Seeing What You Believe or Believing What You See? Belief Biases Correlation Estimation

Cindy Xiong, Chase Stokes, Yea-Seul Kim, Steven Franconeri
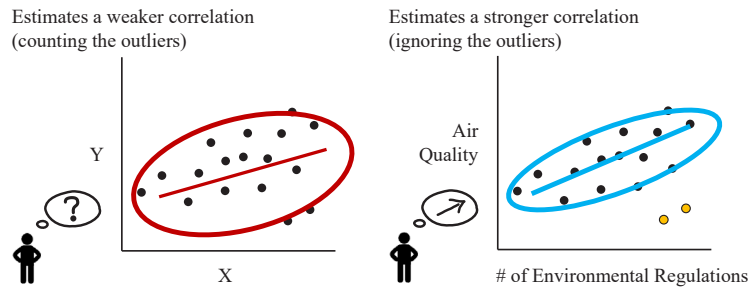

Fig. 1. People might induce themselves to see a correlation as weak or strong, depending on existing beliefs about data.

Abstract—When an analyst or scientist has a belief about how the world works, their thinking can be biased in favor of that belief. Therefore, one bedrock principle of science is to minimize that bias by testing the predictions of one's belief against objective data. But interpreting visualized data is a complex perceptual and cognitive process. Through two crowdsourced experiments, we demonstrate that supposedly objective assessments of the strength of a correlational relationship can be influenced by how strongly a viewer believes in the existence of that relationship. Participants viewed scatterplots depicting a relationship between meaningful variable pairs (e.g., number of environmental regulations and air quality) and estimated their correlations. They also estimated the correlation of the same scatterplots labeled instead with generic 'X' and 'Y' axes. In a separate section, they also reported how strongly they believed there to be a correlation between the meaningful variable pairs. Participants estimated correlations more accurately when they viewed scatterplots labeled with generic axes compared to scatterplots labeled with meaningful variable pairs. Furthermore, when viewers believed that two variables should have a strong relationship, they overestimated correlations between those variables by an r-value of about 0.1. When they believed that the variables should be unrelated, they underestimated the correlations by an r-value of about 0.1. While data visualizations are typically thought to present objective truths to the viewer, these results suggest that existing personal beliefs can bias even objective statistical values people extract from data.

Index Terms—Data Visualization, Visual Analysis, Data Interpretation, Perception, Cognition, Beliefs, Motivated Perception

---

1 INTRODUCTION

Your new data have arrived. With the intent of showing a relationship between the strength of environmental regulations in a region and that region's air quality, you plot your data values in a scatterplot. At first, the correlation is barely noticeable. But you then notice a small set of outliers. If you ignore those, the correlation looks as strong as you expected, as illustrated in Figure 1. This hypothetical example may be more common than one would hope. This type of data analysis is intended to proceed in a purely 'bottom-up' fashion, driven only by the data itself. However, motivation, belief, and experience can bias us across a complex set of 'top-down' interactions [50,53]. When people viewed an ambiguous figure which was carefully balanced to be interpreted either as the letter B or the number 13, those who expected letters were more likely to see 'B', while those who expected to see numbers were more likely to see '13' [4].

Similar top-down biases also seem to appear in data analysis. When viewing a visualization depicting global temperature trends, liberals focused more on the increasing trend, confirming their worries about climate change. Conservatives focused more on the overall flatness of the line and the small effect size of the increase, which were more consistent with doubts about those concerns [50].

In another study, participants were told a 'top-down' story explaining a pattern present across randomly chosen data values in a visualization and were then explicitly told to ignore that story. Yet, later ratings of the 'bottom-up' salience of those values were higher for that subset of values than for others. The same participants even predicted that *other* people, naive to the top-down story, would also find those values more salient [87].

These past studies of top-down influences in data visualizations focus on how top-down information can bias the visual salience of certain values or attributes. In contrast, the present studies provide a quantitative estimate of how powerful this type of influence can be, providing a measurement of how a belief can change an 'objective' correlation estimate. We use scatterplots to show those correlations, as those plots are among the most commonly used visualizations and are especially valuable for depicting correlation [27,41,68], which is a critical attribute that impacts data-driven decisions [19]. Additionally, the perception of correlation has been relatively well-studied. People are fairly accurate at determining the Pearson's r correlation in scatterplots [68], and recent work has begun to explore the underlying visual features that people use as proxies when estimating correlations (e.g., the shape of the ellipse that encloses the scattered dots) [88].

Hypothesis: We hypothesize that people will estimate correlations more accurately when they see scatterplots labeled with a neutral pair of variables that are not associated with existing beliefs or abstracted data labels (e.g., X vs. Y). We also hypothesize that the amount of deviation in people's correlation estimation would be associated with

---

- *Cindy Xiong is with UMass Amherst.*
  E-mail: cindy.xiong@cs.umass.edu.
- *Chase Stokes is with UC Berkeley.*
- *Yea-Seul Kim is with the University of Wisconsin Madison,*
- *Steven Franconeri is with Northwestern University*

the strength of their belief, such that they would report a higher correlation for the same scatterplot when it depicts something that they strongly believe to be correlated and a lower correlation when the plot depicts something that they do not strongly believe to be correlated.

We are careful to state these hypotheses as beliefs affecting what people *report*, as opposed to them affecting what people *see*. Beliefs might affect various stages of processing on the road to a correlation estimate, including what people perceive, what values they preferentially attend, or what values they weigh most heavily when deciding what numeric estimate to report. Depending on a researcher's perspective, these stages can have boundaries that are firm, as some have argued that top-down beliefs and motivations do not directly impact perception [17], or on a continuum, as others have argued for a 'cognitive penetrability of perception' suggesting that top-down effects can directly influence what we see [66, 76].

Experiment Overview: We present two empirical experiments in which participants estimated correlations for identical data sets in scatterplots. The scatterplots have axes either labeled with meaningful variable pairs (e.g., environmental regulation and air quality) or labeled with abstracted variables 'X' and 'Y' (separated by a distractor task). In a separate section, participants reported how strongly they believed there to be a correlation between the variable pairs, either after they saw the scatterplots (Experiment 1a) or before (Experiment 2). In both studies, we found that people who believed the meaningful variable pairs to be correlated overestimated their correlation values significantly, while people who did not hold those beliefs underestimated the correlations significantly, in comparison to the scatterplots labeled with abstracted variables. We additionally conducted a control experiment (Experiment 1b) to test whether the differences in correlation estimate were due to repeated exposure to the same scatterplots (i.e., having seen the same dataset in a scatterplot twice, one with meaningful labels and one with abstracted labels). Our results rule out this possibility, strengthening our argument that belief is the key factor driving estimation bias.

Contribution: These findings shed light on the difficulty that data communicators face when persuading people with data: as definitive and objective as the data is believed to be, our interpretation is often biased by our expectations and beliefs. These findings also have implications for how researchers and data analysts should approach the data analysis process. When approaching an analysis with existing hypotheses, an analyst may put themselves at risk of seeing what they want to see in their data. This relatively low-level bias may become compounded by higher-level ones, such as confirmation bias, which might lead to overly confirmatory follow-up studies or analyses ([60]).

2 RELATED WORK

Although the ideal visual analytic workflow is bias-free, in reality, people can fall victim to a variety of cognitive and perceptual biases when making sense of data [14, 75, 77, 84, 85]. Visualization readers can latch on to salient features when interpreting visualizations and gravitate toward unique colors [1], larger fonts [30, 83], and patterns that are aligned with their beliefs and agendas [7, 87], giving in to letting their intuition drive decision making [35, 62].

2.1 Motivated Perception

Predictive models of perception suggest that what we see is determined by the visual input and prior knowledge [20] and that existing knowledge and beliefs can bias judgments even when there is less ambiguity. For example, in a study where people were asked to report the size of real-world objects in an otherwise dark field of view, their past knowledge of the typical size of those objects affects their perceived size and inferred distance from those objects [22]. Our knowledge of real-world object size can similarly affect the perceived speed of that object in motion [57]. In another example, afterimages that conform to expectations are perceived as more vivid, while those that contradict expectations are less vivid [52].

More recently, data visualization researchers have begun exploring the extent to which existing beliefs or background knowledge can impact how people view data. A visualization contains many patterns to perceive [71, 86, 87], and viewers can perceive the patterns that are more aligned with their existing knowledge to be more salient. For example, when participants were primed with backstories before seeing a visualization, the features mentioned in the backstories were perceived as more visually salient [87]. People can also be biased by their beliefs when looking at visualizations [50, 51]. Those who do not believe in climate change tend to look more at the flat sections, confirming their thoughts on climate change.

While these effects corroborate findings in basic human perception, it is not clear whether objective, low-level visual properties can also be biased by our beliefs. Low-level visual properties, such as brightness, follow Weber's law, where the just-noticeable difference is a fixed proportion of its magnitude [11]. Among these properties, correlation perception is a key property relevant to visual analytics [2, 67]. Thus, as an initial step, we empirically examined how prior beliefs may influence correlation estimations.

2.2 Beliefs and Updating

When we interpret visualized data, our internal representations of relevant knowledge play a critical role [29, 47, 49, 62, 73]. When people encounter new information, they may struggle to properly incorporate past knowledge and existing beliefs to make judgments with regard to the new information, leading to biases in data interpretation [13]. For example, failed belief updating can drive people to be overconfident in their judgment [18, 74], as can confirmation bias [35, 74]. People who are subject to confirmation bias often attend and search only for information that supports prior beliefs and do not consider alternative explanations [46, 61, 81]. They may also over-weigh information congruent with their beliefs while under-weighing information incongruent with their beliefs [45].

Visualization researchers have identified visualization techniques to facilitate adequate belief-updating [54]. For example, effectively showing uncertainty in data, such as using hypothetical outcome plots [33, 37], gradient plots [12], and quantile dot plots [9, 36, 40], has been found to help people properly update their beliefs [42]. However, most visualization practitioners hesitate to incorporate uncertainty information in their designs due to concerns about their ability to precisely represent the uncertainty in a way readers can understand [32]. This motivates us to investigate and model the effect of beliefs with more commonly used scatterplots.

2.3 Scatterplots and Correlation Perception

Current taxonomies in visualizations systematically abstract visual analytic tasks into granular operations (e.g., [8, 23]). To our knowledge, the lowest levels to date include: Retrieve Value, Filter, Compute Derived Value, Find Extremum, Sort, Determine Range, Characterize Distribution, Find Anomalies, Cluster, Correlate [2].

Finding correlation is a commonly used low-level task in visual analytics [48] that has been well-studied in the visualization community [27, 39, 68]. Scatterplots are shown to especially facilitate correlation extraction in visualization [69]. Human participants are sensitive to correlation values and can generally read correlations from scatterplots accurately with little training [68].

Correlation perception is, of course, imperfect and is likely perceived using cues other than correlation per se [67]. Recent work argues such perceptual shortcuts might include using the aspect ratio of the probability density function generated by the dots as a proxy for correlation (usually in the form of an ellipse that describes a constant probability of encountering a dot) [27, 39, 88].

Changes to the design of the visualization can also impact how people extract key statistics [21]. For example, the shape and orientation of the visual encoding marks can impact correlation perception in scatterplots, such that participants are more accurate if the direction of the shapes aligns with the regression line [48]. The size and lightness of the scattered dots can distort visual summaries of data and lead to misjudgments in their weighted average [31]. These findings suggest

that although accurate, there is room for bias to nudge correlation estimations when people read scatterplots. Together with the prevalence of scatterplots and correlation tasks in visual analytics, they make the correlation estimating in scatterplots an ideal test bed to model the potentially biasing effect of belief.

3 GENERAL METHOD

The stimulus, data, and analysis script are available at the Open Science Framework website: `https://osf.io/3cnza/`.

3.1 Stimuli

Stimuli for these experiments consisted of four scatterplots, with values randomly generated to fit a given correlation. In order to provide participants with ample space to over- or under-estimate the correlation, we chose intermediate correlation values of 0.6 and 0.4. Among the four, two scatterplots had a correlation of 0.4, and the other two had a correlation of 0.6. The axes titles on each scatterplot varied within experiments, but the datasets were identical across control and experimental stimuli (see Figure 3). The survey was hosted by Qualtrics [65], and the survey flow for each experiment can be seen in Figure 2. Neither exact stimuli nor the correlation values were shown in the training portion of the surveys.

3.2 Participants

Among all experiments, 748 adults were recruited from an online crowdsourcing site, Amazon's Mechanical Turk (MTurk) [28]. Of these 748, 435 were men, 306 were women, 5 were non-binary, and 2 chose not to disclose their gender. The average age of all participants was 40 years old (SD = 10.62 years). To be included in the analysis, responses had to have the correct answer to all attention check questions and participants must not have seen any of the questions besides demographics in any previous survey. Specific attention check questions are described in Section 3.3.

This led to 27 exclusions in Experiment 1a, 200 exclusions in Experiment 1b, and 32 exclusions in Experiment 2. We recruited 'Master' workers for Experiment 1a and Experiment 2, but due to difficulty collecting responses with 'Master' respondents for Experiment 1b (responses came in at an extremely slow pace), we removed this requirement for Experiment 1b. An additional attention check was used for Experiment 1b to ensure that participants' response and attention quality remained comparable to 'Master' workers. For this reason, the number of exclusions in Experiment 1b was notably high.

Participants completed a Qualtrics survey and were paid between $1.35 and $2.00, depending on the number of sections in the survey, at a rate of $9.47 an hour. This research adhered to ethical guidelines including the legal requirements of the United States as well as the Institutional Review Board.

3.3 Procedure

As shown in Figure 2, in all experiments, participants completed a similarly structured survey.

The first attention check question occurred prior to the training block. The page read "Thank you for participating in our survey. We will ask you to answer short questions about **graphs.** Please read **all** questions and answers carefully before responding." Participants then moved to the next page and were asked to fill in the sentence "We will ask you to answer short questions about ___."

Following this check, all participants were trained to extract correlation from scatterplots through an explanation of correlation, followed by seven examples, ranging from -0.9 to 0.9, labeled with the correct correlation value. They then completed three practice estimations during which they received feedback about any error in estimation.

The second attention check came just before viewing the first experimental stimuli. Participants were informed on the previous page, "Please estimate the correlation of the following graphs. This is different from the previous section, as you will **not** receive feedback upon selecting your answer." They were then asked, "Will you receive feedback after selecting your answer?" and given options of 'yes' and 'no'.
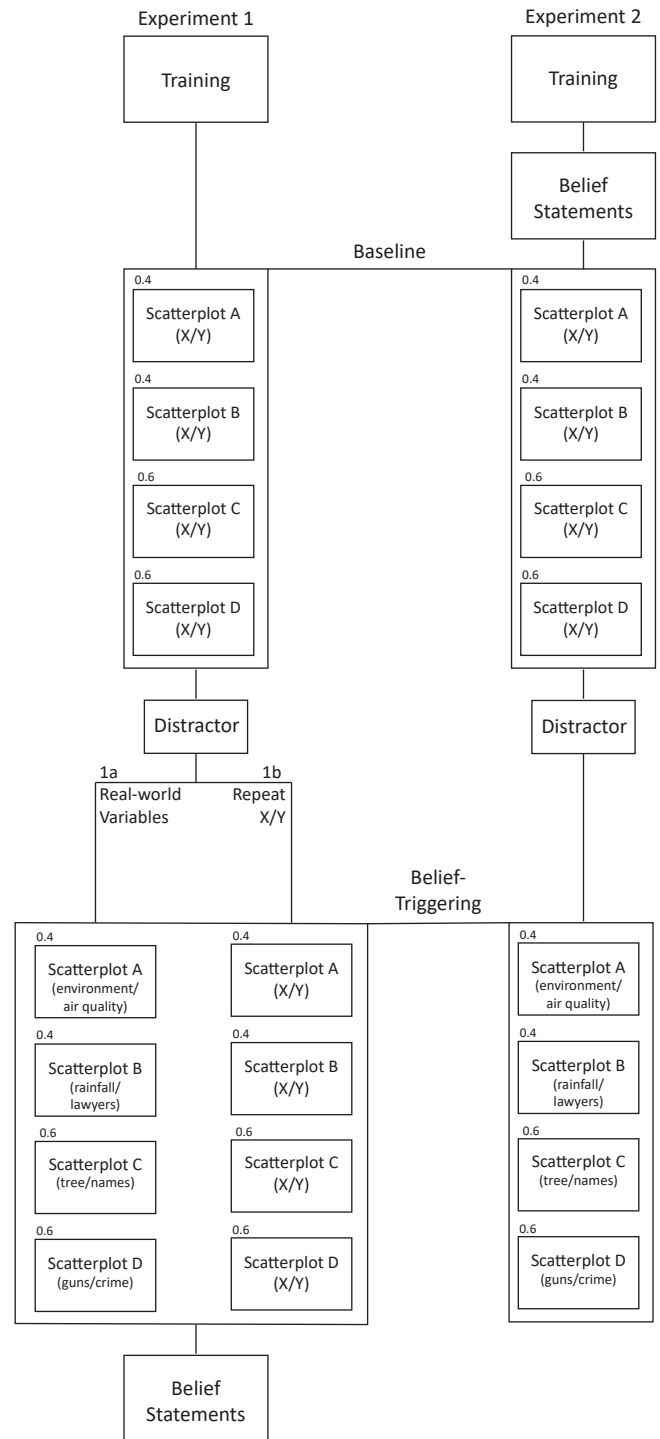

Fig. 2. Flowchart depicting the procedure for each experiment.

After this attention check, participants estimated the correlation of four scatterplots, all labeled with 'X' and 'Y', using a sliding scale from 0 to 1. Two of the scatterplots contained different datasets, both depicting a correlation of 0.4, and the other two depicted a correlation of 0.6. The same four scatterplots were viewed by all participants in a randomized order of display.

After completing the correlation estimations on the first set of four scatterplots, participants completed a distractor task. In Experiment 1a and Experiment 2, participants then went on to complete estimations of the same four scatterplots as the previous estimations, but with real-
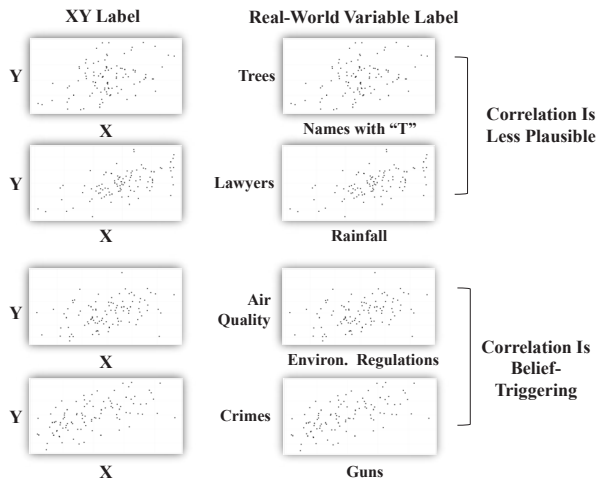
Fig. 3. Scatterplots used in our experiments, either labeled with letters 'X' and 'Y' or with real-world variable names, depicting either a 0.4 or a 0.6 correlation. Note the two scatterplots in the same row contains identical data. The aspect ratio of the scatterplots was approximately 2:1 (1399x712 pixels) and the size of the scatterplots was adjusted to 50% of the width of the participant's screen during the survey. Axes values were not included, as to eliminate the possibility of calculation (or estimated calculation) of correlation.

world variables instead of 'X' and 'Y' on the axes, as shown in Figure 3. The participants always viewed the 'X' and 'Y' labeled scatterplots first, to ensure that they were not primed by real-world variables. This allowed us to treat their estimation in the 'X' and 'Y' labeled scatterplots as an unbiased baseline.

Two of these pairs were topics that were more likely to trigger beliefs (gun ownership and violent crimes; environmental regulations and air quality), and two were neutral topics not usually associated with strong beliefs (the number of trees and people whose names start with 'T;' the number of lawyers and annual rainfall). Including both belief-triggering and neutral topics provided us the ability to examine the effect of belief not only on topics that are belief-triggering but also on those that seem less plausible.

We also counterbalanced the correlation values with the real-world variable labels through two conditions to ensure the chart topic and its correlation values were not confounded. In condition 1, the plot depicting gun ownership and the plot depicting the number of trees had correlations of 0.6. The other two topics were depicted with correlations of 0.4. In condition 2, the correlation values were swapped between the topic pairs.

In Experiment 1a and Experiment 2, participants also rated their level of agreement with statements representing a belief in relationships between the variables depicted in the scatterplots, noted as "Belief Statements" in Figure 2. For example, participants rated the statement, "I believe that as the number of households with guns increases, the number of violent crimes increases," on a 7-point Likert scale. A rating of 1 indicated that the participant strongly disagreed with the statement, while a rating of 7 indicated that the participant strongly agreed with the statement. Similar Likert scales have been used to measure attitudes and beliefs [25, 34, 82]. We were specifically interested in their belief in a correlation between the two variables shown and thus used the same phrasing as the scatterplot axes labels. The plots all presented positive correlations to maximize the comparability of results across examples.

Participants in Experiment 1a made these ratings after completing a second distractor task following the correlation ratings of all scatterplots, not shown in Figure 2, to further the distance between seeing a visual depiction of the scatterplots and reporting their belief on the topic. All participants in Experiment 2 made these ratings in the beginning, after completing the correlation training section but before estimating the correlations for any scatterplots. A subset of partici-

pants in Experiment 2 also completed the same ratings at the end of the survey, as a measure of personal belief for each participant.

In Experiment 1b, participants completed the distractor task after rating the correlation of the first four scatterplots labeled with 'X' and 'Y'. After the distractor task, instead of seeing the same scatterplots with real-world labels, they saw them with a different set of abstracted letter labels, such as 'M' and 'N'. Participants in this experiment did not complete any ratings of belief. There was only one condition of this experiment, with each set of scatterplots shown in a randomized order.

Overall, Experiment 1a and Experiment 2 examined the effect of belief on correlation estimations. Experiment 1b served as a control condition for Experiment 1a, to account for the possibility that differences in correlation ratings between the two sets of scatterplots were a result of time delay.

4 EXPERIMENT 1A MEASURING BELIEF LAST (POSTERIOR)

In this experiment, as shown in Figure 2, participants completed training on estimating correlation, and then estimated the correlation for four scatterplots with 'X' and 'Y' labels on the axes (correlations of 0.4 and 0.6), as shown in Figure 3. Next, they completed a distractor task before rating the same four scatterplots, this time with real-world variables on the axes. After a second distractor task, participants rated their level of agreement with statements depicting belief in the relationship between variables in the scatterplots. In the end, participants reported demographic information such as their age and level of education.

4.1 Participants

Assuming a medium effect size (a $\rho = 0.21$ correlation based on pilot studies) for the effects of belief, power analysis with G*Power [16] suggests that a target sample of 291 participants would give us 95% power to detect an overall effect of the belief factor at an alpha level of 0.05. After applying the exclusion criteria specified previously, we ended up with 295 participants (172 men, 120 women, and three non-binary people) with an average age of 41 years (SD = 9.66 years).

4.2 Ability to Estimate Correlations

We first examine how accurately participants estimated the correlation values in the scatterplots without considering the effects of belief. Overall, participants were quite accurate at estimating correlations in scatterplots, corroborating existing findings.

We compared participant estimations of the XY labeled scatterplot correlations with the ground truth correlations (which is either 0.4 or 0.6). Participants' correlation estimates for the two versions of the 0.4 and 0.6 correlation scatterplots slightly differ from ground truth, but generally, they were quite accurate. They reported an average correlation of 0.415 (SD = 0.19, CI = [0.394, 0.437]) and 0.401 (SD = 0.21, CI = [0.377, 0.425]) for the two versions of scatterplots with a 0.4 correlation. For scatterplots with a 0.6 correlation, participants reported an average correlation of 0.558 (SD = 0.19, CI = [0.537, 0.579]) and 0.625 (SD = 0.18, CI = [0.604, 0.646]).

We compared the correlation estimates of the same scatterplots with and without real-world labels to identify potential bias resulting from labeling the axes with real-world variable pairs. Participants reported significantly different correlation values for the same scatterplot in the presence of real-world variable pair labels. On average, they estimated the correlation for the scatterplots with a 0.4 correlation with real-world variable labels to be 0.385 (SD = 0.19, CI = [0.363, 0.407]) and 0.287 (SD = 0.22, CI = [0.262, 0.312]). They estimated the correlation for the scatterplots with a 0.6 correlation with real-world labels to be 0.561 (SD = 0.19, CI = [0.539, 0.583]) and 0.546 (SD = 0.23, CI = [0.520, 0.572]).

4.3 Self-Reported Beliefs

To examine the effect of participant belief on their correlation estimates, we showed them two plots with belief-triggering variable pairs (gun ownership and violent crimes; environmental regulations and air quality), and two with neutral variable pairs not usually associated with
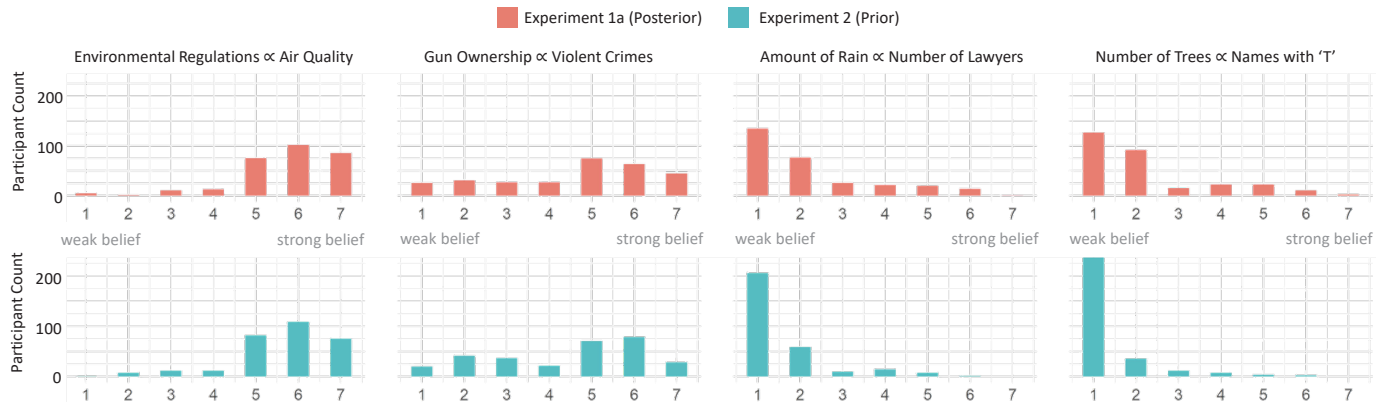
Fig. 4. Distribution for the level of participant agreement with statements supporting a positive correlation between the variables depicted in the scatterplots. This was measured at the end of Experiment 1 (Top) and at the beginning of Experiment 2 (Bottom). The gun/crime and environment/air topics have right-skewed distributions such that more people strongly agreed with the positive correlation statement, while the trees/names and lawyers/rainfall topics have left-skewed distributions such that more people strongly disagreed with the positive correlation statement.

with strong beliefs (the number of trees and people whose names start with "T;" the number of lawyers and annual rainfall).

This analysis verifies that participants more strongly believed that there exists a positive correlation between gun ownership and violent crimes (M = 5.72, SE = 0.07) and between environmental regulations and air quality (M = 4.57, SE = 0.11). They less strongly believed that there exists a positive correlation between the number of trees and people whose names start with "T" (M = 2.19, SE = 0.088) and between the number of lawyers and annual rainfall (M = 2.21, SE = 0.089). Post-hoc pair-wise comparisons between their belief on the four topics suggest that participants most strongly believed in the gun relationship, followed by the environment relationship, followed by the trees/names, and the lawyers/rainfall relationship (their beliefs of which are not statistically different, p = 0.998). Figure 4 shows the belief distributions for the four topics.

4.4 Effect of Belief on Correlation Estimates

To investigate whether labeling the scatterplot with real-world variables biases participants' correlation estimations, we compare the differences between their estimation for the same scatterplot labeled with 'X' and 'Y' and with real-world variables. We demonstrate this difference to be correlated with how strongly they believe the real-world variables are positively correlated.

We constructed a mixed-effect linear model predicting the difference between estimates in the scatterplots labeled with 'X' and 'Y' and with real-world variables using the lmer function in R [6]. We used experiment versions (counterbalanced conditions), topics (the two belief-triggering and two neutral topics), the strength of belief (the degree of agreement with the correlation statements), and the interaction between topics and the strength of belief as predictors, and included a random effect of participants. We also added demographic variables including age, gender, political orientation, level of education, experience with statistics, and experience with doing research as fixed effects to explore potential differences.

We found a significant main effect of belief (χ^2 = 18.39, Est = 0.024, SE = 0.011, t = 2.23, p = 0.023), such that the stronger the participant believes in the variables' correlation, the more they overestimated the correlation when the scatterplot was labeled with that variable pair, as shown in Figure 5. Similarly, the more they disagreed with the variable correlation, the more they underestimated the correlation.

There was no significant main effect of condition (χ^2 = 0.022, p = 0.88), suggesting that the two counterbalanced conditions were not different from one another. There was also no significant main effect of topic (χ^2 = 4.28, p = 0.23), nor a significant interaction between topic and belief (χ^2 = 1.84, p = 0.61). This suggests that regardless of whether the topic was belief-triggering or neutral, the under/overestimation of the correlation was driven by the participants' own beliefs. Figure 6 shows a comparison of the relationship of correlation estimation error and belief strength, between belief-triggering topics and less plausible topics. The two look extremely similar.

We also found no significant effect of any demographic variables. Details can be found in the supplementary materials.

4.5 Ordering Effects and Alternative Explanations

Despite accurately estimating scatterplot correlations when the plots were labeled with 'X' and 'Y', participants became less accurate when estimating the correlation for the same scatterplots but with real-world variable labels. Our model suggests that the deviation in correlation estimation is associated with the strength of participant beliefs, such that strong beliefs predict overestimation and strong disbeliefs predict underestimation.

However, one possible alternative explanation for the over/underestimation could be the factor of time. Participants always saw the real-world labeled scatterplots after they had seen the 'X' and 'Y' labeled scatterplots, at which point they could have become habituated or anchored by previously seen scatterplots. Further, more time has lapsed between training and estimating correlations for the real-world variable labeled scatterplots, compared to the time between training and estimating correlation for the 'X' and 'Y' labeled scatterplots. So it is possible that the effect of training has deteriorated, and thus participants provided less accurate correlation estimations. This motivated us to conduct Experiment 1b, a control experiment where, instead of showing participants the 'X' and 'Y' labeled scatterplots followed by the same plots labeled with real-world variables, we showed them the same 'X' and 'Y' labeled scatterplots twice to account for the potential effect of time.
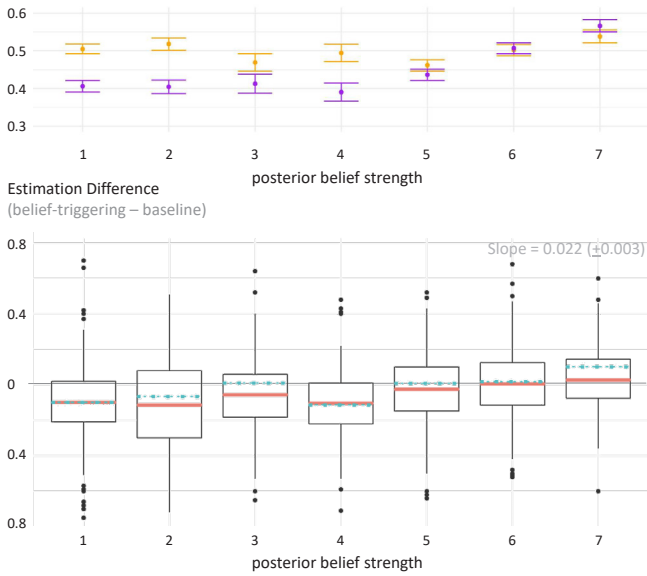
5 EXPERIMENT 1B CONTROLLING FOR THE EFFECT OF TIME

We observed a significant difference in the correlation values participants reported for the scatterplots labeled with 'X' and 'Y', compared to the same scatterplots labeled with real-world variables. We investigate whether the difference is driven by the real-world labels, or simply by the passage of time in this experiment.

5.1 Method

Participants completed training on rating correlation, then rated four scatterplots with alphabetic labels (e.g., X, Y), with correlation values of either 0.4 or 0.6. They then completed a distractor task before rating the same four scatterplots, also with alphabetic labels on each axis. The alphabetic labels are used to strip context away from the scatterplots to ensure that belief does not play a role in our results. Participants then reported demographic information.
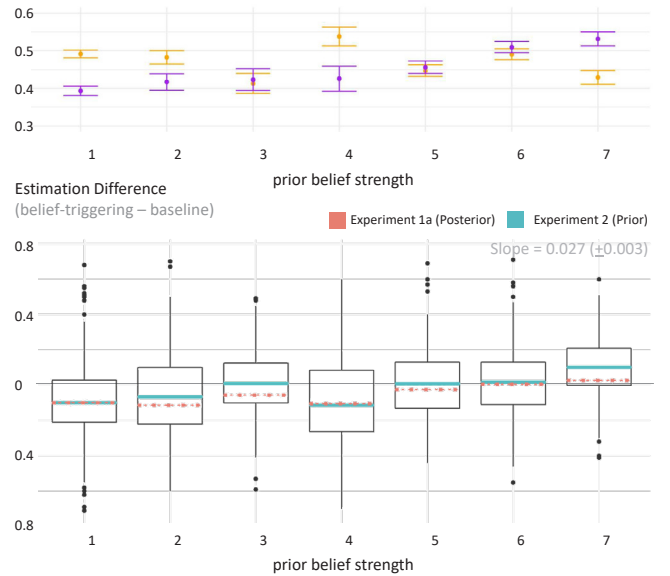
Fig. 5. Top: Shows the raw estimations for scatterplots labeled with baseline (XY) and belief-triggering real-world variables. Bottom: Shows how the difference in correlation estimations between baseline (XY) and belief-triggering plots varies with the strength of belief. Less than 0 means that participants underestimated the correlation. The colored horizontal lines represent the average values per belief strength. Both posterior and prior averages are shown in both charts to facilitate comparisons between them. Left: Experiment 1 where the strength of belief is elicited at the end of the experiment. Right: Experiment 2 where the strength of belief is elicited towards the beginning of the experiment. For both experiments, participants with stronger beliefs more likely overestimated the correlation for the real-world variable labeled scatterplots.
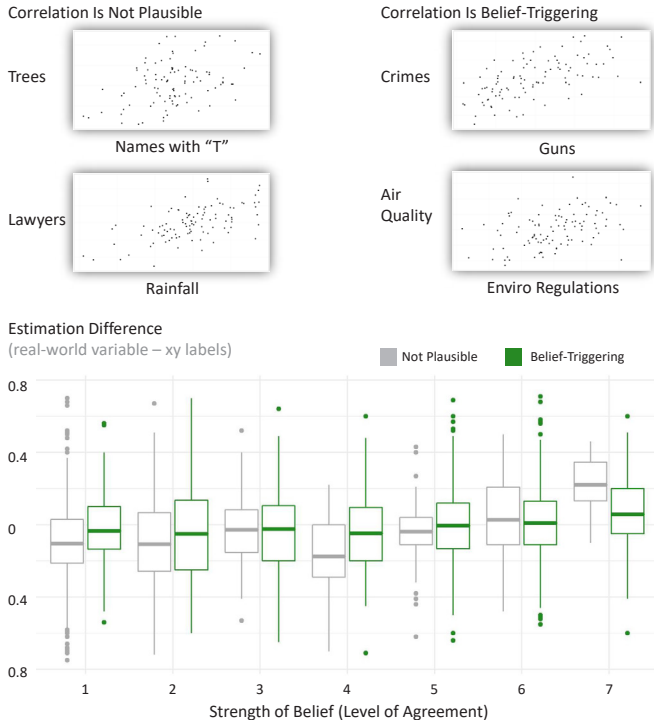

Fig. 6. This belief-driven over and underestimation of correlations happens when people see both not plausible real-world variable pairs (e.g., lawyers and rainfall) and belief-triggering variable pairs (e.g., guns and crimes). Data is consolidated from both Experiments 1a and 2.

5.2 Participants

Because Experiment 1b is intended as a control version of Experiment 1a with no manipulation before and after the distractor task, we recruited participants to match the number of participants in one condition of Experiment 1a, which is equivalent to half the number of participants from Experiment 1a. After applying the exclusion criteria specified previously, we ended up with 153 participants (86 men, 64 women, 1 non-binary person, and 2 who chose not to disclose), with an average age of 38 years (SD = 11.49 years).

5.3 Results

Participants estimated the correlation of scatterplot with alphabetic labels twice (e.g., 'X' and 'Y') in Experiment 1b, which we will refer to as Phase 1 and Phase 2 estimates. In Phase 1, the average correlation they reported for the two versions of the 0.4 correlation scatterplots were 0.423 (CI = [0.395, 0.451]) and 0.354 (CI = [0.395, 0.451]). The average correlation they reported for the two versions of the 0.6 correlation scatterplots were 0.541 (CI = [0.509, 0.572]) and 0.599 (CI = [0.570, 0.628]). In Phase 2 the average correlation they reported for the two versions of the 0.4 correlation scatterplots are 0.442 (CI = [0.413, 0.472]) and 0.372 (CI = [0.339, 0.406]). The average correlation they reported for the two versions of the 0.6 correlation scatterplots are 0.513 (CI = [0.482, 0.545]) and 0.629 (CI = [0.600, 0.656]).

We compared participants' correlation estimates for the real-world variable labeled scatterplots in Experiment 1a to the Phase 2 estimations in Experiment 1b (labeled with alphabetical letters). Between these two sets of estimates, the only difference is how the scatterplots were labeled. If the two sets are similar, that suggests that timing is likely the factor driving any deviation from their estimation of the first set of scatterplots labeled with 'X' and 'Y' (phase 1). If the two are different, it likely suggests that the real-world variable labels, which trigger either belief or disbelief in a positive correlation, affected correlation estimations, rather than it being an effect of time.

We found there to be a significant difference between them. Pairwise comparisons with Bonferroni corrections suggest that the correlation estimates for the two versions of 0.4 correlation plots were significantly different ($p_{v1} = 0.002$, $p_{v2} < 0.001$), and that of the two versions of the 0.6 correlation plots were also significantly different ($p_{v1} = 0.016$, $p_{v2} < 0.001$). We can infer from these comparisons that the deviations in correlation estimates with the scatterplots labeled with

real-world variables in Experiment 1a are not caused by time delays, but rather by the real-world labels and their associated beliefs.

In both Experiment 1a, participants viewed the real-world variable labeled scatterplots before reporting their belief (level of agreement with a positive correlation between the depicted variables). It is possible that these beliefs that people reported are posterior in nature such that they were not their true beliefs, but rather updated beliefs affected by the scatterplots they saw. To demonstrate that participants' prior beliefs have an influence on how they estimate correlations in scatterplots, we conduct Experiment 2 where we elicited their level of agreement on the depicted topics prior to them seeing the real-world variable labeled correlations.

6 EXPERIMENT 2 MEASURING BELIEFS FIRST (PRIOR)

In Experiment 1a, we measured participants' beliefs on relevant real-world variable pairs at the *end* of the experiment, after they had seen scatterplots depicting potential relationships between those variables. To account for the potential effect of these visualizations on participants' beliefs, we conducted Experiment 2 and measured their beliefs on these real-world variable pairs at *the very beginning* of the experiment.

6.1 Method

Similar to both previous experiments, participants first completed training on correlation estimation. Immediately following training, they rated their level of agreement with belief statements regarding the variables previously viewed in the scatterplots. They then rated the correlation of four scatterplots with 'X' and 'Y' labels on the axes (correlations of 0.4 and 0.6). Next, they completed a distractor task before rating the same four scatterplots, this time with real-world variables on the axes. Finally, participants reported demographic information such as their age and level of education.

6.2 Participants

Assuming a medium effect size (a $\rho = 0.21$ correlation based on pilot studies) for the effects of belief, power analysis with G*Power [16] suggested a target sample of 291 participants for 95% power to detect an overall effect of the belief factor at an alpha level of 0.05. After applying the exclusion criteria specified previously, we ended up with 300 participants, with an average age of 40 years (SD = 11.01 years).

6.3 Ability to Estimate Correlations

To gauge how accurate participants were in estimating of 'X' and 'Y' labeled scatterplot correlation, we compared their response with the ground truth correlations. As in Experiment 1a, the correct answers were either 0.4 or 0.6. We observed that participants overall performed well on estimating the correlation. On average, participants reported 0.381 (SD = 0.18, CI = [0.36, 0.40]) for the first 0.4 stimuli, and 0.38 (SD = 0.21, CI = [0.35, 0.40]) for the second. For the stimuli depicting 0.6, participants reported 0.530 (SD = 0.19, CI=[0.51, 0.55]) and 0.62 (SD = 0.18, CI = [0.60, 0.64]), respectively. This was similar to the results of Experiment 1a.

Next, we compared the participants' estimations from when they saw the plots with and without real-world labels. We found that the responses were significantly different. Participants estimated the correlation for the scatterplots with a 0.4 correlation with real-world variable labels to be 0.407 (SD = 0.21, CI = [0.38, 0.43]) and 0.255 (SD = 0.22, CI = [0.23, 0.28]). They reported the correlation for the scatterplots with a 0.6 correlation with real-world labels to be 0.546 (SD = 0.20, CI = [0.52, 0.57]) and 0.546 (SD = 0.25, CI = [0.52, 0.57]).

6.4 Self-Reported Beliefs

Similar to Experiment 1a, we analyzed the influence of the beliefs on estimating correlation. We found that participants more strongly believed in a positive correlation between environmental regulations and air quality (M = 5.66, SE = 1.20) and between gun ownership and violent crimes (M = 4.45, SE = 1.80). They less strongly believed a positive correlation between the number of lawyers and annual rainfall (M = 1.55, SE = 1.03), and between the number of trees and people

whose names start with "T" (M = 1.38, SE = 0.92). We ran post-hoc pair-wise comparisons with corrections between their beliefs on the four topics, and the result shows that belief ratings in the four conditions were reliably different (p < 0.0001). Figure 4 shows the belief distributions.

6.5 Effect of Belief on Correlation Estimates

To understand the influence of the beliefs on the correlation estimation, we ran a mixed effect linear model with the difference between their estimation on the scatterplots with and without the real-world label as a dependent variable, beliefs and the topic of the chart and their interaction as fixed effects, and participants as a random effect. We also added demographic variables including age, gender, political orientation, level of education, experience with statistics, and experience with doing research as fixed effects to explore potential differences. We found that the beliefs had a reliable impact on the difference (Est = 0.025, SE = 0.011, t = 2.28, p = 0.023), as shown in Figure 5. As the average participant's beliefs were 1 unit stronger in a positive relationship of the two variables, they saw 0.02 units of additional correlation compared to when they examined the chart without the real-world labels. There was no reliable effect of the topic nor the interaction. Additionally, consistent with the results from Experiment 1a, there is no significant effect of topic (p = 0.085), condition (p = 0.51), nor the interactions between topic and belief (p = 0.21), as shown in Figure 6. We also found no significant effect of any demographic variables. Details can be found in the supplementary materials.

7 COMPARING EXPERIMENT 1A AND EXPERIMENT 2

We compared our results from Experiment 1a and 2 to examine whether the effect of labeling scatterplots on correlation estimates replicates, and whether asking the belief questions before or after the participants see the scatterplots made a difference.

Ability to Estimate Correlations: We did not observe a difference between participants' ability to estimate the correlation of scatterplots with 'X' and 'Y' labels. Since this step was shown before seeing the scatterplots with the real-world labels, we did not expect the result to differ.

Belief Distributions We compared the participants' beliefs when they were prompted before and after seeing the scatterplots with the real-world labels. We found no difference of their beliefs on environmental regulations and air quality (t = −0.654, p = 0.512) nor gun ownership and violent crimes (t = −0.862, p = 0.389). However, there was a difference between belief ratings for the number of trees and people whose names start with "T" (t = −8.051, p < 0.001) and for the number of lawyers and annual rainfall (t = −5.997, p < 0.001).

Figure 4 shows the belief distributions of participants in Experiments 1a and 2. For the belief triggering topics, there were no differences between people in Experiment 1a and 2 (Environment prior: M = 5.66, SD = 1.201, posterior: M = 5.722, SD = 1.233, Gun prior: M = 4.450, SD = 1.800, posterior M = 4.579, SD = 1.864). However, for the non-belief triggering topics, we observed a difference in people's belief distributions (Rain prior: M = 1.553, SD = 1.035, posterior: M = 2.190, SD = 1.507, Tree prior: M = 1.383, SD = 0.916, posterior: M = 2.214, SD = 1.520).

7.1 Seeing What You Believe

Figure 5 shows the relationship between the belief strength and the correlation estimation difference. We compared the effect of prior and posterior beliefs on participants' correlation estimations when they viewed scatterplots with 'X' and 'Y' labels versus real-world variable pairs. The results are very similar between Experiments 1a and 2. The coefficient of the prior belief factor is 0.022, and the posterior belief factor is 0.023.

7.2 Believing What You See

In the case of non-belief triggering topics, we observed a slight difference in beliefs distribution when participants were asked to indicate their beliefs before and after seeing the scatterplots. This may imply

When the viewers don't believe that the variables are correlated   **Ground truth Data**   **When the viewers strongly believe that the variables are correlated**

Fig. 7. A visualization of the size of the bias based on effect sizes from the experiment. Middle: what the 'objective' data should look like (with Pearson's correlation = 0.5). Left: How the data would look like to a viewer who doesn't believe that the variables are correlated ($r = 0.4$). Right: How the data would look like to a viewer who strongly believes that the variables are correlated ($r = 0.6$).

that seeing the correlated scatterplots led them to believe that there might be a correlation between two variables.

8 GENERAL DISCUSSION

People were biased by their existing beliefs when estimating correlations in scatterplots, and we were able to quantify how much those beliefs can influence the estimates. As shown in Figure 7, people tend to underestimate the correlation of a scatterplot, by approximately 0.1 in Pearson's correlation value, when it depicts a relationship between two variables that they don't believe to be correlated. They tend to overestimate the correlation by approximately 0.1 when it depicts a relationship that they do believe to be correlated. This may explain the difficulty frequently faced when persuading people with data - as objective as the data is believed to be, our interpretations of it are often riddled with bias from our expectations and beliefs.

However, it remains unclear why participants gave differing correlation estimates for the same scatterplot when it was labeled with 'X' and 'Y' compared to when it was labeled with real-world variables. One explanation might be that perceived correlation strength could be impacted by 'top-down' effects such that they paid more attention to scattered dots that would make the correlation look higher or lower (in accordance with their belief), discounting outliers that would strengthen or weaken the correlation against their beliefs [3]. This possibility would be consistent with previous work showing that darker colored dots are weighted more when viewers estimate means in scatterplots [31], and from Luo & Zhao, where viewer decisions and actions can be manipulated by making certain patterns in a visualization more visually salient [50]. In this case, participants may be weighing dots that support their belief more heavily, such as weighing the dots near the regression line more heavily when they more firmly believe in the correlation.

Another possibility is that correlation estimates come with uncertainty intervals around a best guess value. When the scatterplot was labeled with 'X' and 'Y', viewers may be more likely to report the midpoint value in that uncertainty range. But for belief-triggering relationships, they might report a correlation either in the upper or lower range depending on the direction of their belief. Existing work in uncertainty demonstrates that people are more biased when the data shows more uncertainty, supporting this possibility [13, 42]. We discuss potential future research opportunities in Section 10.

9 DESIGN IMPLICATIONS

When an analyst or scientist has a belief about how the world works, their thinking can be biased in favor of that belief. Therefore, one bedrock principle of science is to minimize that bias by testing the predictions of one's belief against objective data. However, we have demonstrated that when people look at scatterplots, their beliefs can interact with the objective data to bias judgments of the objective strength of the relationship between two variables. The stronger the belief (and disbelief), the stronger the bias. Based on this result, when

reading visualizations, if readers can first form an impression of the strength of the relationship based on only the data values (without letting the labels alone drive that impression), they might become less susceptible to belief-driven biases when interpreting the visualization. Similarly, data analysts should thoroughly consider their data filtering and manipulation decisions and proceed with caution once they have already seen the data. If the data is aligned with their beliefs, except for a couple of outliers, as shown in Figure 1, the analysts might be motivated to adjust the exclusion criteria to exclude them or to collect additional data to 'drown the outliers out', which likely will skew the results and introduce biases in their analysis.

10 LIMITATION AND FUTURE DIRECTIONS

We discuss several limitations in our study that provide promising future research directions.

Cognition versus Perception: We did not try to tease apart effects on perception and cognition. Making sense of data, even through a visualization, is a complex process that involves both mechanisms [71]. We demonstrated that belief impacts how data gets interpreted by a viewer, and quantified the extent to which it does.

There is debate in the psychology research community on the extent to which 'top-down' effects (e.g., belief or motivation) affect perception [17, 64]. Some researchers have argued that cognition has the ability to 'penetrate' perception [10, 53, 58], such that existing beliefs or personal goals affect how we perceive the natural world. For example, Balcetis & Dunning demonstrated that wishful thinking could make desirable objects appear closer [5]. For example, a water bottle appears closer to a thirsty person. As mentioned in Section 2, familiarity and prior knowledge can affect size and speed perception [22, 57].

However, some researchers have argued against such an interpretation of these effects, suggesting that although our knowledge and beliefs can affect what we look at or what we pay attention to, the lower-level stages of visual perception are impenetrable to existing beliefs [63, 64]. Recent support for this idea includes a study that demonstrates the influence of object familiarity on perceived distance found in [57] could not be conceptually replicated [59]. Two manipulations differed between the studies: 1) instead of using an adjustment task with unlimited viewing duration, Mischenko et al. used brief presentations to minimize the potential effect of deeper thinking during task completion, and 2) instead of showing participants objects that varied significantly in size (e.g., tennis balls and golf balls), they showed participants different types of coins with similar sizes [59]. Firestone and Scholl, in [17], also note that some experimental designs in traditional 'motivated perception' type studies make the studies less convincing in support of 'top-down' effects. For example, some of these studies capture what participants recognize rather than examining what they actually see, which makes the study more about semantic priming than perception. They also point out the risk of experimenter demands (participants reporting what they believe the experimenter expects them to report) and other forms of response bias. Additionally, researchers do

not always show people the *same* visual stimuli to compare their perception with and without the motivation or additional beliefs. The low-level visual differences between the experimental stimuli can weaken the results [17].

To account for some of the concerns listed by Firestone and Scholl [17], in the current experiment, we tried to detect potential experimenter demands by explicitly asking participants what they thought the study was about and conveying that their answers will not affect their compensation. We also used the same stimuli and compared participants' correlation estimation with or without the belief component. Our participants viewed scatterplots labeled with 'X' and 'Y', and later viewed the same scatterplots again but labeled with real-world variable pairs. They are the same stimuli, so there are no low-level differences, and the differences in correlation estimation we observed should be driven by their attitudes towards the labels alone.

Future research could find stronger evidence for a perceptual root of the present bias by seeking evidence of biases on Just Noticeable Differences (JND) between pairs of correlations, which might be less potentially influenced by cognitive factors compared to the individual ratings used here [68]. For example, if the bias effect is driven by participants reporting estimations on either side in the range of possible correlations due to uncertainty, such that they opt to report correlations in the upper range when they think the variables should be more strongly correlated and vice versa, then experiments that leverage JND comparisons should be far less influenced by this form of bias that does not require the production of correlation estimates.

Mitigating Biases in Data Interpretation: The current results cannot specify *why* participants gave differing correlation estimates for the same scatterplot when it was labeled with 'X' and 'Y' compared to when it was labeled with real-world variables. We discussed some potential explanations in Section 8, and future work can further test the mechanisms driving participants to misestimate the correlations depending on their beliefs. Understanding these mechanisms can help us design tools to potentially mitigate these biases in correlation estimation. For example, a visual analysis tool can monitor patterns of exploration or request objective decision criteria during analysis to mitigate biases [79]. If the effect is driven by participants overweighing points that are more congruent with their beliefs, a visual analytic system might first survey a user's existing beliefs, and then highlight data points that might be under weighed to increase their visual salience. Existing work has shown promising results in this area [78], and future work can test similar mitigation techniques in real-world contexts with more complex data sets and tasks.

Eliciting Beliefs: We elicited beliefs via self-reporting, a survey technique commonly used in existing work to capture people's state of mind [34], following a similar approach in measuring belief in correlations as introduced in [85]. We intentionally asked participants to report their beliefs on a verbal scale instead of the same numerical scale they used to report their correlation estimates to reduce potential carryover effects and experimenter demand. Future work should investigate the effects of reporting beliefs on a verbal versus numerical scale, which can lead to more refined quantitative models of belief elicitation.

There exists a variety of methods to elicit beliefs from multiple disciplines (for a design space on these methods, see [54]). In behavioral economics, researchers often infer belief from a choice tasks where participants indicate their preferences [70], though more recent work has argued against its effectiveness [55]. In visualization work, researchers have experimented with capturing belief through having participants manipulate the slope of a line in a chart [38, 44], specialized elicitation visual interfaces [43, 56], or using a Likert scale where participants rate their agreement with a statement [85]. Griffiths and Tenenbaum [24] demonstrated that we could approximate people's prior beliefs by asking them to make predictions. More recently, Suchow, Pacer, & Griffiths [72] demonstrated that we can also estimate people's priors through transmission chains, such as a game of telephone. Future work should explore how our results might generalize (or not) to other belief elicitation methods, in addition to exploring and evaluating alternative methods to capture people's beliefs.

Belief Updating: Our work also has implications for how people update their beliefs. When participants reported their level of agreement with statements describing the correlation between two variables, when the variables were belief-triggering (gun/violent crime, air quality/environmental regulations), participants' prior and posterior beliefs did not significantly differ. But when the variables were not belief-triggering (tree/names, rain/lawyers), there was a significant difference in participants' prior and posterior beliefs.

This may suggest that when participants hold strong existing beliefs (priors), seeing additional evidence does not change their beliefs. They stick to their strong priors, as it was built on a great deal of data from their past experiences. However, when participants do not have strong existing beliefs about how two variables could be correlated, just seeing one instance of a scatterplot depicting a correlation pushed them to update their beliefs, even though the scatterplot they saw was not designed to be persuasive. The amount of change between prior and posterior beliefs can likely be used as a proxy to determine how much people are persuaded by the new data they saw, which will allow us to reverse engineer the correlation people see in data by comparing the deltas between their prior and posterior beliefs (see [15] for more discussions). Future work can further examine how much seeing a correlation impacts belief, and how much stating a belief impacts subsequent correlation estimation, using a wider range of topics, belief elicitation methods, and correlation values. Additionally, for instances where participants viewed data depicting variables for which they do not have strong priors, future researchers can also examine how that data could influence participants' belief updating through exposure effects [26].

Generalizability: We only tested two levels of correlation in the current experiment using one scatterplot design and elicited participants' correlation estimations through Pearson's r. These decisions were made based on methodologies from prior work to control for the effect of these design-based factors and prevent the combinatorial explosion of having too many conditions [80]. Future work can extend this space of investigation and examine the generalizability of our effects by testing additional levels of correlation, different shapes of the data (e.g., varying the dispersion, the number of outliers, or the range of R-values), and various scatterplot designs (e.g., aspect ratios of the plot area or with regression lines).

Furthermore, future work can test other methods of eliciting participant estimations. For example, instead of asking for an r-value, they could ask participants to indicate which points they think are 'important' or indicate the perceived correlation by drawing an ellipse or regression line. Such measures may be more or less sensitive to bias and may provide a more precise indication of the influence of outliers or other salient points on a trial-to-trial basis. These follow-up experiments can build towards a comprehensive model mapping each component of a scatterplot to its impact on viewer interpretation, potentially pinpointing the factors or mechanisms behind belief-driven misestimation of correlations.

REFERENCES

[1] K. Ajani, E. Lee, C. Xiong, C. N. Knaflic, W. Kemper, and S. Franconeri. Declutter and focus: Empirically evaluating design guidelines for effective data communication. *IEEE Transactions on Visualization and Computer Graphics*, 2021.

[2] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pages 111–117. IEEE, 2005.

[3] M. Bakker and J. M. Wicherts. Outlier removal, sum scores, and the inflation of the type i error rate in independent samples t tests: The power of alternatives and recommendations. *Psychological methods*, 19(3):409, 2014.

[4] E. Balcetis and D. Dunning. See what you want to see: motivational influences on visual perception. *Journal of personality and social psychology*, 91(4):612, 2006.

[5] E. Balcetis and D. Dunning. Wishful seeing: More desired objects are seen as closer. *Psychological science*, 21(1):147–152, 2010.

[6] D. Bates, D. Sarkar, M. D. Bates, and L. Matrix. The lme4 package. *R package version*, 2(1):74, 2007.

[7] M. Bostock, S. Carter, C. Amanda, and K. Quealy. One report, diverging perspectives, 2012.

[8] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE transactions on visualization and computer graphics*, 19(12):2376–2385, 2013.

[9] S. C. Castro, P. S. Quinan, H. Hosseinpour, and L. Padilla. Examining effort in 1d uncertainty communication using individual differences in working memory and nasa-tlx. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):411–421, 2021.

[10] P. Cavanagh. The cognitive impenetrability of cognition. *Behavioral and Brain Sciences*, 22(3):370–371, 1999.

[11] S. Coren, L. M. Ward, and J. T. Enns. *Sensation and perception*. John Wiley & Sons Hoboken, NJ, 2004.

[12] M. Correll and M. Gleicher. Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE transactions on visualization and computer graphics*, 20(12):2142–2151, 2014.

[13] N. F. Dieckmann, R. Gregory, E. Peters, and R. Hartman. Seeing what you want to see: How imprecise uncertainty ranges enhance motivated reasoning. *Risk analysis*, 37(3):471–486, 2017.

[14] E. Dimara, S. Franconeri, C. Plaisant, A. Bezerianos, and P. Dragicevic. A task-based taxonomy of cognitive biases for information visualization. *IEEE transactions on visualization and computer graphics*, 2018.

[15] H. Elhamdadi, A. Gaba, Y.-S. Kim, and C. Xiong. How do we measure trust in visual data communication? In *2022 IEEE Workshop on Evaluation and Beyond-Methodological Approaches to Visualization (BELIV)*. IEEE, 2022.

[16] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner. G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2):175–191, 2007.

[17] C. Firestone and B. J. Scholl. Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and brain sciences*, 39, 2016.

[18] B. Fischhoff, D. Kahneman, P. Slovic, and A. Tversky. For those condemned to study the past: Heuristics and biases in hindsight. *Foundations of cognitive psychology: Core readings*, pages 621–636, 2002.

[19] M. Friendly and D. Denis. The early origins and development of the scatterplot. *Journal of the History of the Behavioral Sciences*, 41(2):103–130, 2005.

[20] W. S. Geisler and D. Kersten. Illusions, perception and bayes. *Nature neuroscience*, 5(6):508–510, 2002.

[21] M. Gleicher, M. Correll, C. Nothelfer, and S. Franconeri. Perception of average value in multiclass scatterplots. *IEEE transactions on visualization and computer graphics*, 19(12):2316–2325, 2013.

[22] W. C. Gogel. The effect of object familiarity on the perception of size and distance. *The Quarterly journal of experimental psychology*, 21(3):239–247, 1969.

[23] D. Gotz and M. X. Zhou. An empirical study of user interaction behavior during visual analysis. *IBM Research RC24525 (W0803-127)*, 2008.

[24] T. L. Griffiths and J. B. Tenenbaum. Optimal predictions in everyday cognition. *Psychological science*, 17(9):767–773, 2006.

[25] A. Hale-Smith, C. L. Park, and D. Edmondson. Measuring beliefs about suffering: development of the views of suffering scale. *Psychological Assessment*, 24(4), 2012.

[26] A. A. Harrison. Mere exposure. In *Advances in experimental social psychology*, volume 10, pages 39–83. Elsevier, 1977.

[27] L. Harrison, F. Yang, S. Franconeri, and R. Chang. Ranking visualizations of correlation using weber's law. *IEEE transactions on visualization and computer graphics*, 20(12):1943–1952, 2014.

[28] J. Heer and M. Bostock. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 203–212, 2010.

[29] M. Hegarty. Diagrams in the mind and in the world: Relations between internal and external visualizations. In *Diagrammatic representation and inference*, volume 8, pages 1–13. Springer, 2004.

[30] M. Hegarty, M. S. Canham, and S. I. Fabrikant. Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(1):37, 2010.

[31] M.-H. Hong, J. K. Witt, and D. A. Szafir. The weighted average illusion: Biases in perceived mean position in scatterplots. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):987–997, 2021.

[32] J. Hullman. Why authors don't visualize uncertainty. *IEEE transactions on visualization and computer graphics*, 26(1):130–139, 2019.

[33] J. Hullman, P. Resnick, and E. Adar. Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering. *PloS one*, 10(11):e0142444, 2015.

[34] A. Joshi, S. Kale, S. Chandel, and D. K. Pal. Likert scale: Explored and explained. *British journal of applied science & technology*, 7(4):396, 2015.

[35] D. Kahneman. *Thinking, fast and slow*. Macmillan, 2011.

[36] A. Kale, M. Kay, and J. Hullman. Visual reasoning strategies for effect size judgments and decisions. *IEEE transactions on visualization and computer graphics*, 27(2):272–282, 2020.

[37] A. Kale, F. Nguyen, M. Kay, and J. Hullman. Hypothetical outcome plots help untrained observers judge trends in ambiguous data. *IEEE transactions on visualization and computer graphics*, 2018.

[38] A. Karduni, D. Markant, R. Wesslen, and W. Dou. A bayesian cognition approach for belief updating of correlation judgement through uncertainty visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):978–988, 2020.

[39] M. Kay and J. Heer. Beyond weber's law: A second look at ranking visualizations of correlation. *IEEE transactions on visualization and computer graphics*, 22(1):469–478, 2015.

[40] M. Kay, T. Kola, J. R. Hullman, and S. A. Munson. When (ish) is my bus?: User-centered visualizations of uncertainty in everyday, mobile predictive systems. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 5092–5103. ACM, 2016.

[41] Y. Kim and J. Heer. Assessing effects of task and data distribution on the effectiveness of visual encodings. In *Computer Graphics Forum*, volume 37, pages 157–167. Wiley Online Library, 2018.

[42] Y. S. Kim. *Designing Belief-driven Interactions with Data*. University of Washington, 2020.

[43] Y.-S. Kim, P. Kayongo, M. Grunde-McLaughlin, and J. Hullman. Bayesian-assisted inference from visualized data. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):989–999, 2020.

[44] Y.-S. Kim, K. Reinecke, and J. Hullman. Explaining the gap: Visualizing one's predictions improves recall and comprehension of data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 1375–1386, 2017.

[45] J. Klayman. Varieties of confirmation bias. *Psychology of learning and motivation*, 32:385–418, 1995.

[46] J. Klayman and Y.-W. Ha. Confirmation, disconfirmation, and information in hypothesis testing. *Psychological review*, 94(2):211, 1987.

[47] J. H. Larkin and H. A. Simon. Why a diagram is (sometimes) worth ten thousand words. *Cognitive science*, 11(1):65–100, 1987.

[48] T. Liu, X. Li, C. Bao, M. Correll, C. Tu, O. Deussen, and Y. Wang. Data-driven mark orientation for trend estimation in scatterplots. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–16, 2021.

[49] Z. Liu and J. T. Stasko. Mental models, visual reasoning and interaction in information visualization: A top-down perspective. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):999–1008, 2010.

[50] Y. Luo and J. Zhao. Motivated attention in climate change perception and action. *Frontiers in psychology*, 10:1541, 2019.

[51] Y. Luo and J. Zhao. Attentional and perceptual biases of climate change.

Current Opinion in Behavioral Sciences, 42:22–26, 2021.

[52] G. Lupyan. Object knowledge changes visual appearance: Semantic effects on color afterimages. *Acta Psychologica*, 161:117–130, 2015.

[53] G. Lupyan. Changing what you see by changing what you know: the role of attention. *Frontiers in psychology*, 8:553, 2017.

[54] S. Mahajan, B. Chen, A. Karduni, Y.-S. Kim, and E. Wall. Vibe: A design space for visual belief elicitation in data journalism. 2022.

[55] C. F. Manski. Survey measurement of probabilistic macroeconomic expectations: progress and promise. *NBER Macroeconomics Annual*, 32(1):411–471, 2018.

[56] P. Mantri, H. Subramonyam, A. L. Michal, and C. Xiong. How do viewers synthesize conflicting information from data visualizations. *IEEE transactions on visualization and computer graphics*.

[57] A. Martín, J. G. Chambeaud, and J. F. Barraza. The effect of object familiarity on the perception of motion. *Journal of Experimental Psychology: Human Perception and Performance*, 41(2):283, 2015.

[58] R. N. McCauley and J. Henrich. Susceptibility to the müller-lyer illusion, theory-neutral observation, and the diachronic penetrability of the visual input system. *Philosophical Psychology*, 19(1):79–101, 2006.

[59] E. Mischenko, I. Negishi, E. S. Gorbunova, and T. Sawada. Examining the role of familiarity in the perception of depth. *Vision*, 4(2):21, 2020.

[60] C. R. Mynatt, M. E. Doherty, and R. D. Tweney. Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, 29(1):85–95, 1977.

[61] R. S. Nickerson. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2):175–220, 1998.

[62] L. M. Padilla, S. H. Creem-Regehr, M. Hegarty, and J. K. Stefanucci. Decision making with visualizations: a cognitive framework across disciplines. *Cognitive research: principles and implications*, 3(1):29, 2018.

[63] L. Pessoa, E. Thompson, and A. Noë. Finding out about filling-in: A guide to perceptual completion for visual science and the philosophy of perception. *Behavioral and brain sciences*, 21(6):723–748, 1998.

[64] Z. Pylyshyn. Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception. *Behavioral and brain sciences*, 22(3):341–365, 1999.

[65] I. Qualtrics. Qualtrics. *Provo, UT, USA*, 2013.

[66] A. Raftopoulos. Is perception informationally encapsulated? the issue of the theory-ladenness of perception. *Cognitive Science*, 25(3):423–451, 2001.

[67] R. A. Rensink. The nature of correlation perception in scatterplots. *Psychonomic bulletin & review*, 24(3):776–797, 2017.

[68] R. A. Rensink and G. Baldridge. The perception of correlation in scatterplots. In *Computer Graphics Forum*, volume 29, pages 1203–1210. Wiley Online Library, 2010.

[69] B. Saket, A. Endert, and Ç. Demiralp. Task-based effectiveness of basic visualizations. *IEEE transactions on visualization and computer graphics*, 25(7):2505–2512, 2018.

[70] P. A. Samuelson. Consumption theory in terms of revealed preference. *Economica*, 15(60):243–253, 1948.

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

[72] J. W. Suchow, M. D. Pacer, and T. L. Griffiths. Design from zeroth principles. In *CogSci*, 2016.

[73] J. G. Trafton, S. B. Trickett, and F. E. Mintz. Connecting internal and external representations: Spatial transformations of scientific visualizations. *Foundations of Science*, 10(1):89–106, 2005.

[74] A. Tversky and D. Kahneman. Belief in the law of small numbers. *Psychological bulletin*, 76(2):105, 1971.

[75] A. C. Valdez, M. Ziefle, and M. Sedlmair. A framework for studying biases in visualization research. In *DECISIVe 2017: Dealing with Cognitive Biases in Visualisations*, 2017.

[76] P. Vetter and A. Newen. Varieties of cognitive penetration in visual perception. *Consciousness and cognition*, 27:62–75, 2014.

[77] E. Wall, L. M. Blaha, L. Franklin, and A. Endert. Warning, bias may occur: A proposed approach to detecting cognitive bias in interactive visual analytics. In *2017 ieee conference on visual analytics science and technology (vast)*, pages 104–115. IEEE, 2017.

[78] E. Wall, A. Narechania, A. Coscia, J. Paden, and A. Endert. Left, right, and gender: Exploring interaction traces to mitigate human biases. *arXiv preprint arXiv:2108.03536*, 2021.

[79] E. Wall, J. Stasko, and A. Endert. Toward a design space for mitigating cognitive bias in vis. In *2019 IEEE Visualization Conference (VIS)*, pages 111–115. IEEE, 2019.

[80] E. Wall, C. Xiong, and Y.-S. Kim. Vishikers' guide to evaluation: Competing considerations in study design. *IEEE Computer Graphics and Applications*, 42(3):29–38, 2022.

[81] P. C. Wason. Reasoning about a rule. *Quarterly journal of experimental psychology*, 20(3):273–281, 1968.

[82] J. L. Wilkins. The relationship among elementary teachers' content knowledge, attitudes, beliefs, and practices. *Journal of Mathematics Teacher Education*, 11(2):139–164, 2008.

[83] J. M. Wolfe. Guided search 2.0 a revised model of visual search. *Psychonomic bulletin & review*, 1(2):202–238, 1994.

[84] C. Xiong, A. Sarvghad, Ç. Demiralp, J. M. Hofman, and D. G. Goldstein. Investigating perceptual biases in icon arrays. *Proceedings of the 2022 ACM Conference on Human Factors in Computing Systems (CHI '22)*, 2022.

[85] C. Xiong, J. Shapiro, J. Hullman, and S. Franconeri. Illusion of causality in visualized data. *IEEE transactions on visualization and computer graphics*, 26(1):853–862, 2019.

[86] C. Xiong, L. van Weelden, and S. Franconeri. Perceiving graphs as ambiguous figures. *Journal of Vision*, 18(10):1327–1327, 2018.

[87] C. Xiong, L. Van Weelden, and S. Franconeri. The curse of knowledge in visual data communication. *IEEE transactions on visualization and computer graphics*, 26(10):3051–3062, 2019.

[88] F. Yang, L. T. Harrison, R. A. Rensink, S. L. Franconeri, and R. Chang. Correlation judgment and visualization features: A comparative study. *IEEE Transactions on Visualization and Computer Graphics*, 25(3):1474–1488, 2018.