The Curse of Knowledge in Visual Data Communication

Cindy Xiong, Lisanne van Weelden, and Steven Franconeri

Abstract—A viewer can extract many potential patterns from any set of visualized data values. But that means that two people can see different patterns in the same visualization, potentially leading to miscommunication. Here, we show that when people are primed to see one pattern in the data as visually salient, they believe that naïve viewers will experience the same visual salience. Participants were told one of multiple backstories about political events that affected public polling data, before viewing a graph that depicted those data. One pattern in the data was particularly visually salient to them given the backstory that they heard. They then predicted what naïve viewers would most visually salient on the visualization. They were strongly influenced by their own knowledge, despite explicit instructions to ignore it, predicting that others would find the same patterns to be most visually salient. This result reflects a psychological phenomenon known as the curse of knowledge, where an expert struggles to re-create the state of mind of a novice. The present findings show that the curse of knowledge also plagues the visual perception of data, explaining why people can fail to connect with audiences when they communicate patterns in data.

Index Terms—Cognitive biases, data communication, expertise, information visualization, perception and cognition.

✦

1 INTRODUCTION

Imagine a scientist showing experimental results at a conference or colloquium, or a data analyst updating the company leadership on recent customer feedback with a dashboard. These people are experts in their respective fields, yet they overwhelm their audiences with overly complex visualizations, delivered too quickly, oblivious to the fact that others do not see what they see. We replicated this phenomenon in the lab, providing empirical evidence for a 'curse of knowledge' in data visualization â once an expert recognizes a given pattern in data as visually salient, the expert assumes that it is also visually salient to naïve observers.

This 'curse of knowledge' is a well-studied psychological phenomenon that appears in many domains. Well-informed decision makers fail to predict the judgments of less-informed decision makers, implicitly allowing their own knowledge to guide those predictions [9]. People given disambiguating information about ambiguous sentences, like "the daughter of the man and the woman arrived," assume that the sentence would no longer be ambiguous to other naïve listeners [26]. When people have access to additional information, e.g. that a message is sarcastic, they tend to perceive ambiguous messages such as "that restaurant was marvelous, just marvelous" as sarcastic – but they also predict that other people would read the same tone [17].

In one particularly powerful demonstration, people were asked to tap the rhythm of a set of well-known songs, such as "Happy Birthday," on a desk, and listeners guessed the songs based on the recorded rhythm of the tappers [34]. Tappers estimated that listeners would identify around 50% of the songs, but in reality, listeners could only identify around 3%, revealing a vast overconfidence in how much information they communicated. The tappers 'filled in' missing information in their own heads, such as the pitches of the 'notes', and it appears impossible to turn off this filling-in process to simulate the experience of others. Taking a naïve perspective can be literally inconceivable [38].

This curse of knowledge has powerful consequences for communication, because people generally do not convey information to others if they assume that it is already shared [18]. Presenters must therefore have an accurate idea of what their audiences know and do not know, so that they can include only the information the audiences still need [20]. Unfortunately, this knowledge is too often not present or not leveraged. Even teachers misjudge their students' abilities and understanding, hindering effective instruction [1], [26], [43].

While the curse of knowledge is well-studied in the psychology of language, decision making and reasoning, there is less direct research on potential consequences for communication with data visualizations. Compared to numerical and textual formats, data visualizations are effective in highlighting the relationships and patterns in data to facilitate understanding [10]. But at the same time, understanding complex visualizations can be similar in time and effort to reading a paragraph [21], [27]. Critically, just as one can read many possible sentences from the paragraph, providing multiple perspectives on a topic, a graph or figure can be seen and interpreted in multiple ways depending on the how they select and interpret visual information over time [32], [40]. The present experiment demonstrates that different experience with a dataset can cause people to adopt a particular perspective, which can substantially change their predictions about what naïve viewers will find salient in a visualization.

Given the primary role visualizations play in the communication of analytic data across science, education and industry [28], [31], focusing on different patterns in the same dataset harbors the potential for miscommunications between the presenters and their audiences [17], [40], [45]. We suspect that the inability to separate

- *Cindy Xiong and Steven Franconeri are with Northwestern University. E-mail: cxiong@u.northwestern.edu, franconeri@northwestern.edu.*
- *L. van Weelden is with Utrecht University, Utrecht 3512, Netherlands. E-mail: l.vanWeelden@uu.nl.*

one's own knowledge and expertise from that of their audience can make visual data communication more difficult and less clear than presenters realize. This means that among the many features and patterns within a visualization, graph viewers could selectively focus on some while ignoring others, and in turn predict that naïve viewers would focus on the same feature and patterns.

Across four experiments, we demonstrate that the 'curse of knowledge' indeed extends to data visualizations. Knowledge, specifically, makes an expert recognize a given pattern in data as more visually salient, and the expert assumes that it is also visually salient to observers that they know to be naïve. Subsets of the present data and descriptive text were presented at the IEEE InfoVis 2017 DECISIVe Workshop [15] and as a poster at IEEE InfoVis in 2017 [44]. The present paper expands upon previous work by introducing two new experiments illustrating that this curse of knowledge effect can be further amplified by the addition of annotations on visualized data.

2 RELATED WORK

Existing work in cognitive psychology shows that the curse of knowledge bias can impact interpersonal communication [18]. The curse of knowledge can have particularly strong effects in children, who have more trouble inhibiting their own knowledge. In the 'Sally-Ann' task, children hear a story about Sally, who put her candy in a box before leaving the room. While she was gone, Ann removed the candy from the box and put it in a basket. Where will Sally look for the candy when she returns? Unable to inhibit their own knowledge of the illicit swap, most 4-year old children will assume that Sally will look in the basket [3], [36]. A modified 'Sally-Ann Task' targeting adults introducing several "boxes" and "baskets," demonstrated that adults also make this error with a more complex scenario and a subtler measure [4].

The curse of knowledge can also occur within a single person [30], in the form of 'hindsight bias'. This bias, studied in business decision making, political strategizing and marketing, is the irrational belief that an outcome was more predictable after it becomes known [38]. People seem unable to recreate the novel and uncertain feelings from their own mind prior to the revelation of the outcome [5], [11], [47].

Visualization researchers have recently become interested in decision biases, for example, the 'attraction effect', which is a cognitive bias where irrelevant information can influence decisions about otherwise equal alternatives, can influence decision making in visualized data [13], [22]. While a perfectly rational memory system should process or remember different types of information equally well, data visualizations can be more engaging and better remembered if they are distinctive, concrete, or look more like real-world objects [2], [6], [7], [8], [19]. Storytelling techniques adapted from journalism can influence the way people extract data from visualizations [23], [24], [33], [35], [39]. Still more work has sought evidence for whether a viewer's history of previously seen data visualizations can sway their perception of a subsequent, but unrelated, visualization [12], [42].

To our knowledge, the present paper provides the first examination of the curse of expertise in data visualization: whether a viewer's background knowledge will affect their predictions about what naïve others will see in a visualization.

3 GENERAL METHOD

Participants completed a Qualtrics [37] survey in which they read a story that conveyed background knowledge about a graph depicting political polling data. They were told that the experimenters will show the same graph they saw to 100 people, along with only the following short description – "in the months before the elections of 2014 in a small European country, a polling organization asked citizens about their voting intentions on a daily basis." They were then asked to predict what uninformed viewers (with no knowledge of the story) would find to be the most visually salient features or patterns in the graph. The participants then predicted a second most salient feature, up to a fifth most salient feature. We intentionally did not specify what types of "features or patterns" the participants should predict, and did not provide them with examples. We defined "saliency" as "the most noticeable and important feature" for our participants. After writing down each feature they predicted, the participants also circled regions on the graph corresponding to each feature on a paper copy of the graph. They then reported how salient (1 = not at all salient, 5 = very salient) they thought their five predicted features were to themselves. Finally, they matched their five predictions as best as possible with five pre-determined features, as shown in Figure 1.

3.1 Design

This within-subject experiment compares individual participant's saliency ratings of primed features (a subset of five critical graph features that were highlighted with a particular story) vs. unprimed features. We introduced three stories to counterbalance the possible primed or unprimed features, and randomly assigned participants to read one of those stories. The critical comparison in this experiment is between the salience ratings that participants assign for primed features vs. unprimed features. The independent variable is therefore whether a feature was primed or not, and the dependent variable the salience ratings for those features. We also measured a second dependent variable of how visually salient each participant rated their predicted features to themselves, on a continuous scale from one (very salient) to five (not at all salient).
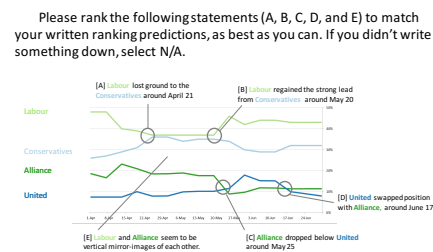

Fig. 1. Matching five pre-determined features in Experiment 1.

3.2 Hypothesis

We hypothesized that participants would:

1) Identify different features to be visually salient.
2) Among the five pre-determined features, they would predict the features that were highlighted in the story version they read to be more visually salient to uninformed viewers than the ones that were not highlighted in the story.
3) Rate the features predicted to be the most/least visually salient to an uninformed viewer to also be the most/least visually salient to themselves.

3.3 Experiments

We tested the hypotheses across four experiments. Experiment 1a provided participants with background knowledge as well as a summarizing graph with visual annotation to emphasize the primed graph features. Experiment 1b tested whether the effects would hold when visual annotations were omitted. Experiment 1c instructionally isolated participant predictions to only certain types of graph features. Finally, Experiment 2 tested the generalizability of this effect with a different type of graph and a different set of stories.

4 EXPERIMENT 1A

4.1 Participants

Eighteen Northwestern University students (10 women) participated in this experiment in exchange for course credits in an introductory psychology class. All participants were asked to bring corrective eye wear if needed.

4.2 Story

The participants read a story highlighting a competition between two out of four political parties, illustrating how citizen voting intentions fluctuated with current events. Figure 2 shows a sample display of the story highlighting the Labour and Alliance party.

According to the story, initially, between the two highlighted parties, one had a healthy lead in the polls. During an initial debate, the leading party lost voters to the less popular party and eventually lost the lead. In a later debate, the originally leading party was able to take back the votes the candidate lost and take the lead back again after a bad debate performance by his opponent. The three versions of the story all describe this same competition over time, but ascribing it to the top two parties (Top-Prime Story), the top and third party (Middle-Prime Story) or the bottom two parties (Bottom-Prime Story), highlighting the corresponding fluctuations. As shown in Figure 3, participants were randomly assigned to read a version of the story and were shown polling data after reading the story. In each pair of lines, the party with the higher line cedes votes to the party with the lower line (initial debate), and then the higher line gains back that ground (later debate). The stories, experimental materials, and results are accessible at:

http://viscog.psych.northwestern.edu/VisualizationCurse2018/

Fig. 2. Snapshot of the story participants read.

4.3 Graph

When participants predicted what an uninformed graph viewer would see as the most visually salient feature on the graph, they were shown an unannotated version of the line graph, depicted
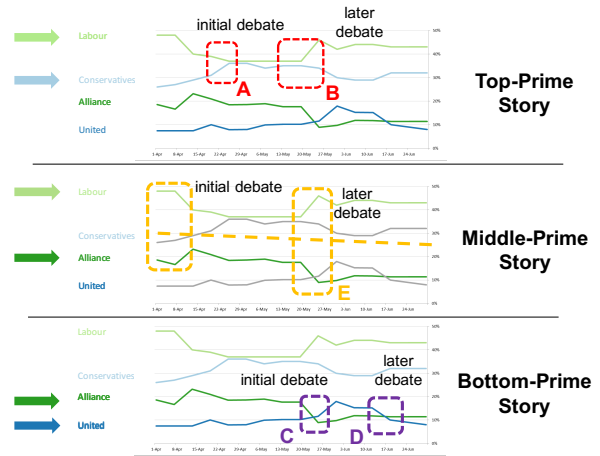

Fig. 3. Three stories highlighting different features in Experiment 1.

in Figure 4. They were told that this unannotated graph (with no story), was all that the uninformed graph viewers would see. Paper copies of this non-highlighted graph were provided to the participants to mark down their five predictions separately. We attempted to construct this graph in a way that balanced the relative salience of several critical features. The bottom two lines were made darker in color to balance the top two lines, which we expected to be more salient as a baseline [31]. We further added two intersections to the bottom two lines to counter the top two lines' natural visual saliency for just being on the top. We worried that the green 'mirror image' lines would form a less salient pattern, so we aligned their major change points to maximize the salience of that pattern. We also conducted several pilot versions of this experiment where we tracked the most salience features regardless of what was primed, and adjusted its appearance to equate those salience values (e.g., by making a peak less sharp, or a color difference stronger).
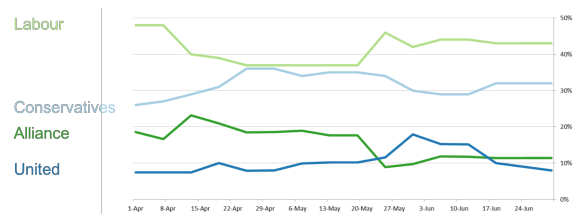

Fig. 4. The unannotated graph of the line graph experiment.

4.4 Matching Features

The participants then matched their own predictions to the five pre-determined features, referring to their markings on the paper copies of the unannotated graph, shown in Figure 4. A subset of the five pre-determined features are highlighted in each of the three stories, as shown in Figure 3. The top-prime version of the story highlighted features A and B on top (describing the two almost-intersections of the top two lines). The middle-prime version of the story highlighted features C and D in the bottom right corner (describing the two intersections of the bottom two lines). The middle-prime version of the story highlighted feature E pointing towards the center section of the graph (describing the mirroring trend of the two green lines). Participants' referred to

their freely identified salient feature drawings and matched them with the five features mentioned above. If the feature they drew did not match any of the five, they indicated it as "N/A." The subsequent quantitative data analysis of the saliency predictions and rankings were done on the rankings of the five pre-determined features. Among the five pre-determined features, 48% matched with the participants' freely identified salient feature drawings, and 56% matched if we only look at the participant's top three predictions. We discuss potential limitations of this approach at the end of this paper. We include the actual freely identified salient feature drawings of the predicted top three salient features in the qualitative results section to provide a fuller picture of the participants' responses in addition to our quantitative analysis.

4.5 Qualitative Results

Examining what the participants marked down on their physical copies of the unannotated graph, we find qualitatively observable differences among the three story versions. Figure 5 shows what the 18 participants who read different versions of the story (6 for each top, middle and bottom-prime story) marked on paper as their predictions of the most, 2nd most and 3rd most salient features to an uninformed viewer.

The top and bottom rows of Figure 5 directly compares the story versions and the respectively highlighted features to the overall predictions participants made. We see that depending on what version of the story participants read, free predictions reflected that they thought other uninformed viewers would see the features highlighted in their particular story as visually salient, even though participants were explicitly told to ignore the story when making their predictions. For example, looking at the bottom row of Figure 5, participants who read the top story identified features highlighting the top two lines to be salient more often than participants who read the bottom prime story and middle prime story. The participants who read the middle-prime story identified global and mirroring features to be salient to other viewers (notice how participants often circled pairs of features spanning a larger area), as opposed to local features identified by participants who read the top and bottom prime story.

4.6 Quantitative Results

4.6.1 Feature Ranking by Story Versions

Using the data from the feature matching section of the experiment, rankings were assigned to the five pre-determined features (ABCDE). The results are shown in Figure 6. For example, if a participant matched their most visually salient feature to uninformed viewer prediction to feature C (which is a bottom feature), feature C would receive a rank of '1' for this participant. The rank '1' would be entered in R for statistical calculations.

We reverse coded the rank in Figure 6, renaming it "saliency prediction," to be more intuitive (e.g. a feature ranked '1' will have a saliency prediction of '5'). For example, if a participant matched their predicted fourth-most feature to feature B (which is a top feature), feature B would receive a rank of '4' and reversely coded as '2' on the 'saliency prediction' axis in Figure 6.

If a participant matched pre-determined features to multiple predictions, then the feature would receive the ranking of the highest rank. For example, if a participant matched their predicted second and third salient features to feature A, then feature A would receive a ranking of two.
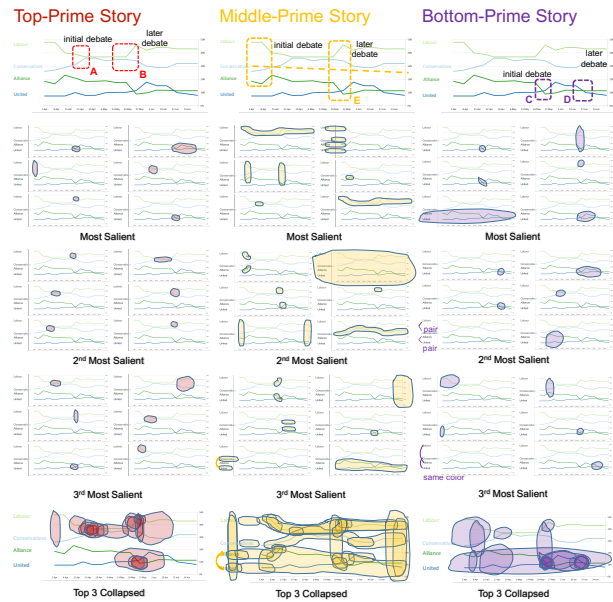

Fig. 5. Summary of Qualitative Results. Each column represents one story version, read by 6 participants who marked their most, 2nd most and 3rd most salient feature predictions.

If a participant did not think any of the five pre-determined features matched to one of their predictions, that specific prediction would be matched to "N/A." The ranking spot of this prediction would be counted as taken. For example, if a participant matched the predicted second most visually salient feature to feature E, the fourth most visually salient feature to feature D, and every other prediction they made did not match to any of the five pre-determined features, feature E would receive a rank of '2' and feature D would receive a rank of '4.' Remaining unranked features (ABC) would take on a rank of '6,' which translate to "saliency predictions" of '0.'

If participants matched two features to a predicted feature, the two features would receive the same rank (e.g., if a participant wrote down a feature to be the second most visually salient feature to an uninformed viewer and matched both feature A and B to it, then both feature A and B would receive a rank of '2.')

4.6.2 Wilcoxon Signed-Ranked Test

We conducted a non-parametric Wilcoxon Signed-Rank Test comparing the participants' saliency rankings of primed and not primed features [25]. Primed feature rankings are rankings of features highlighted in the story the participant read. For example, the middle feature (E) rankings ranked by participants who read the middle-prime version of the story are primed feature rankings. Non-primed feature rankings are rankings of features not highlighted in the story the participant read. For example, top (AB) and bottom (CD) feature rankings ranked by participants who read the middle-prime version of the story are non-primed feature rankings.

The Wilcoxon Signed-Rank test indicates that the overall primed feature rankings, Wilcoxon mean score $= 59.77$, rank mean $= 2.63$, were significantly higher compared to the overall not primed feature ranks, Wilcoxon mean score $= 38.37$, rank mean $= 0.87$, $Z = 4.03$, $p < 0.01$. Primed features were given higher saliency rankings and thus were predicted to be more visually salient to other uninformed viewers than not primed features.
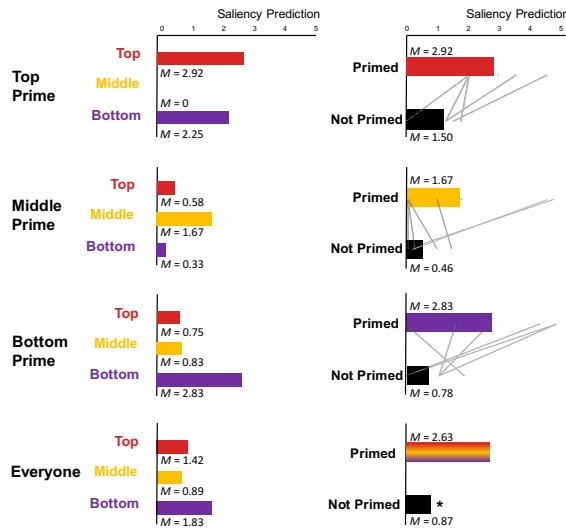
Fig. 6. Ranking details for each story version. The grey oriented lines represent individual participant ratings. The right column shows saliency ratings of primed and not primed features (e.g. in Top Prime, top is primed; middle and bottom are not primed), across the three stories.

4.6.3 Descriptive Statistics

In order to more clearly illustrate the differences in saliency rankings, we visualized their descriptive statistics. Since there are two pre-determined features highlighted in the top-prime and bottom-prime stories, and only one pre-determined feature is highlighted in the middle-prime story, the rankings of the top features (A and B) were averaged to generate a top feature average ranking. Similarly, the rankings of the bottom features (C and D) were averaged. The left column of Figure 6 shows the participant prediction rankings of the top features (AB), middle feature (E) and bottom feature (CD) for the three story versions (no standard deviation is shown because ranking data is nonparametric). The right column of Figure 6 shows saliency ratings of primed and not primed features (e.g. in Top Prime, top is primed; middle and bottom are not primed), across the three story versions.

Overall, most participants rated features that were highlighted in the story (primed), as more visually salient than other features that were not highlighted in the story (not primed). This supports the results of our Wilcoxon Signed-Rank Test. Inspecting the grey lines in the right column of Figure 6, we also see that some participants did not rate the primed features as more visually salient. This might mean that these participants were relatively immune to the curse of knowledge, though the present design cannot distinguish robust individual differences from measurement (or other sources of) noise.

4.6.4 Salience Prediction Ranking

After participants marked down a feature that they predicted other uninformed graph viewers would find visually salient, participants also rated how visually salient that predicted feature was to themselves. We see from Figure 6 that not everyone predicted the story-primed features to be visually salient to others. In the present analysis, we take a different approach here by looking at whether the participants would find features they predicted to be salient to other people also salient to themselves, regardless of whether they were primed features or not.

In Figure 7, 'Saliency to Self' is how salient each participant's predictions were to themselves on a continuous scale, where one means not at all visually salient, and five means very visually salient. Feature Rank is the order of the predictions. For example, 1 corresponds to the feature the participant predicted to be the most salient and 5 corresponds to the feature the participant predicted to be the 5th most salient, to a naïve viewer. Each dot represents one rating from one participant and the three lines are regression lines based on the scattered points.

There was a negative correlation between the Feature Rank and Saliency to Self, showing that regardless of whether the features were primed or not, participants rated the features predicted to be the most/least visually salient to a naïve viewer also to be the most/least visually salient to themselves, suggesting a curse of knowledge where they could not separate their own perspectives from that of another person. Using Spearman's Correlation, we found a moderately strong association ($r_s = 0.55$, $p < 0.001$) between the self-rated salience of a feature, and the predicted salience rating for other naïve observers.
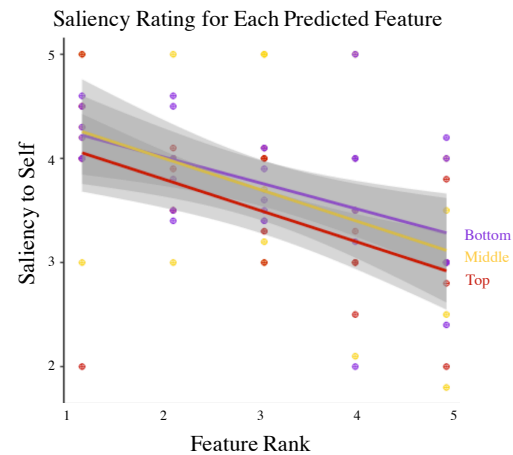

Fig. 7. Regression of predicted saliency and saliency to self in Experiment 1a.

4.7 Discussion

The knowledge the participants obtained by reading the story biased their predictions such that, in general, they saw the features depicted in the story as more visually salient than features not depicted in the story. More importantly, after acquiring this background knowledge, participants were biased to predict that other uninformed graph viewers would rate those features as more visually salient as well.

Both qualitative and quantitative statistical analyses for this experiment were done post-hoc. To ensure the validity of our findings, we conducted two follow up experiments with slight modifications with a new set of participants, and analyzed the data following similar procedures and an identical data analysis.

5 EXPERIMENT 1B (NO ANNOTATIONS)

In Experiment 1a, participants were told the story and then shown a graph visually highlighting the story content before they made their predictions. Experiment 1b hoped to tease apart the priming effect of the visual annotations and that of the story by only including the story and removing the graph visual highlighting the

story. The procedures and data analyses of Experiment 1b were identical to that of Experiment 1a, except we removed the feature cue after viewing the story (see Figure 8). The participants read the story and were presented the same unannotated line graph to draw and predict what other uninformed viewers would see.

We hypothesize that even without the visual cue the participants would be just as biased in predicting what other uninformed viewers would see, thinking they would see the same features as visually salient.
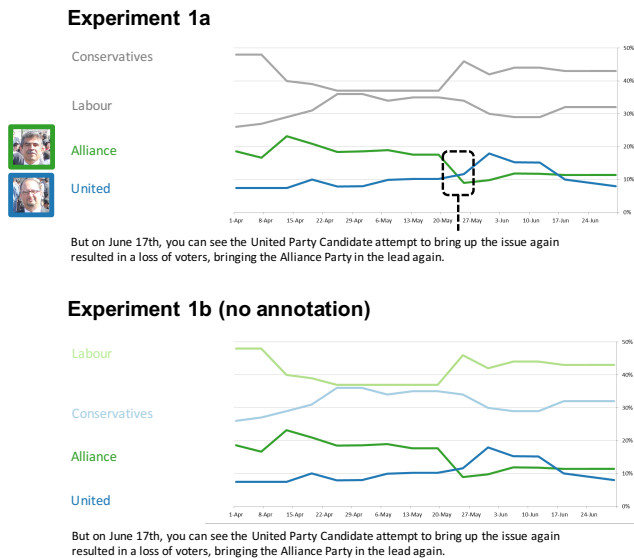

Fig. 8. Comparison between Experiment 1a and 1b annotations.

Twenty-nine Northwestern University students (23 women) participated in this experiment in exchange for course credits in an introductory psychology class or monetary payment. All participants were asked to bring corrective eyewear if needed.

Participants again referred to their freely identified salient feature drawings and matched them with the five features mentioned above. Among the five pre-determined features, 66% matched with participants' freely identified salient feature drawings, and 78% matched if we only look at the top three predictions.

5.1 Quantitative Results

5.1.1 Wilcoxon Signed-Rank Test

The Wilcoxon Signed-Rank Test [25] indicates that the overall primed feature ranks, Wilcoxon mean score $= 83.15$, rank mean $= 2.48$, were significantly higher compared to the overall not primed feature ranks, Wilcoxon mean score $= 67.98$, rank mean $= 1.79$, $Z = 2.13$, $p = 0.035$. Primed features were given higher saliency rankings and were predicted to be more visually salient to other viewers than not primed features, even without visual annotations.

Inspecting the grey lines in the right column of Figure 9, we again see that some participants did not rate the primed features as more visually salient. This might mean that these participants were relatively immune to the curse of knowledge. Compared to Experiment 1a, we see that by taking away the visual annotations, the curse of knowledge effect weakened and the number of people might be immune to the curse of knowledge increased.

We also observed an interesting change in the middle prime saliency prediction from Experiment 1a to 1b, such that the

participants in 1a who were primed with the middle feature rated it slightly more visually salient than participants in 1b. The middle feature â- the mirroring pattern of the two green lines, are more spatially separated than the top and bottom features. Since the participants in 1b only received a story prime without the visual annotation, the more spatially separated middle feature may have become harder for them to see compared to the participants in 1a who were shown clear visual annotations of this spatially separated middle feature. We speculate that while background story and visual annotation both contribute to the curse of knowledge, as shown in Experiment 1a and 1b, for spatially separated features, the visual annotation may play a more influential role in creating a curse of knowledge effect.
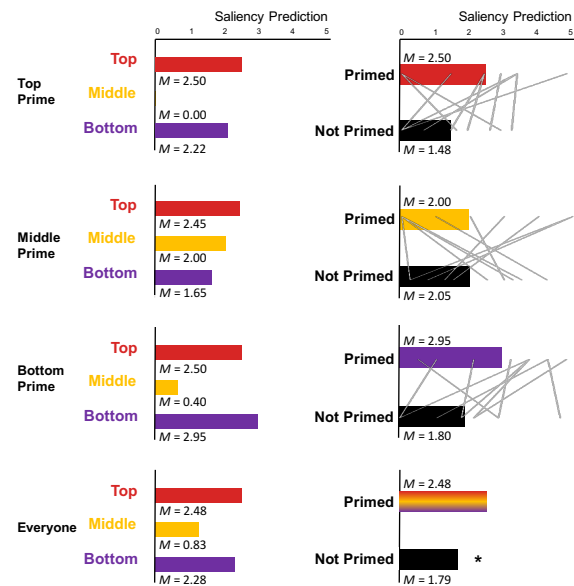

Fig. 9. Saliency prediction ranking for Experiment 1b.

5.1.2 Salience Prediction Ranking

We found a significant relation ($r_s = 0.31$, $p < 0.01$) using Spearman's Correlation between the predicted salience ranking of features for other naïve observers and the self-rated salience of these features, see Figure 10. This indicates that even without the visual annotation cue, the more visually salient a feature participants rated to themselves, the more visually salient they think the features were to a naïve viewer.

5.2 Discussion

We observed a statistically weaker curse of knowledge effect without the visual annotations in the present experiment. However, most participants nonetheless reported features primed by the story to be more visually salient than features not primed by the story, even without visual annotations. This suggests that only having the background knowledge, without any visual annotation cues, is still enough to bias people to predict that other naïve graph viewers would see features primed by the story as more visually salient.

6 EXPERIMENT 1C (INSTRUCTION CHANGE)

We conducted a third follow up experiment on a new set of participants and analyzed the data following the same procedures
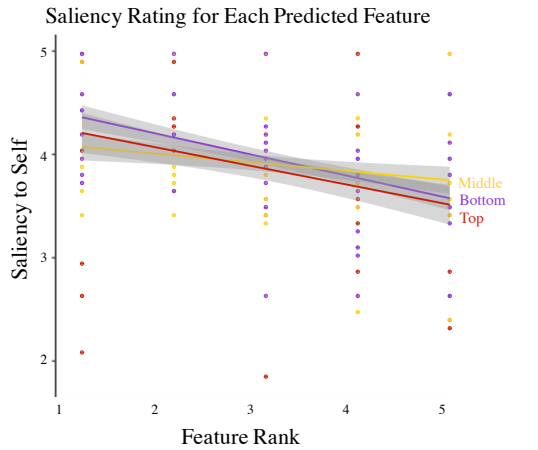
Fig. 10. Regression of predicted saliency and saliency to self 1b.

Fig. 11. Saliency prediction ranking for Experiment 1c.

and data analysis method. Since Experiment 1a and 1b did not specify in the instructions what types of features the participants should be predicting or drawing, we designed Experiment 1c with more specific instructions to maximize the amount of matching between freely identified salient features and the five pre-determined features. This experiment 1c also serves as a conceptual replication of Experiment 1a and 1b.

6.1 Modification

Previously, participants predicted features with no specific restrictions or requirements, leading some to pick out features irrelevant to the study (e.g., one participant circled the entire graph as being visually salient, another circled the y-axis, see Figure 5). To decrease such uninterpretable responses in the feature free-identification stage, participants were instructed to only describe features that involved two or more parties.

Twenty-one Northwestern University students (10 women) participated in this experiment in exchange for course credits in an introductory psychology class. All participants were asked to bring corrective eyewear if needed.

Among the five pre-determined features, 64% matched with participants' freely identified salient feature drawings, which is a 16% increase from Experiment 1a. When we look at the top three predictions, 83% matched in Experiment 1c, which is a 27% increase compared to Experiment 1a and a 5% increase compared to Experiment 1b.

6.2 Quantitative Results

6.2.1 Wilcoxon Signed-Rank Test

The Wilcoxon Signed-Rank test indicated that the overall primed feature ranks, Wilcoxon mean score = 65.93, rank mean = 3.26, were statistically significantly higher than the overall not primed feature ranks, Wilcoxon mean score = 46.54, rank mean = 1.80, $Z = 3.17$, $p < 0.01$. The descriptive statistics are shown in Figure 11. This result is consistent with the Experiment 1a and 1b such that the primed features were given higher saliency rankings and were predicted to be more visually salient to other naïve viewers than unprimed features.
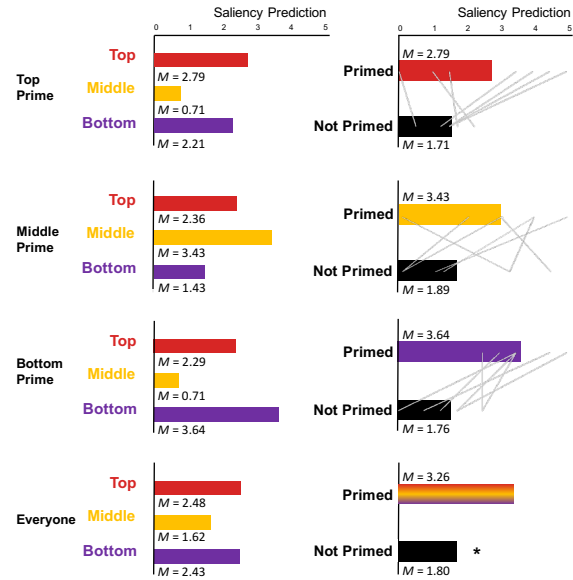
6.2.2 Salience Prediction Ranking

Spearman's Correlation again showed a moderately strong relationship ($r_s = 0.43$, $p < 0.001$) between the self-rated salience of a feature, and the predicted salience rating for other uninformed graph viewers, shown in Figure 12.
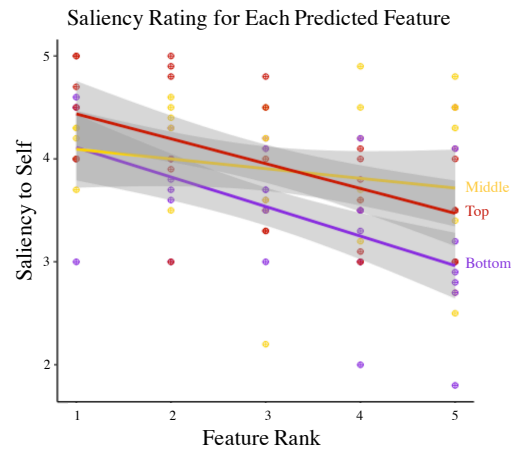

Fig. 12. Regression of predicted saliency and saliency to self 1c.

6.3 Discussion

Both Experiment 1a, where we primed participants with both a story and visual annotations, and Experiment 1b, where we took away the visual annotations, show a curse of knowledge effect where people predict features they themselves see as visually salient to also be salient to naïve viewers. This effect decreased by half in Experiment 1b when we took away the visual annotations, suggesting that both background story and visual annotations contributed to this effect, as shown in Figure 13.

Comparing Experiment 1a and 1c (where we gave the participants more specific instructions on what types of features and patterns to identify), we observed a higher number of matches

between the freely identified features and the pre-determined features. We also see that overall feature saliency for primed and not primed features increased from Experiment 1a to 1c. This instruction phrasing seems to have strengthened the curse of knowledge effect. There was also a decrease in effect size from Experiment 1a to 1c, though not statistically robust. But it is also possible that, by asking participants to predict features that include two or more parties in Experiment 1c, participants were able to match more of their own predictions to the pre-determined features (which involves two parties). This may have increased the likelihood of unprimed features to be included in the participants' predictions, which in turn increased the saliency rating of not primed features and decreased the differences between primed and not primed feature saliency ratings, resulting in a smaller effect size for Experiment 1c.

A comparison of the "Everyone" row across Figure 6, 9 and 11 shows that people gave similar saliency ratings to top and bottom features overall, but slightly lower ratings for the middle features. We speculate this to be due to the middle feature â the mirroring of the two green lines being more spatially separated than the top and bottom features, which makes the middle feature a more difficult feature to see without annotation. Participants still rated this less salient middle feature as the most visually salient to both themselves and other people when they read a story highlighting this feature, supporting the hypothesis that participants predict features they see as more visually salient also visually salient to an uninformed viewer, and that they rated the feature predicted to be the most/least visually salient to an uninformed viewer to also be the most/least visually salient to themselves.

#	% Matching	% Matching (Top 3 Features)	Wilcoxon Z-score	Effect Size (r)	Saliency (spearman's)	Sample Size	Primed Feature Saliency	Not Primed Feature Saliency	Difference between Primed and Not Primed
1a	48%	56%	4.03	0.95	0.55	18	2.75	0.81	1.94
1b	66%	78%	2.13	0.40	0.31	29	2.48	1.79	0.69
1c	64%	83%	3.17	0.69	0.43	21	3.29	1.62	1.67

Fig. 13. Comparison across all three Line Graph Experiments.

7 EXPERIMENT 2 (BAR GRAPH)

To evaluate the generalizability of this specific curse of knowledge effect, we replicated our findings using a novel type of graph, and a new story.

Seventeen Northwestern University students (9 women) participated in this experiment in exchange for course credits in an introductory psychology class. All participants were asked to bring corrective eyewear if needed.

7.1 Design and Procedures

This bar graph experiment followed the same within-subject design and experimental procedures as the line graph experiments. Participants were randomly assigned to read one of three different backstories describing events leading to a presidential election between the Liberal and the Conservative parties.

After reading the story, they were shown public polling data highlighting a key aspect of public opinion that eventually led to the victory of the winning candidate. They were asked to freely identify top five features they predict to be visually salient to a naïve graph viewer on an unannotated graph (Figure 14), rank

the saliency of these predicted features to themselves, and match the freely identified predictions to five pre-determined features, as shown in Figure 15.
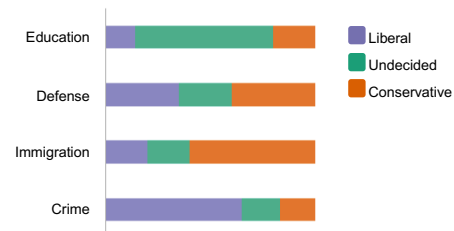

Fig. 14. Unannotated bar graph in bar graph experiment.
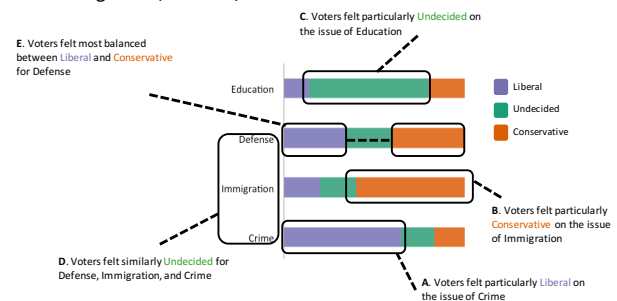

Fig. 15. Matching five pre-determined features.

7.2 Materials

7.2.1 Graph

Figure 14 shows an unannotated version of the bar graph the participants freely drew their predictions on. The stacked bar represents how people with different political stances (e.g., Liberal vs Conservative) view the topics listed, such as education. The length of the bars represents the number of voters.

We pre-determined five features on this graph, as shown in Figure 15. The graph and the features are balanced such that from the top to bottom, the four issues the public polls demonstrate correspond to education, defense, immigration, and crime issues. In the top two bars, the areas of purple and orange bars are the same. Between the bottom two bars, the area of the orange bar on the immigration issues equals the area of the purple bar on the crime issue. Similarly, the area of the purple bar on the immigration equals the area of the orange bar on the crime issue. Additionally, the area of the two undecided bars are equal. Overall, the total area of purple bars equals the total area of the orange bars.

Critically, they were told that this unannotated graph (Figure 14) was all that the uninformed graph viewers had access to, and that there was no background story provided for the uninformed graph viewers. Also, paper copies of this unannotated graph were provided to the participants to mark down their predictions, prior to matching their predicted features to the five pre-determined features, as shown in Figure 15.

7.2.2 Story

There are three versions of the story in this experiment: crime, immigration and education, Figure 16 shows a snapshot of the

stories. The crime story was a story about police brutality toward specific minority groups. The Conservative Party leader supported the police, brazenly stating that people in the minority group deserved such punishment, which was an unpopular position to take. Meanwhile, the Liberal Party advocated for reform in police departments and better treatment of suspected criminals. Participants saw graphs that highlighted the majority's Liberal public opinion of crime, explaining it as the reason behind the Liberal Party's victory, as shown in left most column in Figure 17.

The immigration story described a terrorist attack on the country's bus system two weeks before the election. The Conservative candidate had predicted in the past that immigrants posed a threat to the country's citizens. There was no information whether terrorists were immigrants, but the public was too frightened to care. While the Liberal candidate had laughed at his opponent for being too overly paranoid, the frightened public supported the Conservative view on immigration, leading to the victory of the Conservative candidate at the election. The graph the participants saw corresponded to the story highlighting the majority's Conservative public opinion on immigration, explaining it as the reason behind the Conservative Party's victory, shown in the right-most column in Figure 17.

The education story described a debate between the Liberal and Conservative Parties on the country's education system. They were told that the country had not been performing well compared to other EU countries academically. Neither candidate could come up with a clear vision on how to solve this, and the public was shocked at their incompetence. This opened an opportunity for a third candidate, who was an expert on education (as well as being female, a salient characteristic), in the election. The graph corresponded to the story by highlighting the fact that most people in the country had been undecided (neither Liberal nor Conservatives) on the issue of education, opening the opportunity for the third candidate, shown in the middle column in Figure 17.
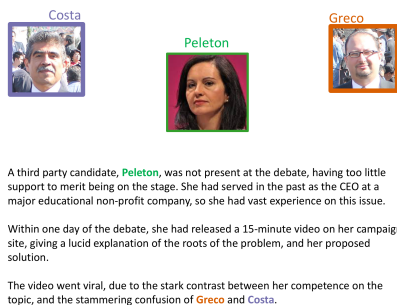

A third party candidate, **Peleton**, was not present at the debate, having too little support to merit being on the stage. She had served in the past as the CEO at a major educational non-profit company, so she had vast experience on this issue.

Within one day of the debate, she had released a 15-minute video on her campaign site, giving a lucid explanation of the roots of the problem, and her proposed solution.

The video went viral, due to the stark contrast between her competence on the topic, and the stammering confusion of **Greco** and **Costa**.

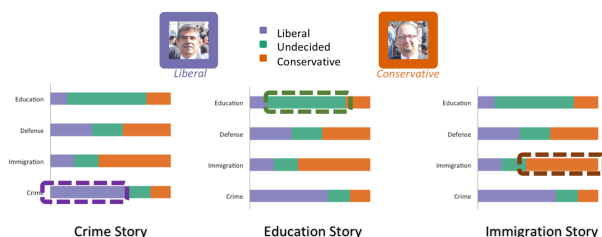Fig. 16. Snap-shot of bar graph experiment story.

Fig. 17. Highlighted feature for three story versions.

7.2.3 Matching Features

The participants referred to their paper copies of the unannotated graph and matched their own predictions to five pre-determined features, shown Figure 15.

Feature A corresponds to the feature reflected in the crime story, highlighting the purple section in the bottom bar representing public opinion on crime issues. Feature B corresponds to the feature reflected in the immigration story, highlighting the orange section in the second to bottom bar representing public opinion on immigration issues. Feature C corresponds to the feature reflected in the education story, highlighting the green section in the top bar on public opinion on education issues. These remaining features (DE) were not directly reflected in any stories, serving as "fillers". Feature D highlighted how the public was equally undecided on the issue of defense, immigration, and crime. Feature E highlighted how the defense issue had equal Conservative and Liberal support.

Among the five pre-determined features, 82% matched with participants' freely identified salient feature drawings, and 94% matched if we only look at the top two predictions.

8 QUALITATIVE RESULTS

Examining what the participants marked down on their physical copies of the unannotated graph, we find observable differences in the order of feature predictions for the three story versions.

In Figure 18, each column represents the responses of participants who read that version of the story. The top row shows the highlighted feature in that story version. Underneath, the first and second rows show the most and second most visually salient predicted features. There are participants who indicated multiple features to be salient for each of the five predictions, therefore the numbers on the graph represent the number of times the highlighted feature was chosen to be visually salient to a naïve viewer. Because the predictions can be overlapped visually across all participants, the darker the shading of a highlighted feature, the more frequently it was chosen to be visually salient to a naïve viewer.

Overall, the participants generally indicated features primed by the story version they read as what others would see as visually salient. Figure 18 compares the story versions and their respectively highlighted features to the overall predictions participants made, supporting our hypothesis.

8.1 Quantitative Results

We analyzed our data using the same method and criteria as the line graph experiments 1a, 1b and 1c.

8.1.1 Wilcoxon Signed-Rank Test

Comparing the feature highlighted in the story (primed feature) with the average rankings of all the features not explicitly highlighted in the story (unprimed features) as shown in Figure 19, across all three stories, descriptive statistics show that participants predicted primed features to be more visually salient than not primed features to naïve viewers. For example, for participants who read the crime story, feature A (the crime feature) was ranked to be more visually salient to naïve viewers than not primed features BCDE.

The non-parametric Wilcoxon Signed-Rank Test indicated that the overall primed feature ranks, Wilcoxon score mean = 34.74,
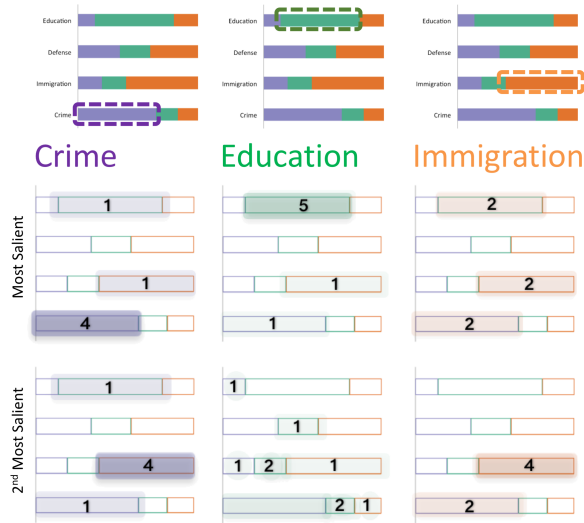
Fig. 18. Qualitative result of bar graph experiment. Heat map shows participants indicated the primed features to be more visually salient to naïve viewers than other features.

rank mean $= 4.29$, were statistically significantly higher than the overall not primed feature ranks, Wilcoxon score man $= 21.63$, rank mean $= 3.38$, $Z = 3.09$, $p < 0.01$. This result adds to the line experiments 1a, 1b and 1c, supporting that the features depicted in the story were given higher priority rankings and predicted to be more visually salient to naïve viewers than features not depicted by the story.

Fig. 19. Prediction rankings break down by story version and primed/not primed in bar graph experiment. The grey lines represent individual participant responses.

8.1.2 Salience Prediction Ranking

We also found strong, significant correlation between predicted features' saliency ranking and self-rated saliency of these features using Spearman's Correlation, $r_s = 0.65$, $p < 0.001$, indicating that participants predicted features which were visually salient to themselves to also be salient to naïve viewers, consistent with previous experiments, see Figure 20.

8.2 Discussion

The significant differences between saliency rankings of the primed and not primed features reveals the curse of knowledge bias in viewing bar graphs. This result is consistent with the line graph experiments, showing that this curse of knowledge can be generalized to bar graphs with different data sets and visual features.
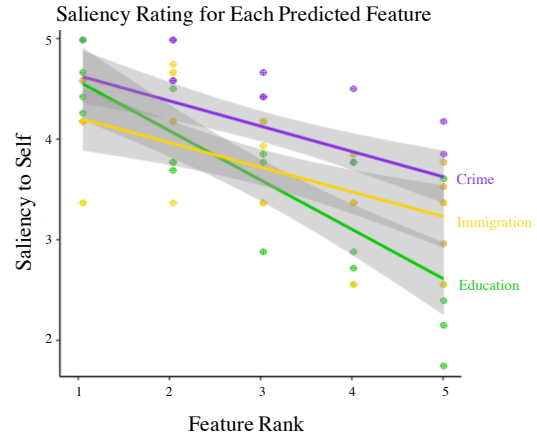

Fig. 20. Regression of predicted saliency and saliency to self.

9 CONCLUSION

Across four experiments and two types of graphs, it is clear that participants are susceptible to a 'curse of knowledge' when asked to simulate what others would see in a visualization. When a participant was told one of three possible background stories, each of which made a particular pattern within a graph visually salient to them, that participant assumed that naïve viewers would also see the same pattern as visually salient. This effect occurred despite explicit instructions to ignore what they knew, and to take a naïve perspective. To our knowledge, this is the first empirical demonstration of the curse of knowledge in the realm of data visualization, and even in the broader realm of visual perception.

This result joins other recent explorations of the influence of perceptual and cognitive biases on interpretations of patterns in data visualizations, many of which cannot be easily mitigated [2], [6], [14], [22], [32], [33], [35], [39]. Some of this research has begun to explore visual designs and interactive decision-making environments that mitigate these biases [13].

9.1 Limitations and Future Directions

The present study collected qualitative data (free drawings) but also required participants to categorize those drawings so that we could statistically evaluate the curse of knowledge in a quantitative fashion. But some drawings picked out features that were not related to the trends of interest, e.g., the participant who circled the entire graph as being a salient feature. This led us to introduce more restrictive instructions in Experiment 1c, where we specifically asked participants to note trends involving 'two or more parties'. We recognize the tension between the need for unrestricted free response in a real-world context, versus the need to qualitatively evaluate the subsequent insights within a more contrived and restricted environment. The robust differences that we see across these four studies suggest that our experimental

design detected a powerful effect, but future work should explore how the curse of knowledge might unfold in new contexts and instruction sets. In addition, all of the participants in these experiments were college students, and the effect would benefit from testing in more realistic domains, from business presentations to science communication.

We also struggled with how to phrase the instructions for what to pick out within the graph: features, trends, patterns? These terms all implicitly bias the participant: features may be local (e.g., a bump), trends may orient people to linear increases, and patterns to high temporal frequency textures or multi-series comparisons. Experiment 1c purposely gave participants a bias to report data patterns involving two lines in the graph, in order to focus their responses on data patterns that were most relevant for the experimental design. Despite this gentle bias introduced by the instructions, the participant's exposure to the background story was strong enough to create a substantial curse of knowledge effect. We similarly struggled with the use of the word 'salience', as opposed to 'most important', 'memorable,' etc. Future work should further explore how results might differ depending on how these instructions are framed.

This experiment simulates the real-world context of focusing on a particular pattern out of multiple possibilities. But if the visualization contained a single dominantly salient pattern (e.g., a downward trend among upward trends), the dominance of that pattern could hide any effect of the curse of knowledge. We attempted to balance the salience of the alternative patterns and the data suggests that these patterns were roughly balanced, according to the 'Everyone' section of Figure 6, 9, and 11, which collapse over the instructional primes (though there is a trend in Experiment 1a for 'bottom' to be more salient). Similarly, we used a set of stories, party names, and pictures, that we hoped would maintain a balance across the experiments. For example, we picked a salient female candidate, and a top position in the visualization, to balance out the presumed lower salience of the green 'Education' bar, which was in the horizontal-center of the graph. Such balancing is critical for finding any experimental effect of a single factor among multiple other factors that potentially compete. However, we have not tested the baseline saliency of the graph in the absence of story primes. Future research could test the robustness of this bias with less balanced visualizations, or more complex visualizations, to more closely emulate real world situations and further explore how stronger baseline salience differences might prevent the curse of knowledge bias.

We recognize that there are many kinds of visual data communication across many types of conversation partners. Communication could be between the creator of the visualization and an audience listening to the creator's story, or between people who did not create a visualization, but are sharing their interpretations with each other. This experiment focused on the later situation where the experts did not create the visualizations themselves. Future research could investigate if the curse of knowledge persists if the communication is between the visualization designer and a naïve audience, perhaps even in more realistic situations instead of lab simulations. We predict the curse to be stronger in these conditions as visualization creators would have richer expertise and deeper understanding of the data pattern and trends, making it even more difficult for them to separate their knowledge with that of their audience.

While most participants predicted primed features to be more visually salient to uninformed others, some participants did not.

Why are some people immune to the curse of knowledge, at least for this case study? Are some people simply better at simulating the thoughts of others, or do they use different strategies? The curse of knowledge can manifest not just from differences in perceived salience, as tested here, but by memorability, context, or impact of the data. Future research could investigate other consequences of the curse and evaluate different methods to discover the manifestation of the curse of knowledge. While these question veers more closely toward the psychology literature (see [16], [46] on discussions of strategy differences in inferring and simulating the perspectives of others), understanding the underlying difference could lead to prescriptions for mitigating the bias.

9.2 Potential Mitigation Strategies

The curse of knowledge may be largely to blame when presenters, paper authors, data analysts or other experts fail to connect with their audiences when they communicate patterns in data. While the following guidelines require empirical testing, we make several speculative suggestions for decreasing its impact.

View data from new angles: Because the design of a visualization influences what comparisons are made (e.g., people are more likely to compare proximal values [40]), so depicting data in a new way may help designers or analysts see patterns with fresh eyes. The change could be as simple as a rearrangement of values in the same visualization (e.g., hitting the 'swap rows and columns' button for a 2D bar graph arrangement in Tableau), or as involved as viewing the data in completely different formats.

Critique is critical: The curse of knowledge is tough to detect and inhibit. Critique provides a feedback loop for what is communicated, and what is not. In a strong case of a curse of knowledge, a set of visualization researchers designed a bus schedule visualization in the style of Mondrian painting, and hung it in a school cafeteria. Only after feedback did they realize that many viewers didn't realize that it was a bus schedule visualization at all, instead assuming that it was artwork [29], [41].

Rely on the wisdom of the crowd: This curse of knowledge bias shows there could be many different percepts for one person looking at one visualization. If multiple people merge different percepts and subsequent interpretations of the same visualized data, they can gain a more complete understanding of patterns in the data.

ACKNOWLEDGMENTS

REFERENCES

[1] D. W. Allbritton, G. McKoon, and R. Ratcliff, "Reliability of prosodic cues for resolving syntactic ambiguity." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 22, no. 3, p. 714, 1996.

[2] S. Bateman, R. L. Mandryk, C. Gutwin, A. Genest, D. McDine, and C. Brooks, "Useful junk?: the effects of visual embellishment on comprehension and memorability of charts," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 2010, pp. 2573–2582.

[3] D. M. Bernstein, C. Atance, G. R. Loftus, and A. Meltzoff, "We saw it all along: Visual hindsight bias in children and adults," *Psychological Science*, vol. 15, no. 4, pp. 264–267, 2004.

[4] S. A. Birch and P. Bloom, "The curse of knowledge in reasoning about false beliefs," *Psychological Science*, vol. 18, no. 5, pp. 382–386, 2007.

[5] H. Blank, V. Fischer, and E. Erdfelder, "Hindsight bias in political elections," *Memory*, vol. 11, no. 4-5, pp. 491–504, 2003.

[6] R. Borgo, A. Abdul-Rahman, F. Mohamed, P. W. Grant, I. Reppa, L. Floridi, and M. Chen, "An empirical study on using visual embellishments in visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2759–2768, 2012.

[7] M. A. Borkin, Z. Bylinskii, N. W. Kim, C. M. Bainbridge, C. S. Yeh, D. Borkin, H. Pfister, and A. Oliva, "Beyond memorability: Visualization recognition and recall," *IEEE transactions on visualization and computer graphics*, vol. 22, no. 1, pp. 519–528, 2016.

[8] M. A. Borkin, A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva, and H. Pfister, "What makes a visualization memorable?" *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2306–2315, 2013.

[9] C. Camerer, G. Loewenstein, and M. Weber, "The curse of knowledge in economic settings: An experimental analysis," *Journal of political Economy*, vol. 97, no. 5, pp. 1232–1254, 1989.

[10] M. Card, *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

[11] G. Cassar and J. Craig, "An investigation of hindsight bias in nascent venture activity," *Journal of Business Venturing*, vol. 24, no. 2, pp. 149–164, 2009.

[12] I. Cho, R. Wesslen, A. Karduni, S. Santhanam, S. Shaikh, and W. Dou, "The anchoring effect in decision-making with visual analytics," in *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, 2017, pp. 116–126.

[13] E. Dimara, G. Bailly, A. Bezerianos, and S. Franconeri, "Mitigating the attraction effect with visualizations," *IEEE transactions on visualization and computer graphics*, vol. 25, no. 1, pp. 850–860, 2019.

[14] E. Dimara, A. Bezerianos, and P. Dragicevic, "The attraction effect in information visualization," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 1, pp. 471–480, 2017.

[15] G. Ellis, E. Dimara, D. Kretz, and A. Endert, "Decisive 2017: 2nd workshop on dealing with cognitive biases in visualizations," in *Workshop at IEEE Transactions on Visualization and Computer Graphics, InfoVis 2017, Phoenix, Arizona*. IEEE.

[16] N. Epley and A. Waytz, "Mind perception," *Handbook of social psychology*, 2010.

[17] T. Gilovich, K. Savitsky, and V. H. Medvec, "The illusion of transparency: Biased assessments of others' ability to read one's emotional states." *Journal of personality and social psychology*, vol. 75, no. 2, p. 332, 1998.

[18] H. P. Grice, P. Cole, J. L. Morgan *et al.*, "Logic and conversation," *1975*, pp. 41–58, 1975.

[19] S. Haroz, R. Kosara, and S. L. Franconeri, "Isotype visualization: Working memory, performance, and engagement with pictographs," in *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. ACM, 2015, pp. 1191–1200.

[20] G. Hart, "The five w's: An old tool for the new task of task analysis," *Technical communication*, vol. 43, no. 2, pp. 139–145, 1996.

[21] M. Hegarty, "Multimedia learning about physical systems," *The Cambridge handbook of multimedia learning*, pp. 447–465, 2005.

[22] I. Herman, G. Melançon, and M. S. Marshall, "Graph visualization and navigation in information visualization: A survey," *IEEE Transactions on visualization and computer graphics*, vol. 6, no. 1, pp. 24–43, 2000.

[23] J. Hullman, E. Adar, and P. Shah, "Benefitting infovis with visual difficulties," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2213–2222, 2011.

[24] J. Hullman and N. Diakopoulos, "Visualization rhetoric: Framing effects in narrative visualization," *IEEE transactions on visualization and computer graphics*, vol. 17, no. 12, pp. 2231–2240, 2011.

[25] D. S. Kerby, "The simple difference formula: An approach to teaching nonparametric correlation," *Comprehensive Psychology*, vol. 3, pp. 11–IT, 2014.

[26] B. Keysar and A. S. Henly, "Speakers' overestimation of their effectiveness," *Psychological Science*, vol. 13, no. 3, pp. 207–212, 2002.

[27] M. Khan and S. S. Khan, "Data and information visualization methods, and interactive mechanisms: A survey," *International Journal of Computer Applications*, vol. 34, no. 1, pp. 1–14, 2011.

[28] C. N. Knaflic, *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons, 2015.

[29] R. Kosara, F. Drury, L. E. Holmquist, and D. H. Laidlaw, "Visualization criticism," *IEEE Computer Graphics and Applications*, vol. 28, no. 3, pp. 13–15, 2008.

[30] M. R. Leary, *The curse of the self: Self-awareness, egotism, and the quality of human life*. Oxford University Press, 2007.

[31] G. McKenzie, M. Hegarty, T. Barrett, and M. Goodchild, "Assessing the effectiveness of different visualizations for judgments of positional uncertainty," *International Journal of Geographical Information Science*, vol. 30, no. 2, pp. 221–239, 2016.

[32] A. L. Michal and S. L. Franconeri, "Visual routines are associated with specific graph interpretations," *Cognitive Research: Principles and Implications*, vol. 2, no. 1, p. 20, 2017.

[33] A. V. Moere, M. Tomitsch, C. Wimmer, B. Christoph, and T. Grechenig, "Evaluating the effect of style in information visualization," *IEEE transactions on visualization and computer graphics*, vol. 18, no. 12, pp. 2739–2748, 2012.

[34] E. Newton, "Overconfidence in the communication of intent: Heard and unheard melodies," Ph.D. dissertation, Department of Psychology, Stanford University, Stanford, CA, 1990.

[35] 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.

[36] R. Pohl and I. Haracic, "Der rückschaufehler bei kindern und erwachsenen," *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, vol. 37, no. 1, pp. 46–55, 2005.

[37] I. Qualtrics, *Qualtrics*, 2013 (accessed 2017-2019), https://www.qualtrics.com/.

[38] N. J. Roese and K. D. Vohs, "Hindsight bias," *Perspectives on psychological science*, vol. 7, no. 5, pp. 411–426, 2012.

[39] E. Segel and J. Heer, "Narrative visualization: Telling stories with data," *IEEE transactions on visualization and computer graphics*, vol. 16, no. 6, pp. 1139–1148, 2010.

[40] P. Shah and E. G. Freedman, "Bar and line graph comprehension: An interaction of top-down and bottom-up processes," *Topics in cognitive science*, vol. 3, no. 3, pp. 560–578, 2011.

[41] T. Skog, S. Ljungblad, and L. E. Holmquist, "Between aesthetics and utility: designing ambient information visualizations," in *IEEE Symposium on Information Visualization 2003 (IEEE Cat. No. 03TH8714)*. IEEE, 2003, pp. 233–240.

[42] A. C. Valdez, M. Ziefle, and M. Sedlmair, "Priming and anchoring effects in visualization," *IEEE transactions on visualization and computer graphics*, vol. 24, no. 1, pp. 584–594, 2018.

[43] A. Ward, L. Ross, E. Reed, E. Turiel, and T. Brown, *Naive realism in everyday life: Implications for social conflict and misunderstanding*. Lawrence Erlbaum Association Hillsdale, NJ, 1997.

[44] C. Xiong, L. van Weelden, and S. Franconeri, "The curse of knowledge in visual data communication," in *DECISIVe: Workshop on Dealing with Cognitive Biases in Visualizations. IEEE VIS, Phoenix, Arizona*, 2017.

[45] A. L. Yarbus, "Eye movements during perception of complex objects," in *Eye movements and vision*. Springer, 1967, pp. 171–211.

[46] H. Zhou, E. A. Majka, and N. Epley, "Inferring perspective versus getting perspective: Underestimating the value of being in another personâs shoes," *Psychological science*, vol. 28, no. 4, pp. 482–493, 2017.

[47] R. Zwick, R. Pieters, and H. Baumgartner, "On the practical significance of hindsight bias: The case of the expectancy-disconfirmation model of consumer satisfaction," *Organizational behavior and human decision processes*, vol. 64, no. 1, pp. 103–117, 1995.