Data Through Others' Eyes: The Impact of Visualizing Others' Expectations on Visualization Interpretation

Yea-Seul Kim, Katharina Reinecke and Jessica Hullman

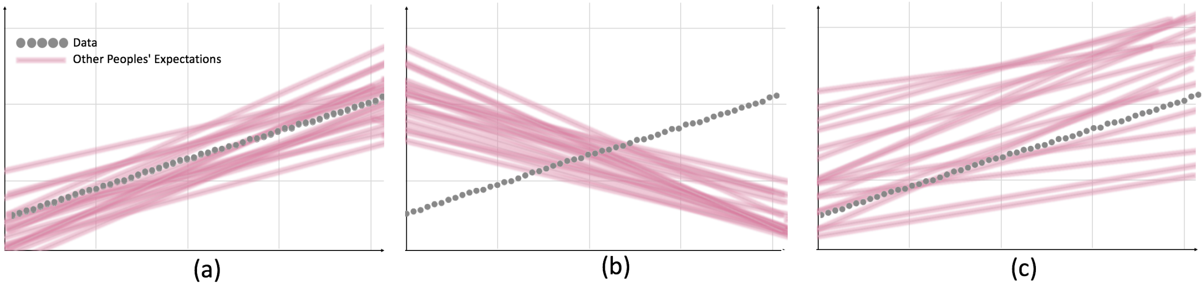

Fig. 1. Different scenarios for integrating social information in a visualization: (a) A dataset where people's expectations (depicted in pink) are aligned with the base data (depicted in gray). (b) A dataset where people's expectations are not aligned with the base data. (c) A dataset where people's expectation are aligned with the base data, but show lower consensus.

Abstract— In addition to visualizing input data, interactive visualizations have the potential to be social artifacts that reveal other people's perspectives on the data. However, how such social information embedded in a visualization impacts a viewer's interpretation of the data remains unknown. Inspired by recent interactive visualizations that display people's expectations of data against the data, we conducted a controlled experiment to evaluate the effect of showing *social information in the form of other people's expectations* on people's ability to recall the data, the degree to which they adjust their expectations to align with the data, and their trust in the accuracy of the data. We found that social information that exhibits a high degree of consensus lead participants to recall the data more accurately relative to participants who were exposed to the data alone. Additionally, participants trusted the accuracy of the data less and were more likely to maintain their initial expectations when other people's expectations aligned with their own initial expectations but not with the data. We conclude by characterizing the design space for visualizing others' expectations alongside data.

Index Terms—Social influence, Social visualization, Data interpretation

1 Introduction

Visualizations often act as social artifacts. The social function of a visualization can be explicit, such as when visualizations support collaborative data analysis by enabling a viewer to share their insights [39, 14], or prompt reflection on visualized social networks [15]. Visualizations may also implicitly embody others' perspectives as a result of editorial choices in the authoring process about which data to present and how to present it [21].

Most visualizations, however, do not directly acknowledge the role of others' opinions in data creation and interpretation. This stands in contrast to the many ways in which social information influences our actions in real life, where we often look to others for cues about what is accurate or correct [10]. Though little prior work has examined the impact of being exposed to others' beliefs about data when we interact with visualizations, exposure to what others think about a visualized data may significantly influence our interpretations of that data. Can knowing what others believe affect how much we trust visualized data, or whether we change our opinion after seeing the data?

- *Yea-Seul Kim is with University of Washington. E-mail: yeaseul1@uw.edu.*
- *Katharina Reinecke is with University of Washington. E-mail: reinecke@cs.washington.edu.*
- *Jessica Hullman is with University of Washington. E-mail: jhullman@uw.edu.*

Manuscript received 31 Mar. 2017; accepted 1 Aug. 2017.
Date of publication 28 Aug. 2017; date of current version 1 Oct. 2017.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TVCG.2017.2745240

Even when we do not agree with others' opinions, being confronted with social information can prompt deeper reflection and more divergent thinking about a topic [7]. This effect is similar to how thinking about our own prior knowledge and expectations about data can prompt deeper processing. For example, recent research indicates that prompting a viewer to predict data before seeing it in a visualization and subsequently asking them to reflect on their own predictions improves their ability to recall the data later [24]. Though some prior work indicates a negative influence of social information on graphical perception [20], incorporating *social information in visualizations as a way to depict others' expectations about data* may, under certain conditions, prompt a viewer to critically engage with the data and its relationship to their prior knowledge, skills that are associated with data and visualization literacy [12].

As an example of how *social information in the form of others' expectations* can be incorporated in a visualization alongside the data, consider the recent New York Times interactive visualization "You Draw It." Data on the relationship between parents' income percentile and the percent of children in the family who attended college in the U.S. is shown alongside a heatmap depicting earlier viewers' estimates of the trend (Fig. 2). Prior to presenting the observed relationship, the interface elicits each viewer's prediction via the interactive drawing interface. Prior viewers' predictions are then presented in aggregate against the data for each new viewer after

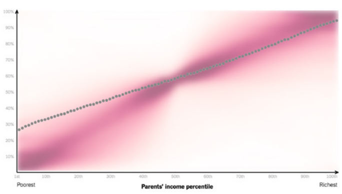

Fig. 2. A New York Times "You Draw It!" interface that visualizes how other people estimate (represented as a heatmap) along with the actual data (dotted line).

they make their own prediction. Such *visualizations with social information* become a tool for understanding the data (e.g., the measured relationship between the parents' income percentile and the percent of children attended college), and also for understanding others' beliefs about the data.

We investigate how presenting others' expectations of data influences people's interpretation of a visualization, including one's ability to recall the data later and how accurate they perceive the data to be.

Our main contributions are as follows. First, through a controlled experiment we demonstrate how social information that exhibits a high consensus among other peoples' expectations improves a participant's ability to recall data points in the visualization. Second, we find that participants trust the accuracy of the data less and are more likely to maintain their initial expectations when other peoples' expectations align with their own but not with the data. Finally, we characterize the design space of *visualizing data with social information* informed by our study.

2 Background

To formulate hypotheses, we surveyed work in social visualization, social influence, and prior knowledge in visualization interaction.

2.1 Social Visualization

Although most visualizations can be viewed as social artifacts, the term *social visualization* has been used to refer to visualizations that are purposefully designed to facilitate social interaction around data. Examples range from social network visualization systems like Vizster, where the data (a social network) is itself an aggregation of social information [15], to collaborative data analysis systems like sense.us [16], Name Voyager, [36], ManyEyes [35] or CommentSpace [39], in which commenting and other representations of social information are supported within an interactive visualization. Interacting with a visualization in the company of others, whether in person or asynchronously through an online visualization, has been shown to lead to benefits for analysis, including encouraging longer interactions [15], promoting collaborative analysis [36], and supporting unique discoveries [38]. However, several studies indicate that people's responses even for basic graphical perception tasks can be subject to biases when social information is available. In an early study, Asch [2] found that people will report the wrong answer to a simple visual perception task based on perceived pressure to align their response with those of others. More recently, Hullman et al. [20] find that when people are motivated to be correct on a graphical perception task, seeing other people's estimates for the same task can bias their responses.

In contrast to our work, these examples present social information next to the visualization, rather than directly integrated in the visualization. In most cases, the social information is in textual comments that require significant effort and motivation on the part of a viewer to process, similar to other types of textual comments on web pages [4]. We are interested in how the process of interpreting a visualization is affected by integrating social information, in the form of others' expectations about the data, directly into the visualization for comparison to the data. A few prior visualization systems integrate abstract visualizations of prior viewers' interactions directly into the interface. For example, Scented Widgets [38] are small visualizations depicting prior visits to views in an interactive visualization system that can be integrated directly into the navigation. As a result of seeing others' interactions, users are more likely to explore the same views that other people visited. Similarly, the BookVoyager system included a feature that greyed out visited data in the interface to encourage users to navigate unexplored views [37].

2.2 Social Influence

Psychological research on social influence spans a large range of topics including conformity, influence, and social comparison (e.g., [6]). We are interested in how conformity—people's tendency to match up their behaviors with other people's for various reasons—influences data interpretation in the context of visualization. Conformity can be driven by an *informational motivation* which arises from a viewer's desire to make an accurate judgement in an uncertain situation [7]. One refers to other people's actions as a source of information on what to do themselves. Alternatively, conformity can arise from a *normative motivation* where the goal is to obtain social approval by conforming to others' behaviors [2, 7].

A viewer's tendency to conform is affected by many aspects of the circumstances in which they are making a judgment, such as the number of people involved or their expertise [6]. We are particularly interested in how various forms of *uncertainty* about the "correct interpretation" of data in the context of a visualization interaction may cause people to be susceptible to social influence. Uncertainty is a prerequisite of influence [6], along with a presumption that others' opinions are as informed or more informed than oneself [7].

2.3 Role of Prior Beliefs and Internal Representation

Studies in psychology have shown that people's prior beliefs significantly influence how they evaluate presented evidence [8, 25, 31]. Various studies in visualization indicate that a viewer's internal representations, including prior knowledge and beliefs, impact how they reason with a visualization [17, 18, 26, 29, 34, 19]. By asking people to make predictions themselves prior to seeing an observed trend, "You Draw It" [1] prompts the viewer to reflect on their own prior knowledge concerning the data. Kim et al. [24] find that this type of explicit visualization interaction with one's prior knowledge can result in better recall of the data later. We are interested in whether seeing *others'* predictions can prompt the same type of reflection and consequently benefits to recall, and how properties of the social information, such as its agreement with the data, may influence this possibility.

3 Hypotheses Development

3.1 Functions of Social Information

Based on the prior work, we formulated hypotheses on how a participant's interpretation of visualized data will be affected by viewing others' expectations about the data. At a high level, we organized our study around different possible impacts that visualized social information might have on how data is interpreted. First, peoples' tendencies to be interested in what others think may cause social information to have a focusing effect, capturing the viewer's attention so that they notice and reflect on aspects of the visualized data that they would otherwise overlook:

- ***Focusing***: Social information causes a viewer to examine the data more closely.

Based on research in conformity, we also expect that a viewer's *expectations about the data*, including their *trust in the accuracy of the data*, may change in several different directions as a result of seeing others' expectations:

- ***Reinforcing***: Social information makes a viewer more likely to believe the data.
- ***Challenging***: Social information makes a viewer less likely to believe the data (i.e., more likely to question the data).

3.2 Formulating Hypotheses

We formulate hypotheses about how social information will play focusing, reinforcing, and challenging roles by identifying combinations of conditions that could arise if social information were integrated in a visualization (Fig. 3(a)). Our first hypotheses describes how simply including social information in a visualization is likely to improve a viewer's *ability to recall the data later* as a result of the focusing effect. This expectation is supported by recent work that indicates that asking a viewer to predict the data in a visualization then showing their prediction against the data can result in improved recall [24].

H1: Participants will more accurately recall the data in a visualization when social information is shown.

Research indicates that the degree of uncertainty a viewer perceives in a situation will affect how susceptible they are to social influence [6]. One way that social information can contribute to the perceived uncertainty of the data is by contradicting the predominant trend in the data (Fig. 3(b)). If the social information is *congruent* with the data, the data may be perceived as more certain; if the social information is *incongruent* with the data, the data may be perceived as less certain. We expect that two measures of a viewer's degree of belief in the data, their trust that the data is accurate and their likelihood of updating their beliefs to match the data, will be affected by congruency:

H2: Participants will report greater trust in the accuracy of the data and will be more likely to update their expectations toward the data when the social signal is congruent with the data signal compared to when it is incongruent.

A second source of uncertainty that may impact a viewer's interpretation of visualized data is the degree to which the social information suggests agreement among multiple people, or suggested *consensus* in the social information (Fig. 3(c)):

H3: The impact of congruency on participants' recall of the data, trust and likelihood of updating their beliefs to match the data will be stronger or weaker depending on the degree of consensus implied by the congruent or incongruent social signal.

Specifically, H3 leads us to expect that:

H3a: When exposed to social information that is congruent with the data signal, participants will report greater trust in the accuracy of the data and will be more likely to update their expectations toward the data when the social information also displays a high degree of consensus. Conversely, when exposed to social information that is incongruent with the data signal, participants will report less trust in the accuracy of the data and be less likely to update their expectations toward the data when the social information also displays a high degree of consensus.

The degree of consensus may also impact a participant's ability to recall the data, since higher variance (low consensus) in the social information is likely to be more distracting to the viewer. We therefore expect that:

H3b: Participants will recall the data more accurately when the social information displays a high degree of consensus (e.g., low variance between others' expectations) compared to when the social information displays a low degree of consensus (e.g., high variance between others' expectations).

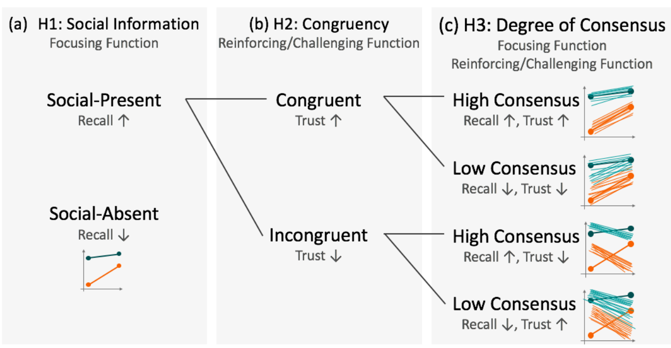

Fig. 3. Hypotheses, study conditions and stimuli for each condition. If participants are assigned to Social-Absent condition, they examine only the data. If the participant is assigned to one of the Social conditions, they examine one of four stimuli combining a level of congruency with a level of the degree of consensus.

3.3 Other Factors Influencing Interpretations

Finally, we expect the extent to which social information *reinforces* or *challenges* the data (H2, H3) to be moderated by the degree to which the viewer's initial expectations align with the data. When initial expectations are closely matched to the data, the situation may be perceived as less ambiguous, such that the social information has a lesser influence [9]. When initial expectations are opposed to the data, however, the viewer may actively search for a signal in the social information to ascertain the correctness of her own views. We speculate that the most important type of agreement is whether the viewer's expectations and the data express the same overall trend (i.e., the lines have the same slope direction), as visualizations tend to emphasize trends over exact numerical values [22]. We therefore include this initial "trend alignment" between the viewer's expectations and the data in analyzing the effects of congruency and degree of consensus.

3.4 Study Conditions

To evaluate our hypotheses, we defined five conditions (Fig. 3).

1. **Social-Absent:** The participant is asked to examine only the base data with no social information.
2. **Social-Congruent-HighConsensus:** The participant is asked to examine the base data with social information that exhibits the same predominant trend as the base data and displays a high degree of consensus among people.
3. **Social-Congruent-LowConsensus:** The participant is asked to examine the base data with social information that exhibits the same predominant trend as the base data and displays a low degree of consensus among people.
4. **Social-Incongruent-HighConsensus:** The participant is asked to examine the base data with social information that exhibits the opposite predominant trend as the base data and displays a high degree of consensus among people.
5. **Social-Incongruent-LowConsensus:** The participant is asked to examine the base data with social information that exhibits the opposite predominant trend as the base data and displays a low degree of consensus among people.

4 Study Design

To evaluate our hypotheses, we conducted a controlled study in which we examined how including other people's expectations about data in a visualization impacts participants' ability to recall the data, their trust in the accuracy of the data, and their beliefs after viewing the data.

4.1 Choice of Datasets

People look to social information to guide their behavior in two situations. First, they must experience some uncertainty about the correct judgment on their own [7]. Second, they must believe that others may be knowledgeable about the correct judgment [7]. These properties suggest testing our hypotheses with datasets where people have some **general familiarity with the domain**, but are not totally unfamiliar nor too familiar. In addition, **the accuracy of the dataset should be perceived as subject to uncertainty** [6] so that other people's expectations are used as a guide to interpret the data accurately [10].

The voting turnout dataset that Kim et al. [24] identified as moderately familiar to a Mechanical Turk population as the base data of the visualization fulfills both of these requirements. This dataset consists of predictions of voter turnout and the percentage who voted for the Republican candidate John McCain in the 2008 Presidential election by different states, ethnicities (e.g., White, Hispanic) and income brackets (under \$75K, over \$75K) [13].

Perceived uncertainty can be affected by multiple factors. From a data perspective, the time frame to which the data pertains (e.g., data about past vs. future events) and the degree to which the data has been transformed (e.g., raw measures vs. aggregated measures vs. output of the model) might affect the viewers' perception. From the viewer's perspective, statistical literacy can affect how critical a viewer is in judging the uncertainty of the data [12]. To make it more likely that all participants in our study perceived uncertainty in the data regardless of their data and statistical literacy, we chose a data framing that would clearly convey potential uncertainty. We framed the dataset as the output of a predictive model that estimates the percentage of voters who will vote for the Republican candidate by ethnicities (White, Hispanic) and income brackets in the 2020 presidential election. We visualized this base data in a line graph following prior work [24].

We hypothesize that the effects of congruency and the degree of consensus in the social information will vary depending on the degree to which a viewer's initial expectations align with the trend of the base data. The dataset we selected allows us to observe participants in cases where initial expectations are aligned and cases where initial expectations are not aligned with the base data. The most salient features of line graphs tend to be the direction of the relationship between the variables plotted on the x-axis and y-axis (e.g., as x increases, y increases) [5, 32, 40]. We can therefore expect that the viewer will perceive a greater alignment (or misalignment) when the slope directions of initial expectations are aligned (or misaligned) with the *slope directions* of the the base data, compared to when the *intercept* of initial expectations is aligned (or misaligned) with the base data. The majority of participants in Kim et al.'s study (77.8% out of 207 participants) expect that people will vote more for the Republican candidate if they are from the higher income bracket, regardless of ethnicity. By selecting two states with opposite trends between income level and Republican voting, we sought to ensure that the enough participants for the both cases where the slope directions of the base data align and do not align with the participants prior expectations. We chose two states (CO, NJ) that exhibited different patterns across income levels (Fig. 4).

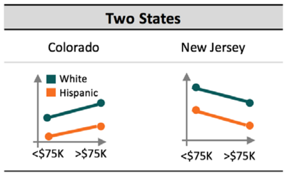

Fig. 4. Participants are asked to examine either the Colorado or New Jersey dataset.

4.2 Creating Social Information Stimuli

To create realistic stimuli for the social conditions (Fig. 6), we conducted a preliminary study on Amazon Mechanical Turk (AMT), in which participants were asked to draw their expectation of voter turnout for White and Hispanic voters in two states (NJ, CO). We recruited 300 participants, rewarding their participation with \$0.20. The average time used to complete the task was 3.1 minutes (SD=2.5).

In creating social stimuli, we aimed to (1) keep the quantity of social information consistent across conditions, (2) control the direction of the predominant trend implied by the social lines (congruent/incongruent to the base data), (3) create realistic stimuli, and (4) represent the raw distribution collected from the 300 participants as much as possible. To do so, we first divided all lines collected through the preliminary survey (see Fig. 6(a)) into two groups based on whether the direction of slope is congruent with the slope of the base data (congruent group, incongruent group) (Fig. 6(b)). Then we calculated the quartiles of intercept and slope of all lines within each group, and categorized them into 16 cells based on their intercept and slope (Fig. 5). To create the low consensus conditions, we sampled 30 lines from all 16 cells, maintaining the original distribution across cells. To create stimuli for the high consensus conditions, we sampled only from the cells where both the intercepts and slopes of the lines fell into mid-range quartiles (Q2 and Q3). To make the stimuli more realistic (i.e., to avoid conveying an unlikely perfect agreement in overall trend), we randomly chose 3 of the lines (10% of the lines from the group) from the other group (e.g., from the incongruent group if the stimuli was intended for one of the congruent conditions). We found through experimenting with different percentages that 10% resulted in realistic stimuli while still conveying a clear overall trend. To select three lines that opposed the overall trend while still being representative of the original distribution, we minimized the KL divergence, a measure of similarity between two probability distributions, between the original set of the lines from 300 participants and sampled 30 + 3 lines until the KL divergence did not improve by 0.001 over a successive iteration (Fig. 6(c), (d)).

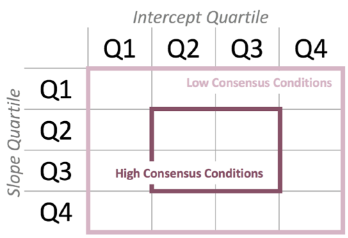

Fig. 5. Quartile criteria to create social stimuli for high consensus and low consensus conditions.

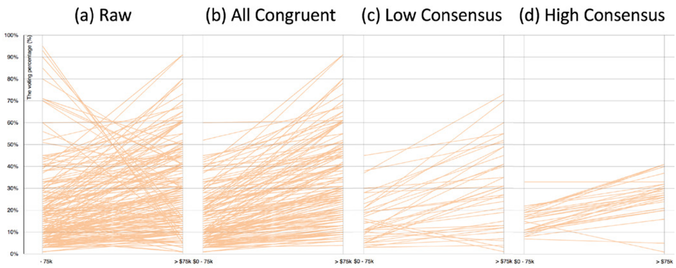

Fig. 6. Creating social stimuli for Hispanic voters in Colorado.

4.3 Participants

To determine the study sample size, we first recruited 10 participants per condition and used this pilot data to perform a prospective simulation-based power analysis. To achieve 80% power under $\alpha = 0.05$, we recruited 26 new participants for each of the eight Social conditions and 62 new participants for each of the two Social-Absent conditions for a total of 432 participants.

We posted the study on Amazon's Mechanical Turk, making it available to U.S. workers to control the quality of responses [33] and to ensure that people had a certain familiarity with the U.S.-specific topic of our dataset as shown in Kim et al. [24]. We instructed workers not to take our study more than once. Participants were randomly assigned to one condition and compensated with \$1.50.

4.4 Procedure

Fig. 7 shows an overview of the study procedure. Participants first read an introduction where we described the domain of the base data as the percentage of voters of two different ethnicities (White, Hispanic) in the assigned state (Fig. 7(a)). Participants then watched a tutorial video describing how to add and adjust a value in the visualization directly.

All participants then were prompted to express their initial expectations of the base data by estimating four data points in the visualization (High income-White, Low income-White, High income-Black Low income-Black) for the assigned state (Fig. 7(b-1)). On the same page, participants were asked to rate how confident they felt in the accuracy of their estimates for the four points they provided, on a scale from 0 to 100 (Fig. 7(b-2)).

On the next page, they were asked to examine the base data. If participants were assigned to one of the Social conditions they saw the corresponding social information; this social information was not shown in the Social-Absent conditions (Fig. 7(c)).

After examining the visualization, participants were asked to express their updated expectations. To reduce error due to participants trying to remember the values as they entered their updated expectations, we asked participants to add their updated expectations directly on top of the stimuli. This enabled participants to input values relative to the values of stimuli (Fig. 7(d-1)). On the same page, participants

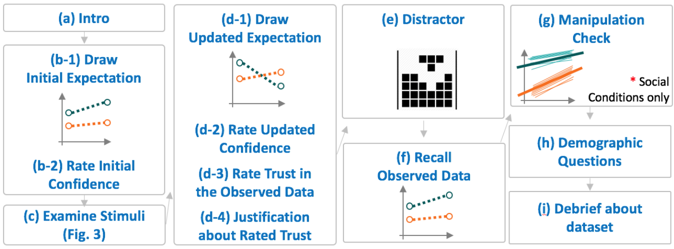

Fig. 7. Overview of the study procedure. Participants are asked to provide their initial expectation of the base data before seeing the data regardless of the assigned condition. After examining the stimuli based on the condition, they are asked to play Tetris for a minute as a distractor task. Then they provide recall values and draw their best estimate of the average line summarizing the social information for each ethnicity if they are in one of the Social conditions.

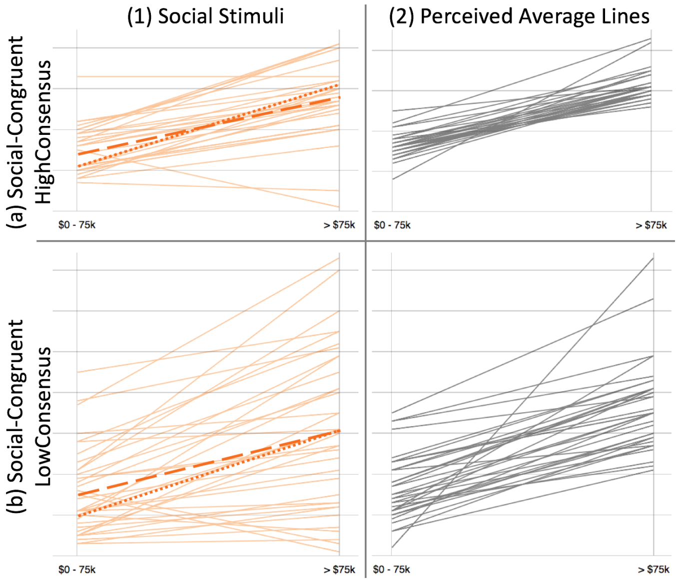

Fig. 8. An example of social information for Hispanic voters in Colorado shown to participants in the Social-Congruent conditions (Column 1) alongside the lines participants drew when asked to draw a single line summarizing others' expectations. A dashed orange line connects the mean value of all data points in the $0-75K income bracket and the mean value of all data points in the >$75K income bracket. The dotted line is the single closest line to every other line.

were asked to rate again how confident they felt in the accuracy of their updated estimates, on a scale from 0 to 100 (Fig. 7(d-2)). Participants' initial confidence value was used to initialize the slider value so that they could consider their updated confidence relative to their initial confidence. Participants were also asked to rate how much they trusted that the base data was accurate, on a scale from 0 to 100 (Fig. 7(d-3)). To encourage thoughtful ratings, we asked them to also type in a text justification of their trust rating (Fig. 7(d-4)).

All participants were asked to play Tetris for one minute as a distractor task (Fig. 7(e)). To enforce the one minute break, participants could only continue to the next page after the full minute had passed. After playing Tetris, participants in all conditions were asked to recall the base data (Fig. 7(f)).

To ensure that participants perceived the congruency of the social information as we intended, we required all participants in the Social conditions to draw a single line that they thought best summarized the social information (Fig. 7(g)).

Participants were asked to respond to demographic questions including age, education, gender and ethnicity (Fig. 7(h)). Lastly, we debriefed participants on the true description of the dataset (Fig. 7(i)).

5 Data Preliminaries

We excluded 11 participants from the analysis (out of 432) who participated in the task multiple times. The average time to complete the task was 15.3 minutes (SD=7.5). There was no difference in task completion time across the conditions (F(11)=1.33, p=.205)). There were also no differences in self-reported demographics (age, education, gender, ethnicity) and relevant experience (e.g., chart experience) across the conditions (see supplementary materials for a detailed breakdown).

To check whether participants in the Social conditions perceived the predominant trend in the social information as we had intended (i.e., whether the slope increased or decreased), we analyzed the slope of the line provided by participants after they were asked to draw a single line that best summarized the social information. Participants correctly identified the direction of the social information in 96.3% of all cases (266 out of 276 participants who were assigned to the Social conditions). The lines drawn by participants in the Social-HighConsensus conditions (e.g., Fig. 8(a-2)) had a lower variance of intercept (M=7.6 voting percent, SD=4.6) and slope (M=6.9, SD=5.0) than the lines drawn by participants in the Social-LowConsensus conditions (intercept: M=16.0, SD=10.8; slope: M=17.4, SD=12.7) (e.g., Fig. 8(b-2)).

5.1 Dependent Variables

To evaluate our hypotheses, we defined three dependent variables, capturing participants' ability to recall the base data, trust in the accuracy of the base data, and whether the trend of participants' updated expectations aligned with that of the base data.

Recall Accuracy: To quantify how accurately participants recalled the base data, we squared the distance between the value the participant recalled and the value of the base data.

Trust: Trust was analyzed using the trust rating, which participants self-reported on a scale from 0 to 100.

Updated Trend Alignment: We used a binary variable to capture whether the trend (increasing or decreasing) of a participant's updated expectations matched that of the base data (0=not matched, 1=matched).

5.2 Independent Variables

We dummy coded the following factors describing conditions:

Social or Social-Absent: Whether participants were assigned to the Social conditions or the Social-Absent conditions. 1=Social, 0=Social-Absent.

Congruency: Whether participants were assigned to the Social-Congruent conditions or the Social-Incongruent conditions. 1=Congruent, 0=Incongruent.

Consensus: Whether participants were assigned to the Social-HighConsensus conditions or Social-LowConsensus conditions. 1=High Consensus, 0=Low Consensus.

Because prior work shows that the values that participants recall can be influenced by their prior expectations when viewing a visualization [24], we added the distances between a participant's expectations and the base data as a covariate when analyzing recall accuracy:

Initial Expectation-Data Gap: The distance between the participant's initial expectation and the base data. We calculated the distance for each data point (out of 4 total) in the visualization.

Updated Expectation-Data Gap: The distance between the participant's updated expectation and the base data. We calculated the distance for each point (out of 4 total) in the visualization.

5.3 Analysis Approach

We conducted a series of regression analyses to analyze the study data. If a dependent variable resulted in a single response per participant (e.g., a participant rated trust in the accuracy of the base data once in the task), we used a linear model to estimate the effect size and the p-value using the lm function in R. We report these results using F-statistics and p-values. If a dependent variable resulted in multiple responses per participant (e.g., a participant recalled four data points (*White-under\$75K income bracket, White-over\$75K, Hispanic-under\$75K, Hispanic-over\$75K*)), we used a mixed-effects model (implemented in R's lme4 package [3]) with participant id as a random effect. To calculate p-values of fixed effects in all mixed-effects models, we used the normal approximation method using the t-value provided by lme4 [3]. If the dependent variable consisted of a binary value, we used the glmer function in the lme4 package and reported the odds ratio (OR).

6 Core Results

6.1 H1: Social Information

Contrary to H1, we observed no overall difference in recall accuracy between participants in the Social conditions and those in the Social-Absent conditions ($t = 1.03, p = .298$, Fig. 9). In other words, the presence of *any* social information was not necessarily enough to improve recall of the data.

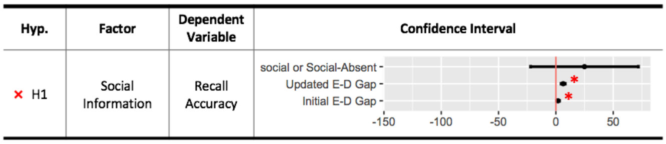

Fig. 9. Confidence intervals for the independent variables that accounted for the participant's recall accuracy. Our results show that H1 is unsupported: Having social information in the visualization did not affect the recall accuracy, but participants' initial expectations and their updated expectations did.

Recall accuracy was however affected by a participant's Initial Expectation-Data Gap and Updated Expectation-Data Gap. In line with the effect of prior knowledge on data recall when viewing a visualization observed by Kim et al. [24], we found a weak positive correlation between the values of a participant's initial expectation and the values they recalled (R^2=0.23, intercept=7.24, slope= 0.75). We also saw a weak positive correlation between the values of a participant's updated expectation and the values they recalled (R^2=0.27, intercept=4.56, slope=0.56).

6.2 H2: Congruency

In line with H2's expectation, participants in the Social-Congruent conditions self-reported a higher trust in the accuracy of the base data by 6.2 (out of 100) compared to those in the Social-Incongruent conditions ($F(1) = 4.22, p < .05$, Fig. 10 (a)).

The trend of the updated expectations of a participant in the Social-Congruent conditions was more likely to align with the trend of the base data than in the Social-Incongruent conditions ($OR = 16.70, t = -2.84, p < .01$, Fig. 10 (b)).

6.3 H3: The Degree of Social Consensus

6.3.1 H3a: Effects on Trust and Updated Beliefs

Contrary to H3a, we did not find an effect of the degree of social consensus on trust and the updated trend alignment (Fig. 11(a), (b)).

We found no difference in participants' trust in the accuracy of the base data ($F(1) = 0.002, p = .967$) nor the likelihood the trend of participants' updated expectations was to align with the trend of the base data ($OR = 2.38, t = 0.68, p = .499$) between the Social-Congruent-HighConsensus and the Social-Congruent-LowConsensus conditions.

We also did not find a difference in participants' trust in the accuracy of the base data ($F(1) = -0.65, p = .420$) nor the likelihood

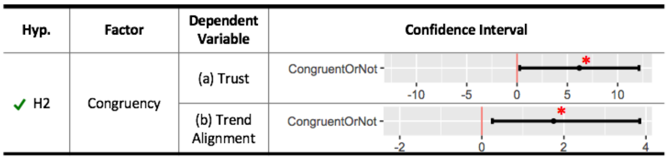

Fig. 10. Confidence intervals show the effect of congruency on the participant's trust in the accuracy of the base data and the likelihood to update the participant's expectations to match with the base data (H2). If others' expectations depicted the same trend as the base data, participants tended to trust the accuracy of the base data more and were more likely to update their own expectations to match to the base data than if their expectations did not follow the same trend as the base data.

the trend of participants' updated expectations was to align with the trend of the base data ($OR = 0.84, t = -0.14, p = .888$) between the Social-Incongruent-HighConsensus and the Social-Incongruent-LowConsensus conditions.

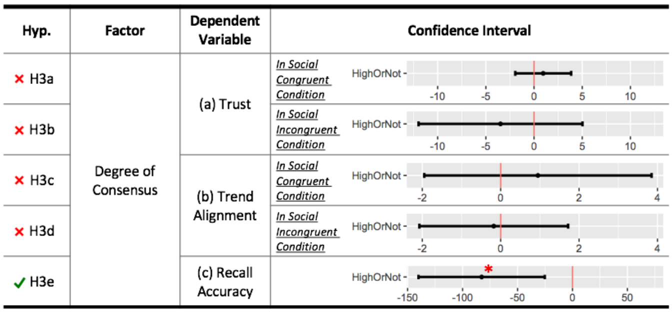

Fig. 11. Confidence interval that show the effect of the degree of consensus (H3) on the participant's trust in the accuracy of the base data and the likelihood to update the participant's expectations to match to the base data. The degree of consensus did not affect how participants trusted the accuracy of the base data nor whether they updated their expectations toward the base data. It did, however, positively affect their recall accuracy.

6.3.2 H3b: Effect on Recall

As hypothesized in H3b, participants in the Social-HighConsensus conditions recalled the base data significantly more accurately (by 3.1 voting percent) than participants in the Social-LowConsensus conditions (Fig. 11(c); $t = -2.81, p =< .01$).

To better understand whether social information plays a different role in each condition, we built two separate mixed-effect models for participants in the Social-HighConsensus and in the Social-LowConsensus conditions. These models show that social information increased a participant's ability to recall the base data when the social information implied high consensus ($t = -2.03, p < .05$), and decreased a participant's ability to recall the base data when the social information implied low consensus ($t = 3.04, p < .01$).

6.4 Alignment of Initial Expectations' with Base Data

The above results indicate that congruency between the social signal and the data signal can have a reinforcing or challenging influence on the base data (H2). To unpack this result further, we investigated how this effect varied by the degree of agreement between participants' initial expectations and the base data, since a participant's own beliefs are likely to influence their interpretation as well. We defined "initial agreement" as "the trend alignment of participants' initial expectations with the base data".

Specifically, we compared the effect of congruency in two cases: those where the participant's initial expectations were *highly aligned* with the base data (Data Agreement), and those where the participant's initial expectations were *not aligned* with the base data (Data Disagreement).

Raw Count (percentage)		(a) Both ethnicities	One of either ethnicity	(b) None of ethnicity
CO		149 (69.6%)	32 (15.0%)	33 (15.4%)
NJ		24 (11.6%)	28 (13.5%)	155 (74.9%)

Fig. 12. The raw count and proportion whose trends of the initial expectations of each ethnicity are aligned with the base data.

We subset the study results by how many of the trends (out of two total lines) of participants' initial expectations aligned with the base data (2 or 0, see Fig 12(a) and (b)). Participants were assigned to a Data Agreement group and a Data Disagreement group, respectively. We excluded those participants whose initial expectations aligned with one trend (either Hispanic or White) of the base data in this analysis (14.3% out of 421).

Participants Initial Expectation	Congruency	(a) Trust in Accuracy of Observed data	(b) Trend Alignment between Updated Expectations and Observed data	(c) Movement of Updated Expectations Toward Observed Data	(d) Changes in Confidence
Data Agreement group	Congruent	No effect of Congruency	No effect of Congruency	No effect of Congruency	No effect of Congruency
	Incongruent				
Data Disagreement group	Congruent	Rated higher by 9.5 (M = 42.65,SD = 25.60)	More likely to align with the trend of the observed data (OR=3.21)	More likely to move toward the observed data (OR=1.11)	Less increased confidence (M = 0.65,SD = 12.72)
	Incongruent	Rated lower by 9.5 (M = 52.17,SD = 25.57)	Less likely to align with the trend of the observed data	More likely to persist in their initial expectations	More increased confidence (M = 6.65,SD = 11.07)

Fig. 13. The effect of congruency based on alignment between a participant's initial expectations and trends in the base data. The data in this figure is based on the Colorado dataset. For participants who are assigned to the New Jersey dataset, the trends in participants' initial expectations in each alignment group, the trend of the base data, and that of the social information are reversed.

6.4.1 Trust

Overall, as expected we found that that whether a participant's initial beliefs aligned with the data impacted trust in the accuracy of the data. Participants who were in the Data Disagreement group reported lower trust in the accuracy of the base data when they were in the Social-Incongruent Conditions, compared to when they were in the Social-Congruent Condition (Fig 13(a)).

Among participants in the Data Agreement group, we found no difference in self-rated trust between participants in the Social-Congruent conditions and those in the Social-Incongruent conditions ($t = 0.49, p = .625$). This result suggests that when one is already predisposed to agree with data, the agreement of others is less influential on one's interpretation.

Among participants in the Data Disagreement group, however, we found that participants in the Social-Incongruent conditions trusted the data significantly less by 9.5 out of 100, compared to participants in the Social-Congruent conditions ($t = 2.08, p < .05$). In other words, if one already disagrees with a data set, seeing that others also disagree may be interpreted as further validation of one's own initial beliefs.

6.4.2 Trend Alignment between Updated Expectations and Base Data

Participants who were in the Data Disagreement group were more likely to update their expectations to be aligned with the trend of the base data when they were in the Social-Congruent conditions, compared to when they were in the Social-Incongruent conditions (Fig 13(b)).

Among participants in the Data Agreement group, we found no difference between participants in the Social-Congruent conditions and the Social-Incongruent conditions in how likely the trend of participants' updated expectations was to align with the trend of the base data ($OR = 3.21, t = 0.68, p = .499$).

Among participants in the Data Disagreement group, however, the trend of Social-Congruent participants' updated expectations was more likely to align with the trend of the base data than that of the Social-Incongruent conditions ($OR = 9.34, t = 2.18, p =< .05$).

6.4.3 Movement of Expectations Toward Base Data vs. Initial Expectations

We analyzed two more dependent variables to corroborate our observations of how participants' selectively updated their beliefs depending on how well their initial beliefs aligned with the data. To observe how likely the participants were to be persuaded by the base data relative to stay true to their own initial expectations, we used a binary variable to capture whether participants' updated expectations were closer to their initial expectations (0) or closer to the base data (1). While the trend alignment between participants' updated expectations and the base data captures how trustworthy participants considered the base data to be, this additional, binary variable measures the relative influence of participants' initial expectation compared to that of the base data on their updated expectations.

Overall, participants who were in the Data Disagreement group were more likely to persist in their initial expectations when the trend of the social information aligned with that of their initial expectations (Social-Incongruent conditions). If the trend of the social information did not align with participants' initial expectations (Social-Congruent conditions), they were more likely to move toward the base data than toward their initial expectation (Fig 13(c)).

Specifically, participants in the Data Agreement group were no more likely to move toward the base data if they were in the Social-Congruent conditions than if they were in the Social-Incongruent conditions ($OR = 0.92, t = -1.59, p = .113$).

Among participants in the Data Disagreement group, however, those in the Social-Incongruent conditions were significantly less likely to move toward the base data compared to participants in the Social-Congruent conditions ($OR = 1.11, t = 2.07, p =< .05$).

6.4.4 Change in Confidence after Exposure to Stimuli

To examine whether participants adjusted their confidence in their own expectations after seeing the social information, we calculated the change in a participant's self-rated confidence before (Fig. 7(b-2)) and after (Fig. 7(d-2)) seeing the social information by subtracting the initial confidence from the updated confidence. By analyzing this variable, we can observe whether viewing social information increases or decreases participants' confidence in the accuracy of their expectations.

We found that participants who were in the Data Disagreement group showed a greater increase in confidence from initial to updated expectations when they were exposed to social information that aligned with their initial expectations (Social-Incongruent conditions), compared to when they were exposed to social information that was not aligned with their initial expectations (Social-Congruent conditions) (Fig 13(d)).

Specifically, in the Data Agreement group, we found no difference in the amount of change in confidence between participants in the Social-Congruent conditions and those in the Social-Incongruent conditions ($t = 0.01, p = .994$).

In the Data Disagreement group, however, participants' increase in confidence in the Social-Incongruent conditions was on average 6.0 points higher, compared to participants in the Social-Congruent conditions ($t = 3.86, p < .001$).

7 Discussion

Our study demonstrates how visualizing others' expectations of data can impact people's ability to recall data, their trust in the accuracy of the data, and their updated beliefs after viewing the data. Visualizing social information led to improved recall when others' expectations exhibited a high degree of consensus (low variance) compared to presenting the base data alone. Contrary to this result, participants who were exposed to others' expectations which exhibited a low degree of consensus (high variance) recalled the data less accurately than participants who were only exposed to the base data. This result suggests that social information can focus or distract a viewer's attention, depending on the visual complexity it adds to a visualization.

Our work also demonstrates the effect of congruency between the predominant trend in the base data and social information, and a viewer's initial expectations, to either reinforce or challenge the validity of the base data. We found that if participants disagreed with the trend of the base data initially, they were more susceptible to believe the social information. If a participant and other people disagreed with the trend of the base data, the participant was less likely to trust its accuracy, was more likely to stick with their initial expectations, and their updated expectations were less aligned with the trend of the base data, though with more confidence. These patterns suggest that social information can serve a tendency toward confirmation bias [28]. On the other hand, if participants disagreed with the trend of the base data but others' expectations agreed with it, they were more likely to trust that the base data is accurate, their updated expectations were more likely to match the trend of the base data, and they were more likely to update their expectations in the direction of the base data, though with less confidence in their expectations.

The influence we observed of social information on participants' beliefs suggests the potential for social information to act as a rhetorical device when integrated in a visualization [21]. In prior work, Pandey et al. [29] found that a person who has a polarized belief is less likely to be persuaded by the visualized data. Our findings similarly account for the role of a user's initial beliefs in examining a visualization, and describe how social information can persuade a user to update their opinion. By manipulating the perceived data uncertainty, the level of social consensus, and the congruency between the base data and the social information, we shift how the data is presented in ways that subsequently impact interpretations. Through further experimentation around these possible rhetorical functions of social information, visualization researchers can work toward ways of helping designers as well as users recognize the power of social information so as to design and interpret visualizations that contain social signals with greater critical awareness. Our findings indicate that designers should be particularly aware of the potential of social information to change beliefs when they expect that what people would predict opposes the main trend that a visualization shows. In cases where the social information opposes the data but is still a desirable part of the design, the designer can leverage other social influences. For example, if the data was collected and analyzed by an expert, emphasizing this expertise may counteract the loss of perceived credibility inspired by the contradictory social information [11]. Similarly, increasing the overall clarity around the dataset (e.g., describing how the data is collected and analyzed) can make the viewer less susceptible to the social information [2].

Overall, our work shows that social information is neither inherently positive or negative but can serve different functions depending on the trend of visualized data and social information, and the views of the person interacting with the visualization. Based on the challenging and reinforcing roles we observed, integrating social information in visualizations could be a promising way to enhance data literacy, by prompting people to think more critically about data and the overall set of evidence behind it. Designing for "socially-aware visualizations" that prompt critical thinking while avoiding negative influence is an exciting area for future work.

7.1 Limitations

We did not observe an effect of the degree of social consensus on people's likelihood to reinforce their opinion or challenge the base data. This result may be due to how we operationalized the degree of consensus. In designing stimuli for the Social-LowConsensus conditions, we mainly sampled social information from each group of lines that shared the predominant trend we intended (i.e., increasing, decreasing), to avoid entangling congruency and the degree of consensus. This process resulted in having many lines with highly varied intercepts but a single major trend. As a result, participants may have perceived the stimuli as having a relatively strong degree of consensus in both the Low and High Consensus conditions.

Our work focused on understanding the influence of social information on how participants interpret the base data. However, beyond confirming that participants perceived the social information in line with our congruency manipulation, we did not measure how participants perceive the social information. For example, is the perceptual averaging that participants engage in when presented with a collection of social information different than that used when the data does not represent others' beliefs? Future work should evaluate how participants interpret social information, how much they trust the accuracy of the social information, and how much they update their expectations toward the social information.

We collected social information from AMT workers and framed social information as "other peoples' expectations" in our study. We did not vary other properties of the social information (e.g., subset by groups) nor personalize the social information (e.g., people from your identified party), which may affect on the participant's perceived distance to others who provide the social information. Future work needs to evaluate how different formulations and framings influence data interpretation and perception.

We intentionally designed a scenario that suggested some uncertainty around the presented dataset. It is possible that our results do not generalize to datasets that imply low uncertainty. Future work should attempt replicating these results with visualizations of data that imply greater inherent validity (e.g., data describing an event that already took place). While some effects we observed in our study may be lessened (e.g., the effect of social information reinforcing or challenging the data) we expect that other effects will emerge that are worth exploring, such as the potential for social information to encourage discussion or prevent belief polarization.

8 Design Space

Prior work provides design considerations for social visualizations [14]. Various systems have also explored particular strategies for integrating comments and the user's interaction history (e.g., [16], [39]). However, the design space for visualizing *social information in the form of others' expectations about data* has not been carefully explored. In visualizing social information alongside the data, a designer must choose *which* social information she wants to present, *how* to visualize the social information within the context of the visualization, and *when* to present the social information. Based on our study and principles from research on social information and social influence [6], we characterize a design space for visualizing data alongside others' expectations.

8.1 Representing Social Information

8.1.1 Types of Expectations

In our experiment, we collected people's expectations of a trend before they had seen the base data to use as social information. However, as an alternative to visualizing peoples' *a priori expectations of data* such as in our study, a designer may instead collect and represent peoples' *updated (posterior) expectations* after the base data has been presented.

The effect of being exposed to others' a priori expectations versus posterior expectations in a visualization may differ. The user may perceive prior expectations of others as the general domain knowledge of others rather than dataset-specific knowledge. In this case, the social information may prompt reflection on one's own beliefs but have lesser influence on one's own updated expectations. Distinctive patterns among people's expectations can inform the user of shared biases about particular datasets or domains.

When others' updated expectations are visualized, the user may consider the social information to be the outcome of others' deeper reflection on the dataset or domain, such that influence is more likely. If the visualization shows others' updated expectations to a user as they view the base data, and before eliciting the user's own updated expectations, and the domain is unfamiliar to many users, the visualized expectations may be less reliable as a record of what people truly believe. Dynamics resembling information cascades, where a few initial opinions take hold across a group (e.g., [20]), can result in the impression of shared biases.

In deciding what form of peoples' expectations to use, a designer may wish to take into account peoples' expected familiarity with a given dataset or domain, as well as the level of objectivity or "facticity" of the data. If the data is highly objective, displaying a priori expectations can highlight how accurate peoples' general knowledge on a topic is, since posterior expectations might not be any more informative than the data itself. If the data has low familiarity, the designer can purposely choose to show a priori expectations to emphasize peoples' incorrect intuitions about the topic. If people are expected to be relatively familiar with the topic but the dataset is not necessarily highly factual (e.g., predictions of which team will win a sports title based on a small amount of prior data), showing posterior expectations may best display "the wisdom of the crowd", by showing what information people can add to the base data.

8.1.2 Data Transformation and Presentation

A designer can choose to present social information with varying **levels of granularity or aggregation**, from the raw data points to a single representative data point. If the designer wants to show the full range of variance observed in collected social information, she can present the raw social information without any aggregation. For example, Fig. 6 (a) presents the raw social data we collected representing expectations of the voting percentage among Hispanic voters who will vote for the Republican Candidate in Colorado in the 2020 election.However, unaggregated social information may also distract the user from perceiving any trends or patterns in the social information or the visualized data itself. As we demonstrate in our study, showing highly varied social information reduces the user's ability to recall the base data later. We also observed that people did not accurately summarize multiple lines in an average line with the high variance (Fig. 8 (b-2)).

An alternative to showing unaggregated social information is to present one or few representative data points. The designer can calculate the mean or median of all data points or select a subset of the social data that suggests a representative pattern: for example, she might identify the data point(s) that minimizes the distance to the rest of the data. Variance can be visualized as an additional layer of context with a representative data point, without necessarily resorting to visualizing raw data. In the New York Times "You Draw It" [1], social information is shown in a heatmap, where the representative pattern is mapped to the position and the variance is mapped to the alpha of gradient.

Whether raw social information is depicted or aggregation is used, the designer faces a choice of whether to visualize or omit **outliers**. Research in social influence indicates benefits to showing minority opinions: specifically, doing so can stimulate divergent and creative thinking [27], if the minority patterns are systematically different from those of the majority.

When presenting social information in an aggregated form, a designer can **cluster the social information**, to emphasize a majority belief or set of more prevalent beliefs (e.g., Fig. 6 (b),(c)). Viewing multiple common beliefs may trigger the user to compare her own views to those represented by each cluster, which may lead a deeper engagement via social comparison [10].

Where user profile information is available, the social information can be **stratified by demographic properties** and visualized. For example, if the designer wants to emphasize the difference in estimation patterns by gender, she can elicit or predict users' gender identities and differentiate the social information using labels. Social psychology research indicates that people are more likely to associate and relate with similar others (homophily) [23], and more likely to adjust their own opinion to match those of others like themselves [10]. Given information about a user, the designer may further emphasize the groups that are predicted to be associated with her, such as by highlighting or annotating to draw attention.

8.2 Interaction between the User's and Others' Expectations

When an interface elicits each user's expectations prior to presenting the base data and social information, there is an opportunity for **personalized interaction through social comparison**. For example, the New York Times example [1] presents the user's expectation against the social information and the base data, along with text annotations such as "Not the worst!" to stimulate comparisons with the group. Personalized feedback can be formulated based on the social information (e.g., "There are 80% people who agree with what you think of the trend should be.", "You think very differently than any other people who estimate this dataset".) or based on the base data and social information (e.g., "You are in the 1st quartile in terms of estimating the base data accurately"). Such feedback can intrinsically motivate people to investigate more data [30].

Alternatively, the designer may choose to present the social information while eliciting a user's expectations. If the social information is presented before the base data, influence is likely to be greatest, as the user will likely experience more uncertainty about the data she is asked to predict.

9 Conclusion

Our work explored the possibility of integrating social information in the form of other peoples' expectations about a dataset into a visualization. Informed by prior work in social influence and data interpretation, we formulated and evaluated a set of hypotheses regarding the effect of such social information on recall, trust in the accuracy of the data, and belief updating. We varied different properties (congruency, degree of consensus) of the social information for further comparison. We observed that social information that exhibits a high degree of consensus leads participants to recall the data more accurately relative to participants who were exposed to the data alone. We also found that people are more susceptible to social information when they initially disagree with the presented data. Our findings suggest new opportunities for integrating a layer of social perspective into data visualizations.

References

[1] G. Aisch, A. Cox, and K. Quealy. You draw it: How family income predicts children's college chances, 2015.

[2] S. E. Asch. Effects of group pressure upon the modification and distortion of judgments. *Groups, leadership, and men*, pages 222–236, 1951.

[3] D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.

[4] S. K. Card, J. D. Mackinlay, and B. Shneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

[5] C. Carswell, C. Emery, and A. M. Lonon. Stimulus complexity and information integration in the spontaneous interpretations of line graphs. *Applied Cognitive Psychology*, 7(4):341–357, 1993.

[6] R. B. Cialdini. *Science and practice*. Pearson, 2001.

[7] M. Deutsch and H. B. Gerard. A study of normative and informational social influences upon individual judgment. *The journal of abnormal and social psychology*, 51(3):629, 1955.

[8] 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.

[9] L. Festinger. Informal social communication. *Psychological review*, 57(5):271, 1950.

[10] L. Festinger. A theory of social comparison processes. *Human relations*, 7(2):117–140, 1954.

[11] J. R. French, B. Raven, and D. Cartwright. The bases of social power. *Classics of organization theory*, 7, 1959.

[12] I. Gal. Adults' statistical literacy: Meanings, components, responsibilities. *International statistical review*, 70(1):1–25, 2002.

[13] Y. Ghitza and A. Gelman. Deep interactions with mrp: Election turnout and voting patterns among small electoral subgroups. *American Journal of Political Science*, 57(3):762–776, 2013.

[14] J. Heer and M. Agrawala. Design considerations for collaborative visual analytics. *Information visualization*, 7(1):49–62, 2008.

[15] J. Heer and D. Boyd. Vizster: Visualizing online social networks. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pages 32–39. IEEE, 2005.

[16] J. Heer, F. B. Viégas, and M. Wattenberg. Voyagers and voyeurs: supporting asynchronous collaborative information visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1029–1038. ACM, 2007.

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

[18] M. Hegarty, S. Kriz, and C. Cate. The roles of mental animations and external animations in understanding mechanical systems. *Cognition and instruction*, 21(4):209–249, 2003.

[19] M. Hegarty and K. Steinhoff. Individual differences in use of diagrams as external memory in mechanical reasoning. *Learning and Individual differences*, 9(1):19–42, 1997.

[20] J. Hullman, E. Adar, and P. Shah. The impact of social information on visual judgments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1461–1470. ACM, 2011.

[21] J. Hullman and N. Diakopoulos. Visualization rhetoric: Framing effects in narrative visualization. *IEEE transactions on visualization and computer graphics*, 17(12):2231–2240, 2011.

[22] S. L. Jarvenpaa. Graphic displays in decision making the visual salience effect. *Journal of Behavioral Decision Making*, 3(4):247–262, 1990.

[23] D. B. Kandel. Homophily, selection, and socialization in adolescent friendships. *American journal of Sociology*, 84(2):427–436, 1978.

[24] 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*. ACM, 2017.

[25] 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.

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

[27] C. J. Nemeth and J. Wachtler. Creative problem solving as a result of majority vs minority influence. *European Journal of Social Psychology*, 13(1):45–55, 1983.

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

[29] 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*, 20(12):2211–2220, 2014.

[30] K. Reinecke and K. Z. Gajos. Labinthewild: Conducting large-scale online experiments with uncompensated samples. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 1364–1378. ACM, 2015.

[31] 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.

[32] P. Shah, R. E. Mayer, and M. Hegarty. Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *Journal of Educational Psychology*, 91(4):690, 1999.

[33] A. D. Shaw, J. J. Horton, and D. L. Chen. Designing incentives for inexpert human raters. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pages 275–284. ACM, 2011.

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

[35] F. B. Viegas, M. Wattenberg, F. Van Ham, J. Kriss, and M. McKeon. Manyeyes: a site for visualization at internet scale. *IEEE transactions on visualization and computer graphics*, 13(6), 2007.

[36] M. Wattenberg. Baby names, visualization, and social data analysis. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pages 1–7. IEEE, 2005.

[37] M. Wattenberg and J. Kriss. Designing for social data analysis. *IEEE transactions on visualization and computer graphics*, 12(4):549–557, 2006.

[38] W. Willett, J. Heer, and M. Agrawala. Scented widgets: Improving navigation cues with embedded visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1129–1136, 2007.

[39] W. Willett, J. Heer, J. Hellerstein, and M. Agrawala. Commentspace: structured support for collaborative visual analysis. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 3131–3140. ACM, 2011.

[40] J. Zacks and B. Tversky. Bars and lines: A study of graphic communication. *Memory and Cognition*, 27:1073–1079, 1999.