A Bayesian cognition approach for belief updating of correlation judgement through uncertainty visualizations

Alireza Karduni, Douglas Markant, Ryan Wesslen, Wenwen Dou

University of North Carolina at Charlotte

Abstract—Understanding correlation judgement is important to designing effective visualizations of bivariate data. Prior work on correlation perception has not considered how factors including prior beliefs and uncertainty representation impact such judgements. The present work focuses on the impact of uncertainty communication when judging bivariate visualizations. Specifically, we model how users update their beliefs about variable relationships after seeing a scatterplot with and without uncertainty representation. To model and evaluate the belief updating, we present three studies. Study 1 focuses on a proposed "Line + Cone" visual elicitation method for capturing users' beliefs in an accurate and intuitive fashion. The findings reveal that our proposed method of belief solicitation reduces complexity and accurately captures the users' uncertainty about a range of bivariate relationships. Study 2 leverages the "Line + Cone" elicitation method to measure belief updating on the relationship between different sets of variables when seeing correlation visualization with and without uncertainty representation. We compare changes in users beliefs to the predictions of Bayesian cognitive models which provide normative benchmarks for how users should update their prior beliefs about a relationship in light of observed data. The findings from Study 2 revealed that one of the visualization conditions with uncertainty communication led to users being slightly more confident about their judgement compared to visualization without uncertainty information. Study 3 builds on findings from Study 2 and explores differences in belief update when the bivariate visualization is congruent or incongruent with users' prior belief. Our results highlight the effects of incorporating uncertainty representation, and the potential of measuring belief updating on correlation judgement with Bayesian cognitive models.

Index Terms—Information visualization, Bayesian modeling, uncertainty visualizations, correlations, belief elicitation

✦

1 INTRODUCTION

Correlation judgement is an important topic and has been recently studied by the data visualization community [15, 26, 47, 57]. Understanding how people perceive correlations from data is necessary for the design of effective visualizations like scatterplots. Visualization researchers have investigated perceptual constraints on correlation judgment, including the use of Weber's Law [15, 47], a log-linear model augmented with censored regression and Bayesian methods [26], and other visual features [57]. While these empirical studies and models provide valuable insights and recommendations for correlation visualization design, they can be expanded to consider other factors that affect people's understanding of variable relationships.

One such factor is a user's prior beliefs when interpreting a correlation visualization. Previous studies often examine the perception of correlations between unnamed variables to avoid the effects of prior knowledge [46, 47] so that participants' beliefs about the variables do not influence their judgements. However, in practice, people rely on prior knowledge when interpreting and learning from correlation visualizations. As a result, it is important to investigate how prior beliefs affect the perception and interpretation of correlations. In addition to prior beliefs, another factor related to correlation judgement that warrants more research is uncertainty communication. Recently, visualization researchers have argued for the importance of uncertainty communication in information visualization [18]. Uncertainty communication techniques like hypothetical outcome plots (HOPs) [21, 25] provide methods to visualize uncertain data for general audiences.

The experiments in this paper build on previous research on correlation judgement by examining the impact of prior beliefs and uncertainty communication. We explore the following research questions: (1) how do prior beliefs impact one's correlation judgement? (2) how do people

adjust their beliefs when the correlation visualization aligns or conflicts with their prior belief? (3) when uncertainty communication is incorporated in a correlation visualization, are users more or less likely to adjust their beliefs based on the conveyed relationship?

We also use Bayesian cognitive modeling [32] to quantitatively model how people interpret newly observed data in light of existing prior knowledge. Bayesian cognitive modeling offers a principled framework to understand how people interpret visualizations in light of prior beliefs [28] and how such beliefs should be updated with new information from a data visualization through Bayesian reasoning [29, 40, 44]. This provides a normative framework for evaluating the effects of visualization on beliefs, including the impact of uncertainty communication on users' interpretations of data [10, 20, 28] and the presence of biases that impair data-driven decision making [29].

Building such Bayesian cognitive models requires an accurate understanding of people's prior beliefs. Existing techniques for eliciting priors about correlations have a number of limitations, including a reliance on expert statistical knowledge related to correlation coefficients and their relationship to data [31, 43, 58]. Our paper first evaluates a novel graphical elicitation method, "Line + Cone", for eliciting beliefs about the correlation between two variables through interactive data visualizations. With the proposed elicitation method, we conducted two experiments to study how people update beliefs about bivariate relationships when seeing correlation visualization with and without uncertainty representation.

This paper bridges several areas of past work on correlation judgment, belief elicitation, and uncertainty visualization, while also drawing on recent methods for modeling belief change using the framework of Bayesian inference. Specifically, this paper's contributions are:

- Study 1: Introduce and validate the graphical "Line + Cone" method for eliciting prior beliefs about bivariate correlations, which is then used in the subsequent studies to measure belief change.

- Study 2: Compare differences in belief updating across correlation visualization with and without uncertainty communication.

- Study 3: Explore differences in users' belief update when the

correlation visualization (with and without uncertainty communication) is congruent or incongruent with their prior beliefs.

Analysis of Study 1 showed that the "Line + Cone" belief elicitation method can be used to estimate peoples' mental representations of the correlation compared to a recent, more labor-intensive approach from cognitive science for measuring subjective belief distributions [52]. Study 2 revealed that participants updated their beliefs more effectively, and felt more confident, after observing visualizations with representations of uncertainty. In Study 3 we found evidence to support the hypothesis that people exhibit less belief change when seeing correlation visualizations that are incongruent with their prior beliefs. These results lay the groundwork for quantitative theories of how visualizations guide, and in some cases distort, how people learn about correlations through data visualization.

2 BACKGROUND

2.1 Correlation perception and the effects of prior beliefs

A common task in visual analytics is assessing the relationship between two or more variables, often as a scatterplot [54]. In statistics, such relationships are typically quantified as correlations. However, statistics like Pearson correlation can be misleading. For example, Anscombe's quartet [1] demonstrates that hidden patterns in the data are obscured by identical statistics. Even for expert data analysts, visual data inspection is an important part of the analysis process. Past psychology studies have considered how perceptual processing of scatterplots can affect an individual's understanding of correlations [39, 46, 47]. Building off that research, InfoVis researchers have identified scatterplots as an effective technique in discriminating correlations [34], testing correlation perception with Weber's law through additional techniques [15, 26], and identifying visual features in correlation perception [57]. However, these studies have not considered how prior beliefs affect individual's perception of variable relationships.

Research in psychology shows that prior beliefs have a strong influence on people's interpretation of uncertain data [4, 9, 30, 50], especially for correlations [2, 3]. A central theory that explains why prior beliefs are important is the dual-process account of reasoning [7, 24]. This theory posits that fast heuristic processes (System 1) competes with slower analytic processes (System 2) that can affect logical decisions. Evans *et al.* [8] suggested that belief bias [7, 9] could occur as "within-participant conflict" between the two systems when participants tend to agree with an argument based on whether or not they agree with the conclusion rather than its logical conclusion. Alternatively, other research focused on theory-motivated reasoning bias based on "congruent" and "incongruent" evidence relative to an individuals' belief systems [30]. These theories motivate design aspects in Study 2 and 3.

2.2 Uncertainty visualizations

Uncertainty visualizations are important as they enable better decision-making by conveying the possibility that a point estimate may vary [20]. More recently, research in InfoVis has provided innovative techniques like Hypothetical Outcome Plots (HOPs) [21, 25], frequency based representations [10, 27], visual semiotics [38], and design guidelines [13] for visualizing uncertainty. Alternatively, other visualization researchers have studied important application aspects of uncertainty visualizations including hurricane prediction through ensemble modeling [36, 49], comparing users' prior beliefs congruence to social data [28], how uncertainty evaluation is prone to error [17], and its potential to improve ones ability to make predictions about replications of future experiments [19].

2.2.1 Eliciting correlation beliefs

Psychologists have used a variety of approaches to elicit beliefs about correlations. Initial research used two-step procedure to elicit participant's correlation belief [3, 22]: (1) determine relationship direction (positive or negative) and (2) rate the strength of the relationship. Later methods expanded on this approach by including Likert Scales, Spearmanss correlation, probability of concordance, and conditional quantile estimates [6, 11, 23, 31, 43]. However, there are several shortcomings

with the previous approaches. Some methods only elicit beliefs about central tendency without capturing degree of uncertainty, while methods which do elicit uncertainty are labor-intensive [58]. Most methods rely on some background knowledge of statistics [31, 43], including how to interpret correlation coefficients, thus limiting their applicability to non-expert populations.

Cognitive scientists have developed a related technique for eliciting subjective belief distributions named Markov Chain Monte Carlo with People (MCMC-P; [52, 53]). Inspired by algorithms for MCMC estimation [11], MCMC-P as an approach to estimate a person's subjective belief distribution through sampling. In Study 1, we use MCMC-P as an elicitation benchmark to our proposed Line + Cone belief elicitation technique and outline this technique in Section 4.

2.3 Bayesian cognitive modeling in data visualizations

Cognitive modeling in visualization initially was studied as a subset of visuospatial reasoning in how individuals derive meaning from external visual representations [55]. Visualization researchers have integrated similar ideas to understand visualization cognitive processes through insight-based approaches [12] and top-down modeling [37, 45]. More recently, InfoVis researchers have used Bayesian models to understand cognitive processing of visualizations [29, 56]. Cognitive scientists have demonstrated the importance of Bayesian modeling to understanding individual decision-making [14, 32]. In this approach, an individual has some prior belief that is updated when the individual consumes additional data, resulting in their posterior beliefs. Bayesian cognition models have been used to understand deviations from optimal belief updating due to conservatism, sample-based inference (approximation) and "resource-rational" interpretations of cognitive bias [35].

To our knowledge only two previous InfoVis studies [29, 56] have combined belief elicitation with a Bayesian cognitive modeling framework. Wu *et al.* [56] examined whether people integrated prior probabilities with data in an optimal manner. They found that priors influenced predictions in a manner consistent with Bayesian inference, although to a lesser extent than predicted by the model. However, a limitation to this study was that participants were given a prior; therefore, prior beliefs cannot be examined. In contrast, Kim *et al.* [29] empirically measured participants' prior beliefs about the a target proportional quantity and used those priors to calculate the normative posterior given the data that was presented. In aggregate, participants' judgments were consistent with predictions derived from Bayesian inference, although less so for large data sets. However, participants expressed greater uncertainty in their judgments than expected from the Bayesian model. Further, the authors connect such Bayesian modeling and belief elicitation with recent research on visualizing uncertainty through techniques like HOPs [21, 25]. Our work extends their framework but considering correlation beliefs rather than proportional values.

## 3 RESEARCH QUESTIONS AND ANALYSIS METHODS

Our primary research question is the effect of providing uncertainty communications on users' belief updating in correlation visualization. In order to address this research question, we conducted a sequence of three experiments with latter ones building on the earlier studies.

A key to understanding users' belief update is the ability to accurately and intuitively capture such beliefs. **Study 1** evaluates the Line + Cone elicitation method relative to Markov Chain Monte Carlo with People (MCMC-P) [52], a belief elicitation method from cognitive science. After validating the Line + Cone method, we apply it in the next two experiments to address the main research question. In **Study 2**, we explore the effect of correlation visualizations with and without uncertainty representation on belief updating. Our primary hypothesis is that visualizations with uncertainty representation will overall lead to less belief updating about the correlation between two variables. Findings from Study 2 provides partial evidence to support the primary hypothesis. To expand on the findings, we are interested in further understanding users' belief update when the data visualization was deliberately manipulated based on users' prior beliefs. Therefore, **Study 3** extends Study 2's design but introduces a treatment that alters the data provided to participants to be either congruent or incongruent with their

A. Line + Cone elicitation
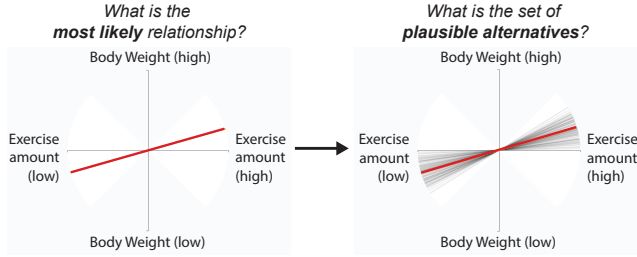
B. MCMC-P elicitation
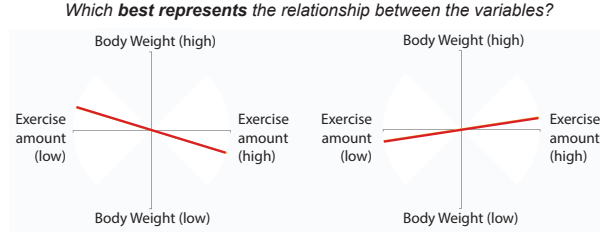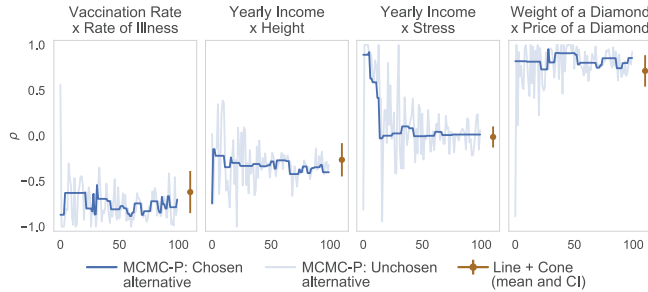

C. Example results

Fig. 1. Elicitation methods in Study 1. **A:** For the Line + Cone elicitation, participants first recorded the belief about the most likely relationship between two variables (red line), then adjusted the set of plausible alternatives based on their uncertainty (gray lines). **B:** For the MCMC-P elicitation, participants responded to a series of two-alternative forced choices in which they judged which of two lines was more likely to represent the true relationship between the variables. **C:** Example comparison of elicitation results for a participant in Study 1. Dark blue lines indicate the chain of chosen alternatives from MCMC-P across 100 trials. Light blue lines indicate unchosen alternatives. The corresponding mean and CI from the Line + Cone elicitation is shown at the right of each plot.

prior beliefs. We then evaluate the degree to which individuals update their beliefs when data provided either conflicts or aligns with their prior and whether the presence of uncertainty visualizations interact with that effect.

To analyze the results of Study 2 and 3, we employ mixed effects models to identify differences between treatments. The mixed effects models control for individual heterogeneity assumed between participants and the datasets (variable pairs) provided to participants. To explain the findings from the mixed effects models, we evaluate whether Bayesian cognitive models can be used to predict users' posterior beliefs under different experiment treatment.

4 STUDY 1: EVALUATING LINE + CONE ELICITATION

Our goal in Study 1 (see preregistration[1]) was to develop and validate the Line + Cone visual interface for eliciting prior beliefs about the correlation between two variables. In selecting our approach, we aimed to measure beliefs about both the *most likely* correlation between variables and the *degree of uncertainty*, without a need for statistics domain knowledge or numerical reasoning (see Section 2.2.1). We assessed

---

[1]http://aspredicted.org/blind.php?x=zp7hr3

the convergent validity of the Line + Cone method by comparing it to a higher resolution, but more labor-intensive, approach to eliciting subjective beliefs: Markov Chain Monte Carlo with People (MCMC-P; [52, 53]). MCMC-P resembles common sampling-based estimation algorithms such as Metropolis-Hastings in which a chain of states are sampled from an underlying probability distribution. In MCMC-P, state transitions are determined by asking participants to make forced-choice comparisons of the likelihood of possible values of the target parameter in many trials (usually in the range of 100 or more).

In our experiment we elicited prior beliefs about five sets of variables using both MCMC-P and the Line + Cone method. We created variable sets to cover a range of plausible correlationspossible divergent prior beliefs. For example, we expected that for the relationship *Weight x Price of diamonds* most participants would believe there is a strong positive correlation, while there may be less consensus about the relationship *Vaccination rate x Rate of illness*. Based on participants' responses we estimated the mean and confidence interval of their subjective prior belief (i.e., the relative likelihood of possible correlations between two variables). We then examined the degree to which the resulting prior means and CIs were correlated across the two methods.

4.1 Study Design

The experiment involved a within-subjects manipulation of elicitation method (Line + Cone vs. MCMC-P). Participants' beliefs were elicited for the same set of five variable sets (Table 1) using each method in a blocked presentation. The order of elicitation methods and variable sets within each block were randomized for each participant.

4.1.1 Line + cone elicitation

We designed a visual interface in which the mean and CI are directly elicited through the user's interaction. Each elicitation involves a two-step procedure (Figure 1A). First, the user selects the orientation of a red line according to their belief about the most likely relationship between the variables. Second, the user adjusts the width of the uncertainty cone. The uncertainty cone was depicted by gray lines which were draws from a Normal distribution centered on the most likely correlation (red line) and truncated at -1 and 1. Participants were instructed to adjust the cone such that the lines captured the range of "plausible alternatives" for the relationship between the variables.

4.1.2 MCMC-P elicitation

Markov Chain Monte Carlo with People (MCMC-P) is used to estimate subjective belief distributions based on a series of choices between two alternatives. In our task, each alternative represents a potential correlation between a pair of variables. For each variable set there were 100 choice trials. On each trial, the participant was shown two lines representing potential correlations (Figure 1B). Participants were instructed to select the alternative which was more likely to represent the true relationship between the variables. On the first choice trial the alternatives were two randomly selected correlations, one positive and one negative. In subsequent trials, the choice set included the alternative chosen on the previous trial and a *proposal* generated from a Normal distribution centered on the previous choice. The width of the proposal distribution was adaptively tuned based on how often a participant accepted new proposals (see [48]). Each block resulted in a chain of alternatives that were chosen by the user (Figure 1C). The prior mean was calculated as the mean of the sampling chain, while the CI was the range between the 2.5% and 97.5% quantiles.

4.2 Participants

$N = 152$ participants were recruited from Amazon Mechanical Turk. Participants earned $2.00 upon completion of the task, which took an average of 25.4 minutes ($SD = 12.2$). Per our pre-registration, we used several measures of task engagement to decide whether to exclude a participant. We excluded 55 participants who failed an attention check question and 36 participants who made nonsensical or incomplete responses to a set of open-ended questions regarding how they would respond to real-world situations. We also excluded 35 participants who
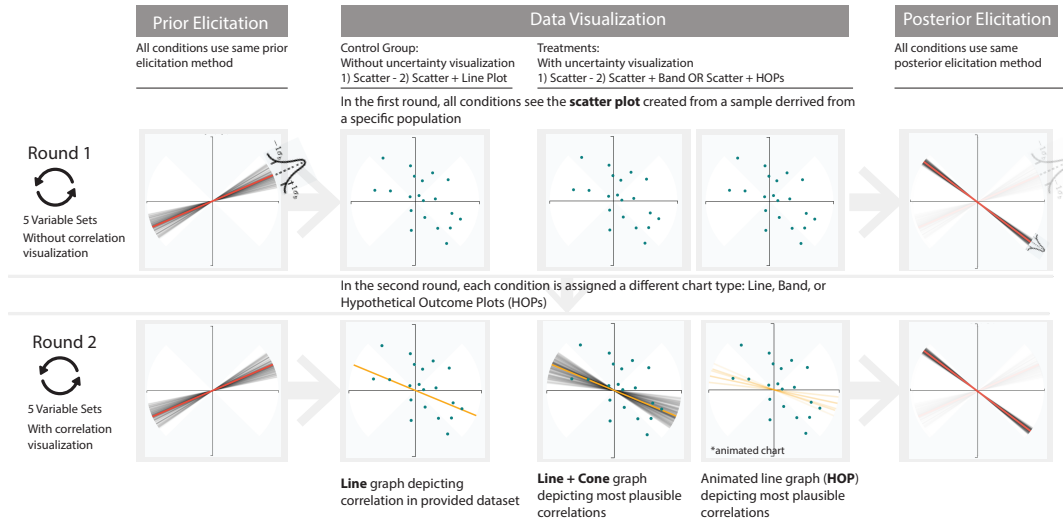
Fig. 2. Study 2 Design. Each user goes through ten variable sets (five variables for two rounds) and elicit their belief before and after seeing data visualizations about each variable set. In Round 1, the user views five variable sets through only scatterplots. In Round 2, the user is randomly assigned to either Line, Cone, or HOP visualization treatments and views the remaining five variable sets.

met pre-specified exclusion criteria based on responses in the MCMC-P elicitation, including response streaks, response alternation, and response time. After accounting for all exclusions, $N = 92$ participants were included in the analysis.

4.3 Results and Discussion of Study 1

Table 1. Correlations between prior means and CIs elicited through Line + Cone and MCMC-P methods in Study 1.

	Prior mean		Prior CI	
Variable set	Pearson r	p-value	Pearson r	p-value
Weight x Price of diamonds	.26	**.012**	.34	**.001**
Exercise amount x Body weight	.37	**< .001**	.29	**.005**
Yearly income x Height	.12	.26	.27	**.010**
Yearly income x Stress	.45	**< .001**	.30	**.003**
Vaccination rate x Rate of illness	.40	**< .001**	.39	**< .001**

Our primary question was whether the belief distributions elicited with the Line + Cone method correlated with those generated using our MCMC-P procedure. We calculated Pearson correlations between the prior means and CIs for each variable set (Table 1). Elicited prior means were significantly correlated for 4 of the 5 variable sets, with the *Yearly income X Height* variable set the only exception. Prior CIs elicited from the two methods were significantly correlated in all 5 variable sets. These results suggest that our visual Line + Cone elicitation method is able to capture variation in beliefs about correlations across different variable sets, including beliefs about the most likely relationship as well as the degree of uncertainty, while being less labor-intensive than MCMC-P and requiring less statistics domain knowledge than existing elicitation methods.

5 STUDY 2: BELIEF UPDATING WITH AND WITHOUT UNCERTAINTY REPRESENTATIONS

In the second study, we applied the Line + Cone elicitation method to examine belief change in the context of correlation visualization. We evaluated whether the type of visualization impacted the degree to which people updated their beliefs. Specifically, our **Study 2 Main Hypothesis** [2] was that correlation visualizations which include representations of the uncertainty in the true population correlation would lead to less belief updating when people's prior beliefs were inconsistent with the presented data. This hypothesis is motivated by research

---

[2]http://aspredicted.org/blind.php?x=39yn5g

on confirmation bias [41, 42] showing that people overweight evidence that is consistent with their prior beliefs. Uncertainty visualizations, by giving credence to a range of possible relationships (including less likely relationships that are more similar to a person's prior belief) may lead to less belief updating compared to visualizations that only represent the most likely a posteriori relationship. As a secondary hypothesis, we hypothesize that datasets with small and moderate correlations lead to less belief updating compared to datasets with stronger correlations.

5.1 Study Design

We employed a mixed design with a between-subjects manipulation of the visualization type (with and without uncertainty representation) and a within-subjects manipulation of the sample correlation of data presented to participants. In each trial participants reported their belief about the relationship between a set of variables, both before and after they experienced a data visualization All participants completed two rounds of five trials. In the first round the datasets were visualized as scatterplots to all participants (**Scatter** condition). In the second round the scatterplots were augmented with a visualization of the predicted population correlation based on the given dataset. Participants were randomly assigned to one of the following conditions (Fig 2):

- **Line**: A line representing the most likely population correlation was superimposed on the scatterplot [3]

- **Cone**: The line appeared with an uncertainty cone which represents the 95% confidence interval for the population correlation

- **HOP**: Hypothetical outcome plots (HOPs, [21, 25]) were used to present animated draws from the 95% confidence interval for the correlation

5.1.1 Datasets

We created two groups of five variable pairs that covered a range of population correlations between -0.9 to 0.9. We then generated 100 random samples for each variable pair based on the population correlation. The participants were told that the dataset is a *a sample of data collected from the real world.* [4] All points were re-centered with a mean of zero on each variable. All participants saw the same

---

[3]Note that the Line condition does not contain an uncertainty representation while the Cone and HOP conditions do.

[4]Due to random sampling, the sample correlations differed slightly from the specified population correlation.

data points for each variable pair. The order of the variable pairs was randomized for each participant.

Note that population correlations were specified for each variable pair based on agreement among the authors (see examples in Figure 3). Our assumptions about the correlations of these variables may not reflect the ground truth relationship, and may differ from participants' beliefs. However, because we measure each individual's prior beliefs, we can assess whether belief updating was affected by any mismatch between their prior and the sample correlation.

5.1.2   Elicitation, attention check procedures, and collected data

Each trial consisted of a prior elicitation, correlation visualization, and posterior elicitation. For both elicitation steps we used the Line + Cone method validated in Study 1 (Figure 1A). Each elicitation resulted in three measurements: the most likely correlation (μ) and the lower and upper bounds of the uncertainty cone (b_{lower}, b_{upper}). All three values were bounded between $\rho = -1$ and $\rho = +1$.

We designed practice questions to familiarize participants with the Line + Cone elicitation. Participants answered test questions to ensure that they understood how to interpret the elicitation interface, including the direction of a correlation and the degree of uncertainty captured with the cone. We also included attention check questions (same as in Study 1) to screen inattentive respondents or other invalid data [5].

In Study 2 and 3, we also collected basic demographic data, duration of each trial, and the error count of users in the instructions section.

5.2   Participants

Participants were recruited from Amazon Mechanical Turk. For all studies we required that participants were located in the U.S. and had a 95% or above approval rating. Participants earned $1.80 upon completion of the task, which took an average of 25.7 minutes ($SD = 14.6$) to complete. Per our pre-registration, we excluded any participants due to: failed attention check questions ($n = 35$); technical errors ($n = 15$); or task completion in less than 5 minutes ($n = 38$). This left $n = 212$ participants for the analysis (Line: 74; Cone: 64; HOP: 74).

5.3   Results

For the analysis we built three mixed effects models using R's `lme4` package for two linear regressions and R's `glmmTMB` for a beta regression. We used the normal approximation to calculate p-values of fixed effects using t-values produced by `lme4`.[5]

Dependent & Independent Variables: We considered three dependent variables (DV): (1) the absolute belief difference, (2) the difference in uncertainty, and (3) belief distance from the model's predicted posterior mean. For our independent variables (IV), we included the Visualization treatment (Line, Cone, HOP, and Scatter) and the absolute correlation of the generated data for the variable sets (see Figure 3)

Model Specification: For each model, we included the visualization treatment and the absolute correlation of the data as fixed effects. For the visualization treatment, the Scatter condition is the omitted reference condition. We treated the sample correlation as a categorical variable and used zero absolute correlation as the omitted reference condition. We included the unique variable set and the participant id as random effects.

5.3.1   Beliefs about variable pairs

We first examined participants' beliefs before and after experiencing the data visualization. Figure 3 displays pre- and post-treatment judgments about the most likely correlation (μ, top row) and uncertainty (CI, bottom row) for five of the ten variable pairs, aggregated across visualization treatments. With respect to the mean correlation μ, prior judgments (green density plots) were largely consistent with the relationship that was designated for each variable pair, such that the modal prior belief was close to the sample correlation. This suggests that the datasets presented were congruent with most participants' prior belief about the relationship between the variables. One notable exception was *Income tax rate X Poverty rate*, where the designated correlation

---

[5]The code used is included in our supplemental materials.
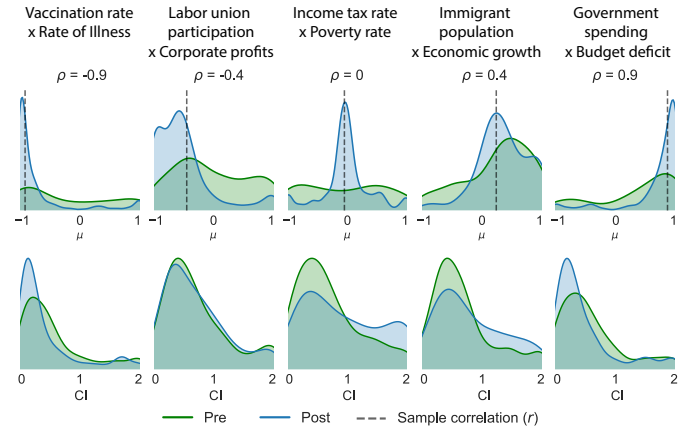

Fig. 3. Density plots of means (top row) and CIs (bottom row) of elicited belief distributions for selected variable sets in Study 2. Dashed lines indicate the sample correlation of the dataset presented to participants.

was $\rho = 0$ but prior beliefs were relatively uniformly distributed from -1 to +1. Post-treatment beliefs about the same variable sets (blue density plots) strongly shifted toward the sample correlation of the observed dataset (dashed lines) for all variable sets. The plots for the CIs reveal that the strength of the sample correlation also affected changes in uncertainty. CIs decreased after seeing strongly correlated datasets ($\rho = \pm 0.9$) but in some cases increased following data visualizations with weaker relationships. We report more detailed analysis of how the uncertainty changed in different treatments in section 5.3.3.

5.3.2   Change in beliefs about most likely relationship

We used linear mixed effects regression to model the effect of visualization conditions and population correlation on the absolute change in beliefs about the most likely correlation ($|\mu_{post} - \mu_{pre}|$. There were no significant effects (Figure 4, left), though the Cone condition showed marginally smaller changes in beliefs compared to the Scatter condition ($\beta = -0.07$ $[-0.14, 0.01]$, $z = -1.779$, $p = 0.075$). Thus, while participants clearly shifted their beliefs about the most likely correlation in response to observed datasets (Figure Figure 3), contrary to our expectations we did not find that the degree of belief change differed by visualization treatment or population correlation.

5.3.3   Change in uncertainty

Mixed effects linear regression was used to model the effects of visualization condition and population correlation on the change in uncertainty ($|CI_{post} - CI_{pre}|$). As shown in Figure 4 right, relative to the Scatter condition, the Cone condition exhibited greater reduction in uncertainty ($\beta = -0.10$ $[-0.16, -0.03]$, $z = -2.782$, $p < .01$). In other
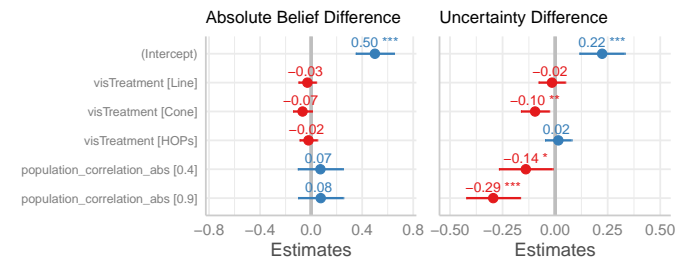

Fig. 4. Study 2 fixed effects coefficients for absolute belief difference (left) and uncertainty difference (right). Error bars indicate 95% confidence intervals. Asterisks indicate statistical significance than zero using p-values: *** 99.9%, ** 99%, * 95%. For visTreatment, the reference category is the Scatter condition.

words, participants assigned to the cone Condition felt less uncertain (more confident) with their input. There was no difference in the Line ($\beta = -0.02$ [$-0.08, 0.05$], $z = -0.468$, $p = 0.640$) or HOP condition ($\beta = 0.02$ [$-0.05, 0.08$], $z = 0.473$, $p = 0.636$). In addition, more extreme sample correlations had a greater impact on belief change: Compared to $\rho = 0.0$, there was a greater reduction in uncertainty for $\rho = .4$ (($\beta = -0.14$ [$-0.27, -0.01$], $z = -2.138$, $p < .05$) and $\rho = .9$ ($\beta = -0.29$ [$-0.42, -0.17$], $z = -4.513$, $p < .001$) variable sets.

5.3.4 Accuracy of posterior beliefs

We examined the accuracy of participants' posterior mean (μ_{post}) compared to the sample correlation of the observed datasets. In the Scatter condition, posterior means were biased to be more extreme for moderately positive and negative sample correlations. Relative to the $\rho = 0$ variable sets, absolute error was higher for $\rho = \pm 0.4$ variable sets ($\beta = .29$ [$.19, .41$], $z = 5.46$, $p < .001$) but did not differ from $\rho = \pm 0.9$ variable sets ($\beta = .002$ [$-.11, .11$], $z = .03$, $p = .96$) The remaining visualization conditions led to more accurate beliefs across the full range of sample correlations. Compared to the Scatter condition, the absolute error was lower in all three visualization conditions (Line: $\beta = -.22$ [$-.34, -.10$], $z = -3.50$, $p < .001$; Cone: $\beta = -.34$ [$-.47, -.21$], $z = -4.96$, $p < .001$; HOP: $\beta = -.25$ [$-.38, -.13$], $z = -3.97$, $p < .001$).

5.4 Bayesian belief updating model
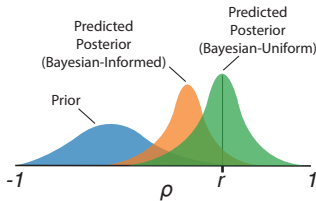
In this section we use Bayesian cognitive modeling to investigate the influence of prior beliefs on the belief updating process. Under the principles of Bayesian inference, people should integrate new evidence about a correlation with their prior beliefs about that relationship. Bayesian models provide a normative benchmark for how beliefs *should* change depending on the strength of the evidence and participants' uncertainty. For instance, a person who is confident that variables are negatively correlated may only shift their beliefs a small amount after seeing a dataset with a positive sample correlation. A second person who is highly uncertain about the relationship, however, may be more strongly influenced by the same data and report posterior beliefs that are closely matched to the sample correlation. This framework also allows us to identify when people systematically fail to adjust their beliefs as predicted by the Bayesian model. Returning to the main hypothesis of Study 2, if uncertainty representations cause smaller adjustments to beliefs, this will correspond to larger divergence between participants' elicited posterior beliefs and the predictions of the Bayesian model compared to other conditions.

Having elicited prior beliefs about each set of variables, we examined whether participants' posterior beliefs (following the data visualization) could be predicted by a normative Bayesian model. The model uses Bayesian inference to predict a posterior belief distributions over possible population correlations, ρ, based on an observed dataset and a particular prior (see [33] for similar model formulation). We evaluated two variants of the model that differed only in their prior. The **Bayesian-Informed** model relied on the participant's elicited prior to calculate the normative posterior distribution after observing a dataset. The prior belief was modeled as a bounded Normal distribution, $\rho \sim BoundedNormal(\mu_{pre}, \sigma_{pre}, [-1, 1])$, where μ_{pre} and σ_{pre} are the mean and standard deviation of the participant's elicited prior. The observed bivariate data X was modeled as having been generated from a standardized multivariate Normal distribution with mean of zero and standard deviation of 1 on each dimension (see [33]),

$$X \sim MultivariateNormal(\begin{bmatrix} 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}^{-1}). \quad (1)$$

Under the **Bayesian-Uniform** model, the prior was a uniform distribution over the correlation coefficient, $\rho \sim Uniform(-1, 1)$. The mean
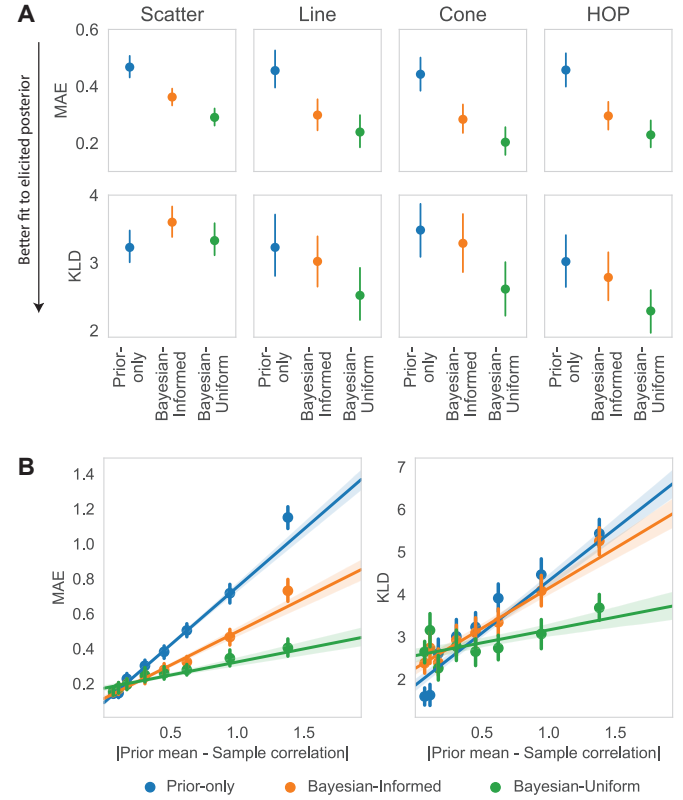

Fig. 5. **A:** MAE and KLD between elicited posterior and predictions of Prior-only, Bayesian-Informed, and Bayesian-Uniform models. **B:** Model performance as a function of the absolute distance between the elicited prior mean and the sample correlation.

and 95% CI of the posterior distribution for this model is equivalent to the values used for the visualizations in the Line, Cone, and HOP treatments. Predicted posterior distributions for ρ were estimated for both models using MCMC with the PyMC3 library [51] with two chains of 20,000 samples and 1000 burn-in iterations. Lastly, we compared the elicited posteriors to the elicited priors, absent any belief updating. We refer to this baseline as the **Prior-only** model in the results below.

The relative fit of the models reflects the weight of prior beliefs in the updating process, with the Bayesian-Informed model representing the normative integration of priors with new evidence. If people relied only on the visualization without accounting for prior beliefs, their elicited posteriors should be best fit by the Bayesian-Uniform model. In contrast, if they did not adjust beliefs upon observing a dataset, the Prior-only model should provide the closest match to posterior beliefs.

5.4.1 Model comparison

Following [29] we evaluated each model's performance with two metrics: mean absolute error (MAE) between the predicted and elicited posterior means; and Kullback-Liebler distance (KLD) (Figure 5A). These measures are complementary in that MAE captures the magnitude of differences in beliefs independently of the amount of uncertainty, while KLD measures correspondence across the entire belief distributions. We used mixed effects linear regression to compare MAE and KLD with model and visualization type as fixed effects and random effects for participants and variable sets.

In terms of MAE there were significant effects of visualization treatment ($\chi^2(1, 3) = 33.17$, $p < .001$) and model ($\chi^2(1, 2) = 333.97$, $p < .001$), but no interaction ($\chi^2(1, 6) = 7.55$, $p = .27$). Pairwise comparisons indicated that MAE was lower under both the Bayesian-Informed and Bayesian-Uniform models than the Prior-only model in all four visualization treatments (all $p < .001$). The Bayesian-Uniform model achieved lower MAE than the Bayesian-Informed model in the

Scatter ($z = -4.52$, $p < .001$), Cone ($z = -2.8$, $p = .01$), and HOP ($z = -2.48$, $p = .04$) conditions, but the two did not differ in the Line condition ($z = -2.22$, $p = .07$). Comparing the best-fitting Bayesian-Uniform model across visualization treatments showed that MAE was higher in the Scatter group than the Cone ($z = 3.14$, $p = .01$) and Line groups ($z = 2.94$, $p = .02$), but not significantly different from the HOP group ($z = 2.54$, $p = .05$).

For KLD there were significant effects of visualization treatment ($\chi^2(1,3) = 67.57$, $p < .001$), model ($\chi^2(1,2) = 31.64$, $p < .001$), and model $\times$ treatment interaction ($\chi^2(1,6) = 32.49$, $p < .001$). In the Scatter condition, KLD of the Prior-only model was lower than the Bayesian-Informed model ($z = -3.32$, $p = .003$), but did not differ from the Bayesian-Uniform conditions ($z = -.87$, $p = .66$). This indicates that the Bayesian model was relatively unsuccessful at predicting the posterior distribution in the Scatter condition, failing to outperform the baseline Prior-only model. In the remaining conditions (Line, Cone, HOP), the Bayesian-Uniform model had lower KLD than both the Prior-only and Bayesian-Informed models (all $p < .022$). As was the case for MAE, the KLD of the Bayesian-Uniform model was higher in the Scatter condition than the other conditions (all $p < .001$), but did not differ among the Line, Cone, and HOP groups. This supports the earlier finding that the accuracy of posterior beliefs was poorer in the Scatter condition compared to the other treatments.

The predictions of the three models diverge most when there is a discrepancy between participants' priors and the sample correlation of the observed dataset. We therefore explored how the fit of each model depended on the absolute distance between the prior mean and the sample correlation (Figure 5B). At small distances the three models have comparable MAE and KLD, while the advantage for the Bayesian-Uniform model grows with increasing distance between the prior and sample correlation. The poorer fit of the Bayesian-Informed model indicates that participants discounted their priors when they observed a dataset with a drastically different correlation. Notably, at small distances KLD was lowest for the Prior-only model. This result suggests that when people observed a dataset that was consistent with their prior, they were less likely to update their beliefs as predicted by either Bayesian model.

5.5 Discussion of Study 2

Results of the regression analysis and cognitive modeling showed that visualizations with representations of the population correlation (Line, Cone, and HOPs) led to greater accuracy in posterior beliefs compared to the Scatter condition. In addition, higher correlations led to larger reductions in uncertainty, potentially because stronger relationships are easier to detect in scatterplots [46,47] and are associated with less uncertainty in the population correlation. We found initial evidence for this updating process using the Bayesian cognitive model, showing that when the sample correlation presented to participants was far from their prior mean, they strongly adjusted their beliefs to reflect the pattern in the data (Figure 5B).

We did not find support for our main hypothesis that uncertainty visualizations would be associated with smaller changes in beliefs. On the contrary, the Cone visualization (with a cone of "plausible alternatives" representing uncertainty in the correlation based on the data) led to greater reductions in uncertainty. This result suggests that the explicit representation of uncertainty provided by the Cone visualization leads to greater confidence about the true relationship compared to the other visualization types. Interestingly, we did not find a similar effect on uncertainty change in the HOP condition, possibly due to the transient nature of the animated uncertainty cone.

There were two shortcomings of the present study that may have limited our ability to detect differences in belief updating between conditions. First, participant's prior beliefs largely aligned with the sample correlation, leading to many cases with little room for participants to adjust their beliefs. Second, the relatively large sample size of the datasets ($n = 100$) meant there was relatively little uncertainty about the population correlation. This may explain why the Bayesian-Uniform model provided the best fit to elicited posteriors, such that the sample correlation had a stronger influence than individuals' priors. **Study 3** was designed to further explore how these factors affect belief change.
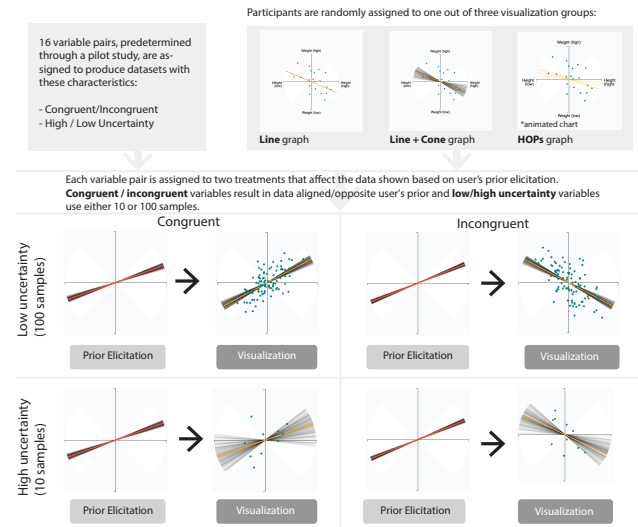

Fig. 6. Study 3 design. Like Study 2, users elicit their beliefs about correlations of variable pairs before and after seeing data visualizations. Users are randomly assigned to Line, Cone, and HOP visualization treatments. The datasets are generated based on users' prior elicitation as either congruent/incongruent and 10 or 100 data points.

We manipulated the data provided to be either congruent or incongruent with the user's elicited prior belief. In addition, we manipulated the amount of data uncertainty by varying the sample size.

6 STUDY 3: HOW CORRELATION CONGRUENCE AND UNCERTAINTY AFFECT BELIEF UPDATING

The hypothesis of Study 2 was that people would exhibit less belief change when they experienced visualizations with representations of uncertainty. The main hypothesis for Study 3[6] extends this further to predict that viewers of uncertainty representations would exhibit smaller changes in beliefs when correlation visualizations are *incongruent* with users' prior belief and when the dataset has a smaller sample size.

6.1 Study Design

For Study 3, we extended the design of Study 2 by explicitly manipulating the congruence of the sample correlation (**factor 1**) with a user's prior belief and the amount of uncertainty (**factor 2**). Both above factors are within-subjects while the visualization treatment remains a between-subject factor. Figure 6 summarizes the design of Study 3. For each variable pair, participants saw datasets that were either congruent or incongruent to their prior beliefs:

- **Congruent** datasets: Random samples were drawn from a multivariate normal distribution with correlation *0.25* away from the prior mean. For example, if a participant's prior mean was 0.85, the data was sampled from a distribution with population correlation of 0.6 (0.85 − 0.25). In this condition a user always saw sample correlations with the same sign as their prior belief.

- **Incongruent** datasets: Random samples from a multivariate normal distribution with correlation value that is *1.0* away from the prior mean. For example, if the prior mean was 0.6, the data was sampled from a distribution with a population correlation of *-0.4*. In this condition, participants saw datasets with the opposite correlation sign from their prior belief.

We also manipulated the number of samples in the datasets for specific variable pairs (10 points vs. 100 points). Datasets with 10 points result in greater uncertainty as measured by the 95% confidence interval. As in Study 2, participants were randomly assigned to visualization

---

[6]http://aspredicted.org/blind.php?x=x7ph2u

conditions of Line, Line + Cone and HOPs. Given Study 2's results that users achieved better accuracy with all three visualization types, we omitted the Scatter condition.

6.1.1 Datasets, elicitation, and attention check procedures

For Study 3 we selected variable pairs from the results of a pilot study. With 50 pilot participants, we elicited prior belief and uncertainty about 30 variable pair candidates, then categorized variables into a 2 X 2 grid of high/low social consensus on correlation and uncertainty. [7] With lessons learned on users' beliefs about the variable pairs from Study 2 (section 5.3.1), we aimed to select pairs that cover a range of distributions of beliefs about the mean correlation and uncertainty. We selected four variables with either high / low correlation consensus and high / low uncertainty. Study 3 used the same elicitation process, instructions, and attention checks as Study 2.

6.2 Participants

Participants were recruited from Amazon Mechanical Turk. Participants earned $1.80 upon completion of the task, which took an average of 22.9 minutes ($SD = 12.28$) to complete. Per our pre-registration, we excluded any participants who: failed attention check questions ($n = 12$); technical errors ($n = 95$); or completed the entire task in less than 5 minutes ($n = 11$). This left $n = 267$ participants for the analysis (Line: 89; Cone: 92; HOP: 86).

6.3 Results
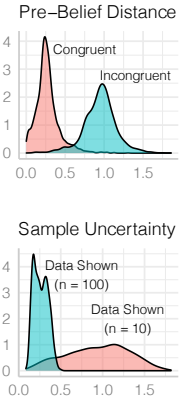
Dependent & Independent Variables: Similar to Study 2, we considered three dependent variables: (1) the absolute belief difference, (2) the difference in uncertainty, and (3) the user's belief distance from the model's predicted posterior. For our independent variables (IV), we created two features based on our variable conditions from Figure 6. First, we defined **pre-belief distance** as the distance between users' prior elicitation and the correlation of the provided sample, which is larger when a participant is provided incongruent datasets. Next, we defined **sample uncertainty** as the size of uncertainty shown to users resulting from the sample size. In doing so, we used continuous IVs ranging from zero to two rather than binary variables. For reference, we provide kernel density plots (left) for the two IV's partitioned by its respective condition categories.

Model Specification: We employed three mixed effects models as in Study 2 (see Section 5.3). For each model, we included the interaction terms between the visualization treatment, the pre-belief distance, and the sample uncertainty as fixed effects. For the visualization treatment, the Line condition is the omitted reference condition.

6.3.1 Change in belief about most likely relationship

For absolute belief difference (Figure 7, left), we found the largest effect to be pre-belief distance ($\beta = 0.73[0.64, 0.81], z = 2.40, p < .001$), indicating that users updated their beliefs more when they viewed incongruent datasets.

There were significant interactions between pre-belief distance and visualization type, such that there were smaller belief changes when the data was incongruent in both the Cone ($\beta = -0.12\ [-0.19, -0.05],\ z = -3.3,\ p < .001$) and HOPs ($\beta = -0.11\ [-0.18, -0.04],\ z = -3.16,\ p < .01$) conditions relative to the Line condition. This finding is in line with our hypothesis that in the incongruent condition, users would show smaller update in their belief when uncertainty representations are present.

Finally, while the HOP condition led to slightly larger changes compared to the Line condition ($\beta = 0.10\ [0.02, 0.17],\ z = 2.400,\ p <$

---

[7]Social consensus was measured as the standard deviation of prior means, while average uncertainty was measured as the mean CI.
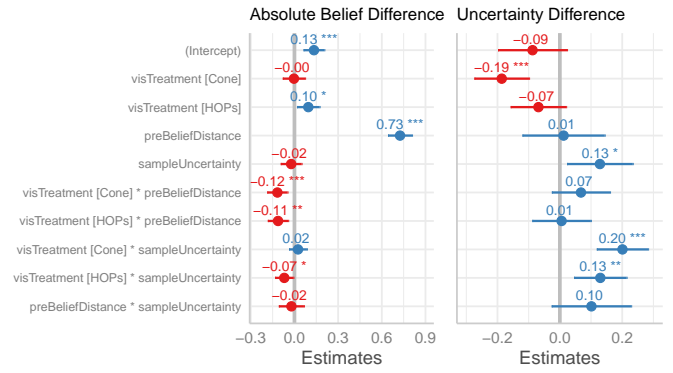

Fig. 7. Study 3 fixed effects coefficients from analyzing absolute belief difference (left) and uncertainty difference (right). The error bars indicate 95% confidence intervals. Asterisks indicate statistical significance than zero using p-values: *** 99.9%, ** 99%, * 95%. For visTreatment, the reference category is the Line condition.

0.05), this condition had a negative interaction with sample uncertainty such that beliefs shifted less after seeing smaller datasets ($\beta = -0.07\ [-0.13, -0.01],\ z = 2.181,\ p < .05$). We did not find corresponding effects for the Cone condition. This difference between the Cone and HOP visualizations might suggest that uncertainty is more evident in larger uncertainty amounts when using the HOP technique. This is potentially due to the lack of a fixed representation of most likely correlation in the HOP technique as opposed to the Cone technique.

6.3.2 Uncertainty change

In our regression of the uncertainty difference (Figure 7, right), we found that users in the Cone condition exhibited more reduction in uncertainty than the Line condition ($\beta = -0.19\ [-0.27, -0.10],\ z = -4.166,\ p < .001$), replicating the effect seen in Study 2. There was not a significant effect in the HOPs condition ($\beta = -0.07\ [-0.16, 0.02],\ z = -1.514,\ p = 0.130$).

Pre-belief distance had no effect on the uncertainty difference in any condition. However, sample uncertainty had a positive effect on changes in uncertainty ($\beta = 0.13\ [0.02, 0.23],\ z = 2.391,\ p < .05$). We also found that the Cone visualization condition had larger effects on the uncertainty difference when interacting with sample uncertainty ($\beta = 0.20\ [0.12, 0.28],\ z = 4.784,\ p < .001$). The HOPs condition also showed a positive interaction with uncertainty difference when interacting with datasets with larger sample uncertainty ($\beta = 0.13\ [0.05, 0.21],\ z = 3.019,\ p < .01$). These findings suggest that participants in the Cone condition showed more overall reduction in posterior uncertainty compared to the Line treatment but the HOP condition did not show similar effects. Interestingly, when dealing with larger uncertainty (10 data points), the presence of an uncertainty representation resulted in an increase in users' uncertainty. This finding suggests that both visualization techniques convey uncertainty when uncertainty amounts are larger, but users' experience of the HOP condition is similar to the Line condition when dealing with datasets with smaller uncertainty. Perhaps this is due to users' inability to perceive small angular movements of the line.

6.3.3 Accuracy of posterior beliefs

We used beta regression to model the effects on the distance of users' posterior beliefs from the true sample correlation. We found that pre-belief distance had the largest positive effect on users' post-belief distance ($\beta = .34\ [0.16, 0.52]\ z = 17.254,\ p < .01$). In other words, posterior beliefs were less similar to the sample correlation when the dataset was incongruent with users' prior beliefs. We also found that compared to the Line condition, the HOP condition had a positive effect on posterior distance when viewing a dataset with more uncertainty ($\beta = 0.29\ [0.14, 0.45],\ z = 2.181,\ p < .01$). This might be due to
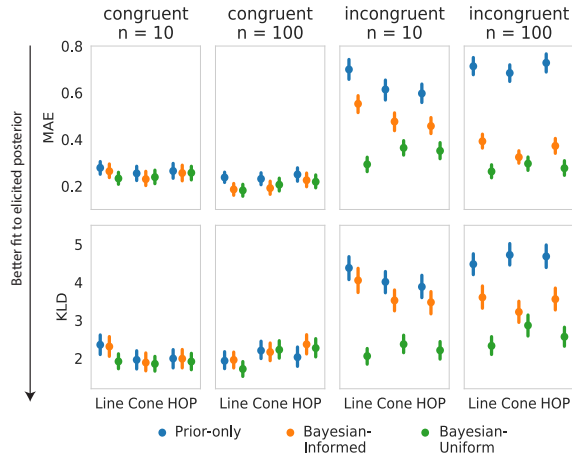
Fig. 8. MAE and KLD by model for Study 3.

the lack of a fixed most-likely correlation representation in the HOPs condition, therefore when sample uncertainty is larger, users are more prone to larger distances (errors) in their judgements.

6.4 Bayesian belief updating model

We used the models from Study 2 to examine how prior beliefs influenced belief updating in Study 3. In general, the best fit to elicited posteriors in terms of both MAE and KLD was achieved by the Bayesian-Uniform model in all conditions (Figure 8). Incongruent trials provide a strong comparison of the Bayesian-Informed and Bayesian-Uniform models because they involve datasets that conflict with participants' prior beliefs. If people integrate new evidence with their elicited prior, they should show smaller shifts in beliefs in Incongruent trials than expected under the Bayesian-Uniform model. However, as was seen in Study 2, posterior distributions were best-fit by the Bayesian-Uniform model, suggesting a stronger influence of the data visualization on posterior beliefs. Notably, the only condition in which the two models performed comparably on Incongruent trials was the Cone treatment, where there were no differences in MAE ($t(722.63) = 1.33$, $p = 0.18$) or KLD ($t(731.98) = 1.78$, $p = 0.08$), indicating that Cone visualizations produced belief updates that more closely aligned with the normative prediction of the Bayesian-Informed model.

6.5 Discussion of Study 3

We predicted that people exposed to uncertainty visualizations (Cone and HOP conditions) would exhibit less belief change compared to those without uncertainty (Scatter and Line conditions). We found strong support for this hypothesis in Study 3 when participants saw data that was incongruent with their prior beliefs. Both the Cone and HOP treatments were associated with smaller belief updates compared to the Line condition which did not represent uncertainty about the correlation. Uncertainty visualizations also affected whether there were shifts in participants' degree of uncertainty. Relative to the Line condition, Cone visualizations led to greater reductions in uncertainty for large datasets, whereas uncertainty did not change when datasets were small. Similar (albeit weaker) effects were present for HOP visualizations.

Finally, we replicated the modeling results from Study 2, showing that posterior beliefs were best-fit by the predictions of the Bayesian-Uniform model. Although this does not imply that participants completely disregarded their prior beliefs, it indicates that the data visualizations tended to have a stronger influence on posterior beliefs than expected from a normative Bayesian perspective. The Cone visualization was the only condition in which the Bayesian-Informed model performed comparably to the Bayesian-Uniform model. This result suggests an alternative interpretation of the smaller degree of belief updating in that condition when faced with incongruent data. Rather

than representing an irrational failure to modify beliefs akin to confirmation bias, the Cone condition may be most effective for striking the appropriate balance between new data and prior beliefs.

7 DISCUSSION, FUTURE WORK, AND CONCLUSION

In this paper, we study the effect of prior belief and uncertainty representations on correlation judgement. In Study 1 we developed the Line + Cone method for eliciting people's beliefs about the correlation between two variables, including their degree of uncertainty. The Line + Cone method serves as a good choice for eliciting users' beliefs about bivariate relationships for future studies of correlation judgement. In addition to capturing users' beliefs about the correlation means (commonly done in previous correlation judgement studies), results from all three studies demonstrate that it is also important to capture users' uncertainties about their judgements. In Studies 2 and 3, we used the Line + Cone method to investigate belief updating in the context of data visualization. We found that visualization conditions with uncertainty communication led to less belief updating compared to visualizations without uncertainty, especially when the presented correlation visualization is incongruent with users' prior beliefs. An important conclusion is that judgements are affected by the existence of uncertainty depictions. How we encode uncertainty (e.g., Cone vs. HOPs), also affects users' belief and uncertainty change. As the visualization community pays more attention to the importance of uncertainty representations and elicitation, it is important to be cognizant to the affects of such techniques on users' judgements.

In our studies we applied a Bayesian cognition framework to understand how people update their beliefs about bivariate correlations with different types of visualizations. Recent studies have applied insights from Bayesian cognitive modeling to understand how people integrate new data with their existing knowledge [32, 35]. The Bayesian framework provides normative benchmarks that can be used to evaluate whether people optimally revise their beliefs given their existing uncertainty and the strength of new evidence conveyed through a visualization [29]. We used Bayesian models to compare participants' posteriors to three benchmarks: no change in beliefs (Prior-only model); the normative posterior when taking into account the elicited prior (Bayesian-Informed model); and the normative posterior when disregarding the prior (Bayesian-Uniform model). In both Studies 2 and 3, elicited posterior distributions were best-described by the Bayesian-Uniform model, suggesting that the characteristics of the visualized dataset had a stronger influence on posterior beliefs than expected under the Bayesian-Informed model.

There are several possible explanations for why posterior beliefs appeared to underweight participants' priors. One possibility is that people have a different interpretation of the cone representation which is used to elicit their uncertainty. In order to minimize demands on numerical or probabilistic reasoning, participants were simply instructed to adjust the cone to capture the range of "plausible alternatives" for the correlation between the variables. In Study 1 we found support for the claim that this method captures participants' uncertainty, but there may nevertheless be a mismatch between the elicited distribution and participants' subjective beliefs such that people are more uncertain than indicated by their elicited priors.

We found other evidence that people updated beliefs in a way consistent with Bayesian inference. In Study 2, users reduced their uncertainty to a greater extent for more extreme sample correlations. In Study 3, uncertainty increased when people saw small datasets ($n = 10$) compared to large datasets ($n = 100$), even in the Line condition which lacked an explicit representation of the correlation uncertainty. Participants also expressed greater uncertainty in the posterior beliefs than predicted by the Bayesian models, echoing the findings of Kim *et al.* [29].

These studies provide the groundwork for investigating how people interpret data that is relevant to strongly-held or favored beliefs. Prior beliefs can distort the perception of new evidence, as is seen in widespread evidence of confirmation bias [16, 41, 42]. Using intuitive, visual belief elicitation methods in conjunction with Bayesian cognitive models offer a promising path toward understanding the causes of such biases in data visualization.

REFERENCES

[1] F. J. Anscombe. Graphs in statistical analysis. *The american statistician*, 27(1):17–21, 1973.

[2] H. Baumgartner. On the utility of consumers' theories in judgments of covariation. *Journal of Consumer Research*, 21(4):634–643, 1995.

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

[4] C. Chinn and W. Brewer. The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of educational research*, 63(1):1, 1993.

[5] M. Chmielewski and S. C. Kucker. An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science*, p. 1948550619875149, 2019.

[6] R. T. Clemen, G. W. Fischer, and R. L. Winkler. Assessing dependence: Some experimental results. *Management Science*, 46(8):1100–1115, 2000.

[7] J. S. B. Evans. In two minds: dual-process accounts of reasoning. *Trends in cognitive sciences*, 7(10):454–459, 2003.

[8] J. S. B. Evans, J. L. Barston, and P. Pollard. On the conflict between logic and belief in syllogistic reasoning. *Memory & cognition*, 11(3):295–306, 1983.

[9] J. S. B. Evans and J. Curtis-Holmes. Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning*, 11(4):382–389, 2005.

[10] M. Fernandes, L. Walls, S. Munson, J. Hullman, and M. Kay. Uncertainty displays using quantile dotplots or cdfs improve transit decision-making. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 144. ACM, 2018.

[11] P. H. Garthwaite, J. B. Kadane, and A. O'Hagan. Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, 100(470):680–701, 2005.

[12] T. M. Green, W. Ribarsky, and B. Fisher. Building and applying a human cognition model for visual analytics. *Information visualization*, 8(1):1–13, 2009.

[13] M. Greis, J. Hullman, M. Correll, M. Kay, and O. Shaer. Designing for uncertainty in hci: When does uncertainty help? In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 593–600. ACM, 2017.

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

[15] 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.

[16] R. J. Heuer. *Psychology of intelligence analysis*. Center for the Study of Intelligence, 1999.

[17] J. Hullman. Why evaluating uncertainty visualization is error prone. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, pp. 143–151, 2016.

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

[19] J. Hullman, M. Kay, Y.-S. Kim, and S. Shrestha. Imagining replications: Graphical prediction & discrete visualizations improve recall & estimation of effect uncertainty. *IEEE transactions on visualization and computer graphics*, 24(1):446–456, 2017.

[20] J. Hullman, X. Qiao, M. Correll, A. Kale, and M. Kay. In pursuit of error: A survey of uncertainty visualization evaluation. *IEEE transactions on visualization and computer graphics*, 2018.

[21] 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), 2015.

[22] D. Jennings, T. M. Amabile, and L. Ross. Judgment under uncertainty: Heuristics and biases. chap. Informal covariation assessment: Data-based vs. theory-based judgments. Cambridge University Press, 1982.

[23] S. R. Johnson, G. A. Tomlinson, G. A. Hawker, J. T. Granton, and B. M. Feldman. Methods to elicit beliefs for bayesian priors: a systematic review. *Journal of clinical epidemiology*, 63(4):355–369, 2010.

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

[25] 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*, 25(1):892–902, 2018.

[26] 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.

[27] 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*, pp. 5092–5103, 2016.

[28] 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*, pp. 1375–1386. ACM, 2017.

[29] Y.-S. Kim, L. A. Walls, P. Krafft, and J. Hullman. A bayesian cognition approach to improve data visualization. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2019.

[30] P. A. Klaczynski. Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: A two-process approach to adolescent cognition. *Child Development*, 71(5):1347–1366, 2000.

[31] B. C. P. Kraan. Probabilistic inversion in uncertainty analysis: and related topics. 2002.

[32] T. L Griffiths, C. Kemp, and J. B Tenenbaum. Bayesian models of cognition. 2008.

[33] M. D. Lee and E.-J. Wagenmakers. *Bayesian cognitive modeling: A practical course*. Cambridge university press, 2014.

[34] J. Li, J.-B. Martens, and J. J. Van Wijk. Judging correlation from scatterplots and parallel coordinate plots. *Information Visualization*, 9(1):13–30, 2010.

[35] F. Lieder and T. L. Griffiths. Resource-rational analysis: understanding human cognition as the optimal use of limited. *Psychological Science*, 2(6):396–408, 2018.

[36] L. Liu, A. P. Boone, I. T. Ruginski, L. Padilla, M. Hegarty, S. H. Creem-Regehr, W. B. Thompson, C. Yuksel, and D. H. House. Uncertainty visualization by representative sampling from prediction ensembles. *IEEE transactions on visualization and computer graphics*, 23(9):2165–2178, 2016.

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

[38] A. M. MacEachren, R. E. Roth, J. O'Brien, B. Li, D. Swingley, and M. Gahegan. Visual semiotics & uncertainty visualization: An empirical study. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2496–2505, 2012.

[39] J. Meyer, M. Taieb, and I. Flascher. Correlation estimates as perceptual judgments. *Journal of Experimental Psychology: Applied*, 3(1):3, 1997.

[40] L. Micallef, P. Dragicevic, and J.-D. Fekete. Assessing the effect of visualizations on bayesian reasoning through crowdsourcing. *IEEE transactions on visualization and computer graphics*, 18(12):2536–2545, 2012.

[41] 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.

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

[43] A. O'Hagan, C. E. Buck, A. Daneshkhah, J. R. Eiser, P. H. Garthwaite, D. J. Jenkinson, J. E. Oakley, and T. Rakow. *Uncertain judgements: eliciting experts' probabilities*. John Wiley & Sons, 2006.

[44] A. Ottley, E. M. Peck, L. T. Harrison, D. Afergan, C. Ziemkiewicz, H. A. Taylor, P. K. Han, and R. Chang. Improving bayesian reasoning: The effects of phrasing, visualization, and spatial ability. *IEEE transactions on visualization and computer graphics*, 22(1):529–538, 2015.

[45] R. E. Patterson, L. M. Blaha, G. G. Grinstein, K. K. Liggett, D. E. Kaveney, K. C. Sheldon, P. R. Havig, and J. A. Moore. A human cognition framework for information visualization. *Computers & Graphics*, 42:42–58, 2014.

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

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

[48] G. O. Roberts and J. S. Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16(4):351–367, Nov. 2001. doi: 10.1214/ss/1015346320

[49] I. T. Ruginski, A. P. Boone, L. M. Padilla, L. Liu, N. Heydari, H. S. Kramer, M. Hegarty, W. B. Thompson, D. H. House, and S. H. Creem-Regehr. Non-expert interpretations of hurricane forecast uncertainty visualizations. *Spatial Cognition & Computation*, 16(2):154–172, 2016.

[50] W. C. Sá, R. F. West, and K. E. Stanovich. The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. *Journal of educational psychology*, 91(3):497, 1999.

[51] J. Salvatier, T. V. Wiecki, and C. Fonnesbeck. Probabilistic programming in python using pymc3. *PeerJ Computer Science*, 2:e55, 2016.

[52] A. Sanborn and T. L. Griffiths. Markov chain monte carlo with people. In *Advances in neural information processing systems*, pp. 1265–1272, 2008.

[53] A. N. Sanborn, T. L. Griffiths, and R. M. Shiffrin. Uncovering mental representations with markov chain monte carlo. *Cognitive psychology*, 60(2):63–106, 2010.

[54] M. Sedlmair, T. Munzner, and M. Tory. Empirical guidance on scatterplot and dimension reduction technique choices. *IEEE transactions on visualization and computer graphics*, 19(12):2634–2643, 2013.

[55] B. Tversky. Visuospatial reasoning. *The Cambridge handbook of thinking and reasoning*, pp. 209–240, 2005.

[56] Y. Wu, L. Xu, R. Chang, and E. Wu. Towards a bayesian model of data visualization cognition, 2017.

[57] 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.

[58] M. Zondervan-Zwijnenburg, W. van de Schoot-Hubeek, K. Lek, H. Hoijtink, and R. van de Schoot. Application and evaluation of an expert judgment elicitation procedure for correlations. *Frontiers in psychology*, 8:90, 2017.