The Impact of Elicitation and Contrasting Narratives on Engagement, Recall and Attitude Change with News Articles Containing Data Visualization

Milad Rogha [ID], Subham Sah, Alireza Karduni [ID], Douglas Markant, and Wenwen Dou

✦

Abstract—News articles containing data visualizations play an important role in informing the public on issues ranging from public health to politics. Recent research on the persuasive appeal of data visualizations suggests that prior attitudes can be notoriously difficult to change. Inspired by an NYT article, we designed two experiments to evaluate the impact of elicitation and contrasting narratives on attitude change, recall, and engagement. We hypothesized that eliciting prior beliefs leads to more elaborative thinking that ultimately results in higher attitude change, better recall, and engagement. Our findings revealed that visual elicitation leads to higher engagement in terms of feelings of surprise. While there is an overall attitude change across all experiment conditions, we did not observe a significant effect of belief elicitation on attitude change. With regard to recall error, while participants in the draw trend elicitation exhibited significantly lower recall error than participants in the categorize trend condition, we found no significant difference in recall error when comparing elicitation conditions to no elicitation. In a follow-up study, we added contrasting narratives with the purpose of making the main visualization (communicating data on the focal issue) appear strikingly different. Compared to the results of study 1, we found that contrasting narratives improved engagement in terms of surprise and interest but interestingly resulted in higher recall error and no significant change in attitude. We discuss the effects of elicitation and contrasting narratives in the context of topic involvement and the strengths of temporal trends encoded in the data visualization.

Index Terms—Belief elicitation, visual elicitation, data visualization, contrasting narratives.

1 INTRODUCTION

IMAGINE one sunny morning, having a freshly brewed cup of coffee by your side and you are ready to catch on up some news. A few headlines caught your eye. You read a couple of articles, scanning over texts and occasionally data visualizations illustrating data trends. The next article (an example shown in Figure 1) spurred your interest because it asked you to draw your guess before revealing the visual and textual content.

Milad Rogha, Subham Sah, and Wenwen Dou are with the Department of Computer Science, University of North Carolina at Charlotte.
Doug Markant is a faculty at the Department of Psychological Science, University of North Carolina at Charlotte.
Alireza Karduni is is an Assistant Professor at Simon Fraser University's School of Interactive Arts and Technology.
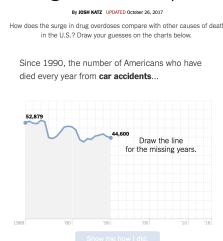

Fig. 1. "You Draw It" published in TheUpshot of New York Times [1].

When reading new articles we receive and evaluate new information. During this process individuals can fall anywhere along an "elaboration continuum", i.e. they can think a lot, a moderate amount, or very little about the information [2]. How do people absorb new information in order to inform their attitudes toward certain issues? Data visualization has played an increasingly important role in presenting new and evolving information to inform the public about data collected on high-stake issues (e.g. the drug overdose epidemic, Covid-19, the president's approval rating) [1], [3]. Given the increased use of data visualizations in news articles [4], [5], is there anything visualization researchers and designers can do to create higher engagement and "nudge" individuals to elaborate more?

More elaboration has been theorized to lead to higher likelihood of attitude change in the Elaboration Likelihood Model of Persuasion (ELM) [6]. More specifically, ELM developed by Petty and Cacioppo as a theory to explain the process involved in attitude change and how these changes maintain themselves overtime [6], captured dual routes to attitude change. The "central" and "peripheral" routes to attitude change are distinguished by the degrees of elaborative thinking [7].

While many studies have provided empirical evidence that existing attitudes are resistant to change [8]–[11], there is a research gap in the field of data visualization on whether the attitude change or the lack thereof is based on an accurate internal representation of data visualizations individuals viewed. In addition, how is attitude change impacted by user engagement, as well as factors related to the individuals such as topic personal relevance in the context of data visualizations? To address this gap, in this paper we investigate whether individuals update their attitudes after viewing a series of news articles related to a central theme, and the relationship between attitude change with recall of the data visualizations and engagement with the news articles.

Inspired by the "You Draw it: Just How Bad is the Drug

Overdose Epidemic?" article by the New York Times [1] (Figure 1), we developed and incorporated the techniques–i.e. (1) visual elicitation by drawing a trend line and (2) contrasting narratives in our experiments. Contrasting narratives show drastic different trends compared to the main charts with accompanying textual descriptions. We aim to investigate the impact of elicitation and contrasting narratives on memory, engagement, and ultimately attitude change.

Our paper makes the following contributions: (1) Inspired by the "You Draw it" article [1], we implement and compare elicitation methods – including Draw trend (Figure 4 A) and Categorize trend (a simple drop-down with fixed-choice categorical answers) to elicit users' prior knowledge on temporal trends/predictions. The Draw trend elicitation method produces a high-fidelity externalization of user's prior knowledge related to a temporal trend; data points and the overall trend were recorded by this elicitation method. To further ascertain whether any impact we observe is due to the elicitation step or the visual nature, we utilize the Categorize Trend elicitation method that asks users to choose a general characterization of the temporal trend they are asked to predict with a drop-down menu.

(2) We offer empirical evidence on the impact of using such elicitation techniques in data communication on attitude change, recall, and engagement by conducting a controlled experiment. We found overall participants showed significant attitude change regardless of elicitation conditions. Comparing across the elicitation conditions, our findings indicate that participants in the draw trend condition reported significantly higher engagement in terms of finding the article content surprising. In terms of recall error, while participants in the draw trend condition exhibited significantly lower error compared to the categorize trend condition, the difference was not significant compared to the control condition.

(3) A follow-up experiment that evaluates the addition of contrasting narratives to accent the main visualization in new articles. The followup study reused the three elicitation treatments in the study 1 and the only change was to add two contrasting narratives per article prior to the visualization shown in study 1. Findings on the impact of elicitation and on attitude change are consistent with study 1. Empirical evidence on contrasting narratives showed they improve engagement in terms of surprise and interest. Interestingly adding contrasting narratives also resulted in higher recall error across all elicitation conditions, possibly due to higher cognitive load required to recall information processed from both the main and contrasting charts.

We connect the dots from all of the above quantitative findings on the impact of elicitation and contrasting narratives in news articles with data visualization. Our discussion explores the relationship between attitude change, recall, and engagement informed by our studies.

2 RELATED WORK

2.1 Effect of elicitation

A few studies in the field of visualization explored belief elicitation as a way to identify belief change and the relationship between belief elicitation and exploration behavior and discovery. Karduni et al. evaluated a visual elicitation technique designed to elicit belief on a linear correlation and its uncertainty [12] and found that the "line+cone" method produced comparable results to the MCMC-P method [13], [14] that requires repeating the elicitation step many times until distribution convergence. The visualization

elicitation technique can be applied before and after an intervention to capture belief change and the associated uncertainty. Mahajan et al. presented VIBE - a design space for visual belief elicitation in data journalism [15]. Many existing belief-driven visualizations from popular news media were discussed within the design space and trends and opportunities were highlighted.

Other studies evaluated the relationship between elicitation and exploration behavior with visual analytics systems. Through evaluating a tool designed to elicit users' beliefs and test their beliefs against data, Koonchanok et al. found that with belief elicitation participants were more likely to attend to discrepancies between their mental models and the data but less likely to engage in exploration, ultimately resulting in fewer discoveries [16]. In a follow-up study evaluating visual belief elicitation on scatterplots, Koonchanok et al. found that visual belief elicitation on correlation led to more accurate inferences and fewer false discoveries [17].

This paper adds to the current literature on evaluating the impact of elicitation on attitude change, recall, and engagement by implementing and designing two temporal elicitation methods, Draw trend and Categorize trend.

2.2 Data visualization and belief and attitude change

Xiong et al. discovered that prior beliefs can bias interpretations of data visualizations, leading viewers to overestimate correlations between variables they believe are related, and underestimate them when they think variables are unrelated [18] . Furthermore, they found that participants exposed to contradictory trends favor the positive one, and prefer the less steep trend when both point in the same direction [19].

People can be reluctant to update their prior beliefs, especially when given a reason not to. Research suggests that people are more conservative in updating their beliefs when the data presented in scatterplots are incongruent with their prior beliefs [8], [18]. Xiong et al. discovered that when the variables are "belief-triggering," there is minimal change between the prior and post beliefs [18]. Kim et al. found that participants were less likely to update their belief when other people's expectations aligned with their own initial expectations but not with the data [20].

Few studies have been conducted to model the effect of prior belief on visualization interpretation. Kim et al. [21] utilized a Bayesian Cognition model to study belief changes in response to visualizations, finding that although people generally update their beliefs rationally, they often deviate from rationality, particularly with large sample sizes. Similarly, Karduni et al. [12] created a Bayesian Cognition model to predict belief updating with uncertainty visualizations. The model revealed no significant belief changes across various uncertainty representations, but indicated a shift in participants' confidence in their judgments. Gupta et al. [22] investigated how beliefs update with new data and the tendency to revert to original beliefs over time. They identified patterns of belief decay (reverting to initial beliefs) and belief persistence (maintaining updated beliefs) in participants.

Different from prior work on belief/attitude update with visual and uncertainty representations, the current studies focus on evaluating the impact of elicitation techniques on attitude change.

2.3 Data visualization and recall

A few studies have focused on recalling from data visualization and explored potential systematic biases in recall.

An empirical study revealed systematic memory bias in participants replicating previously seen bars or dots, with overestimation of higher values and underestimation of lower ones [23]. Additionally, when a reference was present, responses around the 50% mark showed underestimation below and overestimation above this threshold. Similarly, Ceja et al. found that the aspect ratio of the bar marks bias how the position encodings are recalled [24].

Kim et al. [20] studied the effect of presenting social information as others' expectations instead of data alone. They found that participants remember data more accurately when it matches their expectations, and tend to maintain their own expectations and distrust the data when it aligns with others' but differs from the actual data. Kim et al. also evaluated how visually eliciting forms of prior knowledge and presenting feedback on the gap between prior knowledge and the observed data impact recall and comprehension [25]. They found that participants who are prompted to reflect on their prior knowledge by predicting and self-explaining data outperform a control group in recall and comprehension when knowing little about the datasets.

Our studies build on findings from [25] but have many distinctions. Primarily, we explore the impact of elicitation on attitude change which was not studied in [25]. Moreover, the visual elicitation in [25] was to elicit a few (discreet) missing data points while we contributed a visual elicitation technique on temporal trends. In addition to evaluating the impact of elicitation, our second study focuses on evaluating the impact of contrasting narratives on recall, engagement, and attitude change.

2.4 Attitude change and Elaboration Likelihood Model

The Elaboration Likelihood Model (ELM) [6], [7] theorizes how factors like expertise, message content, recipient mood, and context affect attitude change. It outlines a dual-route process—central and peripheral—determining the degree of elaborative thinking [7]. The central route leads to lasting attitude changes resistant to counter-persuasion, whereas the peripheral route results in temporary changes. ELM also identifies three processes where no attitude change occurs despite some elaborative thinking. We use the ELM model as a schema to inform our hypothesis development and as a way to understand and discuss our experiment results.

3 HYPOTHESIS DEVELOPMENT

Two aspects that made the New York Times "You Draw It" article on the topic of drug overdose epidemic particularly interesting to us as visualization researchers are the use of *visual elicitation* and *contrasting narratives* to accent the main data visualization and message focused on rising deaths from drug overdoses over time.

Instead of just viewing the chart, elicitation prompts users to first externalize their belief/expectation on the issue. Prior research showed displaying the real data trend against one's own expectations leads to more attention to the discrepancies [25]. Moreover, Kirsh discussed seven ways that people can "think more powerfully" with external representations [26]. Since users may produce such external representations of their belief/expectations with different resolutions, we designed a non-visual elicitation that simply ask people to categorically characterize the trend compared to the draw-trend visual elicitation method used in the New York Times article [1]. Similarly, contrast may lead to better recall and engagement since they provide a "surprise" factor that accent the severity of the main issue the article is trying to convey
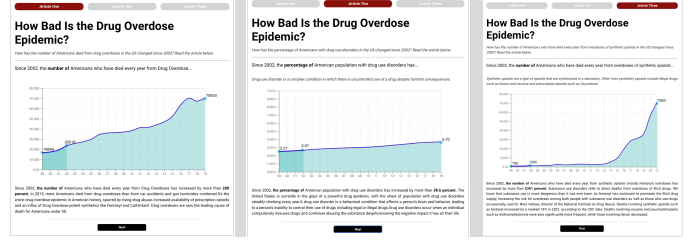

Fig. 2. Three articles participants read during Study 1 regardless of elicitation conditions.

[27], [28]. Connecting to the Elaboration Likelihood Model [6], we contend that elicitation and contrasting narratives may lead to higher amounts of elaborative thinking compared to passive viewing and ultimately result in a higher chance of attitude change.

3.1 Hypothesis formulation

Three main hypotheses were formulated:

H1-Impact of elicitation on recall and engagement. Belief elicitation increases recall and engagement with news articles that contain data visualizations compared to the condition with no elicitation.

Expanding on H1, we expect a difference in the impact of the two forms of elicitation (draw a trend vs. categorize a trend) on recall and engagement. Compared to categorizing a trend as a slight or sharp increase, individuals in the draw a trend condition are asked to draw a line across multiple years, producing data points for each year as well as a trend line. Thus individuals in the "draw a trend" condition generate a higher-fidelity representation of their guess/expectation.

H2-Impact of contrasting narratives on recall and engagement. Adding contrasting narratives to accent the data visualization on the article's focal issue increases recall and engagement with news articles.

H3-Impact of elicitation and contrasting narratives on attitude change. While prior research has demonstrated people's attitudes are notoriously difficult to change, we hypothesize belief elicitation and contrasting narratives increase the likelihood of attitude update on the main issue presented in the news articles.

4 STUDY DESIGN

4.1 Datasets and stimuli

The New York Times "You Draw It" article focuses on rising deaths from the drug overdose epidemic in the US [1]. For our studies, we kept the theme of the drug overdose epidemic and generated two additional articles, with one reporting the percentage of the American population with drug use disorders and the other reporting deaths from a particular type of illicit drug (synthetic opioids). We believe due to the current polarized conversations around the opioids epidemic, this is an appropriate theme to address and learn about how visualizations and narrative might help with attitude change around polarizing topics.

Study 1 - Yearly data for drug overdoses and the percentage of the American population with drug use disorders were collected from *Our World in Data* [29]. Data for the number of Americans who have died every year from overdoses of synthetic opioids was collected from [30]. The stimuli participants saw regardless of elicitation conditions are shown in Figure 2.
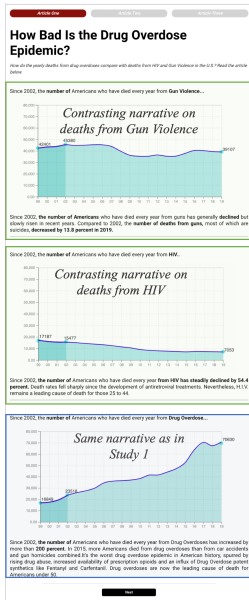
Fig. 3. Contrasting narratives on *gun violence* and *HIV* deaths added to contrast the sharp rise in drug overdose deaths.

Study 2 - Keeping the same stimuli as in study 1, we added two contrasting trends to each of the articles as detailed in the bottom right portion of Figure 4. We refer to these contrasting trends with brief textual descriptions of the trends as *contrasting narratives*. For example, visualizations of deaths from Gun violence and HIV with accompanying text descriptions of the trends (top two cells in Figure 3) were added to article 1 before participants saw the visualization on deaths from drug overdoses (bottom cell in Figure 3). The purpose of adding the contrasting narratives is to accent the last chart in each article related to the drug overdose epidemic in the US. The last chart in each article was the same chart from Study 1. An example of article 1 (in the control condition) is shown in Figure 3.

4.2 Experiment conditions design rationale

In our experiments, the "Draw trend" condition refers to asking participants to complete a timeline by dragging the cursor in the chart area (example shown in Figure 1). The participants can adjust any individual data point to shape the line to their desired trend. This technique is similar to the one used in the drug overdose epidemic NYT article [1]. To ascertain whether the impact of belief elicitation is due to the asking participants to guess the trend first before seeing the data or the visual nature of drawing a trend, we designed the "Categorize trend" condition that asks the participants to select a general characterization of the temporal trend (e.g., slight increase, mostly flat, slight decrease, etc.) from a dropdown menu (Figure 4A). The control condition does not present any form of elicitation; the visualization and accompanying textual description are shown to individuals, similar to the most common way we view news articles.

In summary, we designed three conditions for study 1, which focuses on evaluating the impact of belief elicitation: **Draw trend**, **Categorize trend**, and **No elicitation**.

Building on study 1, we further explore the effect of adding contrasting narratives that accent the main visualization. The three elicitation conditions remain the same in Study 2, and the only difference is adding two contrasting narratives prior to the focal visualization present in Study 1 in each article. The design of Study 2 allows us to explore the potential interaction between elicitation and contrasting narratives.
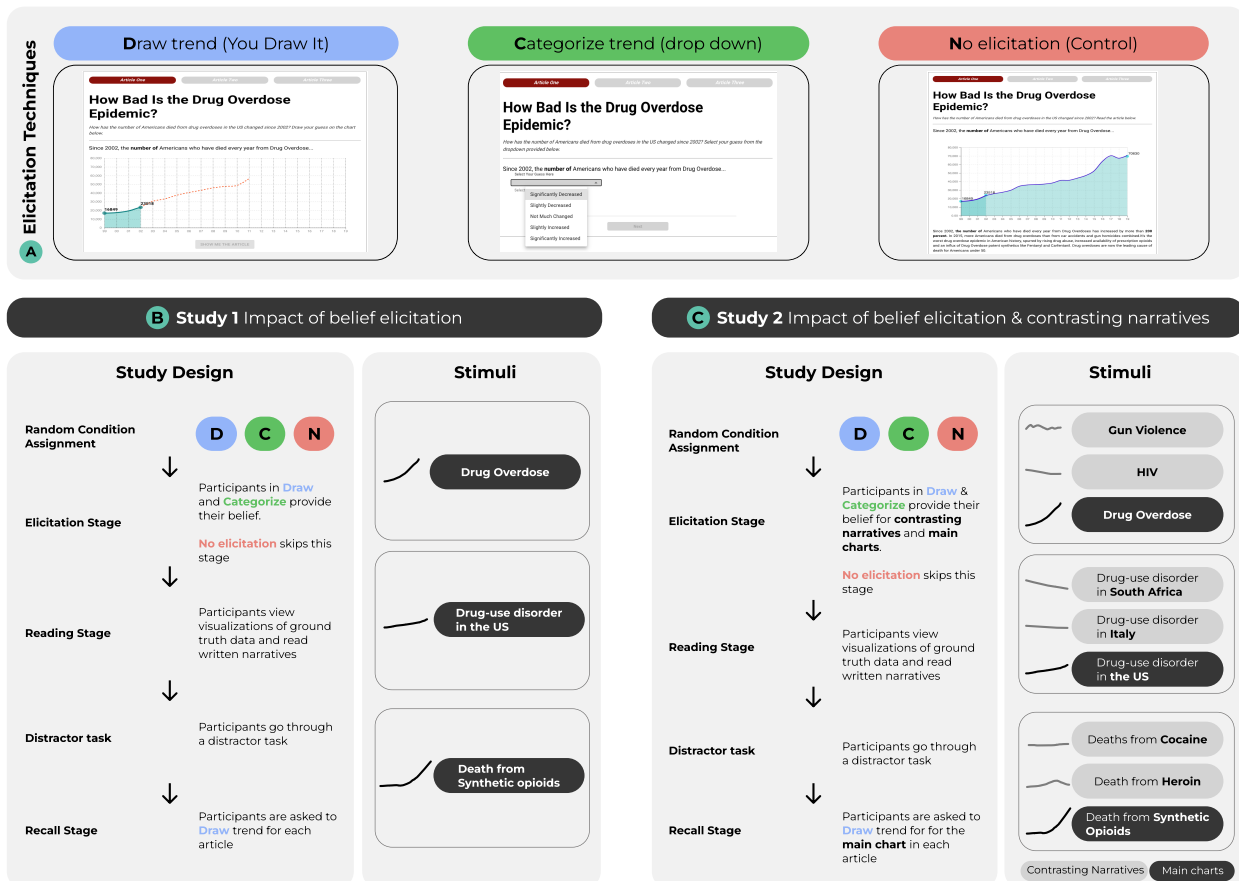

Fig. 4. Study design and procedure. Study 1 focuses on evaluating the impact of belief elicitation while study 2 added contrasting narratives to evaluate its impact on attitude change, recall, and engagement.

4.3 Study procedure

Study 1 - Detailed procedure of study 1 is illustrated in Figure 4B. The goal of Study 1 was to evaluate the impact of elicitation on memory, engagement, and attitude change. **Study 2** - Detailed procedure study 2 is illustrated in Figure 4C. The goal of Study 2 was to evaluate the effect of contrasting narratives on memory, engagement, and attitude change, and potential interactions between elicitation and contrasting narratives. Both studies employed a mixed design structure with 2 fixed and 1 random factors: (F1) elicitation technique (between-subjects) with 3 levels : none [control], draw trend, categorize trend, (F2) article (within-subjects) with 3 levels: drug-overdose, drug-use-disorder, Opioid-death, and (R1) participant levels. Participants were nested within elicitation conditions, and questions were fully crossed with condition. Thus, each participant was randomly assigned to one condition, in which they completed the task with all three articles (in the same order). The pre-registration of study 1 and 2 can be found here [1]

In both studies, participants were instructed to engage in two distractor tasks between the reading phase and the recall phase. Initially, the participants responded to a series of inquiries, after which they engaged with a news article related to the 2023 Super Bowl [2]. Subsequently, they offered a one-sentence summary of the article, before proceeding to the recall phase.

Topic involvement is elicited once per participant prior to seeing any articles using the 4-item questionnaire developed by Liao et al. [31]. Topic involvement measures personal relevance and motivation to learn about controversial topics [31] and was shown in prior research to impact belief and attitude change [8]. We use the measure of topic involvement when analyzing attitude change in our model and interpret results in the context of this factor.

We collected data from 288 participants in Study 1 and 307 participants in Study 2. There was no overlap in participants between the two studies. Participants demographics from both studies share similar distributions. Details on the participants demographics and exclusion of data can be found in section 8.

4.4 Dependent variables

Studies 1 and 2 shared the same sets of dependent variables. Recall accuracy and article engagement are measured at the article-level, resulting in a set of scores for each of the three news articles. Attitude change is measured at the participant-level; i.e. one set of scores from three attitude questions per participant.

4.4.1 Recall error

We calculated the Root Mean Square Error (RMSE) between the participant's recall response and the actual data for each chart to determine the overall recall accuracy for each user. RMSE is a widely used measure of error on numerical predictions [32]. As shown in Equation 1, $\hat{y}$ is the recalled value by a participant for each year, y is the ground-truth value for that year in the dataset and n is the total number of years in each response:

$$RMSE_{overall} = \sqrt{\sum (\hat{y} - y)^2 / n} \qquad (1)$$

4.4.2 Article engagement

We measured participants' engagement with the news articles as a proxy for the amount of elaborative thinking that occurs

in response to each article. We evaluate three components of user engagement (i.e. surprise, recommend, interest) using three questions adapted from the User Engagement Scale (UES) [33]. UES was designed to measure user engagement as a quality of user experience characterized by the depth of an user's investment when interacting with a digital system [34]. The entire UES measures many dimensions including perceived usability and aesthetic appeal. Since not all dimensions apply to our evaluation of news article engagement, we selected a subset of questions to evaluate the following three components of user engagement:

Surprise: *"The content of this article is surprising to me."*

Recommend: *"I would recommend this article to my family and friends."*

Interest: *"I felt interested in this article."*

Responses to each question were on a 5-point scale with the responses "Not at all", "A little", "Moderately", "A lot", and "Extremely."

4.4.3 Attitude change

We designed three questions to measure participants' attitudes regarding the drug overdose epidemic. Responses to these three questions were made on 5-point scale. The three questions and corresponding response scales are:

Q1. To what extent do you think the current rate of drug overdoses in the US is a problem? Responses to this question ranged from "Extremely serious problem" to "Not at all a problem".

Q2. Should the US government make combating drug abuse and overdoses a priority, i.e. by allocating tax dollars to treatment and prevention programs? Responses to this question ranged from "Not a priority" to "High priority".

Q3. What is your opinion on drug legalization and decriminalization in the US? Responses to this question ranged from "Strongly Oppose" to "Strongly Favor."

The first two attitude questions are designed to evaluate the perceived seriousness and whether the participant is willing to direct more resources to address the problem. The first two questions are one directional; we expect responses to indicate the drug overdose epidemic to be more or less serious. With the last question, we were interested in how people think of drug legalization when provided more information on the current drug overdose epidemic in the US.

4.5 Overview of analysis approach

All analyses were conducted in R version 4.2.3. Data and analysis code are available at [3]. We performed Bayesian statistical analyses using *brms* [35]. To evaluate the fitted Bayesian models (e.g. baseline vs. models with additional fixed effects) we used LOO-PSIS [36], a robust method for comparing the performance of Bayesian models in terms of expected predictive accuracy. As described in our preregistration, as a general practice we compared two models for each dependent measure. The baseline model included treatment (Control, Categorize, Draw) and article (Drug Overdoses, Drug-use Disorder, and Synthetic Opioids) as fixed effects and a random intercept for participants (y ~ treatment + article + (1|PID)).[4] We compared that to an alternative model that also included topic involvement and

---

1. Study 1: https://bit.ly/41l856z, Study 2: https://bit.ly/47Y061V
2. http://bit.ly/3RFsn7U

3. the Supplemental Materials is available here: https://bit.ly/3GJoAjE

4. In our preregistration we planned to include article (stimulus ID) as a random effect. However, given the small number of stimulus categories (3 articles) we chose to include it as a fixed effect.

prior attitude as additional fixed effects (`y ~ treatment + article + topicInvolvement + preAttitude + (1|PID)`), based on the expectation that participants' existing views on the topic might influence their engagement in the task and responses to the articles. In all cases, model comparisons based on the LOO-PSIS criterion were either ambivalent or favored the more complex model, so for simplicity we only report the results for the alternative model. In these and all subsequent models, treatment coding was used for categorical predictors with the Control condition and Drug Overdoses article serving as reference levels. Full results for estimated parameters of each model are provided in the Supplementary Material.

In our results we report posterior medians and 95% highest density intervals (HDIs) for estimated parameters and contrasts. For focal effects of interest we also report the *probability of direction (PD)* which ranges from 50–100% and is the proportion of the posterior distribution that is above zero (when the posterior median is positive) or below zero (when the posterior median is negative) [37]. PD thus indicates the overall probability that an estimated quantity is in the same direction as indicated by the posterior median. For ease of interpretation we highlight those effects for which the 95% HDI does not include zero (or another null value). However, we note that under the Bayesian estimation approach the posterior distribution can be directly interpreted as the relative credibility of different parameter values or effect sizes, and a 95% HDI that overlaps with zero can still provide (weaker) evidence for non-zero effects [38].

5 RESULTS ON H1: THE IMPACT OF ELICITATION

5.1 The impact of belief elicitation on recall error

This set of analyses focused on participants' accuracy at the recall stage of the task, when they were asked to recreate the trends for the three articles shown earlier. We normalized the RMSE values by dividing each data point by the maximum of that article. As a result, recall RMSE values can be interpreted as errors as a percentage of the displayed height of the chart. Figure 8 (top) shows the distribution of participants' responses alongside the true trend (black line), and the year-by-year absolute error for each treatment (bottom). Figure 5 shows the overall RMSE aggregated by treatment. For each study we modeled overall recall error (RMSE) using multilevel GLMs with a log-normal link function.

Study 1-Elicitation: The results from pairwise contrasts between treatments are shown in Table 1a. Overall we did not observe credible difference when comparing either belief elicitation method (Draw or Categorize) to the control condition with no elicitation. However, when comparing two methods of elicitation, our results indicated that recall RMSE was lower in the Draw trend elicitation compared to the categorize trend elicitation condition (Draw − Categorize = −.02, 95% HDI [-.04, -.003]), with $PD = 98\%$ probability of a negative effect. This result suggests that participants in the Draw condition were, on average, likely to be more accurate in recreating the trend than participants in the Categorize condition.

Study 2-Elicitation+Contrasts: The elicitation results obtained from Study 2 were consistent with those of Study 1 (Table 1b). Interestingly, the results for Studies 1 and 2 revealed that participants in the Draw elicitation condition were on average more accurate than the Categorize trend condition. However, recall errors from these two conditions were not significantly different than the Control condition. Overall, the magnitude of the effect on

Pairwise Recall RMSE			
Treatment Pairs	**Recall RMSE**	**95% HDI**	**PD**
a: Study 1			
Draw - Categorize	**-0.02**	**[-0.04 , -0.003]**	**98%**
Categorize - Control	0.01	[-0.01, 0.03]	87%
Draw - Control	-0.01	[-0.03, 0.005]	83%
b: Study 2			
Draw - Categorize	**-0.02**	**[-0.04, -0.001]**	**98.6%**
Categorize - Control	0.004	[-0.01, 0.02]	71%
Draw - Control	-0.01	[-0.03, 0.003]	93%

TABLE 1
Pairwise Recall RMSE Difference (including posterior median, 95% High Density Intervals, and Probability of Direction) for Study 1 and 2.
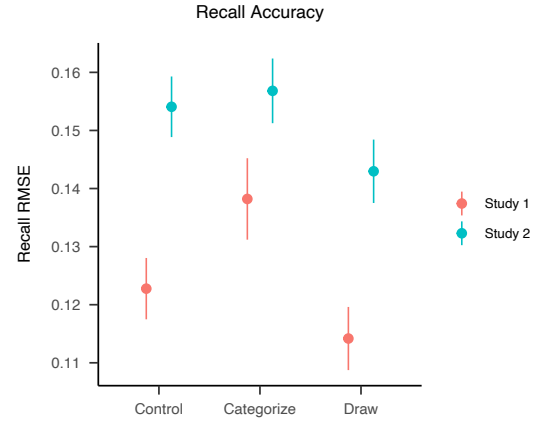

Fig. 5. Recall accuracy (RMSE) by condition for Study 1 and Study 2. Error bars indicate standard error of the mean.

recall remains a question to be explored in future studies. Further discussions of these findings are in section 9.2.

5.2 The impact of belief elicitation on engagement

Next we examined whether elicitation has an impact on participants' responses to the post-recall engagement questions (interest, recommend, and surprise). Responses to each question were on a 5-point scale. We calculated the proportion of participants in each treatment group who gave each response. Figure 6 (top) shows the empirical CDF obtained by calculating the cumulative sum across the five rating categories. Lower curves represent treatments with a greater proportion of high responses on a given question. The blue curve (top of Fig. 6) is lower for both Study 1 and 2, indicating more participants in the Draw condition find the content of the article surprising. Individual responses to the engagement question on "surprise" are plotted under the curves. We will unpack these results for perceived surprise in the next section.

We used ordinal regression to separately model responses for each outcome (surprise, interest, recommend). The models used a cumulative logit link function with separate threshold parameters fit for each rating category.

Study 1: The results indicated that participants in the Draw condition were on average likely to be more surprised compared to the Control and the Categorize conditions, while there was no apparent difference between the Categorize and Control conditions (Table 2a). Looking at the distribution of responses to the

Fig. 6. Top: Cumulative frequency of responses to the interest, recommend, and surprise questions for Study 1 & 2. Bottom: distribution of individual responses across treatments for the "surprise" question.

surprise question, Fig 6 bottom left, fewer participants in the Draw condition found the articles "Not at all" surprising while more participants responded "Moderately" surprising. There were no effects for the perceived interest and recommend questions.

Pairwise Perceived Surprise			
Treatment Pairs	**Surprise**	**95% HDI**	**PD**
a: Study 1			
Draw - Categorize	**0.53**	**[0.15 , 0.99]**	**99.7**%
Categorize - Control	-0.14	[-0.52, 0.23]	78.0%
Draw - Control	**0.38**	**[0.03, 0.77]**	**98.0**%
b: Study 2			
Draw - Categorize	**0.53**	**[0.15, .93]**	**100.0**%
Categorize - Control	-0.14	[-0.52, 0.23]	57.5%
Draw - Control	**0.38**	**[0.02, 0.77]**	**99.9** %

TABLE 2
Results (including posterior median, 95% High Density Intervals, and Probability of Direction) from study 1 and 2 on Perceived Surprise.

Study 2: The results indicated participants in the Draw condition were on average more likely to indicate higher perceived surprise when compared to the Control and the Categorize conditions, while there was no difference between the Categorize and Control conditions, Table 2b. Consistent with the findings from Study 1, there were no effects for the perceived interest and recommend questions. Looking at the distribution of responses to the surprise question (Figure 6, bottom right), more participants in the Draw condition found the articles "extremely" surprising while fewer participants responded "Not at all" surprising.

Overall, the results from both studies showed a consistent increase in *perceived surprise* for the Draw condition compared to both the Control and Categorize conditions, while there was no evidence for differences for the other engagement questions.

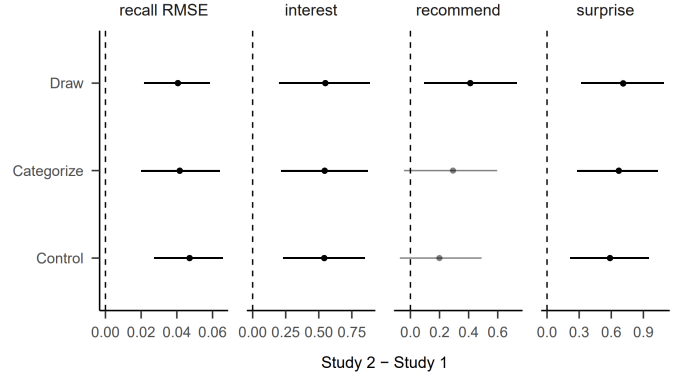# 6 H2: THE IMPACT OF CONTRASTING NARRATIVES ON RECALL AND ENGAGEMENT

Fig. 7. Contrasts between Study 2 and Study 1 for models of recall RMSE and engagement.

In this set of analyses, we examined the effect of contrasting narratives on recall and engagement for the articles by directly comparing the effects from Study 1 and Study 2. Given the similarity of the designs of Study 1 and Study 2, including the same set of outcome measures, we performed an integrative analysis which combined the datasets of the two studies [39]. Integrative data analysis refers to strategies in which two or more independent data sets are pooled or combined into one and then statistically analyzed. It offers advantages over meta-analysis in which *summary statistics* across multiple studies are pooled together [39].

We pooled the data for both studies and fit new Bayesian models for each outcome which included the study (1 vs. 2) and the treatment $\times$ study interaction as predictors. The model
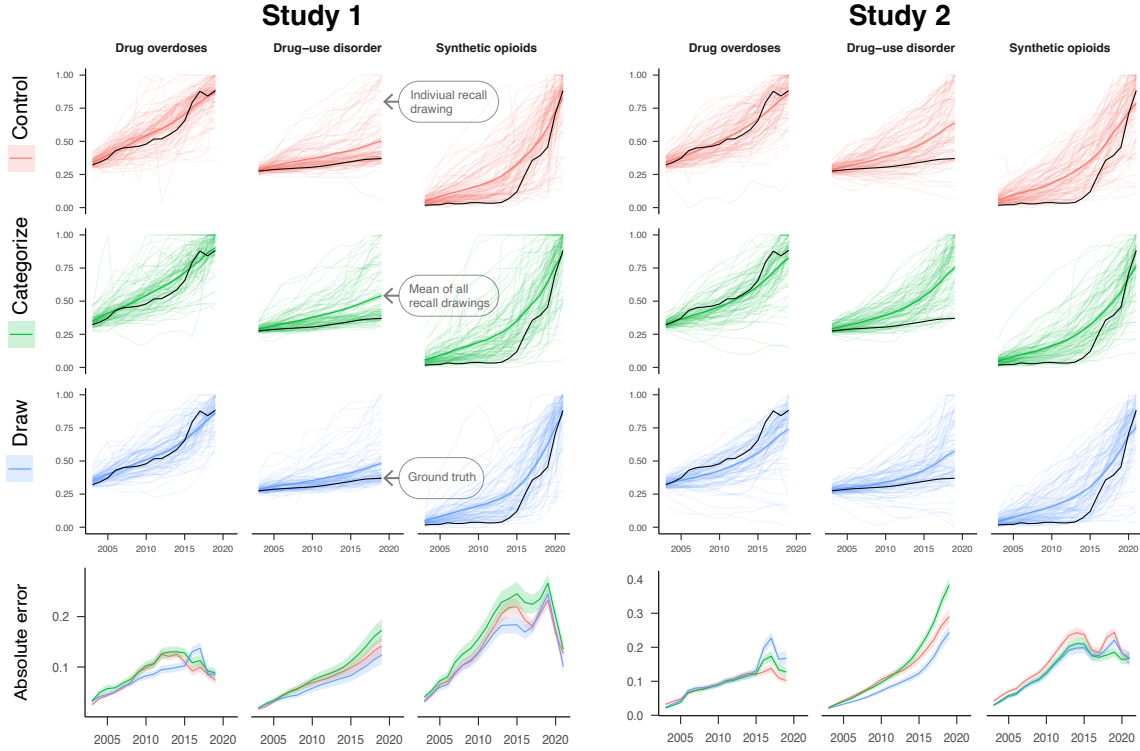
Fig. 8. **Top:** Responses and actual trend (black line) in Study 1 & 2. The trend values are scaled according to how they looked to the participants during the studies. **Bottom:** Yearly absolute error (mean and standard error) in Study 1 & 2.

specifications were otherwise identical to those used for the analyses of the individual studies above, with a baseline model (y ∼ study * treatment + article + (1|PID)) and alternative model (y ∼ study * treatment + article + topicInvolvement + preAttitude + (1|PID)). After fitting the models we calculated, within each treatment, the contrast between studies to assess the impact of contrasting narratives (Study 2) relative to when they are not presented (Study 1) (see Figure 7). It should be noted that there may be other confounding differences between the two studies which naturally limit our conclusions from this analysis, including potential differences in samples that were collected at different timepoints. We discuss these potential limitations further in Section 9.

6.1 Recall error

The results for recall accuracy indicated that participants on average were likely to be less accurate in recreating the trends in Study 2 compared to Study 1, indicating the contrasting narratives had a negative impact on recall accuracy potentially due to being exposed to multiple datasets in a single article. As shown in Figure 5, recall RMSE was higher in Study 2 across all treatment conditions (Table 3). Figure 8 right shows bigger gaps/discrepancies between the mean of recall drawings in each treatment and article compared to corresponding results from Study 1. The aggregate absolute error on the bottom also showed higher error from Study 2 (the error range on the y axis is much larger in Study 2).

6.2 Article engagement

We found substantial evidence of increased interest and surprise across all treatment conditions in Study 2 relative to Study 1 (Table 4a,b), suggesting that the addition of contrasting narratives was

Differences in Recall RMSE Between Study 1 and 2			
Treatment	**Study2 - Study1**	**95% HDI**	**PD**
Control	**0.04**	**[0.02 , 0.06]**	**100%**
Categorize	**0.04**	**[0.02 , 0.06]**	**100 %**
Draw	**0.04**	**[0.02 , 0.06]**	**100%**

TABLE 3
Differences in recall RMSE between Study 1 and 2.

Impact of Contrasting Narratives on Article Engagement			
Treatment	**Study2 - Study1**	**95% HDI**	**PD**
a: Surprise			
Control	**0.52**	**[0.17 , 0.89]**	**99.9%**
Categorize	**0.63**	**[0.25 , 1.00]**	**99.9%**
Draw	**0.80**	**[0.43 , 1.17]**	**100%**
b: Interest			
Control	**0.49**	**[0.22 , 0.76]**	**100%**
Categorize	**0.46**	**[0.16 , 0.75]**	**99.9%**
Draw	**0.67**	**[0.39 , 0.97]**	**100%**
c: Recommend			
Control	0.17	[-.07 , 0.40]	92.3%
Categorize	0.22	[-.04 , 0.49]	94.5%
Draw	**0.52**	**[0.25 , 0.79]**	**100%**

TABLE 4
The impact of contrasting narratives on article engagement.

more likely to create a positive impact on participant engagement with the articles. Responses to the Recommend question were higher in Study 2 for the Draw condition, but there was less
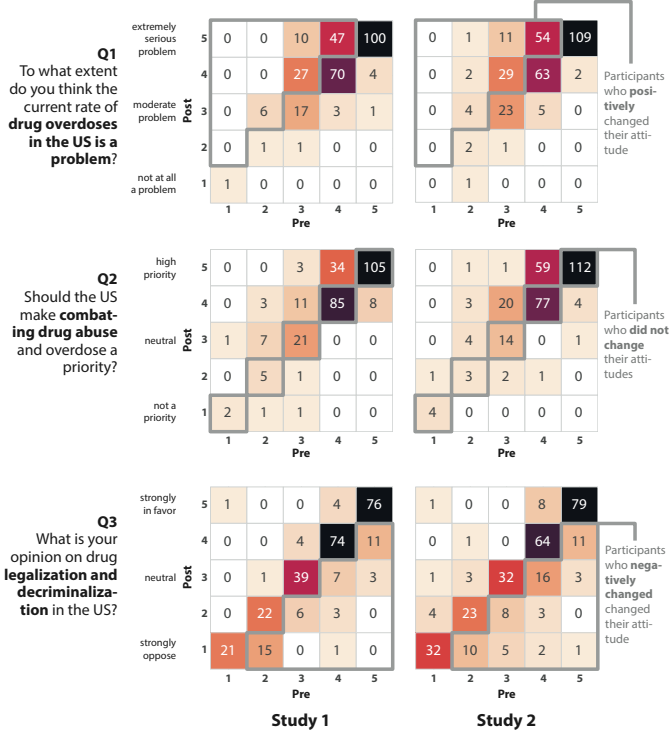
Fig. 9. Heatmap visualization of participants' attitude responses before and after completing the study **Left:** Pre and post response to attitude questions in study 1. **Right:** Pre and post response to attitude questions in study 2.

evidence of a credible difference between studies in the Control and Categorize conditions (Table 4c).

7 H3: RESULTS ON ATTITUDE CHANGE

7.1 Overall attitude change

To measure attitude change, we calculated the difference between responses to the three attitude questions before and after completing the tasks for each study. Since attitude change scores were calculated as the difference between ratings on 5-point scales the resulting distribution was discrete and highly skewed. We therefore fit ordinal regression models (with cumulative logit link functions) to attitude change scores for each question. For both studies, we first fit a baseline model with elicitation treatment as the only additional predictor besides the thresholds between response categories. We also explored a more complex model in which we added topic involvement and average responses to the article engagement questions (interest, recommend, and surprise) as predictors. Model comparisons with LOO-PSIS favored the more complex model. Figure 10 (bottom) shows the posterior median and 95% HDIs of the focal parameters for the attitude change model in both studies.

For Q1 (*To what extent do you think the current rate of drug overdoses in the US is a problem?*) and Q2 (*Should the US government make combating drug abuse and overdoses a priority?*), positive changes in responses suggest increased perceived importance of the drug overdose issue. For Q3 (*What is your opinion on drug legalization and decriminalization in the US*), positive changes indicate increased support for legalization/decriminalization while negative changes indicate decreased

support.[5] Overall, we see mostly positive attitude change for Q1, with almost 1/3 of participants in both Study 1 and 2 shifting their attitude to perceiving drug overdose as a more serious problem. In comparison, a smaller proportion of participants shifted their opinion on Q2 and Q3. Figure 10 shows model predictions for attitude change and Figure 9 shows the pre and post-responses for each attitude question from both studies.

To report the results on the impact of elicitation and contrasting narratives on attitude change, we first elaborate on the questions used to measure attitude change and show summary statistics on the overall attitude change.

While more participants overall did not change their attitudes, the directions of shift for the ones who did were systematic on all three questions. For questions 1&2, aside from participants with the most extreme pre-attitude choices with no room for positive attitude change (i.e., extremely serious problem for Q1 and high priority for Q2), the majority of participants had a positive attitude change (See Figure 9 top and middle rows). For question 3, where overall participants' pre-attitudes were more spread out, we observe many participants shifting attitudes towards opposition to drug legalization (See Figure 9 bottom row). While we do believe a general shift in attitudes for participants who had room to change in the provided scale is interesting, it is important to note that although we do observe general shifts in attitudes for many participants, these changes at different parts of the scale might not be equal. For example, a shift from somewhat opposed to neutral might have a completely different weight in comparison to one from neutral to somewhat in favor. Future studies should dive deeper into the nuances of attitude change.

We contend the overall observation on attitude change is related to our selection of a real-world topic that people have thought of and may have already formed an opinion. In comparison to the uncontentious topics chosen in other studies [40], we instead picked a topic that resembles many issues we wrangle with in our daily lives where most of us have formed our opinions and see if there is room for data visualizations to persuade. Moreover, We intentionally refrained from eliciting participants' attitudes after each article to minimize fatigue and because all articles covered the same topic (Drug abuse) and therefore this study cannot shed light on whether introducing only one article might produce similar effects.

7.2 Impact of belief elicitation on attitude change

Figure 10 (top) shows the expected posterior median and 95% HDI for the amount of attitude change for each question. Overall changes in responses to questions 1 and 2 were broadly positive in both Study 1 and Study 2, while changes were in the negative direction for question 3, leaning more against drug legalization. This indicates that there were small but reliable changes in attitude about the topic as a result of viewing the charts.

We first used the estimated model to examine whether the amount of attitude change differed between treatment conditions. For each attitude question (Q1, Q2, and Q3) we calculated the pairwise contrasts between treatment conditions for each study (Table 5). For **Study 1-Elicitation**, we did not observe an impact

5. We pre-registered to derive a compound score for measuring attitude change. Upon examination, we agreed that Q3 captures more complex policy-related aspects of attitude towards drug overdose compared to the first two questions. Therefore, we report the analysis of the three attitude questions separately.

Treatment Pairwise Attitude Differences

a: Study 1

Questions	Treatment Pairs	Attitude Change	95% HDI	PD
Q1	Draw - Control	0.05	[-0.13, 0.22]	70%
	Draw - Categorize	0.02	[-0.16, 0.20]	57.8%
	Categorize - Control	0.03	[-0.15, 0.20]	63.1%
Q2	Draw - Control	0.07	[-0.07, 0.24]	83.3%
	Draw - Categorize	-0.01	[-0.18, 0.17]	53%
	Categorize - Control	0.08	[-0.07, 0.24]	85.6%
Q3	Draw - Control	-0.06	[-0.22, 0.11]	66.1%
	Draw - Categorize	-0.09	[-0.25, 0.08]	85.3%
	Categorize - Control	0.03	[-0.12, 0.18]	66.1%

b: Study 2

Questions	Treatment Pairs	Attitude Change	95% HDI	PD
Q1	Draw - Control	-0.12	[-0.28, 0.04]	92.8%
	Draw - Categorize	-0.17	[-0.33, 0.01]	98.1%
	Categorize - Control	0.05	[-0.11, 0.23]	73.2%
Q2	Draw - Control	-0.13	[-0.28, 0.01]	96.8%
	Draw - Categorize	**-0.19**	**[-0.35, -0.05]**	**99.3%**
	Categorize - Control	0.06	[-0.09, 0.22]	77.6%
Q3	Draw - Control	-0.03	[-.21, 0.13]	64.3%
	Draw - Categorize	0.15	[-0.04, 0.33]	94.2%
	Categorize - Control	**-0.17**	**[-0.35, -0.01]**	**97.6%**

TABLE 5
Treatment pairwise attitude difference in Studies 1 and 2.

Impact of contrasting narratives on Attitude Change

Questions	Treatment	Attitude Change	95% HDI	PD
Q1	Draw	-0.06	[-0.23, 0.11]	75.6%
	Categorize	0.08	[-0.10, 0.25]	80.7%
	Control	0.05	[-0.11, 0.22]	75.5%
Q2	Draw	0.01	[-0.13, 0.17]	59.8%
	Categorize	0.13	[-0.03, 0.29]	94.1%
	Control	**0.16**	**[0.01, 0.30]**	**98.5%**
Q3	Draw	0.05	[-.14, 0.22]	71%
	Categorize	-0.16	[-0.34, 0.00]	97.1%
	Control	0.05	[-0.14, 0.22]	76%

TABLE 6
Impact of contrasting narratives on attitude change between studies.

of elicitation on the amount of attitude change for Q1, Q2, or Q3. The results for **Study 2-Elicitation+Contrasts** were generally consistent with Study 1 in showing similar amounts of attitude change across treatments (Table 5b). There were no differences in attitude change for Q1. For Q2, attitude changes were smaller in the Draw condition compared to the Categorize condition, while for Q3 attitude change was more negative in the Categorize condition compared to the Control condition. In sum, while there were some indications of small differences in attitude change between treatments, no consistent effects of elicitation method were observed.

7.3 Impact of engagement on attitude change

The estimated model also indicated effects of topic involvement and engagement with the articles on attitude change (Figure 10, bottom). First, greater topic involvement was associated with smaller changes in attitude for questions 1 and 2. This is likely because topic involvement was positively correlated with initial responses to these questions, implying that participants who expressed high involvement already gave the maximal rating to these questions. Second, we found that average perceived surprise to the articles was associated with positive changes in responses to Q1 (Studies 1 and 2) and Q2 (Study 1), while the likelihood of recommending the article was associated with more negative changes in responses to Q3 (Study 2). Together, these results suggest that participants who were less involved in the topic and who were more surprised by the trends were more likely to subsequently express greater concern compared to the beginning of the study.

7.4 Impact of contrasting narratives on attitude change

In order to assess the effect of contrasting narratives on attitude change we combined the datasets from Studies 1 and 2 and fit a

new Bayesian ordinal regression model for each attitude question. The model included study (1 vs. 2), treatment (Draw, Categorize, Control) and their interaction as fixed effects. For each question (Q1, Q2, and Q3) we then calculated the expected difference in attitude change between the two studies within each treatment condition. For Q1 and Q3 there was no evidence for differences between Study 1 and Study 2. For Q2 there was evidence for greater attitude change in the Control condition for Study 2 compared to Study 1, but not the other conditions, see Table 6. Thus, these comparisons suggest that while participants in study 2 exhibited overall attitude changes, the addition of contrasting narratives in Study 2 did not lead to a significant amount of attitude change relative to Study 1.
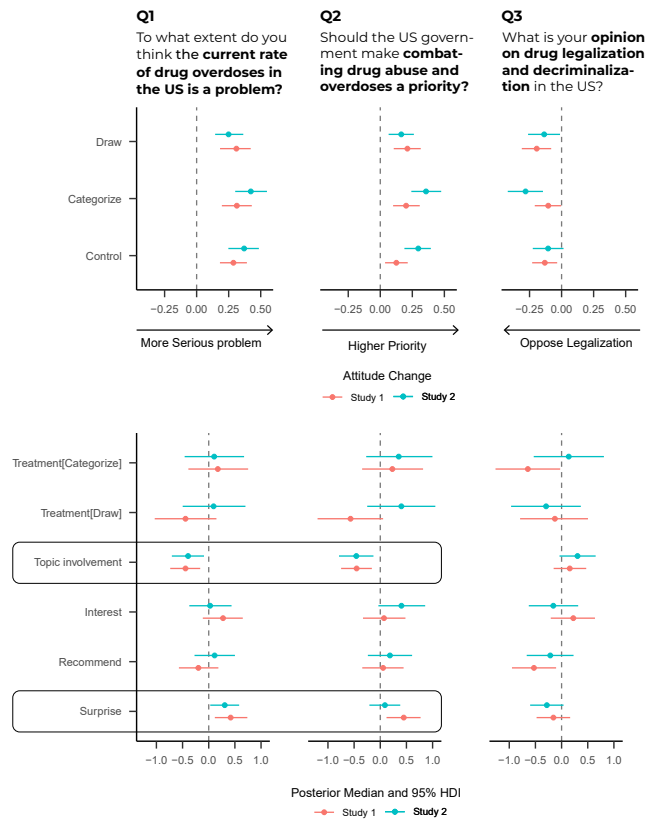

Fig. 10. **Top:** Attitude change measured by 3 questions for each treatment condition (predicted means and 95% HDIs). **Bottom:** Posterior medians and 95% HDIs for attitude change model coefficients.

8 PARTICIPANTS

We recruited all participants from Prolific, limited to participants in the United States and above 95% job approval rate.

8.1 Study 1: Elicitation

We recruited 306 participants from Prolific. Participants earned $4.5 for completing the task, which on average took 13.8 minutes (SD = 7.8). Per our preregistration, we excluded 1 participant who did not complete the full study. We planned to exclude any participants who gave non-sensible or inappropriate responses to the open-ended questions for the CRT, but no participants met this criterion. We followed our pre-registration and excluded participants who complete the study below or above two standard deviations of average completion time.

After excluding responses based on incompleteness and time, there were $N = 288$ participants included in the analyses (Control: 106; Categorize: 91; Draw: 91). The average age was 35.6 years, range: 18–78). 110 identified as female, 166 identified as male, 8 as other, and 4 chose 'prefer not to say'. 194 were White, 26 were Black/African-American, 39 were Asian, 17 were Hispanic, 1 was Middle Eastern, 8 were another race/ethnicity, and 3 chose 'prefer not to say'. 143 participants reported a college degree, 87 with a high school degree, 36 with a Masters's degree, 7 with a Doctorate degree and 15 chose either "other" or "prefer not to say".

8.2 Study 2: Elicitation+Contrasts

We recruited 320 participants from Prolific. One response was excluded due to incompleteness. Participants earned $6 for completing the task, which on average took 19.1 minutes (SD = 11.9). Per our preregistration, we planned to exclude any participants who gave non-sensible or inappropriate responses to the open-ended questions for the CRT, but no participants met this criterion. We also pre-registered to exclude participants whose completion time is above or below 3 standard deviations. This criterion resulted in 12 exclusions. The completion times ranges from 5.3 to 82 minutes, with a median of 15.5 minutes.

Among $N = 307$ participants for study 2, 110 were in the Control condition, 98 were in the categorize trend condition, while 99 were in the draw trend condition. The average age was 38.4 years, range: 18–79). 148 identified as female, 151 identified as male, 3 as other, and 5 chose 'prefer not to say'. 224 were White, 19 were Black/African-American, 20 were Asian, 22 were Hispanic, 4 were American Indian or Alaska Native, 11 were another race/ethnicity, and 7 chose 'prefer not to say.' 130 participants had obtained a college degree, 109 reported a High School degree, 36 with a Masters's degree, 7 with a Doctorate degree, and 25 chose either "other" or "prefer not to say".

9 DISCUSSION, LIMITATIONS, AND FUTURE WORK

9.1 The persuasive power of journalistic articles with data visualization

Journalistic visualizations often aim to influence attitudes and behaviors. We examined how belief elicitation and charts with contrasting trends affect attitude change. We created interactive articles similar to the NYTimes article "You Draw It: Just How Bad Is the Drug Overdose Epidemic?" [1]. We observed systematic changes in participants' attitudes across three dimensions: After viewing the visualizations, participants expressed more concern about drug overdoses, more support for government intervention, and more opposition to drug legalization/decriminalization. While these changes in attitude were relatively small, they are nevertheless notable since past work suggests that achieving attitude change is often difficult [8]–[11] and because many participants' preexisting attitudes were already strong, leaving relatively little room for change. In most cases, participants already believed that the drug epidemic was a problem and should be a priority; after interacting with the visualizations, they believed it was even more important (Figure 9).

Persuasion and attitude change regarding specific policies are complex and vary depending on the topic. For example, research in social psychology suggests that inter-individual differences can impact attitude change by affecting one or more underlying processes by which variables induce persuasion [41]. In addition, research demonstrates that attitudes are more amenable to change when new information aligns with their original basis, whether cognitive- or affect-based [42]. In our study, we included a question that focused on attitudes toward the specific drug legalization policy. While we observed small shifts in attitudes against drug legalization, it is worth noting that there are ongoing arguments for and against this policy in policy circles (as discussed in this New Yorker article: [43]). Due to the absence of specific claims or evidence for or against drug legalization in our articles, participants had to rely on their interpretations. This underscores the necessity for visualization and article designers to provide clear, direct evidence linking data to policy when advocating for a policy through visualization. This situation presents intriguing research possibilities on the complexities and effectiveness of data visualization in promoting contentious policy agendas.

9.2 Differences between belief elicitation techniques

We observed several effects of using belief elicitation techniques in the articles. First, in both studies, visual elicitation (Draw condition) resulted in a more accurate recall of the presented data compared to the categorization elicitation technique (Categorize condition). After drawing a trend, participants may pay more attention to differences between their drawing and the presented data [16], consistent with the idea that when combined with data visualization, visual elicitation provides direct feedback to participants [25]. This enhanced attention to finer-grained details of the data (e.g., inflection points or short-term fluctuations) may then lead to better recall when asked to reproduce the chart. In contrast, participants in the Categorize condition were asked to identify the overall trend using loosely defined categories like "slightly increased" or "significantly increased." While the displayed chart still provides feedback, participants' evaluation of the chart may be at the same level of detail, e.g. checking whether the overall trend confirms their selection. This may lead to less scrutiny of specific details and poorer recall later on. One potential confound on drawing a trend leading to better recall compared to categorize a trend is the matching of elicitation and recall technique. It might be interesting to explore in a future study to develop a recall technique that doesn't match any of the elicitation techniques (draw and categorize) yet still captures participants' memory in high fidelity.

It is important to note that neither the Draw nor Categorize elicitation conditions led to noticeable differences in recall error from the Control condition with no elicitation. One possibility is that the Control condition represents variability in unprompted

attention to fine-grained visual details vs. evaluation of the overall direction of a trend, while the Visual and Categorize direct attention in one way or the other. Moreover, we did not control for the impact of y-axis limits on participants' recall errors for each individual chart. Instead, we opted to keep the y-axis limit constant for each question for consistency. For example, the chart on deaths from HIV (Fig 3 middle) shows a 50 percent decrease but looks flatter than other charts. This might produce some issues with chart recall that should be addressed in future studies. In the future, it is important to assess whether elicitations have long-lasting effects on recall.

We also observed that in Study 2, participants, on average, had larger recall errors. Although we hesitate to draw strong conclusions about the cause of this difference (see next section), we suspect it might be caused by cognitive load due to the increased number of elicitations in the articles. We suspect that recollection of trends in real-world settings where consumers rapidly consume information might lead to even worse recollections. It is important to note, based on the findings from our studies, that attitude change is almost independent of recall accuracy. In other words, participants potentially found the articles persuasive enough to update their attitude without being able to recreate the trends they saw in the articles accurately. Our findings raise interesting questions about the affordances and differences between different belief elicitation techniques.

Moving beyond recall, we observe other potential benefits of using visual elicitation techniques. Notably, we observe that in both studies, participants expressed higher levels of surprise in the Draw condition (See figure 6). This is interesting because surprise is known to be an important factor in persuasion and trust [44]. One way to explain this response is through the elaboration likelihood model. It is possible that when participants are asked to draw a pattern before being exposed to it, they may elaborate more on the discrepancies and thus report feelings of higher surprise.

Although we do observe impacts of elicitation techniques on surprise, compared to control, we do not observe an effect of elicitation on attitude change. It is possible that the provided narratives and visualizations that are shared between all conditions provide a stronger impact on attitude change for this topic, and the potentially smaller impact of belief elicitation might go unnoticed. In future studies, the impact of elicitation with and without provided narratives might shed light on this point.

9.3 The surprisingly effective contrasting narratives

In Study 1, each article presented only one visualization to directly convey the main message. However, the "You Draw It" article by the NY Times [1] took a different approach. It included two visualizations with relatively flat trends that sharply contrasted the increasing trend of the last visualization, which is related to the article's main topic of the drug epidemic. This contrast between the first two datasets and the main visualization might amplify the level of surprise at the rapid recent acceleration of the drug overdose crisis and lead to stronger feelings of engagement.

Our combined analysis of Studies 1 and 2 provide initial evidence that contrasting narratives (Study 2) have the potential to increase engagement in terms of feelings of surprise and interest compared to when the same data is presented in isolation (Study 1). As we noted earlier, however, our conclusions from this integrative analysis [39] are limited by a number of potential confounds. In order to present contrasting narratives in Study 2,

participants were necessarily shown a greater number of charts. Heightened engagement could result from presenting additional information separately from the specific nature of the contrast in each series of charts. In addition, since the studies were run at different timepoints, there might be other differences between the samples which could contribute to the observed effects. In future work, manipulating the content of the charts to control the degree of contrast could better isolate the role of contrasting narratives in driving enhanced surprise and interest.
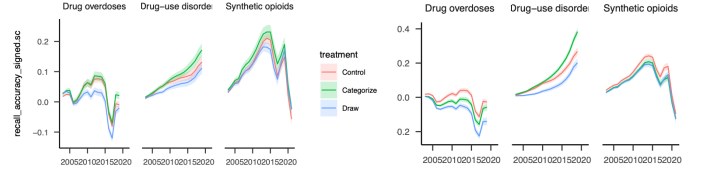

Fig. 11. Signed recall error (mean and standard error) in Study 1 & 2.

With those caveats in mind, it is interesting to note that the combined analysis also suggested that participants had poorer memory for the visualizations when adding contrasting narratives. For instance, comparing the average recall results from Study 2 (Figure 8 right) to Study 1 (Figure 8 left), we found that participants consistently underestimated the trends on Drug Overdoses and Synthetic Opioids, as shown by the enlarged gaps between the average guessed trend to the ground truth. In contrast, participants systematically overestimated the trend on Drug-user disorders. In other words, participants underestimated the more drastic trends while overestimating the "milder" trends. Figure 11 shows such systematic under and over estimation with the signed recall error, which is calculated with by subtracting the mean of responses ($\bar{\hat{y}}$) from the mean of true trend ($\bar{y}$) for each chart. Note that the range of the absolute error (y axis) is larger in Study 2 compared to Study 1. This effect may have been caused by the contrast introduced by the first two visualizations in the each article. These unintended consequences warrant further exploration to better understand their underlying causes. As designers of interactive articles, the choice to orchestrate multiple datasets to convey meaning or elicit a reaction is an interesting and under-explored aspect of visualization research that deserves further exploration.

10 CONCLUSION

Our work is inspired by the "You Draw It" article published by New York Times [1]. Two techniques we found particularly interesting are eliciting users' prior beliefs expressed by drawing a timeline, and the effective use contrasting narratives to accent the main data visualization on a specific issue. We designed two experiments to evaluate the impact of elicitation and contrasting narratives on attitude change, recall, and engagement. In the first study that focus on belief elicitation, we compared two different forms of belief elicitation with a control condition with no elicitation. Our findings revealed an overall attitude change and that the draw trend elicitation leads to higher engagement in terms of feelings of surprise. In the second study, we evaluate the impact of contrasting narratives. Compared to the results on study 1, we found that contrasting narratives increased attitude change, and improved engagement in terms of surprise and interest, but interestingly resulting in higher recall error. Our discussion further highlights the significance of our findings and many future research opportunities.

ACKNOWLEDGMENTS

REFERENCES

[1] K. Josh. You draw it: Just how bad is the drug overdose epidemic? https://www.nytimes.com/interactive/2017/04/14/upshot/drug-overdose-epidemic-you-draw-it.html.

[2] B. C. Wagner and R. E. Petty, "The elaboration likelihood model of persuasion: Thoughtful and non-thoughtful social influence," *Theories in social psychology*, pp. 96–116, 2011.

[3] A. Bycoffe, E. Mejía, D. Mehta, N. Silver, and M. Radcliffe. How unpopular is joe biden? https://projects.fivethirtyeight.com/biden-approval-rating/?cid=rrpromo.

[4] FiveThirtyEight. fivethirtyeight.com. https://fivethirtyeight.com.

[5] Upshot. the new york times. https://www.nytimes.com/section/upshot.

[6] R. Petty, J. Priester, and P. Briñol, *Mass Media Attitude Change: Implications on the Elaboration Likelihood Model of Persuasion*. Lawrence Erlbaum Associates, 01 2002, pp. 155–198.

[7] R. Petty and D. Wegener, *The Elaboration Likelihood Model: Current status and controversies*. Guilford Press, 01 1999, pp. 41–72.

[8] D. Markant, M. Rogha, A. Karduni, R. Wesslen, and W. Dou, "When do data visualizations persuade? the impact of prior attitudes on learning about correlations from scatterplot visualizations," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 1–12.

[9] J. Heyer, N. K. Raveendranath, and K. Reda, "Pushing the (visual) narrative: The effects of prior knowledge elicitation in provocative topics," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–14.

[10] A. V. Pandey, A. Manivannan, O. Nov, M. Satterthwaite, and E. Bertini, "The persuasive power of data visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 2211–2220, 2014.

[11] J. Liem, C. Perin, and J. Wood, "Structure and empathy in visual data storytelling: Evaluating their influence on attitude," *Computer Graphics Forum*, vol. 39, no. 3, pp. 277–289, 2020.

[12] 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*, vol. 27, no. 2, pp. 978–988, 2021.

[13] A. Sanborn and T. Griffiths, "Markov chain monte carlo with people," in *Advances in Neural Information Processing Systems*, J. Platt, D. Koller, Y. Singer, and S. Roweis, Eds., vol. 20. Curran Associates, Inc., 2007.

[14] A. N. Sanborn, T. L. Griffiths, and R. M. Shiffrin, "Uncovering mental representations with markov chain monte carlo," *Cognitive Psychology*, vol. 60, no. 2, pp. 63–106, 2010.

[15] S. Mahajan, B. Chen, A. Karduni, Y.-S. Kim, and E. Wall, "Vibe: A design space for visual belief elicitation in data journalism," *Computer Graphics Forum*, vol. 41, no. 3, pp. 477–488, 2022.

[16] R. Koonchanok, P. Baser, A. Sikharam, N. K. Raveendranath, and K. Reda, "Data prophecy: Exploring the effects of belief elicitation in visual analytics," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–12.

[17] R. Koonchanok, G. Y. Tawde, G. R. Narayanasamy, S. Walimbe, and K. Reda, "Visual belief elicitation reduces the incidence of false discovery," *arXiv preprint arXiv:2301.12512*, 2023.

[18] C. Xiong, C. Stokes, Y.-S. Kim, and S. Franconeri, "Seeing what you believe or believing what you see? belief biases correlation estimation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 1, pp. 493–503, 2022.

[19] 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*, vol. 29, no. 1, pp. 1005–1015, 2023.

[20] Y.-S. Kim, K. Reinecke, and J. Hullman, "Data through others' eyes: The impact of visualizing others' expectations on visualization interpretation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 760–769, 2018.

[21] 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*, 2019, pp. 1–14.

[22] S. Gupta, A. Karduni, and E. Wall, "Belief decay or persistence? a mixed-method study on belief movement over time," *Computer Graphics Forum*, vol. 40, no. 1, pp. 207–217, 2021.

[23] C. M. McColeman, L. Harrison, M. Feng, and S. Franconeri, "No mark is an island: Precision and category repulsion biases in data reproductions," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 2, pp. 1063–1072, 2020.

[24] C. R. Ceja, C. M. McColeman, C. Xiong, and S. L. Franconeri, "Truth or square: Aspect ratio biases recall of position encodings," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 2, pp. 1054–1062, 2021.

[25] 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*, ser. CHI '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 1375–1386.

[26] D. Kirsh, "Thinking with external representations," *AI Soc.*, vol. 25, no. 4, p. 441–454, nov 2010.

[27] G. Brod, M. Hasselhorn, and S. A. Bunge, "When generating a prediction boosts learning: The element of surprise," *Learning and Instruction*, vol. 55, pp. 22–31, 2018.

[28] L. K. Fazio and E. J. Marsh, "Surprising feedback improves later memory," *Psychonomic Bulletin & Review*, vol. 16, no. 1, pp. 88–92, 2009.

[29] H. Ritchie and M. Roser. (1999) Drug Use. https://ourworldindata.org/drug-use.

[30] N. C. for Health Statistics, C. for Disease Control, and Prevention. (2021) Drug Opioid. https://nida.nih.gov/research-topics/trends-statistics/overdose-death-rates.

[31] Q. V. Liao and W.-T. Fu, "Beyond the filter bubble: Interactive effects of perceived threat and topic involvement on selective exposure to information," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 2359–2368.

[32] D. Christie and S. P. Neill, "8.09 - measuring and observing the ocean renewable energy resource," in *Comprehensive Renewable Energy (Second Edition)*, T. M. Letcher, Ed. Oxford: Elsevier, 2022, pp. 149–175.

[33] H. L. O'Brien, P. Cairns, and M. Hall, "A practical approach to measuring user engagement with the refined user engagement scale (ues) and new ues short form," *International Journal of Human-Computer Studies*, vol. 112, pp. 28–39, 2018.

[34] H. O'Brien and P. Cairns, *Why engagement matters: Cross-disciplinary perspectives of user engagement in digital media*. Springer, 01 2016.

[35] P.-C. Bürkner, "Brms: An R package for Bayesian multilevel models using Stan," *Journal of statistical software*, vol. 80, pp. 1–28, 2017.

[36] A. Vehtari, A. Gelman, and J. Gabry, "Practical bayesian model evaluation using leave-one-out cross-validation and waic," *Statistics and computing*, vol. 27, pp. 1413–1432, 2017.

[37] D. Makowski, M. S. Ben-Shachar, S. H. A. Chen, and D. Lüdecke, "Indices of Effect Existence and Significance in the Bayesian Framework," *Frontiers in Psychology*, vol. 10, p. 2767, Dec. 2019.

[38] J. K. Kruschke and T. M. Liddell, "The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective," *Psychonomic Bulletin & Review*, vol. 25, no. 1, pp. 178–206, Feb. 2018.

[39] P. J. Curran and A. M. Hussong, "Integrative data analysis: The simultaneous analysis of multiple data sets." *Psychological Methods*, vol. 14, no. 2, pp. 81–100, 2009.

[40] 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*, vol. 27, no. 2, pp. 978–988, 2020.

[41] P. Briñol and R. E. Petty, "Individual differences in attitude change," *The handbook of attitudes*, pp. 575–615, 2005.

[42] G. Haddock and G. R. Maio, "Inter-individual differences in attitude content: Cognition, affect, and attitudes," in *Advances in experimental social psychology*. Elsevier, 2019, vol. 59, pp. 53–102.

[43] B. Wallace-Wells. Is there a case for legalizing heroin? https://www.newyorker.com/news/annals-of-populism/is-there-a-case-for-legalizing-heroin.

[44] R. E. Petty, M. A. Fleming, J. R. Priester, and A. H. Feinstein, "Individual versus group interest violation: Surprise as a determinant of argument scrutiny and persuasion," *Social Cognition*, vol. 19, no. 4, pp. 418–442, 2001.

11 BIOGRAPHY SECTION
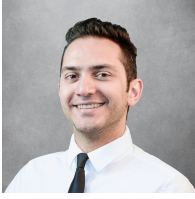
Milad Rogha holds a bachelor's degree in Electrical Engineering and a master's degree in Architecture. He is working toward a Ph.D. in Computer Science at the University of North Carolina at Charlotte. His research interests include human-AI interaction and the impact of visualization on human decision-making. His research is dedicated to enhancing artificial intelligence applications that bolster human decision-making processes.
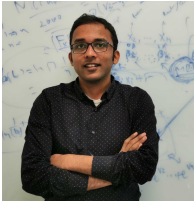
Subham Sah holds a Dual Master's of Science degrees in Software and Information Systems ,and Architecture from the University of North Carolina Charlotte. He loves building products that can engage and educate people. His area of interest is creating Visual Analytics user interfaces that assist users in becoming more conscious of their analytical behaviours and directing them towards achieving their goals.

Doug Markant is a cognitive scientist who examines human learning, memory, and decision making through behavioral experiments and computational modeling. He is currently an Assistant Professor in the Department of Psychological Science at UNC Charlotte, where he is a core faculty member of the Health Psychology Ph.D. program and an affiliate of the Ribarsky Center for Visual Analytics. He also serves as the director of the interdisciplinary program in Cognitive Science.

Alireza Karduni is an Assistant Professor of Human-Centred Computing at Simon Fraser University's School of Interactive Arts and Technology. His work focuses on designing interactive visual tools that help people understand and make decisions based on new information. Recognizing the socially and politically situated nature of information ecosystems, he researches how our existing views might influence our receptivity to new (mis)information and how such dynamics might influence societal interactions.

Wenwen Dou is an associate professor at the University of North Carolina at Charlotte and co-director of the Ribarsky Center for Visual Analytics. Her research interests include Visual Analytics, Text Mining, and Explainable AI. Dou has worked with various analytics domains in reducing information overload and providing interactive visual means to analyzing unstructured information. She has experience in turning cutting-edge research into technologies that have broad societal impacts.