The Curious Case of Combining Text and Visualization

Alvitta Ottley[1], Aleksandra Kaszowska[2], R. Jordan Crouser[3], and Evan M. Peck[4]

[1] Computer Science and Engineering, Washington University in St. Louis
[2] Department of Psychology, Tufts University
[3] Department of Computer Science, Smith College
[4] Department of Computer Science, Bucknell University

Abstract

*Visualization research has made significant progress in demonstrating the value of graphical data representation. Even still, the value added by static visualization is disputed in some areas. When presenting Bayesian reasoning information, for example, some studies suggest that combining text and visualizations could have an interactive effect. In this paper, we use eye tracking to compare how people extract information from text and visualization. Using a Bayesian reasoning problem as a test bed, we provide evidence that visualization makes it easier to **identify** critical information, but that once identified as critical, information is more easily **extracted** from the text. These tendencies persist even when text and visualization are presented together, indicating that users do not integrate information well across the two representation types. We discuss these findings and argue that effective representations should consider the ease of both information identification and extraction.*

Test bed: Any venue, setup, etc. used for experimentation

CCS Concepts

• ***Human-centered computing*** → ***Visualization;***

1. Introduction

Decades of research have demonstrated the value of visualization for communicating complex information in a concise and compelling way. As a result, news organizations and scientific platforms have increasingly used visualization as a tool for promoting understanding alongside their written content. Despite this trend, the benefits of combining visual and textual representations are not always clear, and observing visualization's use in complex domains has not provided a straightforward answer. For example, when communicating Bayesian statistics for medical decision-making, some studies demonstrate that adding visualizations aids comprehension [Bra09, GRH13, GRC17], while others demonstrate no significant difference [MDF12, KBGH15, OMHC12, OPH*16]. Moreover, recent research suggests that users may struggle to integrate the information across text and visualization, indicating a possible interference between the two representations [MDF12, OPH*16].

In this paper, we use eye-tracking technology to lend insight into this problem. Rather than capture measures that aggregate behavior over the entirety of interaction (speed, accuracy), eye-tracking offers a window into decision-making and explicit attention during comprehension. Fixation counts have been shown to inversely correlate with efficiency in information search [GK99], whereas fixation durations are indicative of difficulty in comprehending and processing information [Irw04]. We situate our study in the aforementioned Bayesian reasoning problem - one that is simultaneously critical to the medical domain, and has received conflicting behavioral results for decades. Using Bayesian reasoning as a testbed, we investigate how users extract information from visualizations and examine how these extraction patterns differ when the same information is presented as text. We inspect users' interaction patterns when text and visualization are presented independently and combined into a single representation. Our study confirms the findings of prior work [MDF12, OMHC12, OPH*16], demonstrating that visualizations did not lead to any measurable increase in accuracy over text representations when solving Bayesian reasoning problems. We found that participants could **identify** critical information more effectively when using a visualization, but that it may have been easier for them to **extract** this information from text. These tendencies persist when text and visualization are presented side-by-side, suggesting that participants do not integrate information across the two representations.

In the context of a Bayesian reasoning problem, we make the following contributions toward the understanding of how users interact with information in text and visual forms:

- We demonstrate that textual and visual problem representations likely have distinct differences, which are not well-captured by traditional behavioral measures such as speed or accuracy, but are evident in eye movements.
- These eye-tracking patterns suggest that users' interactions with the textual and visualization components did not change based on whether they were presented independently or together.
- We show that users likely did not integrate information across

Fixation count: the number of fixations within an area of interest

Fixation: the amount of time that the eye stops on a focal point, allowing the brain to process information

Source: https://thearf.org/councils/creative-glossary/

DOI: 10.2312/evs.20191181

representations when text and visualization were presented together. Eye fixations suggests that people primarily use the text portion of the representation and seldom switch between the two.

In the following sections, we describe the experimental results, and discuss the potential broader implications of these findings in the design of visual reasoning aides.

2. Experiment

To explore the benefits and trade-offs of combining text and visualization, we designed a study to investigate participants' eye-movements and interaction with Bayesian reasoning problems. We adopted 9 problems from Gigerenzer and Hoffrage [GH95], revised to mitigate framing effects [OPH*16], and ensure consistency in phrasing, structure, and readability. Each scenario began with a short description of the problem's context (Figure 1A), followed by a data presentation in the form of text (Figure 1B), visualization (Figure 1C), or both. Questions (Figure 1D) were present on screen throughout the interaction. We adopted a within-subjects experimental design, wherein each participant interacted with all 9 Bayesian reasoning problems: 3 presented as a visualization, 3 presented as text, and 3 presented as both visualization and text. The order of problems and conditions was randomized.

We used icon arrays to visualize each problem, similar to those used by Ottley et al. [OPH*16] and Brase [Bra09]. Icon arrays are widely tested and are generally viewed as one of the best techniques for visualizing Bayesian statistics [GRH13, KCF07, MDF12, OMHC12, OPH*16, SG01], and are widely used in the medical community to communicate risk. To assess understanding, each participant answered 2 questions about the presented data (see Figure 1D). These questions were designed to be consistent with work by Ottley et al. [OPH*16] and Brase [Bra09]. For each question, participants provided the true positive rate as natural frequency.

2.1. Materials

We recorded eye movement data using SMI Red500 remote eye tracker (SensoMotoric Instruments, Inc.), at 500Hz binocular tracking. Stimuli were presented on a 22" computer screen and were distributed over 53% of screen area (centered), with 19.2px font size (0.51 visual angle degrees) and 16px dot size (0.42 visual angle degrees). Our strict data quality criteria (rejecting datasets with calibration offset exceeding 0.5 degrees on x- and y- axes) allow us to make reliable inferences on how participants interacted with relatively small details within the stimuli.

2.2. Participants

32 participants took part in the study, of which 3 were rejected due to eye tracking calibration offset exceeding 0.5 degree on both x- and y-axis. The remaining 29 participants (21 female, $age_{mean} = 20$, $age_{SD} = 3$) are included in the final analysis. Average recorded eye tracker calibration offset was 0.41 degrees for x-axis ($SDx = 0.25$) and 0.39 degrees for y-axis ($SDy = 0.22$). Participants' vision was tested on-site and all participants had normal vision or corrected-to-normal vision using contact lenses.

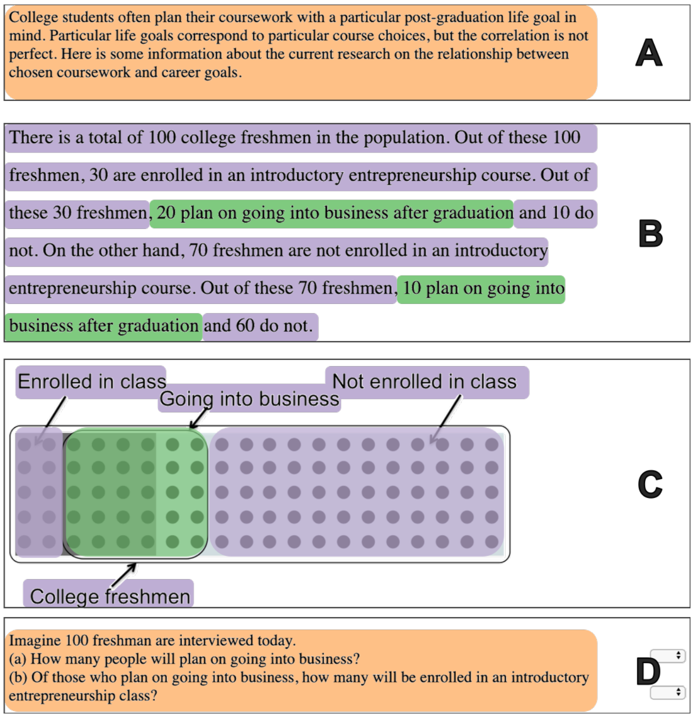

Figure 1: *The experimental stimuli. All participants received the problem framing (A), then were presented with either a text-only representation of the data (B), visualization-only representation, (Figure 1C), or text+visualization (B & C). Finally, participants answered two questions that relied on Bayesian reasoning (D). In this diagram, green areas denote critical AOI, purple areas denote irrelevant AOI, and orange areas denote other semantically meaningful AOI (context of each scenario, questions, and answer drop-down menus) that were not included in the analyses.*

2.3. Procedure

Participants consented to completing the study in accordance with Tufts University's IRB guidelines. Their vision was tested, and they were then instructed to sit comfortably in front of the computer (approximately 60 cm away from the screen) but minimize head movements. The eye tracking system was then calibrated. Participants completed the 9-question Bayesian reasoning problem set while having their eye movements monitored. There was no time limit for responding, and participants received no feedback on their accuracy. Answers were chosen from a drop-down menu of numbers from 0 to 100 using a mouse. Last, participants completed a demographic questionnaire and surveys for cognitive ability.

3. Data

In our analysis, we separate each trial into two stages. Participants first interact with the stimulus to gain a general understanding of its content (**encoding**). After reading the questions, the interaction goal changes to extracting information relevant to answering the question (**recognition**). Our analysis focuses on the recognition stage - the visual behavior *after* participants read the question, when they return to the text and/or visualization to seek information. We therefore only analyze eye-tracking data *from the first fixation on a question onward.*

3.1. Areas of Interest (AOI)

We performed area of interest (AOI) analysis on the data presentations for each type to assess how users' interaction patterns are mediated by the presentation itself. We defined AOIs with regard to their semantic composition: containing information *critical* or *irrelevant* (Figure 1) to answering the questions. For example, to answer both questions correctly for the problem posed in Figure 1, users must locate information about 30 students going into business after graduation: 20 enrolled in the class, and 10 not enrolled in the class. The AOIs that contain corresponding representations are therefore termed *critical*, while the remaining AOIs are termed *irrelevant*. Fixations falling within the scenario (Figure 1A) and questions (Figure 1D) were excluded from the analysis.

3.2. Measures

We use a combination of behavior and eye-tracking to examine three dimensions of interaction:

1. *Response accuracy (boolean):* a problem is solved successfully if the participant answered both questions correctly.
2. *Information search:* the content-normalized ratio of the fixation count for each AOI to the total fixation count within a particular problem. Later, we discuss whether this can be interpreted as proxy of cognitive cost of information search.
3. *Information extraction:* the average fixation duration within each AOI. Later, we discuss whether this can be interpreted as proxy of cognitive cost of information acquisition.

4. Results

Participants were generally successful at answering the Bayesian reasoning questions regardless of the data's representation. *We found no significant effect of presentation on accuracy*, observing near-identical proportions of correct responses for each presentation style ($Text = 55/87$ correct, $Vis = 56/87$ correct, $Text + Vis = 59/87$ correct). These findings are in line with prior work that demonstrated an overall accuracy of 63% [OPH*16].

4.1. Fixation Count: Searching for Critical Information

Users interacting with any representation format must successfully identify all information pertinent to their reasoning. To investigate information search behavior, we compared AOI fixation counts with visualization and text problem presentations. AOI fixation counts were normalized and weighted according to how many objects (words in text and dots in visualization) were contained within a specific AOI.

A two-way ANOVA with AOI relevance (critical, irrelevant) as a within-subject factor and presentation format (text-only, visualization-only) as between-subjects factors revealed significant main effect of AOI relevance ($F(1,109) = 11.357, p = .001, \eta_p^2 =$

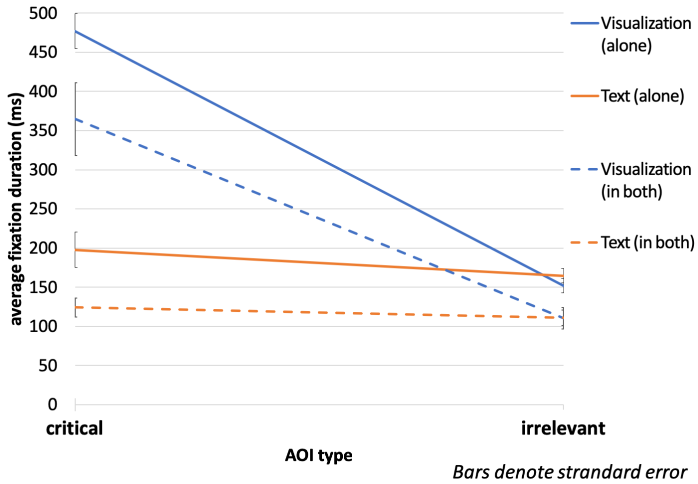

Figure 2: *Duration of fixations deployed towards critical and irrelevant information in* ***correctly answered questions****. On average longer fixations were deployed towards critical information.These patterns held even when text and visualization were side-by-side.*

$.094$), but no significant interaction effect of relevance on presentation format for correct trials. This implies that *participants who answered correctly fixated more frequently on critical information as compared to irrelevant information, regardless of presentation*.

For incorrect trials, both the main effect of relevance and the interaction effect between relevance and presentation format were statistically significant ($F(1,61) = 34.208, p < .001, \eta_p^2 = .359$). *When interacting with text in incorrect trials, participants fixated more frequently on information that was irrelevant to the problem.* However, the reverse was true for interacting with a visualization. *For incorrect answers within the visualization-only condition, participants fixated more frequently on critical information.*

4.2. Combining Text and Visualization: No Benefit?

We obtained similar results when comparing fixations in the condition in which text and visualization were presented together. Participants who answered incorrectly deployed more frequent fixations towards irrelevant information within text as compared to irrelevant information within visualization ($F(1,27) = 21.887, p < .001, \eta_p^2 = .448$). In correct trials, higher fixation frequency was deployed towards critical information in visualization as compared to irrelevant, with little difference between critical and irrelevant information within the text portion ($F(1,58) = 8.288, p = .006, \eta_p^2 = .125$). These interactions mirror results obtained from text and visualization presented separately, *suggesting that participants do not integrate information across representations when text and visualization are presented together.*

4.3. Fixation Duration: Cost of Extracting Information

Turning our attention to fixation duration, we ran a two-way ANOVA with AOI relevance (critical, irrelevant) as a within-subject factor and presentation format (text, visualization) as between-subjects factors.

Our analysis revealed significant main and interaction effects for both correct and incorrect trials. *In correct trials of the visualization-only condition, regions with critical information prompted longer fixation durations than regions with irrelevant information* ($F(1,109) = 82.669, p < .001, \eta_p^2 = .431$, Fig. 2). This effect did not persist within the text condition, where we found no significant difference in average fixation duration between areas with critical vs. irrelevant information. In accordance with previous research on fixation duration [Irw04], our results *suggest an overall lower cost of extracting information with the text-only condition.*

Analyzing incorrect trials yielded similar results ($F(1,61) = 12.556, p = .001, \eta_p^2 = .171$). Average fixation duration for fixations deployed over irrelevant information was remarkably similar between two representation formats. These effects persist when visualization and text were presented together (correct trials: $F(1,58) = 22.803, p < .001, \eta_p^2 = .282$, Fig. 2; incorrect trials: $F(1,27) = 17.406, p < .001, \eta_p^2 = .392$), providing further evidence that *participants do not take advantage of distinct text and visualization affordances when presented with both.*

4.4. Integrating Information across Text and Visualization

For our final analysis, we investigate whether participants integrate information across text and visualization when they are presented together. A paired-samples t-test reveals a significant difference ($t(94) = -5.793, p < .001$) between the number of fixations deployed towards text ($M = 67.51, SD = 49.956$) and visualization ($M = 33.41, SD = 28.662$). *Participants deployed more fixations toward text when both text and visualization were available.*

However, a Mann-Whitney test indicated no effect of accuracy on the number of fixations on text ($U = 905, p = .488$) and on visualization ($U = 907.5, p = .5$). *Participants relied more on text regardless of answer accuracy*, suggesting that integrating information across representations was not critical to successfully solving the problem. On average, participants switched between interacting with text and visualization 7 times ($SD = 5.793$). A Mann-Whitney test ($U = 920, p = .564$) found no significant difference between the number of switches for correct trails and incorrect trials, providing further evidence that *answer accuracy is not moderated by successful information integration across different representations.*

5. Discussion

The results of our study demonstrate how eye-tracking technology can reveal important differences between representations beyond traditional measures of speed and accuracy. While we found no significant difference in accuracy across the three conditions, an analysis of participants' eye-tracking data revealed the distinct benefits of text and visualization.

When using a visualization, we found that participants were more likely to attend to information that was critical to solving the problem than to irrelevant information. Although potentially influenced by our design choices, the relationship held even as AOIs were rescaled. This suggests that in this particular problem scenario, visualization may be better suited for aiding information retrieval. Conversely, in our text conditions, we found that participants deployed more fixations towards irrelevant information, suggesting that information retrieval with our textual representation was a more difficult task.

Our analysis of fixation duration (see Section 4.3) revealed potential limitations of our visualization representation. Given previous correlations between fixation duration and processing difficulty [Irw04], it is possible that this data reflects challenges in extracting visual information. In our particular case, carefully chosen labels (with values) may have alleviated these difficulties. However, by pushing our conditions to these highly-controlled vis-only vs. text-only scenarios, it allows us to begin understanding the critical nature of these small design decisions, and how they might impact information processing.

In our stimuli, text occupied a similar screen area (10.7%) to vis (10% with labels) but contained fewer items (words) on average than visual element (dots). However, a closer look at interaction patterns with the combined presentation revealed that participants were more likely to attend to text than visualization, supporting findings from prior work [BBK*16]. They also seldom switched between the two representations. This reliance on text persisted for both successful trials (subjects provided the correct answers) and unsuccessful trials. Further research is needed to decode whether the observed behavior is in fact a preference, or is driven by people simply being more familiar with using text than visualization.

Overall, we found that neither representation alone was effective at facilitating both information retrieval and information processing for Bayesian reasoning. Effective representations should take advantage of the affordances of both text and visualization, i.e. representations should accommodate both ease of information retrieval and ease of information processing.

6. Conclusion and Future Work

In our study of Bayesian reasoning problems, we found that visual representations excelled in identifying the location of important information, but may not have been as effective in helping people extract that information. We found no impact on our measures when text and visual representations were integrated together.

This study is only a first step in understanding how interaction with text and visualization and we hesitate to broadly generalize these results. It is clear from our experiment that naively pairing text with visualization does not categorically lead to improved reasoning. Yet, we currently do not have more sophisticated guidelines that maximize the impact of these two representations. As a result, more investigation into methods for combining text and visualization are necessary. One possible avenue that may already be bridging the cognitive gap between the two representations is interactive visualizations. Used carefully, interactive visualizations have the potential to avoid the pitfalls of both text and visual forms. However, further studies would need to validate this hypothesis. Future work is also necessary to provide a careful articulation of *if*, *when*, and *how* text should be added to visualizations.

Acknowledgements

This project was supported by the National Science Foundation under Grant No. 1755734.

References

[BBK*16] BORKIN M. A., BYLINSKII Z., KIM N. W., BAINBRIDGE C. M., YEH C. S., BORKIN D., PFISTER H., OLIVA A.: Beyond memorability: Visualization recognition and recall. *IEEE transactions on visualization and computer graphics 22*, 1 (2016), 519–528. 4

[Bra09] BRASE G. L.: Pictorial representations in statistical reasoning. *Applied Cognitive Psychology 23*, 3 (2009), 369–381. 1, 2

[GH95] GIGERENZER G., HOFFRAGE U.: How to improve bayesian reasoning without instruction: frequency formats. *Psychological Review 102*, 4 (1995), 684. 2

[GK99] GOLDBERG J., KOTVAL X. P.: Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics 24*, 6 (1999), 631–645. 1

[GRC17] GARCIA-RETAMERO R., COKELY E. T.: Designing visual aids that promote risk literacy: A systematic review of health research and evidence-based design heuristics. *Human factors 59*, 4 (2017), 582–627. 1

[GRH13] GARCIA-RETAMERO R., HOFFRAGE U.: Visual representation of statistical information improves diagnostic inferences in doctors and their patients. *Social Science & Medicine 83* (2013), 27–33. 1, 2

[Irw04] IRWIN D. E.: Fixation location and fixation duration as indices of cognitive processing. *The Interface of Language, Vision, and Action: Eye Movements and the Visual World* (2004), 105–133. 1, 4

[KBGH15] KHAN A., BRESLAV S., GLUECK M., HORNBÆK K.: Benefits of visualization in the mammography problem. *International Journal of Human-Computer Studies 83* (2015), 94–113. 1

[KCF07] KELLEN V. J., CHAN S., FANG X.: Facilitating conditional probability problems with visuals. In *Human-Computer Interaction. Interaction Platforms and Techniques*. Springer, 2007, pp. 63–71. 2

[MDF12] MICALLEF L., DRAGICEVIC P., FEKETE J.-D.: Assessing the effect of visualizations on bayesian reasoning through crowdsourcing. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2536–2545. 1, 2

[OMHC12] OTTLEY A., METEVIER B., HAN P. K., CHANG R.: Visually communicating bayesian statistics to laypersons. In *Technical Report*. Tufts University, 2012. 1, 2

[OPH*16] OTTLEY A., PECK E. M., HARRISON L. T., AFERGAN D., ZIEMKIEWICZ C., TAYLOR H. A., HAN P. K., CHANG R.: Improving bayesian reasoning: the effects of phrasing, visualization, and spatial ability. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (2016), 529–538. 1, 2, 3

[SG01] SEDLMEIER P., GIGERENZER G.: Teaching bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General 130*, 3 (2001), 380. 2