EARLY PROCESSING OF SCATTERPLOTS: THE IMPACT OF OUTLIER AND CONTENT ON TREND-LINE ESTIMATION

EMRE ORAL

BOĞAZİÇİ UNIVERSITY

2021

EARLY PROCESSING OF SCATTERPLOTS: THE IMPACT OF OUTLIER AND CONTENT ON TREND-LINE ESTIMATION

Thesis submitted to the

Institute for Graduate Studies in Social Sciences

in partial fulfillment of the requirement for the degree of

Master of Arts

in

Psychology

by

Emre Oral

Boğaziçi University

2021

DECLARATION OF ORIGINALITY

I, Emre Oral, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature ....

Date 16.02.2021 ....

ABSTRACT

Early Processing of Scatterplots: The Impact of Outlier
and Content on Trend-Line Estimation

Recently, there has been a growing interest in integrating data visualization research with basic visual perception findings. Following this line of reasoning, in this thesis, I investigated how our understanding of ensemble perception, our visual system's ability to accurately and rapidly extract summary information of briefly presented set of objects that are spatially or featural similar, can contribute to our understanding of scatterplot processing. Across two experiments, I sought to answer two separate yet related questions. One, I investigated whether the presence of an outlier could influence how viewers extract best-fits in scatterplots. Two, I investigated whether familiar content and the presence of trend-consistent outliers influenced best-fits extracted in scatterplots. In both experiments, I briefly presented participants with scatterplots that varied outlier presence. Then, participants drew their best-fit estimates by using the mouse. Comparing their responses with possible best-fit alternatives, I found that outliers are equally weighed with the remaining points in trend-line estimates when there was not any context. However, when the relationship depicted in the scatterplot was familiar, and the outlier point represented a trend-consistent position, viewers were more likely to generate best-fits that overweighed those outliers in their responses, demonstrating that prior beliefs could influence our trend-line estimates from briefly presented scatterplots.

ÖZET

Dağılım Grafiklerinin Erken İşlenmesi: Uç Değerin ve İçeriğin Uyum Doğrusu Kestiriminde Etkisi

Son dönemde veri görselleştirme araştırmalarına temel görsel algı bulgularını entegre etme yönünde artan bir ilgi var. Bu düşünceyi takip ederek, bu tezde, özet temsil algısı, görsel sistemimizin oldukça kısa süreli gösterilen ekranlardaki birbirine uzaysal ya da niteliksel bakımdan benzeyen nesne topluluklarının özet bilgilerini hızlı ve isabetli çıkarabilme becerisi, kavrayışımızın dağılım grafiklerinin işlenmesi kavrayışımıza nasıl katkıda bulunabileceğini inceledim. İki deney boyunca birbirine bağlı iki farklı soruyu cevaplamak için çabaladım. İlk olarak uç değere sahip noktanın izleyicilerin en-iyi-uyum-doğrusu kestirimlerine etki edip edemeyeceğini soruşturdum. İkinci olarak ise tanıdık bir bağlam ve ilişki yönünde uyumlu bir uç değerin varlığının izleyicilerin en-iyi-uyum-doğrusu kestirimlerini etkileyip etkileyemeyeceğini soruşturdum. İki deneyde de katılımcılara uç değer varlığının değiştiği dağılım grafiklerini kısa bir süre için gösterdim. Sonrasında katılımcılar en-iyi-uyum-doğrusu kestirimlerini boş dağılım grafiğinde fare ile çizdi. Katılımcıların çizimlerini alternatif en-iyi-uyum-doğrusu çözümleri ile kıyasladığımda, herhangi bir bağlamın yokluğunda, uç değere sahip noktanın uyum doğrusu çiziminde geriye kalan noktalarla eşit seviyede hesaplamaya katıldığını gözlemledim. Fakat, tanıdık bir bağlam ve ilişki yönünde bir uç değer varlığında, izleyicilerin en-iyi-uyum-doğrusu çizimlerinde uç değere sahip noktanın ağır bastığını gözlemledim. Bu sonuç önceki inanışlarımızın kısa süreli gösterilen dağılım grafiklerinde bile en-iyi-uyum-doğrusu çizimimizi etkileyebileceğini gösterdi.

ACKNOWLEDGMENTS

I would like to thank Dr. Ayşecan Boduroğlu for her endless support, encouragement, and guidance throughout my masters education.

I want to thank Dr. Mehmet Yiğit Gürdal and Dr. Priti Shah for having served in my thesis committee, and for their valuable comments and advice on future directions in my research.

I would like to thank all my professors, colleagues, and friends in the Psychology Department of Boğaziçi University.

I want to thank TÜBİTAK (Scientific and Technological Research Council of Turkey) for providing financial support.

Lastly, I want to give my most special thanks to my parents, brother and sister-in-law for believing in me and providing their endless support.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

THE EFFECT OF OUTLIER PRESENCE ON TREND-LINE ESTIMATES IN SCATTERPLOTS

1.1 Literature review

In today's world, data visualizations are frequently used to present data, even in articles written for the general public. For example, it is not surprising to encounter an icon array while reading a news article about the spread and the risk of getting a particular disease, or a line graph that shows the change of market prices. Since the beginning of the COVID-19 pandemic, there has been an abundance of visualizations trying to communicate the pandemic spread, current status, and future projections to masses (Zacks, & Franconeri, 2020; "Coronavirus in Charts: The Fact-Checkers Correcting Falsehoods", 2020). Given the abundance of graphs in both academic and non-academic contexts, it is no surprise that researchers have been studying graphical processing for many decades (Freedman, Shah, 2002; Hegarty, 2011; Kossyln, 2012; Pinker, 1990; Shah, & Freedman, 2011; Shah, Freedman, & Vekiri, 2005), with numerous guidelines for better visualizations (Cleveland, & McGill, 1984; Hullman, Adar, & Shah, 2011; Kossyln, 2012; Lewandowsky, & Spence, 1989). During the last decade, there has been a renewed interest in visuospatial displays, mainly focusing on how cognitive and vision science perspectives can improve design features and identify individual difference variables that would allow for efficient processing of various graphical displays (for a review, see Hegarty, 2011). In this regard, there has been cross-fertilization between ensemble perception and graph perception literature (e.g., Correll, & Heer, 2017; Gleicher, 2018; Nothelfer, Gleicher, & Franconeri, 2017; Padilla, Ruginski, Creem-Regehr, 2017;

Sarikaya, Gleicher, & Szafir, 2018; Szafir, Haroz, Gleicher, & Franconeri, 2016). Ensemble perception refers to the extraction of statistical summary information of multiple individual objects with similar featural or spatial characteristics (for a review see Whitney, & Yamanashi, 2018). Recent research in ensemble perception has shown that viewers do not merely extract the mean of a set of features but they also rapidly extract information on the variance and range of visual properties (Khayat, Hochstein, 2018; Jeong, & Chong, 2020; Semizer, & Boduroglu, in press), and detect outliers (Alvarez, 2011; Avci, & Boduroglu, in press; Hochstein, Pavlovskaya, Bonneh, & Soroker, 2018). Considering these findings, in this thesis, I investigated whether ensemble mechanisms can contribute to the perception of trends in scatterplots. Specifically, I investigated how the presence of outliers, which are believed to be detected very rapidly through ensemble perception mechanisms, impact perceived trends in scatterplots (Experiment 1). Further, I also investigated the effect of familiar content on viewers' trend-line estimates (Experiment 2).

I believe that this translational approach integrating ensemble perception findings to the scatterplot perception domain could be instrumental because, perceptually speaking, the area of each scatterplot contains an ensemble of points. In this regard, these displays, apart from their axes, highly resemble displays used in studies of spatial working memory, configural processing, and contextual cueing (e.g., Boduroglu, & Shah, 2009; Boduroglu, & Shah, 2014; Brady, Konkle, & Alvarez, 2009; Jiang, Olson, & Chun, 2000; Mutluturk, & Boduroglu, 2014). It is known that viewers can quickly summarize the visual and spatial information depicted in such displays via ensemble perception mechanisms (e.g., Alvarez, & Olivia, 2008; Boduroglu, & Yildirim, 2020).

Scatterplots are graphic displays in which the numerical relationship between two variables is presented; each data point represents the corresponding values for variables depicted in the x and y axes. Thus, they typically accompany correlational analyses. It is not surprising that the studies on scatterplots generally has been focusing on the accuracy and precision of perceived correlation magnitude for decades (Doherty, Anderson, Angott, Klopfer, 2007; Harrison, Yang, Franconeri, & Chang, 2014; Lewandowsky, & Spence, 1989; Meyer, & Shinar, 1992; Meyer, Taieb, & Flascher, 1997; Pollack, 1960; Rensink, 2012; Rensink, 2017; Rensink, & Baldridge, 2010; Sher, Bemis, Liccardi, Chen, 2017; van Onzenoodt, Huckauf, & Ropinski, 2020; Yang, Harrison, Rensink, Franconeri, & Chang, 2019). There is a consensus that viewers generally underestimate the correlation magnitude (Doherty et al., 2007; Rensink, 2017; Strahan, & Hansen, 1978; but also see Cleveland, Diaconis, & McGill, 1982; Meyer, & Shinar, 1992). Researchers have also investigated the effect of several display features on perceived correlations: the presence of a trend-line (Meyer, & Shinar, 1992; Meyer et al., 1997), the density of the points (Doherty et al., 2007; Lauer, & Post, 1989; Rensink, 2012; Rensink, 2017), dispersion of points around the major axis of imaginary ellipsis that encapsulated the point cloud (Meyer, & Shinar, 1992; Meyer et al., 1997, Rensink, 2012; Rensink, 2017), aspect ratio (Cleveland et al., 1982; Rensink, 2017), geometric scaling (Wei et al., 2020) and the effect of individual difference measures such as expertise (Meyer, & Shinar, 1992; Meyer et al., 1997). These studies have yielded several different models of correlation magnitude perception (Cleveland et al., 1982; Harrison et al., 2014; Jenning, Amabile, & Ross, 1982; Meyer et al., 1997; Rensink, 2017). For instance, Meyer et al. (1997) found that viewers' correlation estimates could be explained by the mean geometrical distance (i.e., perpendicular distance) of points across the regression line. Although Meyer and colleagues constructed the model based on the mean perpendicular distance of points to the regression line, they also suggested that it is not likely that viewers computed the distance of each point and summed all the results. In fact, they suggested that viewers were somehow able to immediately extract mean distance by perceptual mechanisms. Today, these abilities are broadly referred to as ensemble perception, and as Rensink (2017) suggested, they are likely to play a role in correlation magnitude estimation.

Ensemble coding can also impact scatterplot perception by influencing outlier detection and trend estimation (Szafir et al., 2016). Recently, Correll and Heer (2017) investigated viewers' perception of trends in scatterplots. Specifically, they asked participants to adjust the probe trend-line slope to specify their trend-line estimate for a given scatterplot that depicted 100 points. Their results showed that viewers' estimates were accurate regardless of the complexity of the trend, i.e., linear or higher-order trend-lines. They also found that viewers' accuracy became worse as the residuals bandwidth increased, and that suggested viewers' performance in the trend-line estimation task decreased with decreasing physical correlation. Furthermore, they argued that the presence of outliers might influence viewers' trend-line estimates, and they expected viewers to down-weight outliers while estimating the trend-line. To investigate this, they presented viewers with scatterplots that had either an outlier cluster with 5, 10, or 15 points or not. They also controlled for the location of the outlier cluster by positioning those clusters at the very beginning, middle, or end of the display. As they expected, results showed that viewers' trend-line estimates were closer to a robust ordinary least squares (OLS) trend-line, which discounted the outlier cluster, than it was compared to an OLS solution including the outlier cluster. However, in Correll and Heer's (2017) study, scatterplots were always perceptually available, and rather than presenting participants with a single outlier, they showed an outlier cluster. I argue that the effect of a single outlier could be different from a cluster because the latter could be perceived as a separate group and more easily be excluded from trends. On the other hand, a single outlier point may be detected and yet not be excluded from trend estimates.

In this thesis, I briefly presented participants with scatterplots that either had a single outlier point or not and asked viewers to draw their trend-line estimates to test the impact of a single outlier on trend estimates. I expected trend-line estimates to be more accurate as the correlation magnitude depicted in scatterplots increased. More critically, based on Correll and Heer's study (2017) and research in ensemble perception (e.g., Epstein, Quilty-Dunn, Mandelbaum, & Emmanouil, 2020; Haberman, & Whitney, 2010) that showed that viewers discount outlier in their summary representations, I tested whether a single outlier is similarly discounted in trend-line estimates. I also explored the effect of the physical correlation valance and the outlier position on viewers' trend-line estimates.

1.2 Experiment 1

1.2.1 Participants

Twenty-one Boğaziçi University undergraduates (11 male; mean age = 22.05 ± 2.52) were recruited with the exchange of one course credit. All participants had normal or corrected to normal vision. Based on prior criteria, two participants who had less than 80% valid trials[1] for any condition excluded.

1.2.2 Materials

Each trial began with a screen that showed the x and y axes of a scatterplot and a green fixation cross at the center of the screen that remained on for 500 ms. Then the fixation cross turned red for another 500 ms, alerting viewers to the onset of the data cloud during which they were expected to maintain fixation. Afterward, the red fixation cross disappeared, and the viewers were presented with data points in the scatterplot for 250 ms. After the offset of the data points, participants drew their trend-line estimate using a mouse on a blank scatterplot.

Overall, in the main experiment, there were 360 trials. I equated the number of trials for two trend direction (i.e. negative, and positive), three correlation magnitude intervals (i.e. [.2, .4), [.4, .6), [.6, .8)), and four outlier types (i.e. NO-, X-, Y-, and XY-outlier). Figure 1 shows some example trials from the experiment; here on the figure at the top, you can see three no-outlier trials for all possible correlation magnitude, and at the bottom, there are examples of three types of scatterplots with an outlier.

Each scatterplot was presented at the center of the screen, extending a region of 16.38° x 13.14°. The 20 dots were inside a 10.28° x 10.28° square area; marked with black (#000000) squares on a white (#FFFFFF) background. The centroid of the 20 dots was at the center of the 10.28° x 10.28° square area for all displays. Throughout the displays, the standard deviation of 20 dots was the same, 1.55°.

[1] Valid trials mean that average absolute error is calculable (e.g., it is not calculable when a participant drew a single point) and not extreme (i.e., is not away from the mean of the condition more than 3 standard deviation).

While the no outlier dots could vary up to 2 standard deviation on both dimensions, outlier dots separated from the other dots on x, y, or both axes based on the type of the outlier, while its location was somewhere 2.5 to 3 standard deviation away from the centroid on the axis or axes depending on the type of the outlier.

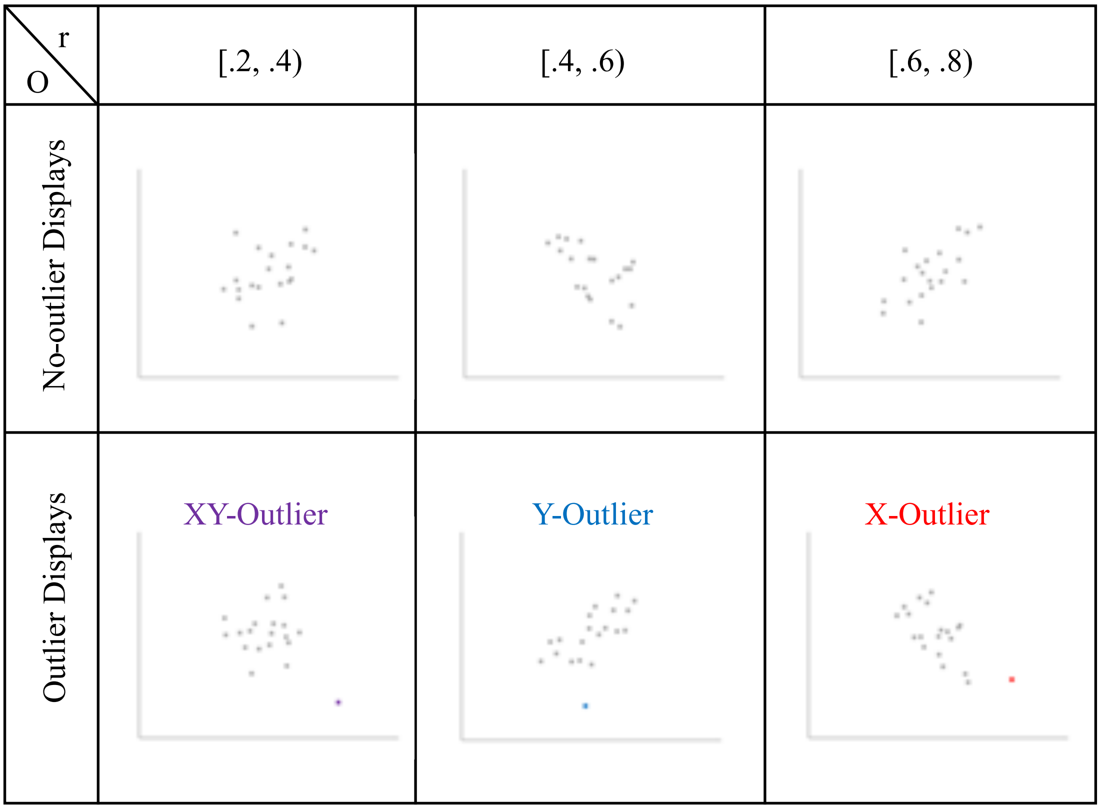


Figure 1. Type of scatterplots in Experiment 1

1.2.3 Procedure

The experiment took place in a well-lit room. First, participants provided informed consent, and then they read instructions that explained what a scatterplot is and how to draw a trend-line estimate in detail. After the instructions, participants first completed the training phase consisted of 36 trials, and all had no outlier points. Unlike the main experiment, in the training phase, each participant was shown a visual feedback screen on which both participant's response and OLS trend-line of 20 points were presented after they had drawn their trend-line estimates.

In the main part of the experiment, participants completed the trend-line estimation task for 360 trials in which there were several types of scatterplot that mentioned under materials title. After participants completed the training and experimental phases of the experiment, they filled the demographic form. At the end of the experiment, they were thanked and debriefed. All procedure took approximately 45 minutes.

1.2.4 Apparatus

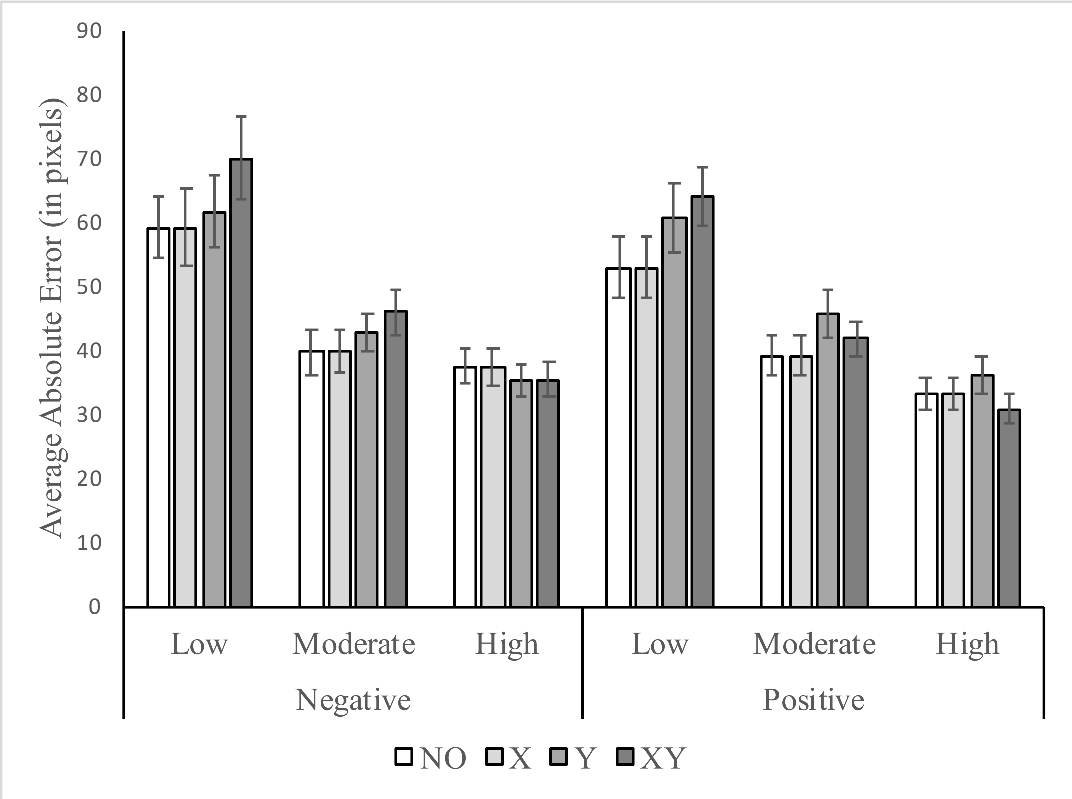
A computer with an Intel Core 2 Duo processor, an ATI Radeon X300/X550/X1050 Series graphics card was used. Stimuli were shown on a 17-in. CRT Philips 107S6 monitor. The screen resolution was set to 1280 x 1024 pixels, with a refresh rate of 60 Hz (Refresh duration = 16.67 ms). The experiment was programmed in MATLAB, using the Psychtoolbox (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007).

1.2.5 Results and discussion

I calculated the average absolute error (AAE) for each trial. To calculate AAE between participants' response and OLS solution of 20 points, I calculated the absolute pixel difference between y values for each integer x value between 0 and 640, that was the boundaries of the x-axis by pixels. Then, I summed all absolute difference scores and averaged them to get the AAE.

First, to determine if the presence of outliers and the correlation magnitude had any effect on trend-line estimates, I conducted a 2(Trend Direction: Negative, Positive) x 3(Correlation Magnitude: Low, Moderate, High) x 4(Outlier Presence:

NO-, X-, Y-, XY-outliers) repeated measures ANOVA (see Figure 2); the trend direction was included in the analyses as a control variable.



Note. Error bars represent ±1 standard errors.
Figure 2. The results of 2(Trend Direction: Negative, Positive) x 3(Correlation Magnitude: Low, Moderate, High) x 4(Outlier Presence: NO-, X-, Y-, XY-outliers) ANOVA

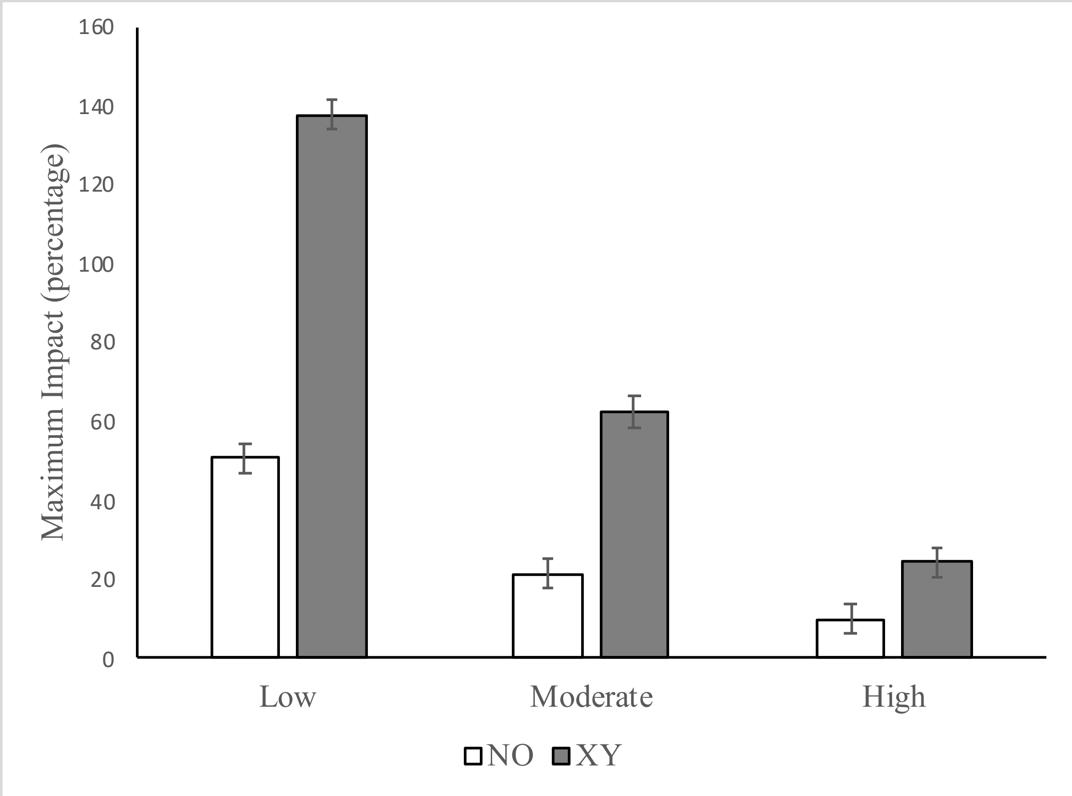
I found that there was no main effect of trend-direction, $F < 1$. As I expected, I found a main effect of correlation magnitude, $F(1.066, 19.194) = 28.970$, $p^2 < .001$, $\eta_p^2 = .617$. That suggested viewers' trend-line estimates got more accurate as correlation magnitude increased; High ($M = 34.494$, 95% CI [29.871, 39.117]), Moderate ($M = 42.266$, 95% CI [36.388, 48.145]), Low ($M = 60.178$, 95% CI [49.915, 70.440]). I also found a main effect of outlier presence, $F(1.783, 32.087) = 4.011$, $p = .32$, $\eta_p^2 = .182$. Post-hoc analysis on the main effect of outlier presence showed that viewers' estimates were closer to the 20 OLS solution in the NO ($M = 44.070$, 95% CI [38.253, 49.887]) and X ($M = 43.573$, 95% CI [36.846, 50.299]) outlier conditions

[2] All *p*s are Bonferroni corrected.

compared to XY (M = 47.943, 95% CI [42.260, 53.626]) outlier condition, p = .002, p = .003, respectively. These main effects were further qualified by significant trend-direction X outlier presence, and outlier presence X correlation magnitude interactions, F (2.783, 50.093) = 8.762, p < .001, η_p^2 = .327, F(4.405, 79.290) = 8.331, p < 0.001, η_p^2 = 0.316, respectively. No other interactions were significant. The post-hoc analysis on trend-direction and outlier presence interaction showed that there was no difference between outlier conditions in positive trends; NO (M = 45.377, 95% CI [39.397, 51.357]), X (M = 41.704, 95% CI [35.008, 48.400]), Y (M = 47.499, 95% CI [39.812, 55.186]), XY (M = 45.475, 95% CI [40.060, 50.891]). But, viewers' trend-line estimates were more accurate in the No (M = 42.764, 95% CI [36.512, 49.016]) and X (M = 45.441, 95% CI [38.103, 52.779]) outlier conditions compared to XY (M = 50.410, 95% CI [43.699, 57.122]) outlier condition while the depicted trend was negative, respectively p < .001, p = .01. Furthermore, the post-hoc analysis on outlier presence and correlation magnitude interaction pointed out that while there was no difference between outlier conditions in high and moderate levels of correlation magnitudes, in low correlation magnitudes trend-line estimation errors were more in XY (M = 66.964, 95% CI [56.137, 77.791]) outlier condition compared to NO (M = 56.533, 95% CI [46.865, 66.201]) and X (M = 56.033, 95% CI [44.799, 67.267]) outlier conditions. That suggested that the effect of outlier presence was mainly driven by the disproportionate errors in XY-outlier condition for scatterplots that depicted low correlation magnitudes.

I decided to explore what might have led to the disproportionate levels of error in the XY-outlier condition for the low correlation magnitudes. I thought that the impact of the outlier point on the correlation magnitude of the remaining 19 points could be a factor that might explain these inflated error levels. Thus, for each scatterplot, I calculated how much each point, regardless of outlier status, changed the correlation magnitude by calculating the absolute difference between the correlation of all points and all except that particular point. Then, I calculated what percent of the correlation of 20 points equal to each point's impact in the previous sentence. That gave me a percentage impact score for each point in each display. I chose the maximum impact for each display. The XY outlier had the maximum impact for all XY-outlier displays. After those calculations, I ran a 3(Correlation Magnitude: Low, Moderate, High) x 2 (Outlier Presence: NO-, XY-outliers) ANOVA with the maximum impact score as the DV. I only focused on the NO- and XY-outlier conditions for this analysis because the critical difference in the previous analysis was observed only across these two conditions. What I wanted to explore in that analysis was whether there was a difference between the NO- and XY-outliers displays on the maximum correlation change impact. This particular analysis was solely based on the displays and it showed a significant main effect of correlation magnitude and outlier presence on maximum impact, $F(2, 174) = 218.807$, $p < .001$, $\eta_p^2 = .716$; $F(1, 174) = 238.874$, $p < .001$, $\eta_p^2 = .579$, respectively (see Figure 3). The main effect of the correlation magnitude suggested that the maximum impact was the least for the High ($M = 17.283$, 95% CI [12.027, 22.539]) correlation magnitudes and was the most for the Low ($M = 94.438$, 95% CI [89.181, 99.694]) correlation magnitudes, all ps < .001. Moreover, the maximum impact was larger for the XY-outlier ($M = 75.018$, 95% CI [70.727, 79.310]) displays than the NO-outlier ($M = 27.491$, 95% CI [23.199, 31.783]) displays, $p < .001$. More critically, I found a significant interaction between the correlation magnitude and outlier presence, $F(2, 174) = 47.628$, $p < .001$, $\eta_p^2 = .354$. Although the maximum impact was the biggest for XY-outlier displays for all correlation magnitude levels, the difference between maximum impacts was the most for low correlation scatterplots (see Figure 3). The most important value in Figure 3 was that the percentage maximum impact score for XY-outliers in low correlation magnitudes was more than 100%, which means, for those displays, XY-outlier change the direction of the relationship between the remaining 19 points. Therefore, I argue that the disproportionate levels of error for XY-outliers in low correlation magnitudes was most likely due to the impact of XY outlier for the low correlation displays.

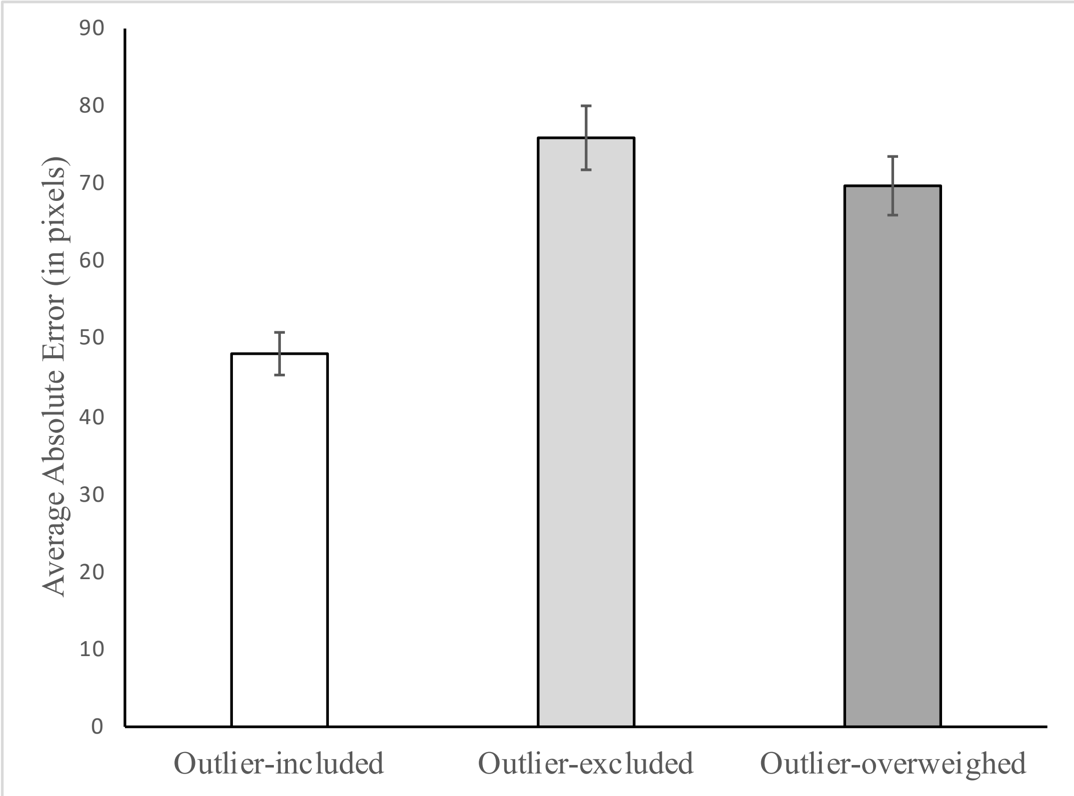
Another exploratory line of analyses investigated the possibility that trend-line estimates were closer to another trend-line compared to the OLS solution of 20 points. Here two extreme alternatives for the OLS solution of 20 points were that of 19 points excluding the outlier point and an outlier-overweighed trend-line, which passed through the centroid of the remaining 19 points and the outlier location. If viewers' estimates were closer to the 19-point OLS solution, that would suggest viewers were excluding outliers in their trend-line estimates. On the other hand, if their estimates were closer to an outlier-overweighed trend-line, then that would suggest that viewers' trend-line estimates were influenced by a pull from the outlier point, revealed in responses getting paid more getting closer to the outlier-overweighed trend-line. To test these possibilities, I conducted a 2 (Trend Direction: Negative, Positive) x 3(Correlation Magnitude: Low, Moderate, High) x 3(Error Type: Outlier-included, Outlier-excluded, Outlier-overweighed) for XY-outlier displays. Since I found an effect of outliers for XY-outlier displays, I only analyzed responses for the XY-outlier condition. The results showed a main effect of trend-direction, $F(1, 18) = 11.405$, $p = .003$, $\eta_p^2 = .388$; and correlation magnitude, $F(1.116, 20.084) = 136.274$, $p < .001$, $\eta_p^2 = .883$. More critically, I found a main effect of error type, $F(1.036, 18.640) = 14.112$, $p = .001$, $\eta_p^2 = .439$.



Note: Error bars represent ±1 standard errors.
Figure 3. The difference between maximum impact scores for NO- and XY- outlier displays in low, moderate, and high correlation magnitudes

Post-hoc analyses showed that viewers' trend-line estimates were the closest to outlier-included trend-line (M = 47.943, 95% CI [42.260, 53.626]) compared to outlier-excluded (M = 75.826, 95% CI [67.222, 84.429]) and outlier-overweighed (M = 69.726, 95% CI [61.879, 77.573]) trend-lines (see Figure 4). Thus, I concluded that viewers do not exclude and over-weigh the outlier points. Rather, they weigh them equally with the remaining points. There were also a two-way interaction between correlation magnitude and error type, and a three-way interaction between trend direction, correlation magnitude and error type; $F(1.559, 28.054) = 7.252$, $p = .005$, $\eta_p^2 = .287$, $F(1.675, 30.151) = 3.715$, $p = .043$, $\eta_p^2 = .171$, respectively. Those interactions suggested that despite viewers' estimates were always significantly the closest to outlier-included trend-line. Besides, estimates converged numerically more to either outlier-excluded or outlier-overweighed trend-lines in the levels of trend direction and correlation magnitude, although those results were not statistically significant.

To summarize, in this experiment, I found that trend-line estimates in scatterplots were more accurate when the depicted correlation magnitude was stronger, replicating Correll and Heer's (2017) findings. More critically, regarding the impact of outliers, I found that in all cases except for the XY outliers in the low-magnitude scatterplots, viewers integrated outliers in their trend-line estimates. This was revealed by the similar levels of errors across the outlier and no-outlier conditions. However, when the scatterplot depicted weaker correlations, and there was an XY-outlier, then errors were disproportionately higher than in the no-outlier condition. In fact, in these cases, viewers' trend-line estimates were pulled even closer to the outlier-weighed trend-estimate, in contrast to earlier findings of outlier exclusion from summaries (e.g., Correll & Heer, 2017; Haberman & Whitney, 2010). However, in the current study, what constituted an outlier was quantitatively and qualitatively different from the previous studies. For example, Correll and Heer (2017) presented participants with displays consisting of 5, 10, or 15 outlier points (among 100 points), and the size of these outlier clusters could have increased the distinctiveness of these outlier points from the remaining data cloud. Thus, I suggest that the difference between my results and Correll and Heer's (2017) is caused by the difference between the saliency of outliers as a distinct group in their case. This interpretation is consistent with recent research from our group that shows that an outlier's perceived distinctiveness moderates its exclusion from the broader feature summary in a display.



Note: Error bars represent ±1 standard errors.
Figure 4. The main effect of error type on AAE

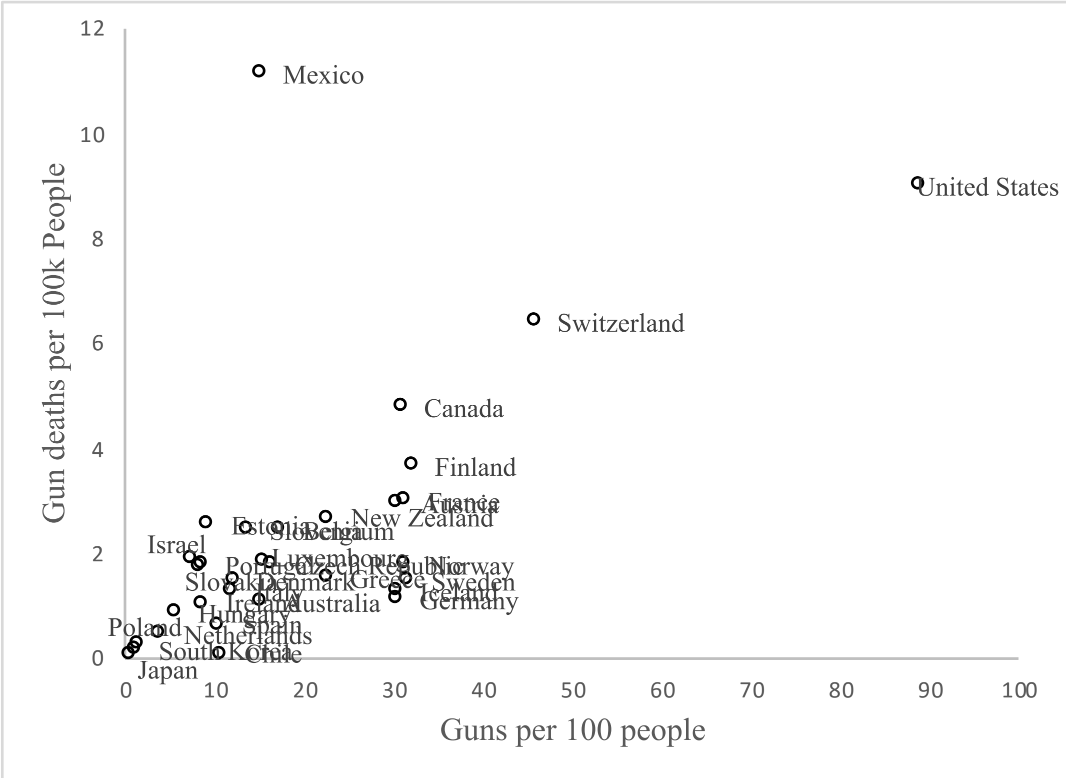
While this experiment demonstrated that outliers are not excluded from trend-estimates, I must note that these findings may be limited in that the lack of meaningful context accompanying scatterplots (i.e., no explicitly identified variables on the x and y axes) may have prevented us from observing top-down influences in trend-estimation. In Experiment 2, I directly addressed this issue.

CHAPTER 2

THE EFFECT OF CONTENT ON TREND-LINE ESTIMATES IN SCATTERPLOTS WITH OUTLIERS

2.1 Literature review

In real life, scatterplots depict meaningful context. When a meaningful pair of variables are plotted against one another, while the spatial outlier item may seem perceptually similar to an outlier in a no-context scatterplot, its impact and meaning may be differently perceived. Consequently, a single outlier may have a different effect on the trend-line estimates when there is a meaningful context compared to when there is none. For instance, the outlier may represent either an intuitively expected case or one that is less so. Suppose one were to look at the relationship between gun ownership and gun-related deaths. In that case, as can be seen in Figure 5 the issue of general trends and intuitively expected and unexpected outliers may be better understood. For instance, in Figure 5 the data reveal a positive trend between gun ownership and gun-related deaths with a correlation magnitude of .62. In this scatterplot, two countries- the US and Mexico are placed at two visually distinct positions that are approximately equidistant to the remaining cluster of countries, with the highest level of gun-related deaths. However, these two countries differ on gun ownership. The US, with its highest level of gun ownership and the highest level of gun-related deaths, and given the positive trend, may represent an extreme and yet intuitively expected case. However, the gun-ownership in Mexico is reported to be approximately one fifth that of that in the US and at similar levels to the average gun ownership in the rest of the countries depicted in this graph, yet it has the highest level of gun-related deaths. Someone who did not know anything about Mexico may find this pattern counterintuitive.



Note: The data downloaded from http://mark.reid.name/blog/gun-deaths-vs-gun-ownership.html.
Figure 5. The relationship between gun ownership per 100 people and gun-related death per 100K people across countries

In my current thesis, I specifically chose to focus on intuitively possible outliers to test how context impacts trend-line estimates as a function of outlier presence. The impact of context may likely be stronger for intuitively expected outliers because once people consider the context and variables of interest, they may also anticipate where an outlier may be present on the scatterplot. For instance, in the above gun ownership-gun-related deaths example, if we remove the US from the scatterplot, the correlation depicted by the remaining data is .43. In the scatterplot without the US, Mexico is the only point that is saliently distinct from the other points. When we compare the correlation between all points, including Mexico (r = .43), to the correlation between all points excluding Mexico (r = .72), one would realize that Mexico represents an outlier point that weakens the positive relationship between all other points. On the other hand, the inspection of the data point representing the US reveals an opposite pattern. If the only outlier point in the scatterplot was the one represented by the US (i.e., Mexico excluded the comparison of the correlation magnitude when the outlier was included (r = .86) and excluded (r = .72) illustrates that the US case is an outlier that strengthens the positive relationship depicted in the scatterplot. This example suggests that the presence of intuitive (trend-consistent) outliers is likely to strengthen existing relationships, while unexpected outliers are likely to reduce the magnitude of existing relationships. I specifically focused on expected outliers because their presence is likely to increase absolute correlation magnitude. When the content is familiar, and the depicted relationship is in a typically expected direction (e.g., the scatterplot on Figure 5 without Mexico), an intuitive outlier point (e.g., the US) may further strengthen viewers' perception of that relationship; furthermore, the outlier point may be overweighed as trends are extracted, resulting in the outlier acting as a partial anchor for the trend. Given these possibilities, I thought it was essential to investigate the link between outlier processing and trend extraction when the context was familiar and the outlier was expected.

As I indicated in the previous paragraph, the aim of the current thesis was not only investigating the effect of an outlier in isolation of a context but also incorporating the presence of the outlier with familiar content to further investigate the role of context in the early processing of outliers in scatterplots. It is known that graph processing is not purely based on bottom-up perceptual characteristics of the display; indeed, prior knowledge of graph schema and the conceptual content of the visual display interact with the perceptual features and constitute the graph comprehension (for a review, see Shah, Freedman, & Vekiri, 2005). Thus, I aimed to observe the effect of graph context even in the early processing of outliers in scatterplots. In subsequent paragraphs, first, I summarized the viewpoint of previous research on graph comprehension. I then mentioned the results of studies which investigated how prior knowledge influence covariation judgments in scatterplots.

A camp of graph comprehension research focused on low-level perceptual features, while another camp combined the perceptual characteristics of a display with factors such as prior knowledge of the content presented on a visual display and graph knowledge per se (for a review, see Shah et al., 2005). For example, Pinker (1990) suggested a model of graph comprehension in which perceptual features and top-down factors interact. According to the model, viewers first translate the sensory image, which is encoded with the help of bottom-up attentional mechanisms, to a visual description. Properties of the visual system (e.g., gestalt law) and some individual difference factors such as working memory capacity and graph knowledge intervene in this translation process. After the extraction of visual description, that description is translated into a conceptual message, and finally, viewers answer a conceptual question based on the conceptual message. Freedman and Shah extended and differed from Pinker's model by suggesting concurrent activation of prior knowledge and expectations with visual chunks and their top-down influence on subsequently formed interpretations (Freedman, & Shah, 2002; also see Hegarty, 2011; Padilla, Creem-Regehr, Hegarty, & Stefanucci, 2018).

Covariation judgment was one of the most studied tasks in which the effect of viewers' prior knowledge on their estimates investigated while the data presented in a tabular format or a scatterplot (Anderson, 1995; Anderson, & Kellam, 1992; Baumgartner, 1995; Billman, Bornstein, Richards, 1992; Karduni, Markant, Wesslen, Dou, 2020; Wright, & Murphy, 1984). Those studies generally yielded that prior beliefs could bias viewers' judgment in a direction that favors prior intuitive theories (e.g., Anderson, 1995; Billman et al., 1992, but also see Wright, & Murphy, 1984). However, Wright, and Murphy (1984) showed that having a prior theory could be better than having none. They found that people who had either high or low expectations about a relationship between two variables responded congruous with their prior expectations. However, the accuracy of people who had not had a prior belief was worse than those who had a prior theory. Thus, they concluded that having a theory could facilitate viewers' performance by urging them to look for further evidence in presented data. Moreover, they also found that, regardless of the outlier compliance with prior expectations, viewers discounted outliers in their covariation judgments, especially when the outlier dramatically weakened the correlation. On the other hand, Billman and colleagues argued that the previous works did not distinguish the meaningfulness of variables from prior beliefs (Billman et al., 1992). For example, in Wright and Murphy's study (1984), they did not show any pair of variables to participants in no-theory condition. That means participants in no-theory conditions observed abstract numerical data without any content. Then, Billman et al. (1992) suggested that the positive effect of having a prior theory confounded by a third variable, that is, the meaningfulness of individual variables. They piloted to find variable pairs for four belief conditions; positive belief, negative belief, zero correlation belief, and no-belief. Although participants did not have any expectation about the possible relationship between two variables in the no-belief condition, those two variables individually meant something to participants. The study results showed that viewers' discrimination of different correlation magnitude was the best when they were agnostic (Billman et al., 1992). Those results suggested that prior belief does not necessarily facilitate viewers' performance; it could bias viewers' judgment in a systematic manner. A recent study also supported the idea that prior beliefs could bias viewers' judgment, despite the presence of graphically presented data (Karduni et al., 2020, see also Valdez, Ziefle, & Sedlmair, 2018). Their study measured viewers' belief change upon the presentation of a scatterplot with/without uncertainty information and the congruency/incongruency of the presented data with prior beliefs. They found that while viewers observed a display in which the data is consistent with their prior beliefs, viewers were less likely to update their prior beliefs. Nevertheless, viewers discounted their prior beliefs to some extent while posterior correlation judgments were closer to the posterior estimate of a Bayesian-Uniform model than that of a prior-only model. That finding also supported the fact that neither viewers ignore the data presented nor disregard their prior knowledge (see Shah et al., 2005). Based on those previous findings, in the present thesis, I wanted to investigate the effect of an expected outlier, which is consistent with prior beliefs on viewers' trend-line estimates. I thought expected outliers had special importance here. Previous research on ensemble perception consistently showed that viewers discount outlier objects in their summary inferences. In parallel with those findings, I also predicted that viewers would discount outliers in their trend-line estimates in the absence of content. However, as the literature on the effect of prior knowledge suggested, viewers tended to ignore data, especially the deviant ones, when the deviant ones weakened the presented relationship (e.g., Wright, & Murphy, 1984). Thus, I asserted that an outlier like the US case could be overweighed while the presence of that outlier fosters the prior belief. That assertion is also compatible with the finding that viewers are hesitant to update their prior beliefs while the data is consistent with their prior beliefs (Karduni et al., 2020). For example, suppose a viewer ignores an expected outlier. In that case, the posterior belief will become less congruent with the prior belief, which was not the case in Karduni and colleagues' findings. Therefore, in the present thesis, I expected to find that an expected outlier in a familiar context could bias viewers' trend-line estimates to an outlier-overweighed trend-line solution.

2.2 Experiment 2

2.2.1 Participants

One-hundred fifty-two Boğaziçi University undergraduates (92 female and 55 male, mean age = 20.76 ± 2.54) participated in the experiment in exchange for course credit. Participants were randomly assigned to one of two sets; half of the scatterplots in each set had outliers. However, the sets varied on whether a scenario was presented along a scatterplot containing an outlier or not. All participants had normal or corrected to normal vision. I had two prior exclusion criteria, but none of the participants violated these criteria and were therefore not excluded. These criteria were (1) excluding participants who mismatched axes and variables on more than 20% of trials in the axis-matching phase of the trend-line estimation task (this phase is described in greater detail under the materials), and/or (2) excluding participants who had less than four valid[3] responses in at least one of the conditions.

2.2.2 Materials

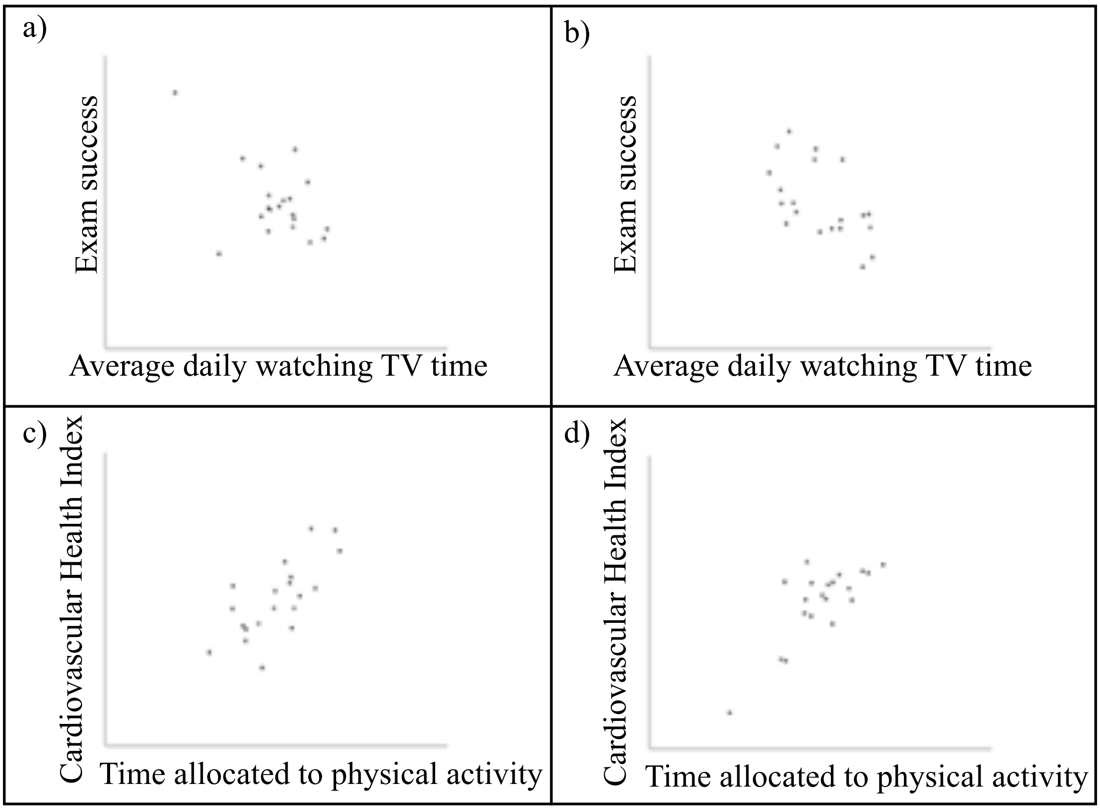
For the trend-line estimation task, participants first read a brief description of a fictitious study investigating the relationship between two variables. The description stated that the results were plotted in a scatterplot and identified which of the two variables were plotted on the x and y axes, respectively. To ensure that participants

[3] If a participant drew a dot as opposed to a trend-line as a response, I tagged that trial as an invalid trial.

paid attention and thought about the variables before seeing the results, I asked them to drag and drop the variable names to the label boxes along the corresponding axes on an empty plot. For both correct and incorrect responses, they received feedback. Then, in the next phase of the trial, participants were presented with the body of a graph with the axis correctly labeled. The graph extended a 15.06° x 12.07° region at the center of the screen. At the center of the plot, a green (hex code: #00FF00) fixation sign was presented for 500 ms. Then, the fixation sign turned red (hex code: #FF0000) and stayed on the screen for another 500 ms, indicating that the data set would be revealed. Upon the offset of the red fixation sign, the data set that consisted of 20 dots appeared and was visible for 250 ms. Then the data disappeared, leaving behind an empty plot with the axes labeled. On this scatterplot, participants drew their trend-line estimates. I recorded the coordinates of two points at which viewers started to draw and finished drawing.

Overall, in the main part of the experiment, participants completed 30 trials involving 30 separate scenarios; of all trials, 24 were included to test main hypotheses, and the remaining six were filler scenarios depicting neutral relationships randomly distributed among negative and positive scenarios. Of the 24 trials, on half of them, the scatterplots contained an outlier point. In addition, while half of the participants saw a scatterplot with an outlier point (from here onward XY-outlier) for a scenario, the other half observed a scatterplot without an outlier (from here onward NO-outlier); thus, from the 24 scenarios, I created two separate sets. In total, each participant was shown 12 XY-outlier and 12 NO-outlier scatterplots. In the XY-outlier displays, the outlier was at a typically expected position (see Figure 6). The fictitious scenarios and outlier positions were determined based on two extensive pilot studies in which, in total, 581 participants reviewed a set of 54 unique scenarios and identified the ones with the most homogeneous answers. Specifically, I extracted the distribution of correlation magnitude estimates for each scenario and picked the ones with the least mean and median differences based on the pilot work (for details, see Appendix A). Participants estimated the magnitude of the relationship between each pair of variables; I chose pairs of variables expected to have moderate-to-high correlation magnitudes (absolute mean expected correlation range: [.38, .86], $M = .63$, $SD = .15$). For each fictitious context, participants also had rank-ordered the plausibility/expectedness of the four possible XY-outlier cases (top-left, top-right, bottom-left, bottom-right). Based on this pilot work, I identified 24 scenarios and determined the most expected outlier in each case. Then, I matched each scenario with scatterplot displays with and without an outlier. Importantly, the depicted correlation magnitude on scatterplots was matched with the mean expected correlation magnitude from the previous surveys.

Each scatterplot was shown at the center of the screen, extending a region of 15.06° x 12.07°. The 20 dots were inside a 9.45° x 9.45° square area, marked with black (hex code: #000000) squares on a gray (hex code: #808080) background. The centroid of the 20 dots was at the center of the 9.45° x 9.45° square area for all displays. Throughout the displays, the standard deviation of 20 dots was the same, 1.56°. While the no outlier dots could vary up to 2 standard deviations on both dimensions, outlier dots were 2.5 to 3 standard deviations away from the centroid on both axes.



Note: Variable names are identical in the upper (a & b) and bottom (c & d) row; however, the scatterplot either contains an outlier (a & d) or does not (b & c). The magnitude of the overall pattern for each context is similar across scatterplots. Participants saw one of each pair (i.e., a and c, or b and d). Different participants saw one of each one of these pairs.
Figure 6. Example scatterplots depicting positive and negative trends with and without outliers

2.2.3 Procedure

The experiment took place in a well-lit room. First, participants provided informed consent and completed a training phase. The training consisted of two components. In the first phase, participants were given a brief description of what trend-lines are and were familiarized with trend-line drawing. I presented them with six successive scatterplots without any variable name on the axes. Those scatterplots consisted of random points that were in the area enclosed by x and y axes, and they specifically did not include any outliers. This part of the training was just to teach participants how to draw a trend-line by using a keyboard touchpad. In the second phase of the training, participants were familiarized with how the actual experimental trials would proceed. After receiving those instructions, participants completed a practice trial that described a study that showed the relationship between ice cream sales and temperature. They could repeat this practice trial if needed, otherwise, they continued with the actual experimental trials. Even though the scenarios were fictitious, participants were told that these scenarios were summaries of actual published studies. After completing the 30 trend-estimation trials, participants filled a demographic form, and then they were thanked and debriefed. Overall, the experiment took approximately 35 minutes.

2.2.4 Apparatus

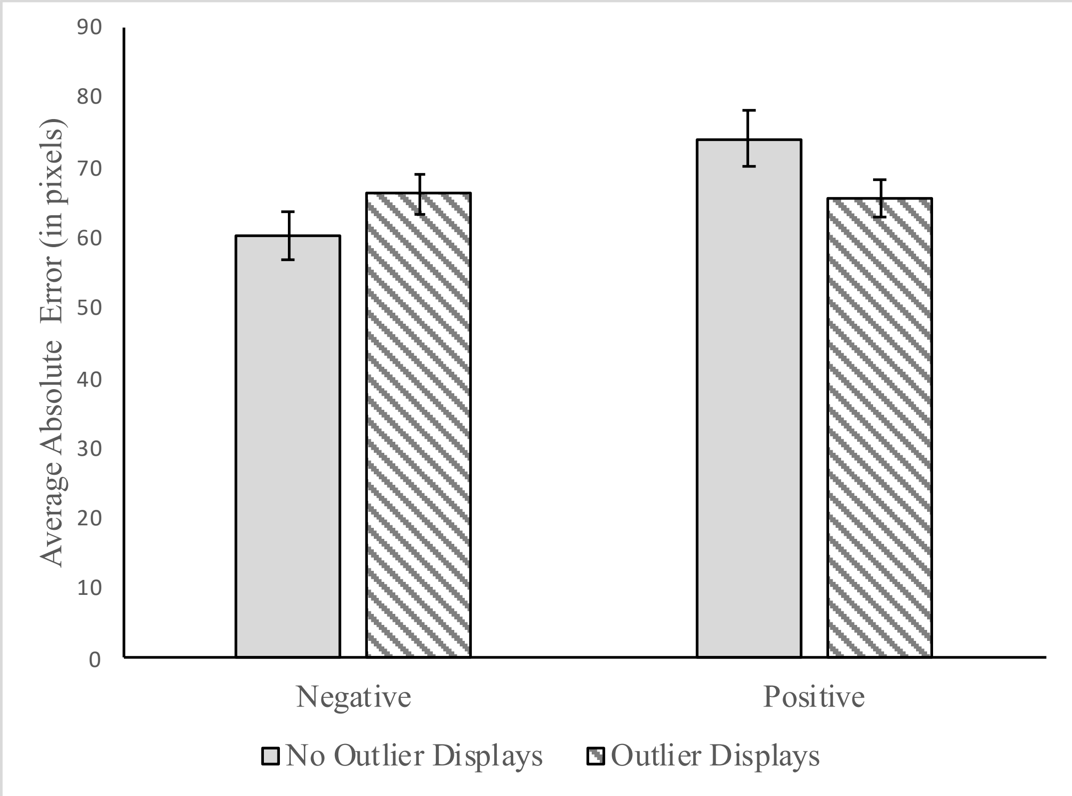
HP ProOne 600 G1 All-in-One PC was used to run the experiment. That model has integrated Intel HD Graphics 4600, and discrete AMD Radeon HD 7650A graphics card, and Intel 4th generation core i7 processor those are embedded into the monitor. The monitor has a LED-backlit display of which size is diagonally 21 in. I set the display resolution to 1920 x 1080 pixels, with a refresh rate of 60Hz. The experiment was programmed in MATLAB, using the Psychtoolbox (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007).

2.2.5 Results and discussion

For each participant, I first excluded the trials in which they did not provide a trend-line. Then for each participant, I separately calculated their AAE on each trial. The AAE is the average absolute pixel difference between two lines. Here AAE was the average absolute pixel difference between viewers' trend-line estimate and the OLS solution including all 20 dots. Then, for each participant, I separately calculated the average AAE for each outlier condition for positive and negative trends.

First, I conducted a 2(Group: Set1, Set2) x 2(Outlier Presence: No outlier displays, Outlier displays) x 2(Trend Direction: Negative, Positive) mixed ANOVA. The results pointed out that between-subjects factor group did not have a main effect ($F < 1$), also the two- and three-way interaction(s) that included group as factor were not significant (i.e. $F < 1$ for two-way interactions; and $F(1, 150) = 1.418$, $p = .236$, $\eta_p^2 = .009$ for the three-way interaction between group, outlier presence, and trend direction). Therefore, I do not discuss the effects of different sets any further. I found no main effect of outlier presence on AAE, $F < 1$, replicating my earlier findings with moderate-to-high correlations in Experiment 1. However, I found a small marginal main effect of trend direction, $F(1, 150) = 3.422$, $p = .066$, $\eta_p^2 = .022$, that suggested when the depicted relationship was positive ($M = 69.900$, 95% CI [63.710, 76.090]) trend-line estimates deviated more from the 20 points OLS solution compared to when it was negative ($M = 63.212$, 95% CI [57.301, 69.124]). Critically, I found a significant interaction between outlier presence and trend direction, $F(1, 150) = 19.005$, $p < .001$, $\eta_p^2 = .112$. Specifically, in no outlier displays viewers' trend-line estimates deviated significantly more from 20 points OLS solution while the depicted trend was positive ($M = 74.167$, 95% CI [66.551, 81.783]) compared to when it was negative ($M = 60.164$, 95% CI [53.316, 67.012]), see Figure 7. This may be partly due to the axis constraining responses more when the trend is negative compared to when it is positive (i.e., the upper right-hand quadrant of a scatterplot is not as constrained). The fact that such a difference does not emerge for outlier trials may be linked to how outlier points themselves constrain responses.[4]

[4] To eliminate a power-based explanation of the above resulted null findings, I also conducted Bayesian repeated measures ANOVA (Appendix B) on JASP (JASP Team, 2020) with group as a between-subjects factor, outlier presence and trend direction as within-subjects factors. The results supported my findings on the previous analysis (see Appendix B, Table B1 for the Bayesian analysis results).



Note: Error bars represent ±1 standard errors.
Figure 7. Mean AAE of displays with and without outlier for negative and positive trends

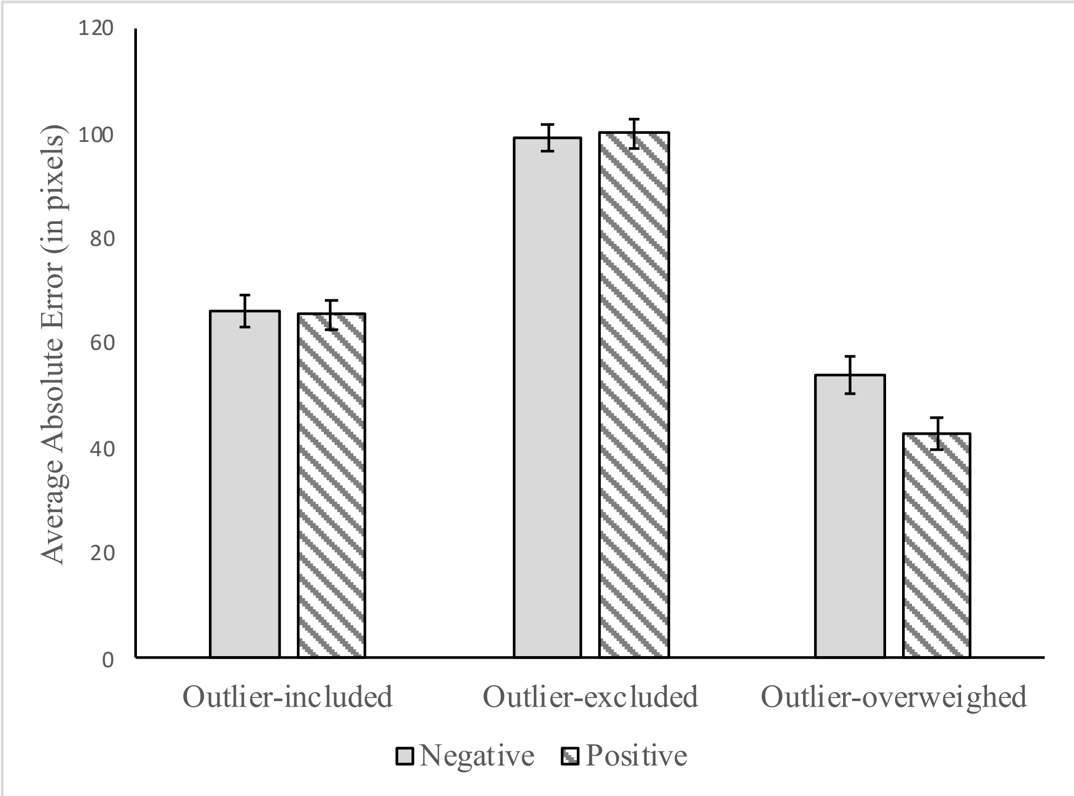
To determine whether viewers' trend-line estimates were closer to another trend line rather than the 20 points OLS solution, I also conducted a 2(Group: Set1, Set2) x 2(Trend Direction: Negative, Positive) x 3(Error Type: Outlier-included, Outlier-excluded, Outlier-overweighed) mixed ANOVA for outlier displays. Here error type indicates the line to which viewers' estimate was compared. The outlier-included line was the OLS solution for 20 dots. The outlier-excluded line was the OLS solution for 19 dots excluding the outlier point. Finally, the outlier-overweighed line was a trend-line that passed through the centroid of 19 points in the main data cloud and the outlier point. As before, there was no main effect of group (i.e. set), $F < 1$, $p = .504$, $\eta_p^2 = .003$, and all two-way and three-way interactions that included group variable were not significant (F = .616, $F = .289$, $F = .81$, respectively for group X trend direction, group X outlier presence, and group X outlier X trend direction interactions). There was no main effect of trend direction, $F(1, 150) =$

1.305, $p = .255$, $\eta_p^2 = .009$. Critically though, there was a main effect of the error type that showed viewers' trend-line estimations were closest (*M* = 48.421, 95% CI [43.408, 53.434]) to the outlier-overweighed trend-line, and they were the farthest from the outlier-excluded trend-line (*M* = 99.686, 95% CI [95.133, 104.239]), $F(1.055, 158.196) = 381.023$, $p < .001$, $\eta_p^2 = .718$. This main effect was further qualified by a trend direction and error type interaction, $F(1.089, 163.374) = 16.847$, $p < .001$, $\eta_p^2 = .101$. As shown in Figure 8, the AAE from outlier-included and outlier-excluded trend-lines did not change as a function of trend-direction. However, viewers' trend-line estimates were closer to the outlier-overweighed trend-line more when the depicted trend was positive (*M* = 42.822, 95% CI [37.129, 48.514]) compared to it was negative (*M* = 54.021, 95% CI [46.812, 61.229]), hinting at the different levels of perceptual saliency of outliers[5] for positive and negative trends.[6] To summarize, I found that outlier presence did impact how responses deviated from trend-line solutions including either 19 or 20 dots. Interestingly, I found that responses approached outlier-weighed solutions more than the two other solutions, suggesting that expected outliers influenced trend-estimates. Additionally, the impact of expected outliers was even more substantial when the depicted relationships were positive, suggesting that outliers may vary in salience based on their position to the main data cloud.

---

[5] The effect of outlier on correlation magnitude did not vary for two sets, for both sets the outlier in average increased the correlation between the remaining 19 points .27. Also, the impact of the outlier did not change for negative and positive trends. For the Set 1, the outlier increased the correlation between remaining dots .29 and .25 respectively for negative and positive trends. It was .27 and ,24 for the Set 2.

[6] To further support the ANOVA results I conducted another Bayesian repeated measures ANOVA (Appendix B) on JASP (JASP Team, 2020) while the group was the between-subjects factor, trend direction and error type were within-subjects factors. In general, Bayesian repeated measures ANOVA supported my results (see Appendix B, Table B2 for the Bayesian results in detail).



Note: Error bars represent ±1 standard errors.
Figure 8. Mean AAE of responses compared to outlier-included, outlier-excluded, and outlier-overweighed trend-lines

CHAPTER 3

GENERAL DISCUSSION

In this thesis, I studied the effect of outlier presence on trend-line estimates and further extended these results by investigating if viewers' prior beliefs influenced the effect of outlier presence. In Experiment 1, I found that viewers integrate outliers in their trend-line estimates. However, with the decreasing level of correlation magnitude, viewers' estimation error increased disproportionately in the XY-outlier condition. Exploratory results showed that this result was caused by the impact of XY-outliers in low correlations, those type of XY-outliers even changed the direction of the depicted relationship on the scatterplot. Yet, viewers' estimates were closest to an OLS solution that weighed the outlier equally with the remaining points. Those findings demonstrated that viewers did not discount a single outlier in their trend-line estimates in the absence of content on the scatterplot. In Experiment 2, I investigated the effect of an expected outlier in a familiar context on trend-line estimates. Results showed that when the outlier's location matched prior beliefs and the outlier was trend-consistent, viewers overweighed the outlier, and their estimates were biased to an outlier-overweighed trend-line solution. In the following paragraphs, I discuss the present findings in the context of prior findings in the literature.

I found correlation magnitude is a significant factor that impacts trend-line estimates. As Correll and Heer (2017) demonstrated, I also found that viewers' accuracy in trend-line estimation got worse as the depicted correlation got weaker. It is possible that when the depicted correlation was weaker, there was additional noise in the perceived display making it harder to rely on a well-known schema (e.g., the x = y line) to extract a trend-estimate (Pinker, 1990). This finding was also similar to findings on correlation estimation, where viewers are more accurate when estimating stronger as opposed to weaker correlations (e.g., Meyer, & Shinar, 1992; Rensink, 2017).

Of more critical interest was the effect of the outlier on trend-line estimates. Contrary to previous findings that suggested that outliers were discounted from trend-estimates (e.g., Correll & Heer, 2017), I found that viewers did not discount outliers; in fact, they were pulled towards the outlier point, especially when there was familiar content. There could be two possible explanations for this finding. It is possible that the outlier point was not visually distinctive resulting in viewers treating the outlier point as the remaining 19 other points on the scatterplot. However, the disproportionate level of estimation error for XY-outliers in low correlation magnitude compared to that for NO-outlier condition suggested that viewers were sensitive to the presence of outliers, at least under some conditions. The reason that I did not observe that effect on the higher level of correlation magnitude could be caused by the disproportionate impact of the outlier on overall correlation magnitude of the display. That is how much the outlier changes the correlation magnitude of the remaining 19 points. That impact was the highest for the lower level of correlation magnitude, and on those displays, the outlier even changed the direction of the relationship between the 19 points. The second explanation could be that although viewers were able to detect the outlier, they preferred to equally weigh that outlier with the remaining dots. That could be the case since, except the lower level of correlation magnitude, the outlier was always at a location that was trend-consistent and increased the correlation magnitude between the remaining 19 points. In more detail, even though these outliers were mathematically and spatially distinct from the remaining points, they have a nature that decreases the overall noise since they increase the magnitude of the existing relationship. Thus, viewers could prefer not to discount those outliers. Nevertheless, while the saliency is a critical factor that affects outlier detection and discount on subsequent processes (Avci, & Boduroglu, in press), I admitted that having only outliers that strengthen the existing relationship is a limitation. Because the direction of the effect of the outlier on overall correlation could change its saliency regardless of its location, further studies that include more systematic manipulation of outlier saliency could shed light on that discussion.

The current thesis further extended findings by showing the effect of the outlier in scatterplot on trend-line estimates by incorporating a familiar content into the scatterplot. As I expected, viewers overweighed the outlier in their trend-line estimates when there were familiar content and a trend-consistent outlier in the scatterplot. That suggested that prior beliefs could influence our perception of trend-lines in scatterplots in a direction that favors those beliefs (Anderson, 1995; Billman et al., 1992). The results also supported the findings of Karduni and colleagues (2020), which pointed out that viewers' do not update their prior belief when the data is congruent with their expectations. In a similar vein, the current findings showed that viewers draw a trend-line estimate which promotes their prior beliefs by overweighing the outlier.

Surprisingly, in Experiment 2, I found a small marginal effect showing that error was larger when there were no outliers, and the trend was negative. More critically, when I compared trend-line estimates with the outlier-bias trend line, estimates were closer to the outlier-bias trend-line only when there was a positive trend. The fact that the outlier acted as a stronger anchor when the depicted relationship was positive may be linked to how such an outlier may serve as an anchor point, limiting the possible paths the estimate could follow in the upper right-hand corner of the graph. In other words, because scatterplots are only constrained on the left and the bottom with the y and x axes, respectively, participants may have no clear anchor to place their estimates through when there is no outlier. An intuitively expected outlier presented along with a positive trend may serve such an anchor, constraining the response space. These findings are in line with arguments that spatially coded data is open to perceptual biases. For example, to investigate whether spatial coding of data could generate a perceptual bias on viewers' estimate, in a recent study, Xiong and colleagues presented participants with bar and line graphs and asked them to report their average positions (Xiong, Ceja, Ludwig, & Franconeri, 2020). They found that viewers systematically, under-, and over-estimated the position of lines and bars, respectively.

This thesis showed that presenting people with familiar content could make them more prone to outlier-bias when the outlier position is congruent with viewers' prior belief. This tells us that graph perception is not purely bottom-up; instead, previous knowledge and low-level perceptual features interact and establish a whole perception. In their Construction-Integration (CI) model proposal for graph comprehension, Freedman and Shah (2002) suggested that graphical comprehension is analogous to text comprehension, and both perceptual and cognitive mechanisms play a role in whole graph comprehension. They also stated that graphical perception is the first step of graph comprehension, and in that phase, viewers construct a coarse mental model of the display. Based on growing ensemble perception research, I argue that ensemble mechanisms utilize the extraction of summary statistics of the display in the graphical perception phase. More importantly, I also established here that those summary statistics somehow interact with the viewers' previous knowledge, and together they create a whole perception.

However, I still do not know how this summary statistics and previous knowledge interact, and I do not know how presenting an outlier of which location is incongruent with viewers' expectation and/or trend-inconsistent could influence the results. In addition, encoding time could also possibly affect viewers' trend-line estimates. For example, Epstein et al. (2020) demonstrated that viewers reduce the noise iteratively, which means the impact of an outlier on summary statistics decreases with the extended time interval. If I had increased the encoding time of scatterplot in a similar vein, viewers' estimate could have been closer to the outlier-excluded trend-line in the absence of content. On the other hand, I speculate that the outlier-overweigh effect could have remained while there is a familiar content and the outlier was expected. That is because viewers have a propensity to discount the data while they have prior beliefs. Lastly, changing the aspect ratio of the axes could have impacted trend-line estimates. However, I think that would be a factor only if there were a content since Rensink's (2017) research showed the invariance of correlation magnitude perception to the aspect ratio of the display for scatterplot without a content. Thus, while it is likely that correlation magnitude and trend-line perception are somehow related to each other, it is not expected to find an effect of aspect ratio on trend-line estimates. However, when there is content, a higher aspect ratio could lead to more over-weigh of the outlier while making the trend-line steeper. So, future studies are needed to answer these open questions.

In sum, in this thesis, I demonstrated that the presence of an outlier on scatterplots impacts how trend-lines are estimated, especially when there is a meaningful context, the outliers have a greater impact on trend-line estimates. This suggests that top-down expectations and bottom-up ensemble perception mechanisms work in tandem to shape graph perception from a very early point onwards.

APPENDIX A

SCENARIO NORM SURVEY

A group of graduate and undergraduate students, all members of the lab, identified pairs of variables that one might encounter in real life, and we created possible scenarios (i.e., bivariate variables) for the subsequent experiment. After some internal group discussions, we determined the final 34 scenarios. Those scenarios consisted of a pair of variables everyone could encounter in daily life, for example, degree of education and aggression. Then to learn Boğaziçi University undergraduates' thoughts about these relationships, we asked undergraduates to estimate the correlation magnitude between these variable pairs. We also asked them to rank the likelihood of encountering each one of four potential outlier locations (i.e., both variables are too high; both variables too low; and while one of them is too high, and the other one is too low). Each participant responded to these questions for each of 34 scenarios in an online survey.

A.1 Survey 1

A.1.1 Participants

Six-hundred and eighty-six Boğaziçi University undergraduates took part in the online survey for one course credit. Before analyzing the data, I discarded participants who completed the survey more than once and only kept their first completion. Subsequently, for the remaining data, I excluded ones who answered less than 80% of 34 scenarios (less than 28 scenarios) or used a number more than a single time while sorting the potential outlier locations. After the cleaning of data, 276 (at least 251 for a scenario, red meat consumption and the risk of colon cancer) participants remained.

A.1.2 Material and design

A total of 34 scenarios were presented to each participant in random order. For each scenario, participants first stated their correlation estimation for that scenario by using a slider from -1 to +1 in 0.1 intervals. Then they completed a sorting task for potential outlier locations by choosing a number from 1 to 4, while 1 indicated the most possible outlier location, 4 indicated the least possible one.

A.1.3 Procedure

First, each participant was informed about the following tasks, and in a single training trial, they were taught to use a slider while estimating correlation magnitude. Then they completed correlation magnitude estimation and potential outlier location sorting tasks.

A.1.4 Results

I extracted descriptive statistics of each scenario and analyzed the possibility of potential outlier locations in participants' minds by conducting a one-way ANOVA for each scenario; while the four potential outlier locations were IV, participants' answers between 1 and 4 to indicate the order of that location was DV. The results showed that for the negative scenarios, while the most possible outlier location was at the bottom-right of the screen, the least possible one was at the top-right of the screen, and it was vice versa for the positive scenarios.

A.2 Survey 2

When I analyzed the participants' responses in the previous survey, I figured out that there were only two negative scenarios. Since I planned to control the direction of scenarios for the subsequent experiment, we focused on creating some negative scenarios in our laboratory. Everyone in our laboratory group proposed their negative scenarios, and after an inner-group discussion, we determined 20 negative scenarios. Those scenarios consisted of a pair of variables that anyone could encounter as the scenarios we created earlier, for example, age and speed of walking. To avoid using only negative scenarios that create a bias on participants' responses, I added 14 positive scenarios from the first survey to those 20 negative ones. Then conducted the same survey with another group of Boğaziçi University undergraduates by using 34 scenarios.

A.2.1 Participants

Five-hundred and forty-one Boğaziçi University undergraduates took part in the online survey for one course credit. Before analyzing the data, I ran the same data cleaning procedure as in the previous survey. After the cleaning of the data, 305 (at least 291 for a scenario, caffeine consumption and concentration) participants remained.

A.2.2 Material and design

It was the same as the previous survey.

A.2.3 Procedure

It was the same as the previous survey.

A.2.4 Results

Likewise, in the previous survey, first, I extracted the descriptive statistics of each scenario and then analyzed the possibility of potential outlier locations in participants' minds by conducting a one-way ANOVA for each scenario. I replicated the results in the first survey pointed out that for the negative scenarios, while the most possible outlier location was at the bottom-right of the screen, the least possible one was at the top-right of the screen, and it was vice versa for the positive ones.

A.3 Choosing scenarios for the subsequent experiment

Since I used scatter plots on which the absolute correlation magnitude was between .4 and .8 in Experiment 1, my first criterion was to determine scenarios whose absolute mean estimated correlation magnitude is between .4 and .8. Additionally, I also wanted to determine scenarios in which the mean and median difference is as less as possible. Because the more mean and median difference means the more heterogeneous thoughts about a scenario, which I do not want since I would not be able to classify participants' side in the subsequent experiment. By using the previous two criteria, I determined 12 negative and 12 positive scenarios.

APPENDIX B

BAYESIAN ANALYSES

To compare different models, we can compare the different effects. The Analysis of Effects (Table B1) gives Bayes factors for the inclusion of each effect that appears in at least one model. The table showed that while there was weak evidence in favor of the inclusion of trend direction effect ($BF_{incl} = 2.750$), there was moderate evidence for the inclusion of trend direction and outlier interaction effect ($BF_{incl} = 7.468$). For the other effects, there was moderate evidence against the inclusion (see pg. 14 Stevens, 2019 for BF cutoff values).[7]

In conclusion, I showed that moderate evidence in favor of the inclusion of trend direction and outlier presence interaction and weak evidence for the inclusion of trend direction predicts the AAE between viewers' trend-line estimates and 20 dots OLS solution as I demonstrated on the previously mentioned 2(Group: Set1, Set2) x 2(Outlier Presence: No outlier displays, Outlier displays) x 2(Trend Direction: Negative, Positive) mixed ANOVA. Importantly, for other main effects, and two- and three-way interaction(s), the Bayesian repeated-measures ANOVA yielded moderate evidence against the inclusion of those effects that suggested those effects do not predict the AAE.

In a similar vein, to further support the results of 2(Group: Set1, Set2) x 2(Trend Direction: Negative, Positive) x 3(Error Type: Outlier-included, Outlier-excluded, Outlier-overweighed), I conducted another repeated measures Bayesian ANOVA (for the model comparison see Table B2). There was weak to extreme evidence against the

[7] Beware that the interpretation could change based on which BF score that we are looking at. But the magnitude (e.g. being extreme, or weak) of effect is the same.

Table B1. Analysis of Effects

Effects	P(incl)	P(incl|data)	BF_{incl}
Trend Direction	0.263	0.477	2.750
Outlier Presence	0.263	0.061	0.101
Group	0.263	0.142	0.171
Trend Direction x Outlier Presence	0.263	0.340	7.468
Trend Direction x Group	0.263	0.017	0.137
Outlier Presence x Group	0.263	0.009	0.141
Trend Direction x Outlier Presence x Group	0.053	1.829e -4	0.206

Note. Compares models that contain the effect to equivalent models stripped of the effect. Higher-order interactions are excluded. The analysis is suggested by Sebastiaan Mathôt (Mathôt, 2017)

inclusion of all effects except error type, and trend direction, and error type interaction. There was weak evidence in favor of inclusion for the trend direction and error type interaction (BF_{incl} = 1.434). Importantly, the evidence for the inclusion of error type was extreme (BF_{incl} = 2.815e+85). Like the first Bayesian analysis, this analysis also further supports my results in the main mixed ANOVA.

Table B2. Analysis of Effects

Effects	P(incl)	P(incl|data)	BF_{incl}
Trend Direction	0.263	0.219	0.497
Error	0.263	0.667	2.815e +85
Group	0.263	0.172	0.215
Trend Direction x Error	0.263	0.329	1.434
Trend Direction x Group	0.263	0.024	0.240
Error x Group	0.263	0.006	0.029
Trend Direction x Error x Group	0.053	1.879e -5	0.050

Note: Compares models that contain the effect to equivalent models stripped of the effect. Higher-order interactions are excluded. The analysis is suggested by Sebastiaan Mathôt (Mathôt, 2017).

REFERENCES

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. In *Trends in Cognitive Sciences* (Vol. 15, Issue 3, pp. 122–131).

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention: Research article. *Psychological Science*.

Anderson, C. A. (1995). Implicit Personality Theories and Empirical Data: Biased Assimilation, Belief Perseverance and Change, and Covariation Detection Sensitivity. *Social Cognition*, *13*(1), 25–48.

Anderson, C. A., & Kellam, K. L. (1992). Belief Perseverance, Biased Assimilation, and Covariation Detection: The Effects of Hypothetical Social Theories and New Data. *Personality and Social Psychology Bulletin*, *18*(5), 555–565.

Avci, B., & Boduroglu, A. (in press). Contributions of ensemble perception to outlier presentation precision.

Baumgartner, H. (1995). On the Utility of Consumers' Theories in Judgments of Covariation. *Journal of Consumer Research*, *21*(4), 634.

Billman, D., Bornstein, B., & Richards, J. (1992). Effects of expectancy on assessing covariation in data: "Prior belief" versus "meaning." *Organizational Behavior and Human Decision Processes*, *53*(1), 74–88.

Boduroglu, A., & Shah, P. (2009). Effects of spatial configurations on visual change detection: An account of bias changes. *Memory and Cognition*.

Boduroglu, A., & Shah, P. (2014). Configural representations in spatial working memory. *Visual Cognition*.

Boduroglu, A., & Yildirim, I. (2020). Statistical summary representations of bound features. *Attention, Perception, and Psychophysics*, *82*(2), 840–851.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in Visual Working Memory: Using Statistical Regularities to Form More Efficient Memory Representations. *Journal of Experimental Psychology: General*.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*.

Cleveland, W. S., Diaconis, P., & Mcgill, R. (1982). Variables on scatterplots look more highly correlated when the scales are increased. *Science*.

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, *79*(387), 531–554.

Coronavirus in charts: the fact-checkers correcting falsehoods. (2020). *Nature*.

Correll, M., & Heer, J. (2017). Regression by eye: Estimating trends in bivariate visualizations. *Conference on Human Factors in Computing Systems - Proceedings*, *2017-May*, 1387–1396.

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

Epstein, M. L., Quilty-Dunn, J., Mandelbaum, E., & Emmanouil, T. A. (2020). The Outlier Paradox: The Role of Iterative Ensemble Coding in Discounting Outliers. *Journal of Experimental Psychology: Human Perception and Performance*.

Freedman, E. G., & Shah, P. (2002). Toward a model of knowledge-based graph comprehension. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *2317*, 18–30.

Gleicher, M. (2018). Considerations for Visualizing Comparison. *IEEE Transactions on Visualization and Computer Graphics*, *24*(1), 413–423.

Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception, and Psychophysics*.

Harrison, L., Yang, F., Franconeri, S., & Chang, R. (2014). Ranking visualizations of correlation using Weber's law. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 1943–1952.

Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science*, *3*(3), 446–474.

Hochstein, S., Pavlovskaya, M., Bonneh, Y. S., & Soroker, N. (2018). Comparing set summary statistics and outlier pop out in vision. *Journal of Vision*.

Hullman, J., Adar, E., & Shah, P. (2011). Benefitting InfoVis with visual difficulties. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2213–2222.

JASP Team. (2020). JASP (Version 0.13.1). In *[Computer software]*.

Jennings, D., Amabile, T. M., & Ross, L. (1982). Informal covariation assessment: Data-based vs. theory-based judgments.

Jeong, J., & Chong, S. C. (2020). Adaptation to mean and variance: Interrelationships between mean and variance representations in orientation perception. *Vision Research*, *167*(January), 46–53.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of Visual Short-Term Memory. *Journal of Experimental Psychology: Learning Memory and Cognition*.

Karduni, A., Markant, D., Wesslen, R., & Dou, W. (2020). A Bayesian cognition approach for belief updating of correlation judgement through uncertainty visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 1–1.

Khayat, N., & Hochstein, S. (2018). Perceiving set mean and range: Automaticity and precision. *Journal of Vision*, *18*(9), 1–14.

Kleiner, M., Brainard, D. H., Pelli, D. G., Broussard, C., Wolf, T., & Niehorster, D. (2007). What's new in Psychtoolbox-3? *Perception*.

Kosslyn, S. M. (2012). Graph Design for the Eye and Mind. In *Graph Design for the Eye and Mind*.

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

Lewandowsky, S., & Spence, I. (1989). Discriminating strata in scatterplots. *Journal of the American Statistical Association*.

Mathôt, S. (2017). Bayes like a Baws: Interpreting Bayesian repeated measures in JASP. *Retrieved June*, *12*, 2018.

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.

Mutluturk, A., & Boduroglu, A. (2014). Effects of spatial configurations on the resolution of spatial working memory. *Attention, Perception, and Psychophysics*.

Nothelfer, C., Gleicher, M., & Franconeri, S. (2017). Redundant encoding strengthens segmentation and grouping in visual displays of data. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(9), 1667–1676.

Padilla, L. M., Creem-Regehr, S. H., Hegarty, M., & Stefanucci, J. K. (2018). Decision making with visualizations: a cognitive framework across disciplines. *Cognitive Research: Principles and Implications*, *3*(1).

Padilla, L. M., Ruginski, I. T., & Creem-Regehr, S. H. (2017). Effects of ensemble and summary displays on interpretations of geospatial uncertainty data. *Cognitive Research: Principles and Implications*, *2*(1).

Pelli, D. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies in: Spatial Vision Volume 10 Issue 4 (1997). *Spatial Vision*.

Pinker, S. (1990). A theory of graph comprehension. *Artificial Intelligence and the Future of Testing*.

Pollack, I. (1960). Identification of visual correlational scatterplots. *Journal of Experimental Psychology*.

Rensink, R. A. (2012). Invariance of Correlation Perception. *Journal of Vision*, *12*(9), 433–433.

Rensink, Ronald A. (2017). The nature of correlation perception in scatterplots. *Psychonomic Bulletin and Review*, *24*(3), 776–797.

Rensink, R. A., & Baldridge, G. (2010, June). The perception of correlation in scatterplots. In *Computer Graphics Forum* (Vol. 29, No. 3, pp. 1203-1210). Oxford, UK: Blackwell Publishing Ltd.

Sarikaya, A., Gleicher, M., & Szafir, D. A. (2018). Design Factors for Summary Visualization in Visual Analytics. *Computer Graphics Forum*, *37*(3), 145–156.

Semizer, Y., & Boduroglu, A. (in press). Variance leads to overestimation of mean estimates in ensemble perception.

Shah, P., & Freedman, E. G. (2011). Bar and line graph comprehension: An interaction of top-down and bottom-up processes. *Topics in Cognitive Science*, *3*(3), 560–578.

Shah, P., Freedman, E. G., & Vekiri, I. (2005). The Comprehension of Quantitative Information in Graphical Displays. In *The Cambridge Handbook of Visuospatial Thinking* (pp. 426–476). Cambridge University Press.

Sher, V., Bemis, K. G., Liccardi, I., & Chen, M. (2017). An Empirical Study on the Reliability of Perceiving Correlation Indices using Scatterplots. *Computer Graphics Forum*, *36*(3), 61–72.

Stevens, J. R. (2019). Introduction to Bayes factors.

Strahan, R. F., & Hansen, C. J. (1978). Underestimating Correlation from Scatterplots. *Applied Psychological Measurement*.

Szafir, D. A., Haroz, S., Gleicher, M., & Franconeri, S. (2016). Four types of ensemble coding in data visualizations. *Journal of Vision*, *16*(5).

Valdez, A. C., Ziefle, M., & Sedlmair, M. (2018). Priming and Anchoring Effects in Visualization. *IEEE Transactions on Visualization and Computer Graphics*, *24*(1), 584–594.

van Onzenoodt, C., Huckauf, A., & Ropinski, T. (2020). On the perceptual influence of shape overlap on data-comparison using scatterplots. *Computers and Graphics (Pergamon)*.

Wei, Y., Mei, H., Zhao, Y., Zhou, S., Lin, B., Jiang, H., & Chen, W. (2020). Evaluating perceptual bias during geometric scaling of scatterplots. *IEEE Transactions on Visualization and Computer Graphics*.

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble Perception. *Annual Review of Psychology*, *69*, 105–129.

Wright, J. C., & Murphy, G. L. (1984). The utility of theories in intuitive statistics: The robustness of theory-based judgments. *Journal of Experimental Psychology: General*, *113*(2), 301–322.

Xiong, C., Ceja, C. R., Ludwig, C. J. H., & Franconeri, S. (2020). Biased Average Position Estimates in Line and Bar Graphs: Underestimation, Overestimation, and Perceptual Pull. *IEEE Transactions on Visualization and Computer Graphics*, *26*(1), 301–310.

Yang, F., Harrison, L. T., Rensink, R. A., Franconeri, S. L., & Chang, R. (2019). Correlation judgment and visualization features: A comparative study. *IEEE Transactions on Visualization and Computer Graphics*.

Zacks, J. M., & Franconeri, S. (2020). Reading the pandemic data. *Sciencemag.*