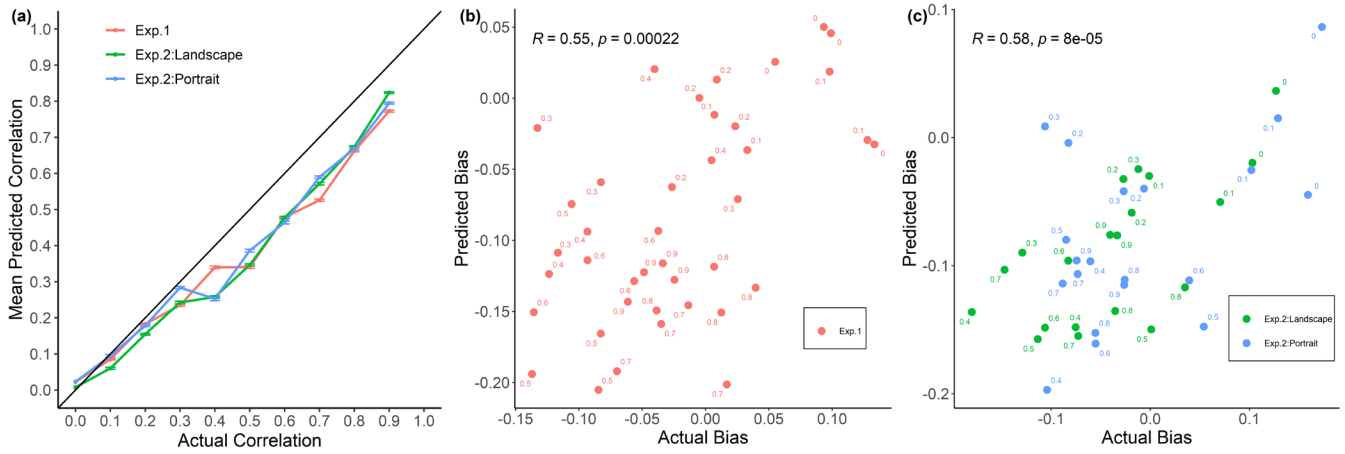

Figure 5: Predicted trial-level correlation estimation, setting $\alpha = 0.1$ and $\beta = 10$. (a) Predicted correlations at different levels of actual correlations; (b) Trial-level actual and predicted biases in Experiment 1; (c) Trial-level actual and predicted biases in Experiment 2. Error bars denote standard errors.

Discussion

This paper studies correlation judgment from scatterplots. We propose that decision makers perform statistical learning from the observed points via visual input. In two eye-tracking experiments, we first replicated two classic behavioral findings in correlation judgment from scatterplots: Subjects underestimated the correlation, especially at the intermediate levels, and the underestimation bias was stronger in the Landscape condition than in the Portrait condition. Critically, we obtained participants' visual sampling data using eye tracking and our Bayesian learning model using simply the visual information sampling data as input replicated these behavioral regularities. Further, our model was able to account for trial-level variations in judgment biases at high accuracy. Our results suggest that the behavioral biases in correlation judgment from scatterplots may arise from visual information sampling biases.

Future research can investigate the factors that guide these information sampling patterns. An immediate interesting question is how participants' search patterns interact with the visual features in the graph. For example, as we show in Experiment 2, the landscape/portrait manipulation changes gazing strategies. But the cognitive or biological mechanisms remain unknown. Further, it has been argued that humans are able search in a way that maximizes information gain in motion detection, object perception, recognition, and decision making (Wei & Stocker, 2015; Yang et al. 2016; Callaway et al. 2021). It is possible that similar processes are also at play in high-level cognition such as correlation judgment. If so, subsequent research need explain why optimal visual information search processes give rise to biased correlation judgments.

Our study also has important implications for data visualization and statistical communication. Our work is in line with the recent call of using visualization as stimuli for studying cognitive system (Rensink, 2021). While previous research stresses the visual features in the graph (Meyer & Shinar, 1992; Ma et al., 2018), our study highlights the important role of the visual system, particularly the interaction between visual information sampling and geometric features in the graph. It sets out to provide theoretically grounded guidelines for data visualization that facilitates information communication (Harold et al., 2016; Franconeri et al., 2021) and reduces perceptual biases (Ciccione et al., 2021). This line of work may be further applied to investigate how different people interact with different visual features, identifying individuals with high visualization literacy (Boy et al., 2014; Ludewig et al., 2020).

In conclusion, our study paves a new way to test the Bayesian models for judgment and decision making. We did so by integrating computational modeling with eye-tracking data. We use the process-level data to constrain the potentially overly flexible cognitive models, leveraging the explanatory power of formal computational and algorithmic models in understanding human cognition.

Acknowledgements

The authors acknowledge two anonymous reviewers for helpful comments. We thank Yufan Tian and Yutong Zhang for their assistance in data collection. This research was supported by the National Natural Science Foundation of China [grant number: 72101156] and Shanghai Pujiang Program [grant number: 2020PJC102].

References

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in cognitive sciences, 15*(3), 122-131.

Bobko, P., & Karren, R. (1979). The perception of Pearson product moment correlations from bivariate scatterplots. *Personnel Psychology, 32*(2), 313-325.

Boy, J., Rensink, R. A., Bertini, E., & Fekete, J. D. (2014). A Principled Way of Assessing Visualization Literacy. *IEEE transactions on visualization and computer graphics, 20*(12), 1963-1972.

Boynton, D. M. (2000). The psychophysics of informal covariation assessment: perceiving relatedness against a background of dispersion. *Journal of Experimental Psychology: Human Perception and Performance, 26*(3), 867-876.

Callaway, F., Rangel, A., & Griffiths, T. L. (2021). Fixation patterns in simple choice reflect optimal information sampling. *PLoS computational biology, 17*(3), e1008863.

Chater, N., Zhu, J.-Q., Spicer, J., Sundh, J., León-Villagrá, P., & Sanborn, A. (2020). Probabilistic Biases Meet the Bayesian Brain. *Current Directions in Psychological Science, 29*(5), 506-512.

Ciccione, L., & Dehaene, S. (2021). Can humans perform mental regression on a graph? Accuracy and bias in the perception of scatterplots. *Cognitive Psychology, 128*, 101406.

Ciccione, L., Sablé-Meyer, M., & Dehaene, S. (2022). Analyzing the misperception of exponential growth in graphs. *Cognition, 225*, 105112.

Cleveland, W. S., Diaconis, P., & McGill, R. (1982). Variables on scatterplots look more highly correlated when the scales are increased. *Science, 216*(4550), 1138-1141.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*(4899), 1668-1674.

Denrell, J. (2005). Why most people disapprove of me: experience sampling in impression formation. *Psychological Review, 112*(4), 951-978.

Doherty, M. E., & Anderson, R. B. (2009). Variation in scatterplot displays. *Behavior Research Methods, 41*(1), 55-60.

Doherty, M. E., Anderson, R. B., Angott, A. M., & Klopfer, D. S. (2007). The perception of scatterplots. *Perception Psychophysics, 69*(7), 1261-1272.

Erlick, D. E. (1966). Human estimates of statistical relatedness. *Psychonomic Science, 5*(10), 365-366.

Franconeri, S. L., Padilla, L. M., Shah, P., Zacks, J. M., & Hullman, J. (2021). The Science of Visual Data Communication: What Works. *Psychological Science in the Public Interest, 22*(3), 110–161.

Friendly, M., & Denis, D. (2005). The early origins and development of the scatterplot. *Journal of the History of the Behavioral Sciences, 41*(2), 103-130.

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: exploring representations and inductive biases. *Trends in cognitive sciences, 14*(8), 357-364.

Haberman, J., & Whitney, D. (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. *From perception to consciousness: Searching with Anne Treisman*, 339-349.

Harold, J., Lorenzoni, I., Shipley, T. F., & Coventry, K. R. (2016). Cognitive and psychological science insights to improve climate change data visualization. *Nature Climate Change, 6*(12), 1080–1089.

Lauer, T. W., & Post, G. V. (1989). Density in scatterplots and the estimation of correlation. *Behaviour Information Technology, 8*(3), 235-244.

Lifchits, G., Anderson, A., Goldstein, D. G., Hofman, J. M., & Watts, D. J. (2021). Success stories cause false beliefs about success. *Judgmen Decision Making, 16*(6), 1439-1463.

Ludewig, U., Lambert, K., Dackermann, T., Scheiter, K., & Möller, K. (2020). Influences of basic numerical abilities on graph reading performance. *Psychological Research, 84*(5), 1198–1210.

Mahrholz, G., Belin, P., & McAleer, P. (2018). Judgements of a speaker's personality are correlated across differing content and stimulus type. *PLoS One, 13*(10), e0204991.

Meyer, J., & Shinar, D. (1992). Estimating correlations from scatterplots. *Human Factors, 34*(3), 335-349.

Meyer, J., Taieb, M., & Flascher, I. (1997). Correlation estimates as perceptual judgments. *Journal of Experimental Psychology: Applied, 3*(1), 3-20.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*: Oxford University Press.

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin, 68*(1), 29-46.

Rensink, R. A. (2017). The nature of correlation perception in scatterplots. *Psychon Bull Rev, 24*(3), 776-797.

Rensink, R. A. (2021). Visualization as a stimulus domain for vision science. *Journal of Vision, 21*(8):3, 1-18.

Rensink, R. A., & Baldridge, G. (2010). *The perception of correlation in scatterplots.* Paper presented at the Computer Graphics Forum.

Sanborn, A. N., & Chater, N. (2016). Bayesian brains without probabilities. *Trends in cognitive sciences, 20*(12), 883-893.

Shefrin, H., & Statman, M. (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *The Journal of Finance, 40*(3), 777-790.

Strahan, R. F., & Hansen, C. J. (1978). Underestimating correlation from scatterplots. *Applied Psychological Measurement, 2*(4), 543-550.

Ting, J.-A., Vijayakumar, S., & Schaal, S. (2016). Locally Weighted Regression for Control. In *Sammut & Webb* (eds.), *Encyclopedia of Machine Learning*.

Trolier, T. K., & Hamilton, D. L. (1986). Variables influencing judgments of correlational relations. *Journal of Personality and Social Psychology, 50*(5), 879- 888.

Venables, W., & Ripley, B. (2002). Exploratory multivariate analysis. In *Modern applied statistics with S* (pp. 301-330): Springer.

Wei, X. X., & Stocker, A. A. (2015). A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts. *Nature Neuroscience, 18*(10), 1509-1517.

Yang, F., Harrison, L. T., Rensink, R. A., Franconeri, S. L., & Chang, R. (2018). Correlation Judgment and Visualization Features: A Comparative Study. *IEEE Trans Vis Comput Graph, 25*(3), 1474-1488.

Yang, S. C., Lengyel, M., & Wolpert, D. M. (2016). Active sensing in the categorization of visual patterns. *Elife, 5,* e12215.