Effect of Adaptive Guidance and Visualization Literacy on Gaze Attentive Behaviors and Sequential Patterns on Magazine-Style Narrative Visualizations

OSWALD BARRAL, SÉBASTIEN LALLÉ, ALIREZA IRANPOUR, and CRISTINA CONATI,
The University of British Columbia

We study the effectiveness of adaptive interventions at helping users process textual documents with embedded visualizations, a form of multimodal documents known as Magazine-Style Narrative Visualizations (MSNVs). The interventions are meant to dynamically highlight in the visualization the datapoints that are described in the textual sentence currently being read by the user, as captured by eye-tracking. These interventions were previously evaluated in two user studies that involved 98 participants reading excerpts of real-world MSNVs during a 1-hour session. Participants' outcomes included their subjective feedback about the guidance, and well as their reading time and score on a set of comprehension questions. Results showed that the interventions can increase comprehension of the MSNV excerpts for users with lower levels of a cognitive skill known as visualization literacy. In this article, we aim to further investigate this result by leveraging eye-tracking to analyze in depth how the participants processed the interventions depending on their levels of visualization literacy. We first analyzed summative gaze metrics that capture how users process and integrate the key components of the narrative visualizations. Second, we mined the salient patterns in the users' scanpaths to contextualize how users sequentially process these components. Results indicate that the interventions succeed in guiding attention to salient components of the narrative visualizations, especially by generating more transitions between key components of the visualization (i.e., datapoints, labels, and legend), as well as between the two modalities (text and visualization). We also show that the interventions help users with lower levels of visualization literacy to better map datapoints to the legend, which likely contributed to their improved comprehension of the documents. These findings shed light on how adaptive interventions help users with different levels of visualization literacy, informing the design of personalized narrative visualizations.

CCS Concepts: • **Human-centered computing** → *User studies*; *Empirical studies in visualization*; *HCI design and evaluation methods*;

Additional Key Words and Phrases: Narrative visualization, adaptive visualization, guidance, visualization literacy, eye-tracking, gaze metrics, scanpath, pattern mining

28

The reviewing of this article was managed by special issue associate editors Cristina Conati, Nava Tintarev, Davide Spano.
Oswald Barral and Sébastien Lallé contributed equally, alphabetical order.
This research was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC grants RGPIN-2016-04611 and RTI-2019-00711).
Authors' address: O. Barral, S. Lallé, A. Iranpour, and C. Conati, Department of Computer Science, The University of British Columbia, Vancouver, BC, Canada; emails: {obarral, lalles, alireza.iranpour, conati}@cs.ubc.ca.

2160-6455/2021/08-ART28 $15.00
https://doi.org/10.1145/3447992

ACM Reference format:
Oswald Barral, Sébastien Lallé, Alireza Iranpour, and Cristina Conati. 2021. Effect of Adaptive Guidance and Visualization Literacy on Gaze Attentive Behaviors and Sequential Patterns on Magazine-Style Narrative Visualizations. *ACM Trans. Interact. Intell. Syst.* 11, 3-4, Article 28 (August 2021), 46 pages.
https://doi.org/10.1145/3447992

1 INTRODUCTION

Visualizations embedded into narrative text (Figure 1), also called **Magazine-Style Narrative Visualizations (MSNVs)** [62], is one particular type of narrative visualizations widely used to tell story with data in real-world documents (e.g., newspaper, textbook, blogs, science articles). However, it is known that processing multimodal documents that, like MSNVs, combine text and visual information can be challenging to some readers due to the additional visual search processes involved in integrating the two modalities [2, 8, 23, 47, 63]. In particular with MSNV, mapping the information provided in the text to the corresponding data points in the visualizations is a well-known difficulty [23, 33, 34, 62, 67, 72], and is even exacerbated in users with low levels of several cognitive abilities (e.g., reading proficiency, **visualization literacy (viz literacy)**) [71]. To alleviate this difficulty, researchers have designed adaptive guidance to help users integrating the textual information to the relevant data points in the visualizations.

Steinberger et al. [67] investigated how to guide user attention in MSNV by displaying, upfront, a set of lines linking words in the text to the corresponding information in the visualization, leading to reduced visual search time. However, providing all of the links upfront does not scale to visualizations that contain a large number of links between the text and the visualization, as it may overly clutter the visualization. Zhi et al. [77] proposed an alternative approach that avoids cluttering by allowing the users to willingly trigger the guidance mechanism, when they like. Parts of the text that refer to datapoints in the visualization are underlined, and the user can click on each of these parts to have the relevant elements of an accompanying visualization highlighted. This approach, however, did not result in improved comprehension, possibly because users did not use the self-triggered guidance in an effective manner.

In recent work [42], we attempted to overcome these issues by proposing a novel gaze-driven adaptive mechanism that dynamically highlights the relevant parts of the visualization when users read the corresponding part in the text (Figure 2 and sample online video[1]). We specifically focused on MSNVs featuring bar charts, one of the most ubiquitous visualizations found in real-world MSNV documents [16], and the guidance is displayed by dynamically thickening the border of the bars in black. Eye-tracking is used to recognize in real-time what sentence the user is currently reading so as to trigger the adaptation. Two user studies with 97 participants compared the outcomes of participants who received the adaptive guidance while reading a set of 14 excerpts of real-world MSNV, versus those of participants who underwent the same tasks without guidance (i.e., a control group), while controlling for a set of user abilities that might influence the outcomes. We found in earlier work [42] that the adaptive guidance led to improved comprehension of the MSNV excerpts for users with low levels of *vis literacy*—that is, the ability to use common data visualizations in an efficient and confident manner [7]. Furthermore, most users found the guidance to be useful and easy to use, although half of them also reported that it can be moderately distracting due to its dynamic nature. In this article, we leverage the eye-tracking data collected in earlier work [42] to further investigate whether and how the adaptive guidance influences processing of the MSNV in a way that can explain the aforementioned effects on comprehension, perceived usefulness, and distraction.

[1] https://www.cs.ubc.ca/cs-research/lci/research-groups/human-ai-interaction/small_bar_higlight.avi.

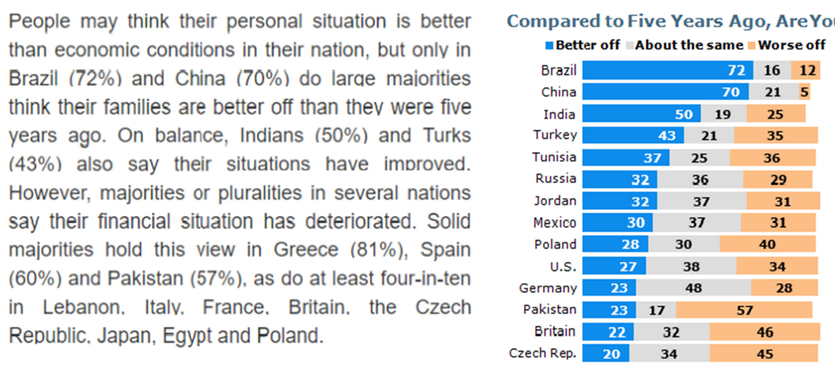


Fig. 1. Sample narrative visualization studied in this article, extracted from a real-world document.

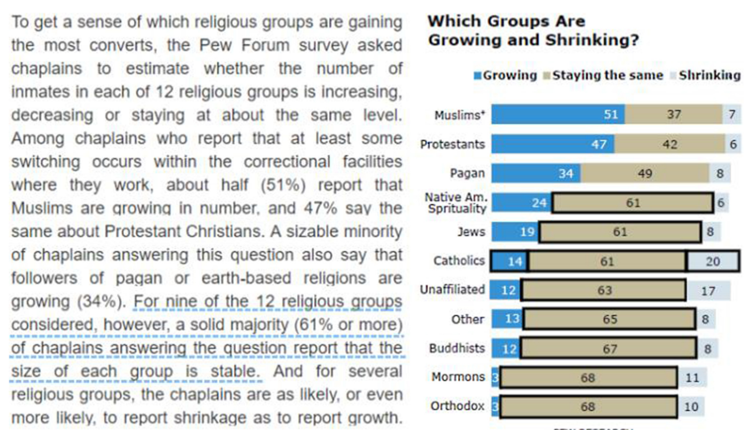


Fig. 2. Sample narrative visualization with the adaptive bar highlighting mechanism from Lallé et al. [42] (fully described in Section 3.2). The blue underlines have been added for clarity to indicate the sentence that is inked to the highlighted bars but was not displayed to the participants.

This work is driven by previous work that leveraged eye-tracking to evaluate how users process visualizations, by examining their underlying attentional behaviors (e.g., [10, 56, 73]). Furthermore, previous work has showed that eye-tracking can reveal how vis literacy influences gaze behaviors during processing of non-adaptive visualizations, showing, for instance, that low vis literacy users focus less on important regions of the visualizations compared to high literacy users [40]. However, previous work did not investigate in depth how users' attentional gaze behaviors are modified by effective adaptive visualizations, as a way to explain how and why such adaptation improves user performance in visualization tasks.

In this article, we start filling this gap by leveraging eye-tracking to evaluate how our gaze-driven adaptive guidance described earlier [42] support processing of narrative visualizations, depending on the user's levels of vis literacy. Specifically, we address the following research questions:

- *RQ1*: How does adaptive guidance in the MSNV snippets influence users' gaze behaviors compared to no guidance?
- *RQ2*: What are the changes in gaze behaviors of users with low vis literacy that can explain their better comprehension when provided with adaptive guidance?

We address these research questions with several complementary gaze-based analyses. The first analysis was presented in Barral et al. [3], in which we analyzed summative gaze metrics over specific **areas of interest (AOIs)** to unveil changes in the overall user processing of the different components of the visualization. In this work, we further extend this analysis by investigating whether and how promptly users process newly triggered interventions, as a way to understand how effective the interventions are at guiding the user's attention to relevant data points in the visualizations. We also look at whether MNSV order affects attention to the adaptation, to check for a possible effect of habituation.

The second analysis, which is new in this article, complements the one from Barral et al. [3] by investigating the users' scanpaths (i.e., entire sequences of eye movements) to better understand how the adaptive interventions influence gaze processing. Unlike the summative metrics of fixation data used in earlier work [3], the scanpaths retain the sequential nature of eye-tracking patterns, providing additional insights about how the users integrate different components of the narrative visualizations [24, 65]. We do so by applying differential sequence mining to the scanpaths and show that this approach was able to uncover many additional changes in gaze behaviors when receiving the adaptive interventions.

Finally, we look at whether there are changes in gaze behaviors of users who reported that the adaptive guidance can be distracting, as compared to those of users who did not report distraction, given that distraction is considered one of the main pitfalls of adaptive interventions (e.g., [19, 29, 30]).

We present four main contributions. First, we show that the interventions are successful at guiding the user's attention, as we found that a majority of the users process most of the interventions after they appear, which is encouraging because it shows that the interventions do not go unnoticed and are not overly ignored by the users.

Second, we uncover the underlying eye movement changes that are induced by the target adaptive interventions. As discussed, although previous work provided such guidance in narrative visualizations [42, 67, 77], it is still unknown how exactly users process and benefit from them. We fill this gap by showing that these adaptations lead to increased integration of the key elements of the visualization (labels, legend, datapoints), as shown by the summative gaze metrics and the scanpath analysis. Namely, the summative metrics reveal an increased amount of transitions between these key elements, and the scanpath analysis indicates that the adaptive interventions generate increased back-and-forth integration between the labels and the relevant bars. Furthermore, the summative analysis shows that the interventions lead to more processing of the relevant datapoints and of the legend of the visualization. The scanpath analysis indicates that the interventions induce increased transitions from reading a reference sentence to processing the relevant bars. We found no other salient behaviors among the participants who did not receive the interventions, despite the fact that the reading time was the same on average regardless of whether the participant received the interventions. This indicates that without the interventions, there was no salient attentive patterns in the gaze behaviors of the participants, possibly due to more diverse processing.

Third, we reveal how this adaptive mechanism affects users differently depending on their levels of vis literacy. As said earlier, our previous work [42] found that users with low vis literacy were the ones who benefited the most from the adaptive mechanism, but it was still unknown why and how the guidance modified the way they process MSNV. We show that receiving the adaptations leads these users to transition more often between the relevant datapoints and the legend in the visualization, indicating improved information processing of these key elements. This constitutes the first analysis that investigates the relationship between users' individual differences and their gaze behaviors when processing visualizations with and without dynamic adaptations.

Fourth, we found that the distraction reported by about half of the participants who received the guidance translated into a lower amount of fixations on the relevant sentences in the text, possibly because their attention was pulled away from the text by the guidance, whereas the participants who were not distracted were able to focus more on the text. Apart from this effect, we found no other distinctive gaze patterns among distracted and non-distracted participants, suggesting that the perceived distraction does not deeply modify MSNV processing. This would be consistent with previous findings in earlier work [39] that the distraction did not lead to lower comprehension nor lower perceived usefulness, indicating that whereas adaptive interventions can cause distraction, this may not necessarily interfere with their effectiveness.

Based on these findings, we discuss how to further improve the adaptive guidance, namely via personalization to the users' vis literacy.

2 RELATED WORK

2.1 Guidance in Multimodal Documents

Previous research has studied how to guide the user attention to relevant parts of multimodal instructional material consisting of text and accompanying diagrams or pictures (but not visualizations), by means of visual cues—that is, visual prompts that guide user attention (see the work of Van Gog [74] for an overview). These cues have been so far based on the "brushing & linking" interaction technique [9], in which the reader selects a subsets of the data of interest in one modality (*brushing*), which triggers cuing of the corresponding data in the other modalities (*linking*). In particular, color coding matching parts of the text and the graphics, and visual links connecting these parts, were found to increase comprehension [31, 55]. These cues were provided either upfront [55] or at the user's request when clicking on a specific paragraph referring to the graphics [31]. Brushing & linking is also frequently used in interfaces that combine multiple visualization types—for example, to show how selecting or modifying datapoints in one visualization affects the other ones [35, 60].

There has been recent interest in studying cuing for supporting the processing of MSNV [62]. Cuing in MSNVs is intended to guide the user attention to relevant datapoints in the visualization—that is, datapoints that are described (or referred to) in the narrative text. One approach for delivering these cues is by displaying them up front in the MSNV, as done by Steinberger et al. [67], who drew colored lines over the document to link words in the text to the corresponding datapoints in the visualization. A preliminary evaluation showed that the cues can reduce task time in simple search tasks. However, providing all cues up front is hard to scale to MSNVs with a large number of references between the text and the visualization, as is often the case in real-world documents (e.g., Pew Research documents on public policy can include up to 30 references [34]), because the many cues can visually clutter the document and overwhelm the users [1, 21, 27]. Other work allowed the users to display the cues themselves. Specifically, Zhi et al. [77] highlighted relevant datapoints in the visualization when the users select a reference in the text, and vice versa. Although they found that users extensively use such on-demand cues, it did not result in improved reading comprehension, possibly because users did not use the on-demand cues in an effective manner, or because not all users can effectively process them. Metoyer et al. [48] proposed a similar approach for sports narratives with visualizations, albeit with no evaluation. Latif et al. [43] proposed an authoring tool to ease the implementation of visual cues in MSNVs triggered by hovering the mouse over references, but not based on the user's gaze.

As an alternative to cues displayed up front or on-demand, in recent work [42] we leveraged eye-tracking to guide the user attention in an adaptive way, based on user reading behaviors captured by an eye-tracker (see Figure 2). This guidance is also derived from the aforementioned brushing & linking interaction technique, with a main difference compared to previous visualization research: the brushing part (selection of the bar to be highlighted) is performed implicitly via gaze tracking rather than explicitly with the mouse. We found two main sets of results in the work of Lallé et al. [42].

First, the gaze-driven adaptive guidance significantly improved comprehension performance of some users, depending on their levels of vis literacy. Specifically, low vis literacy users achieved significantly higher comprehension when reading MSNVs with the adaptive guidance compared to non-adaptive MSNVs, whereas there was no such difference for high vis literacy users. In fact, low vis literacy users even outperformed high vis literacy users thanks to the adaptive guidance. This difference in comprehension came at no expense of slower reading, as the guidance did not have an impact on reading time. These findings are intriguing because they indicate that gaze-driven guidance can improve the performance of some users; however, the reasons and the underlying processing behaviors that led to this increase in performance are still unknown. As mentioned in Section 1, in this work we leverage eye-tracking to gain a more fine-grained understanding of how the gaze-driven adaptation that we proposed in earlier work [42] impacts MSNV processing, particularly for low vis literacy users who do benefit from the guidance. Such analysis is also important to understand why users with higher vis literacy do not benefit from the adaptation, by examining whether they exhibit suboptimal gaze behaviors when processing the adaptation compared to low vis literacy users.

Second, most users reported that they found the interventions useful and easy to use (rate of 5 and above on a 7-point Likert scale); however, half of them also reported that they can be distracting at times (rate of 4 and above on a 7-point Likert scale) due to their dynamicity. This distraction remained moderate and did not escalate into hindering comprehension or perceived usefulness. Still, it is important to study the negative impacts of this distraction and understand how to mitigate them. As mentioned in Section 1, as a first step toward this direction, we examine in this work whether and how this distraction affected how users processed the MSNVs.

2.2 Eye-Tracking for Adaptation

Eye-tracking has been used to guide the delivery of adaptation in different HCI applications (see the work of Lallé et al. [40] for an overview), for instance, by triggering prompts to refocus student attention in educational software when they look away from the screen [22], or by adapting the content of online advertisements based on what information users look at in e-commerce webpages [2].

In InfoVis, recent work has also used eye-tracking to drive adaptive guidance in stand-alone (non-narrative) visualizations, although no evidence was found that these adaptations can improve the user performance. Specifically, Göbel et al. [22] used eye-tracking to deliver adaptive support in maps by dynamically placing the legend of the map next to where the user is looking, and highlighting in the legend the symbols that lie in the area of gaze location. Although they found that users could process the adaptive legend faster than the non-adaptive one, this did not translate into improved user performance in map reading tasks. Silva et al. [64] used eye-tracking to recommend relevant patterns in line charts showing time-series signal, based on where the user look at in the visualization system. A preliminary analysis revealed that users extensively look at the recommended patterns. However, this work did not include a control group to formally evaluate the effects of the adaptation. Unlike these works, we leverage eye-tracking to perform an in-depth evaluation of gaze-driven adaptation, which has been shown to improve the performance of some users, to elicit the gaze behaviors and patterns that may explain this improvement. We further extend the previous work by providing an analysis of the users' entire scanpaths, whereas other works [22, 64] focused solely on how extensively the users processed the different parts of the visualizations. We also extend the type of visualizations studied in other works [22, 64] by focusing on narrative visualizations and further examine how vis literacy affects processing of the gaze-driven guidance.

2.3 Eye-Tracking for Modeling Individual Differences

There is extensive evidence that user performance in non-adaptive HCI and visualization tasks is impacted by individual differences, such as cognitive abilities, personality traits, and expertise (see other works [37, 51, 54, 76] for overviews). To understand these differences in performance, researchers have leveraged eye-tracking to investigate if and how individual differences influence processing of stand-alone visualizations. Most of this research has focused on comparing eye-tracking behaviors of different groups of users, either by generating summative features of eye-tracking or by analyzing the user's scanpaths, as detailed next.

Summative features of eye-tracking (total number of gaze fixations, mean of gaze durations, number of transitions between two areas, etc.) are meant to capture overall behaviors over the different parts of the visualization. Such features have been extensively used to compare the gaze behaviors of experts and novices in different visualization tasks, namely map reading [53], visual information search [36], and processing of scientific charts [26, 68]. A few studies have investigated how cognitive abilities and skills influence gaze behaviors in low-level analytic tasks with bar and radar charts [70, 73], in decision making tasks with maps and deviation charts [40], and in tasks involving understanding medical information with bar charts and line charts [52]. Results showed, in particular, that eye-tracking can unveil suboptimal gaze behaviors exhibited by low vis literacy users, such as performing few visual comparisons among visualizations [40] and not focusing on important regions of the visualizations [52]. Cognitive styles were also linked to how users visually process a visualization-based bibliographic retrieval system [69] and perform visual decision-making tasks [58]. So far, no work has looked into the relationship among adaptive guidance, gaze behaviors, and individual differences in visualization tasks, as we do in this work. Vis literacy has been shown to play an important role in visualizations tasks [45]. For example, low vis literacy was found to hinder user experience during decision-making tasks supported by maps and deviation charts [40], as well as during processing of network visualizations in science museums [6] and during visualization of medical data [52]. However, so far, few works have examined how vis literacy influences gaze behaviors in such tasks, with results showing that low vis literacy users do not integrate well multiple visualizations [40], or spend less time processing the key component of the visualization compared to high literacy users [52]. Another work examined whether vis literacy influences processing of visual cues displayed upfront in visualizations but found no such influence [33]. We extend these works by providing insights into how vis literacy influences the effectiveness of gaze-driven adaptive guidance in narrative visualizations.

Most of the proceeding works on eye-tracking and individual differences have not examined so far the context of multimodal documents, but a few works have leveraged such documents [39, 71]. Specifically, Lallé et al. [39] explored whether and how eye-tracking data relate to the student's motivational goals during processing of instructional material that combines text and pictures (but not visualizations). In the work of Toker et al. [71], we studied how gaze behaviors is affected by individual differences in MSNVs, still with no guidance. We identified that low levels of several cognitive abilities generate longer suboptimal processing behaviors that result in longer reading time, such as spending longer time processing non-relevant datapoints. Here, we extend this previous work by showing that vis literacy also influences gaze behaviors when receiving adaptive guidance, and we provide insights into how these behaviors explain the performance of the users depending on their levels of vis literacy.

Users' scanpaths have the advantage of retaining the sequential nature of the gaze over the entire interaction. Several approaches to examine the users' scanpaths have been studied in previous HCI

Table 1. Summary Statistics of the References and Relevant Bars Across the 14 Target MSNVs

MSNV Property	Mean	SD	Min	Max
Length (total words)	91	50	43	228
References	2.6	1.8	1	7
Relevant bars	10.1	7.8	2	24

studies. For instance, distance metrics have been proposed to measure the dissimilarity of scanpaths between experts and novices in specialized tasks [11, 13, 20], and between neuro-typical and neuro-atypical users during web browsing and information search tasks [16, 46]. Such approach, however, requires comparing the entirety of the scanpaths, which can be computationally costly especially in longer and open-ended tasks. As an alternative, other studies have focused on identifying salient subsequences in the users' scanpaths. For instance, the scanpaths of male and female were compared in web search tasks by identifying the longest common subsequence in each group [17]. Pattern-mining algorithms have been used to identify salient gaze patterns in the scanpaths of experts versus novices in programming tasks [50] and radiology image reading tasks [44].

In InfoVis, differential sequence mining was used to identify gaze pattern differences among users with high and low levels of several cognitive abilities during processing of grouped bar charts and radar charts [66]. For instance, this study found that users with low levels of perceptual speed exhibit significantly less frequently patterns involving processing the labels and values of both the bar and radar charts, as compared to their counterparts with higher levels of perceptual speed. In this work, we use an approach similar to that of Steichen et al. [66], but to investigate how vis literacy influences processing of static versus adaptive visualizations. We also extend their work [66] by applying differential sequence mining to a user's scanpaths when processing narrative visualizations.

3 USER STUDIES

The eye-tracking dataset used in this work to evaluate adaptive MSNVs was previously collected in two separate user studies that we conducted, together conforming a between-subject design. In the first study (fully reported in the work of Toker et al. [71]), participants read a set of MSNVs with no adaptive interventions (*control group*), whereas in the second study (fully reported in the work of Lallé et al. [42]), participants read the same MNSVs with adaptive interventions (*adaptive group*). The two studies used the exact same task and procedure, as summarized next.

3.1 MSNV Set

Both the control and the adaptive study used the same set of 14 bar-chart-based MSNVs that were derived from an existing dataset of MSNVs extracted from real-world sources (e.g., *Pew Research Center, The Guardian*) [71]. Each MSNV in the dataset consists of "snippets" of larger source documents whereby each snippet includes a self-contained excerpt of the original text and one accompanying bar chart (see Appendix B and Figures 3 through 6). We use this format to more easily control for different factors of complexity of the MSNVs that might impact their processing. In particular, the 14 MSNVs were selected to include a balanced variety of document length, and number of referenced datapoints, as shown in Table 1. We focus on bar charts because they are one of the most ubiquitous visualizations in real-world MSNVs [49], and we leverage different types of bar chart types (four simple, six stacked, four grouped). Including visualizations types beyond bar charts would have required many more repetitions in the experimental design, making the study exceedingly complex and taxing for users. The selection process of the MSNV is fully described

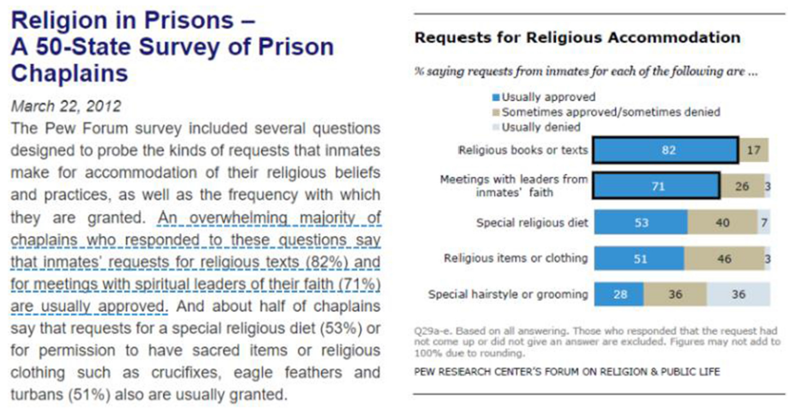


Fig. 3. Sample MSNV with one active intervention, namely the two highlighted bars at the top of the bar chart. The blue underlines (added for clarity but not displayed to the participants) indicates the corresponding reference sentence.

in the work of Toker et al. [71]. In each MSNV, the mapping between sentences in the text that refer to the visualization (called *references* from now on), and the specific referenced datapoints in the bar chart (called *relevant bars* from now on), was identified via a rigorous coding process detailed by Kong et al. [34]. Datapoints in the bar chart that are never referred to in the text are called *non-relevant bars*.

3.2 Adaptive Interventions in MSNVs

For the adaptive group, we proposed in earlier work [42] the gaze-driven adaptive interventions meant to drive the user's attention to the relevant data in the visualization of the MSNV when it is the most relevant (i.e., when the user is attending to that piece of information in the text). To this end, we devised an eye-tracking mechanism that dynamically highlights the set of bars in the MSNV visualization corresponding to the reference sentence that the user is reading. These highlights are displayed by thickening the border of the bars in black. For instance, in Figure 3 and Figure 4, the two bars at the top are highlighted with a thick black border when the user reads the sentence underlined in blue in the text, as this sentence directly describes these bars. Notice that as mentioned previously, the blue underlines in the text in Figure 3 and Figure 4 have been added here for clarity but would not have been displayed in the interface.

As the user reads though the text, highlights corresponding to each newly read reference sentence are cumulatively added to the previously highlighted bars. To help the user distinguish the most recent highlighting from the previous ones, previous black outlines are desaturated so that they become grey. This is shown in Figure 5 and Figure 6, where the bars thickened in black correspond to the underlined reference sentence, whereas the desaturated bars at the top correspond to the previously triggered interventions.

Importantly, we opted to not implement the opposite mechanism—that is, highlight a reference sentence when the user starts by processing the corresponding bars in the visualization. We did so based on preliminary analysis in our control group (see Section 3.3) that a very large majority of our participants (>90%) reads the text first. Thus, as a starting point, we examined the value of guiding attention to the chart as the users read through the text, and future work can build on our work to experiment with other forms of interventions.

This gaze-driven adaptive mechanism has been carefully designed, pilot tested, and evaluated, as fully reported in Lallé et al. [41, 42], with most participants in the pilot and main user study

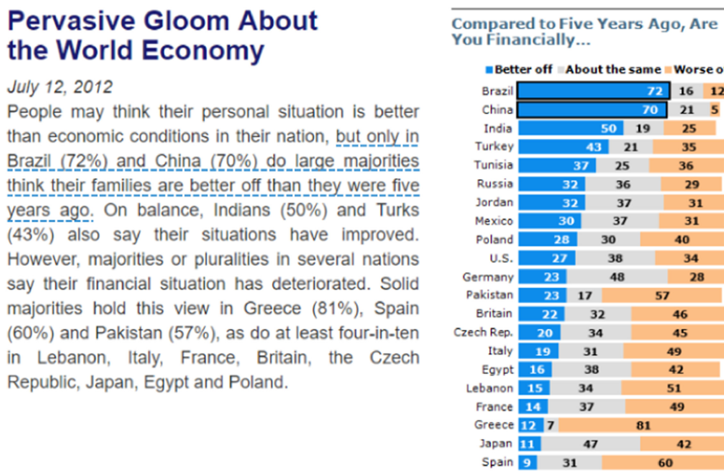


Fig. 4. Sample MSNV with one active intervention, namely the two highlighted bars at the top of the bar chart.

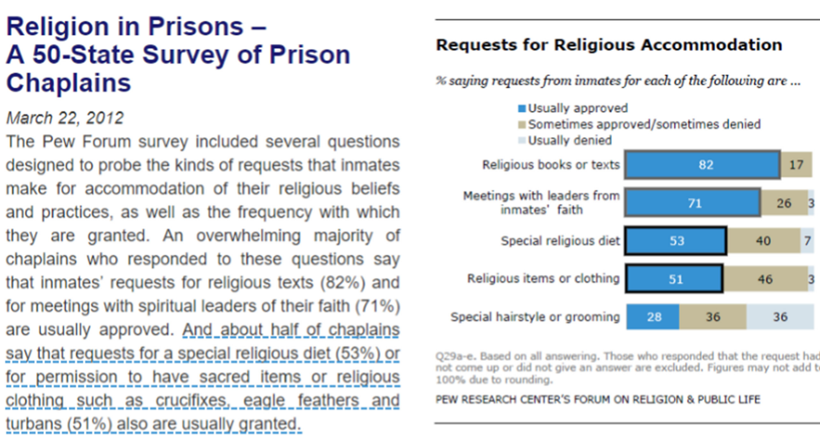


Fig. 5. Sample MSNV with one active intervention (bottom two highlighted bars) and a desaturated one (top two bars).

reporting that they found the interventions to be useful and intuitive. The experiment platform used to trigger the adaptive interventions is publicly available on GitHub for reproducibility.[2] A demo video of the delivery mechanism is also available online.[3]

3.3 Procedure for Data Collection

The control study included 56 subjects (32 female), age ranging from 19 to 69 (M = 28, SD = 11), whereas the adaptive study included 63 participants (34 female), age ranging from 18 to 59 (M = 25, SD = 8). In both studies, about 60% of participants were university students. The study procedure involves a single session lasting about 60 minutes. The session starts with the participant filling out a consent form and undergoing calibration with the eye-tracker, a Tobii T-120 camera-based remote eye-tracker embedded in a display of 1280 × 1024 pixels, with a sampling rate of

[2]https://github.com/ATUAV/ATUAV_Experimenter_Platform/.
[3]https://www.cs.ubc.ca/cs-research/lci/research-groups/human-ai-interaction/small_bar_higlight.avi.

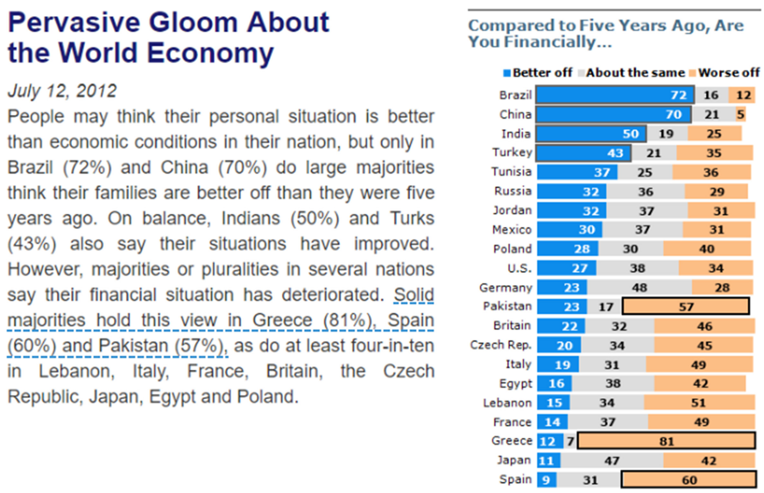


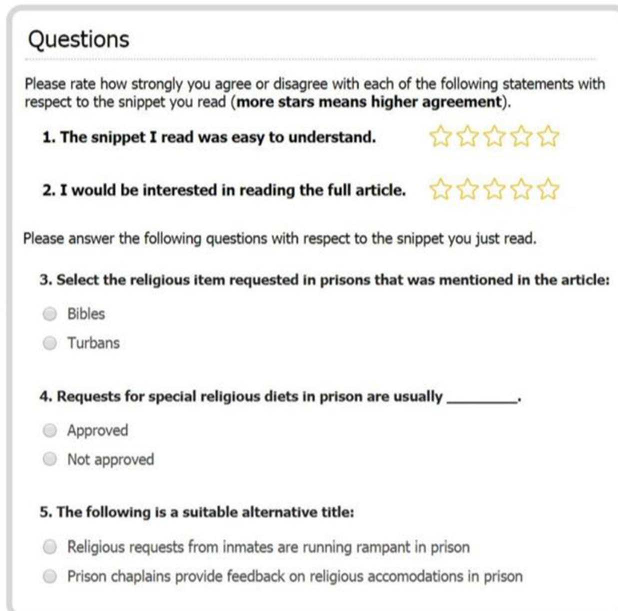
Fig. 6. Sample MSNV with one active intervention (bottom three bars thickened in black) and desaturated ones (top four bars).

120 Hz. Participants were instructed to read each of the 14 MSNVs on the computer screen. The MSNV order was fully randomized, and there was no time limit to read the MSNVs, nor any sort of training, to mimic how users read MSNVs in their daily life. After reading each MSNV, they were presented with a set of comprehension questions meant to evaluate their understanding of the MSNV they just read. These questions were based on the work by Dyson and Haselgrove [78], where they provide five different types of multiple choice questions for evaluating users' comprehension of *National Geographic* articles. We designed our questionnaire based on two of their question types that we could apply to all MSNVs in our dataset. Namely, the questionnaire include the following:

- One *title question* that asks to select a suitable alternative title for the MSNV (see question 5, bottom of Figure 7) and provides a simple way to ensure that the user had a grasp of the general document narrative.
- One or two (depending on the length of the snippet) *recognition questions* asking to recall specific information from the MSNV—that is, recognizing the name, magnitude, and/or directionality of an entity discussed in the text and named/labeled in the bar chart (e.g., questions 3 and 4 in Figure 7).

In both the control and adaptive studies, the users took on average about 20 minutes to read through all 14 MSNVs and answer the comprehension questions. Next, the *Bar Chart Visualization Literacy Test* was administered to collect the levels of vis literacy of the participants [7]. This test involves the participant answering data analysis questions based on data distributions shown in bar charts, such as identifying the min, max, average, and trend of the distribution.[4] The participant scores on the vis literacy test were retrospectively compared among the control and the adaptive group with a Mann–Whitney U test, with results showing no significant differences among the two groups ($U = 1458.5$, $p = 0.87$, $r = 0.01$). Last, the participants filled out a questionnaire about their experience with the adaptive interventions, and an interview was conducted to further elicit their response to the questionnaire. This questionnaire particularly asked participants to rate on the

[4]Link to the test: http://peopleviz.gforge.inria.fr/trunk/vLiteracy/home/tests/bc/.


Questions

Please rate how strongly you agree or disagree with each of the following statements with respect to the snippet you read (**more stars means higher agreement**).

1. The snippet I read was easy to understand. ☆☆☆☆☆

2. I would be interested in reading the full article. ☆☆☆☆☆

Please answer the following questions with respect to the snippet you just read.

3. Select the religious item requested in prisons that was mentioned in the article:

- Bibles
- Turbans

4. Requests for special religious diets in prison are usually _________.

- Approved
- Not approved

5. The following is a suitable alternative title:

- Religious requests from inmates are running rampant in prison
- Prison chaplains provide feedback on religious accomodations in prison

Fig. 7. Sample comprehension questionnaire used in the user studies.

Table 2. Summary Statistics of Collected Measures

Type of Measure	Measure	Control Group		Adaptive Group	
		Mean	SD	Mean	SD
Performance	Accuracy on comprehension questions	71.9%	3%	74.4%	4%
	Time-on-task	56.3 sec	32 sec	60.1 sec	33 sec
Perception of the adaptive guidance	Useful (scale: 1-7)	—	—	4.51	1.54
	Easy to use (scale: 1-7)	—	—	4.8	1.42
	Confusing (scale: 1-7)	—	—	3.1	1.66
	Distracting (scale: 1-7)	—	—	4.3	1.83

7-point Likert scale whether they found the adaptive guidance to be useful, intuitive, confusing, and distracting. The study procedure is fully reported in the work of Toker et al. [71]. We report in Table 2 the summary statistics of the measures collected during the studies, across the control and adaptive groups.

3.4 Data Quality

To ensure accurate analysis, we investigated the quality of the eye-tracking data, as the triggering mechanism requires to accurately capture series of fixations over the reference sentences in the text to display the interventions. This data quality check is extensively detailed in earlier work [42] and is summarized next. For the purpose of this work:

— We examined the proportion of invalid gaze samples as reported by the eye-tracker, and discarded nine users who exhibited more than 25% of missing gaze samples, as missing fixations may result in delayed or un-triggered interventions even if the participants read the corresponding reference. This 25% threshold is commonly used in eye-tracking analysis to ensure a high-quality dataset.

— We examined the attention map of the participant over the MSNV layout in the first, middle, and last task, and found that the attention map was substantially misaligned with the MSNV layout for eight additional participants. We discarded these participants, as this technical issue means that the participant's fixations will not be aligned with the reference sentences.

As a result, we retained 46 participants (73%) with high-quality data in the adaptive group. These participants triggered on average 93% of the interventions (SD = 7%), thus demonstrating that the triggering mechanism was adequate to deliver the interventions to almost all participants. The same data cleaning process was applied to the control group to ensure a fair comparison among the two groups, which resulted in 4 participants being discarded from the control group. This produced a dataset of 98 participants with high data quality (52 control, 46 adaptive).

4 OVERVIEW OF ANALYSES

To address how the adaptive interventions influence users' gaze behaviors (RQ1), and the specific changes in users with low vis literacy that can explain their improved performance (RQ2), we present the following analyses:

1. We investigate whether and how promptly the participants in the adaptive group processed the newly delivered adaptive interventions, to ascertain whether the interventions were successful in guiding the user attention (*intervention processing analysis*, Section 5).
2. We leverage summative eye-tracking metrics to understand whether and how the interventions influenced the MSNV processing, as compared to receiving no interventions (*gaze metrics comparative analysis*, Section 6).
3. We investigate the users' scanpaths by mining the salient fixation patterns in the adaptive and control groups, to understand whether and how the interventions influence how users sequentially process the different parts of the MSNV (*scanpath comparative analysis*, Section 7).
4. We leverage summative eye-tracking metrics as well as scanpaths to unveil whether there are specific eye movements that are characteristic of users that reported being distracted by the interventions, to identify how we could better support these users (*distraction analysis*, Section 8)

In each analysis, we account for the role of vis literacy. As stated in Section 1, the second analysis was previously reported in our previous conference paper [3], whereas the first and the third analyses are new in this work.

5 PROCESSING OF THE ADAPTIVE INTERVENTIONS

We examine whether, how quickly, and for how much time the participants in the adaptive group processed the interventions, depending on their levels of vis literacy. We do so by examining the participants' fixations after each intervention trigger and compute the following metrics for each intervention:

- The proportion of triggered interventions for which the participant fixated at least once on the corresponding highlighted bars, as captured by the eye-tracker (*proportion of processed intervention*).
- The elapsed time, in seconds, between the intervention trigger and the first fixation on the corresponding highlighted bars (*time to look at intervention*).

Table 3. Summary Statistics of Usage of the Interventions

Metric	Intervention State	Mean	SD
Proportion of processed intervention	Active only	49%	16%
	Desaturated only	35%	15%
	All	84%	18%
Time to look at intervention	Active only	10.87 sec	6.34 sec
	Desaturated only	29.47 sec	11.23 sec
	All	18.40 sec	9.05 sec

We also look at whether attention to the interventions is affected by MNSV order, to ascertain whether there is an effect of habituation on intervention processing.

As described in Section 3.2, previous interventions are desaturated in grey when a new intervention is triggered, so as only the most recently triggered intervention is highlighted with the thick black border at a time. Thus, we compute the preceding metrics for two distinct sets of interventions, depending on the state of the interventions when the participant processed them for the first time. Namely, interventions that were processed when highlighted in black are grouped into the *active state* set. Interventions that were never processed while active, but only once they were desaturated, are grouped into the *desaturated state* set. Distinguishing between the two states is meant to provide complementary insights into how users processed the interventions. Namely, the active state captures whether the interventions caught the participants' attention once appeared, whereas the desaturated state reveals whether the participants ended up processing the interventions only later on, after they were desaturated. Table 3 reports the summary statistics for the three metrics in each state, as well as for both states combined (active and desaturated), over the 46 participants in the adaptive condition.

Intervention processing. Table 3 (top) shows that users looked at most of the triggered interventions (average 84%), and furthermore they processed on average about half of the interventions while they were active. This is encouraging, because the primary purpose of the intervention was to draw attention to the relevant bars in a timely manner (i.e., when the user reads the corresponding piece of information). Additionally, on average, 35% of the interventions were processed only after they were desaturated, and 16% of them were never processed, which suggests that some participants either did not see some of the interventions when they appeared or chose to temporarily ignore them, perhaps because they intended to finish reading a paragraph or even the entire text first. Although further analysis could focus on explaining this behavior, overall, the proportion of processed interventions is substantial enough to proceed with our analysis of the impact of the intervention on the users' gaze behaviors.

Table 3 (bottom) shows that participants took, on average, about 10 seconds to look at the active interventions after they were triggered, and that this time varies across participants, with a standard deviation of ~6 seconds. Naturally, this time increased for interventions that were looked at after they had been desaturated (around 30 seconds on average). This result suggests that interventions do not take participant's attention away from the text immediately as they are triggered (in fact, only 11.5% of the active interventions across all participants and tasks were looked at in under 1 second after they appeared); rather, participants tend to take a few seconds before processing the highlighted bars, perhaps to finish reading the reference sentence, or to process other key elements of the MSNV such as the bar labels and legend first.

Role of task order. As we did not include a training phase in the experimental protocol, we investigate whether we can observe any noticeable learning effect in the way users process the adaptive visualizations over time. To do so, we run a set of **Linear Mixed Models (LMMs)** over the metrics in Table 3. We use mixed models because they can handle more than one random effect at once (i.e., within-subject and within-document correlation, as each user performed several tasks, and each of the 14 MSNVs were presented to every user). Specifically, we select the metrics listed in Table 3 one at a time as the dependent variable, with *task order* as the fixed effect, and *Participant-ID* and *MSNV-ID* as random effects (i.e., repeated measures). We used the *lmerTest* package in R [38] and account for the multiple models run (six metrics) by adjusting all p-values with the Benjamini–Hochberg procedure to control for the **false discovery rate (FDR)** [5]. Results show a significant[5] main effect of *task order* on the *proportion of ALL processed intervention* ($\chi^2(1) = 9.37$, $p = .002$, $r = .45$) with earlier tasks presenting a higher *proportion* of *processed intervention* than later tasks, representing an average of 87% of triggered interventions being processed over tasks 1 through 7, whereas this proportion decreases to 80% on average over tasks 8 through 14. Given that we did not include a training session, this could be related to the learning effect naturally occurring during the interaction (i.e., loss of initial surprise effect). The lack of significant effect of *task order* on *time to process intervention* and *time to return to text* indicates that this learning effect does not significantly impact the way in which users process these interventions which, together with the fact that the proportion of processed interventions remains at 80% even in the later tasks, indicates that participants keep using and processing the interventions, suggesting that they do not get bored or annoyed by them over time.

Role of vis literacy. To investigate whether the participant's levels of vis literacy influence processing of the interventions, we run a set of LMMs over the metrics in Table 3. Similar to the preceding task order, we select the metric listed in Table 3 one at a time as the dependent variable, with *vis literacy* (low, medium, high) as the fixed effect, and *Participant-ID* and *MSNV-ID* as random effects (i.e., repeated measures). For vis literacy, we divide the participants into three bins based on a three-way-median split of their vis literacy scores: low ($N = 31$, 18 control), medium ($N = 34$, 16 control), and high ($N = 33$, 18 control), following the same approach as in our earlier work [42]. We do so because the previous work [42] found that the adaptive interventions substantially helped users in the low bin, so much so that they significantly outperformed users in the high bin in the adaptive group, as well as users in the control group. We reuse the same bins throughout this entire article, so as to elicit this previous finding, which is the goal of our research question RQ2 as stated earlier.

Results indicate no significant main effect of vis literacy in any of the mixed models. This indicates overall that participants processed a similar amount of interventions, and with a similar timing, regardless of their levels of vis literacy. This finding already shows that the low vis literacy users did not benefit from the interventions solely because they processed more of them, and the next analyses will be focused on examining in a more fine-grained fashion how differently the adaptation influenced the participants' processing behaviors depending on their levels of vis literacy.

6 COMPARATIVE ANALYSIS ON USERS' GAZE METRICS OVER SPECIFIC AOIS

In this section, we describe the eye-tracking analysis on gaze metrics that we performed to address our research questions. We first describe the AOIs and gaze metrics used to capture users' gaze behaviors followed by the statistical analysis. We then discuss results pertaining to RQ1, namely the main effect of groups found in the statistical analysis that capture changes in gaze behaviors

[5]We report in this article statistical significance at $p < .05$ after FDR correction, as well as effect sizes as high for $r > .5$, medium for $r > .3$, and small otherwise.

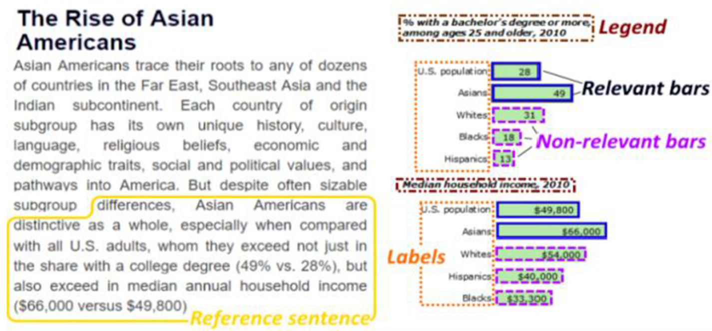


Fig. 8. AOIs defined in the MSNVs.

among the adaptive and control group. Next, we examine RQ2 by discussing the effects of group that are qualified by an interaction with vis literacy and discuss their potential implications in terms of improved comprehension for users with low vis literacy.

6.1 AOIs and Gaze Metrics

To analyze the differences in gaze behaviors with and without the adaptive interventions, we define a set of AOIs over the key regions of the MSNVs. We focus on these AOIs because in earlier work [71], we showed that they can reveal users' suboptimal gaze patterns within non-adaptive MSNVs. Here we aim to understand whether the adaptive interventions modify how users attend to them. AOIs are shown in Figure 8 and described next:

- *Reference sentences* (Ref) contain the combined area of all sentences that directly refer to the visualization of the MSNV. For instance, in Figure 8, there is one reference sentence, constituting the AOI with plain yellow border. We expect to capture more transitions from this AOI to the visualization side, given that the adaptive interventions were designed to guide the attention of users toward the visualization when they read a reference sentence.
- *Relevant bars* (R-Bar) and *Non-Relevant bars* (NR-Bar) contain the combined area of all bars in the visualization that are described (R-Bar) or not (NR-Bar) by any of the references. We expect users to focus more on the R-Bar when provided with adaptive interventions, since the interventions specifically highlight these bars, and expect the inverse for NR-Bar that would not be highlighted.
- *Legend* and *Labels* respectively contain the legend and the combined area of all labels in the visualization. These are key elements to integrate the information conveyed by the bars, and we expect them to be accessed more from the R-Bar, which are highlighted by the adaptive interventions, and less from the NR-Bar.

To evaluate how users process the MSNVs, we leverage gaze metrics meant to capture how users allocate their attention to each of the AOIs, as well as how they integrate these AOIs by transitioning between them.

Attention metrics. We compute the *number of fixations* and the *average fixation duration* within each AOI. These two measures are complementary in understanding how users processed the target AOI. In particular, the total number of fixations gives a sense of attention allocation to that AOI [28, 57]. The average fixation duration gives a sense of how users process the AOI. In particular, longer fixations on average can indicate more cognitive processing dedicated to that AOI [59] or that the AOI is more engaging to the users [15, 57].

Table 4. Gaze Metrics for Which a Significant Main Effect of Group Was Found

AOI	Gaze Metric	Main Effect of Group
R-Bar	Number of fixations	More in adaptive $\chi^2(1) = 6.83$, $p = .0089$, $r = .26$
	Transitions to Label	More in adaptive $\chi^2(1) = 6.43$, $p = .0011$, $r = .26$
Label	Transitions to Legend	More in adaptive $\chi^2(1) = 9.44$, $p = .0021$, $r = .31$
Legend	Transitions to R-Bar	More in adaptive $\chi^2(1) = 9.70$, $p = .0018$, $r = .31$
	Avg. fixation duration	Longer in adaptive $\chi^2(1) = 5.58$, $p = .0181$, $r = .24$

Transition metrics. We investigate the shift in attentional focus between the different regions of the visualization by computing, for each AOI, the *number of transitions* to all other AOIs. Namely, a transition is defined as a series of two fixations that land in two different AOIs. Such transitions have been used in particular to understand the processing strategies of users—that is, how users integrate different components of a visual interface [25, 57, 61]. In our case, transitions are important to evaluate not only to what extent users process certain AOIs but also how they sequentially access these areas to integrate the different parts of the visualization.

In total, we generated 6 gaze metrics for each of the five AOIs, resulting in 30 gaze metrics.

6.2 Statistical Analysis

We evaluate differences in the users' attention allocation with and without adaptations using Mixed-Effect Models, as done for the intervention processing analysis in Section 5. Specifically, we fit one model for each of the 30 *gaze metrics* defined earlier by selecting each of them one at a time as the dependent variable, with *group* (adaptive, control) as the independent variable, *vis literacy* (low, medium, high) as the fixed effect, and *Participant-ID* and *MSNV-ID* as random effects (i.e., repeated measures). For *average fixation duration*, we fit an LMM using the lmerTest package in R [38], as *average fixation duration* follows a near normal distribution. For the *number of fixations* and *number of transitions* between AOIs, we fit **Generalized Mixed Models (GLMMs)** for the negative binomial family using the *glmer.nb* function in the lme4 package in R [4], which is suitable for discrete count distribution. We account for multiple comparisons within each AOI family (gaze metrics within the same AOI) and report results significant after applying the Benjamini–Hochberg FDR procedure [5]. For interaction effects of group with vis literacy, we run post hoc, pairwise contrast comparisons using the *emmeans* package in R[6] and apply again the Benjamini–Hochberg FDR procedure to account for the multiple pairwise comparisons.

6.3 Results

Table 4 reports the significant main effects of group that were found on several of the gaze metrics that we evaluated showing that the adaptive interventions influence the users' processing behaviors in several ways. Table 5 indicates the significant interaction effect we found between group and vis literacy, as well as the corresponding significant pairwise comparisons.

[6] https://CRAN.R-project.org/package=emmeans.

Table 5. Gaze Metrics for Which a Significant Interaction Effect of Group and Vis Literacy Was Found

AOI	Gaze Metric	Group × Vis Literacy	Significant Pairwise Contrasts
R-Bar	Transitions to Legend	$\chi^2(1) = 15.18$, $p = .0005$, $r = .39$	More in adaptive for low vis literacy $z = 4.27$, $p < .0000$, $r = .43$ In control, less for low than for medium $z = -2.9$, $p = .01$, $r = .29$

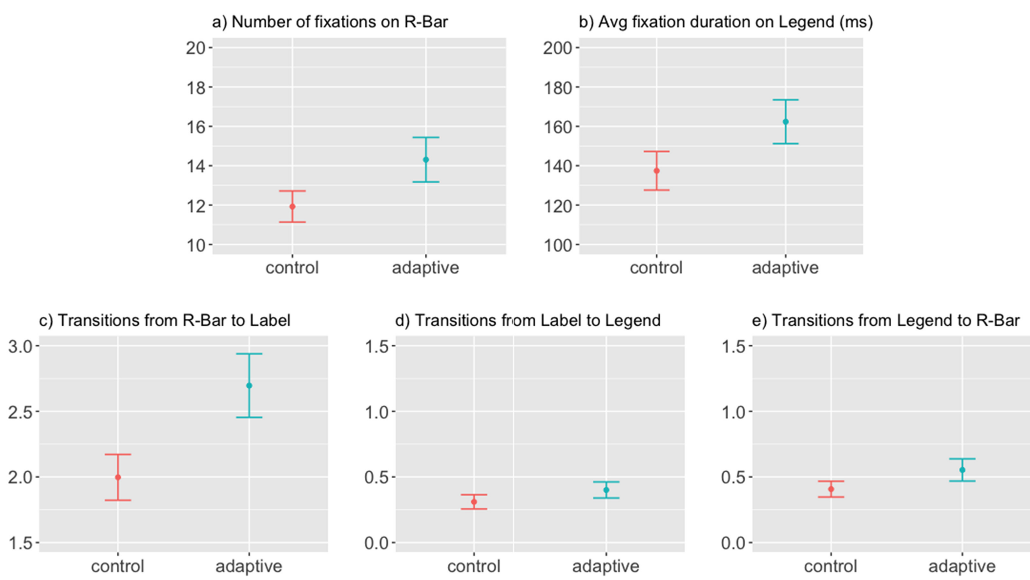


Fig. 9. Gaze metrics for which a main effect of group was found. Error bars show 95% confidence intervals.

All of the main effects reported in Table 4 follow the same direction, indicating that users in the adaptive group present higher values than users in the control group. These main effects are shown in Figure 9 and discussed next, along with the description of the interaction effect in Table 5.

Main effect of number of fixations on R-Bar. Figure 9(a) shows that users in the adaptive group fixate on average 20% more on the relevant bars than users in the control group. This indicates that the adaptive highlights helped the users in the adaptive group focus their attention on the relevant bars, which was the main purpose of the adaptations (Section 3.2). Noteworthy, previous analysis on the same dataset did not report significant difference in overall MSNV processing time among groups [42]. This indicates that this additional processing of the relevant bars comes at no expense of the overall information processing time.

Main effect of average fixation duration on Legend. Figure 9(b) shows that users in the adaptive group presented on average 18% longer fixations on the legend than users in the control group. Longer fixations can indicate higher engagement in processing the target AOI [28, 57], here the legend. This is consistent with the rest of the main effects described in the following, as users transition more to the legend thus may pay closer attention to the legend to better understand the meaning of the highlighted bars and their labels.

Main effect of Transitions from R-Bar to Label, from Label to Legend, and from Legend to R-Bar. These main effects reveal several interesting findings:

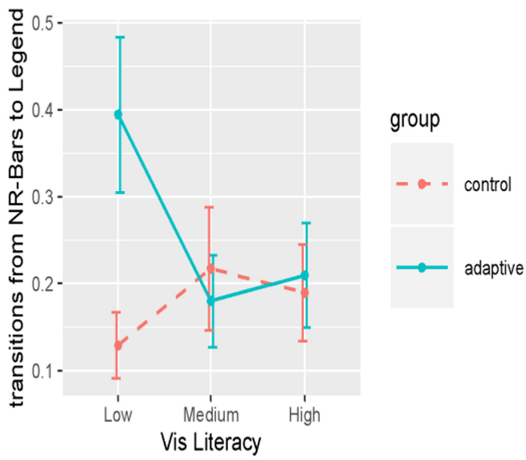


Fig. 10. Interaction effect of vis literacy with group on *Transitions* from Relevant Bars to Legend. Error bars show 95% confidence intervals.

- Users in the adaptive group transition on average 35% more often from the relevant bars to the labels (Fig. 9(c)), most likely to understand to what label(s) the highlighted bars are mapped to.
- Users in the adaptive group transition on average 29% more often from the labels to the legend (Figure 9(d)), and 36% more often from the legend to the relevant bars (Figure 9(e), indicating that the adaptive interventions may be prompting users to integrate better the key components of the bar charts.

These findings indicate that users are overall proactive in processing the bar charts once their attention is directed to the relevant bars, which is very encouraging in terms of the effectiveness of the adaptive interventions. In particular, the adaptive interventions seem to encourage users to integrate the relevant bars with their labels and the legend, which is a key aspect of understanding the information conveyed by the bar charts.

Interaction effect on number of transitions from R-Bar to Legend. This interaction effect reported in Table 5 is shown in Figure 10. The pairwise comparisons indicate that users with low levels of vis literacy perform significantly more of these transitions in the adaptive group compared to the control group, representing an increase of 181% of these transitions on average (left side of Figure 10). There is no such statistical difference for users with higher levels of vis literacy (middle and right side of Figure 10). This finding is interesting because our previous work on the same dataset [42] showed that low vis literacy users performed significantly better in the adaptive group compared to the control group, in terms of comprehension of the information conveyed by the MSNVs. Here, our finding indicates that the adaptive interventions prompted specific gaze behaviors in low vis literacy users, related to integrating the relevant bars and the legend, which possibly helped these users make better sense of the MSNVs. This is consistent with the main effect discussed earlier, that the adaptive interventions, overall, induce similar transitions in the opposite direction (i.e., from the legend to the relevant bars). Here, our results suggest that low vis literacy users, guided by the adaptive interventions, go one step further in the integration of the legend and the relevant bars by transitioning back and forth between these two AOIs.

The pairwise comparisons showed, in addition, that users low in vis literacy within the control group performed significantly less of these transitions compared to users with medium levels of vis literacy, representing 77% fewer transitions on average (left side of dashed red line in

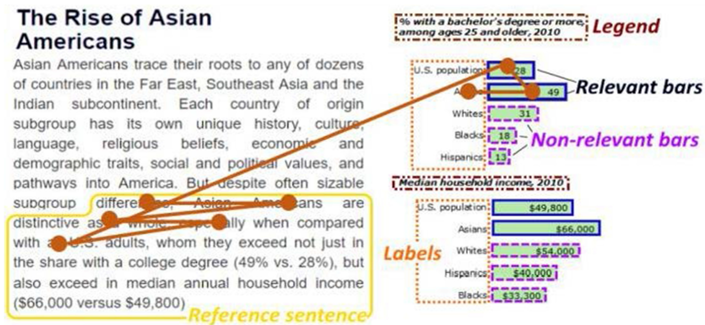


Fig. 11. Example fixation scanpath of a given task. In the process of pattern mining, this scanpath is mapped to the AOI sequence Ref(×5)-RBar(×2)-Lab.[7]

Figure 10). This result indicates that the adaptive interventions help low vis literacy users catching up with users with higher levels of vis literacy in terms of transitions from relevant bars to legend.

Most of the significant differences in eye movements involve the relevant bars and/or the legend. Although the relevant bars were the target of the interventions and were highlighted in the adaptive condition, and thus changes in how users process them could be anticipated, the legend and labels were not affected by the interventions, and thus these changes were less expected. The fact that we found longer fixations on the legend and increased transitions from/to the legend suggests that the legend is a critical component to contextualize the relevant bars. Furthermore, a majority of the significant effects we found earlier are on gaze metrics that capture transitions between AOIs, suggesting that the interventions influence how users integrate the different parts of the MSNVs. We further explore this finding in Section 7 by leveraging the entire user scanpaths, so as to better contextualize these transitions within longer sequences of fixations.

7 COMPARATIVE ANALYSIS ON THE USERS' SCANPATHS

In this section, we leverage the user scanpaths (i.e., the entire sequence of fixations within a task) to further understand how the adaptive interventions influenced the processing of the MSNVs. We first describe the methodology we used to mine the scanpaths, as well as the statistical analysis performed. We follow with results on the sequences of eye movements that change when users are provided with adaptive interventions as compared to the control group (RQ1), as well as how vis literacy may influence the scanpaths (RQ2).

7.1 Pattern Mining and Metrics

For each user task, we extract the scanpath from the Tobii eye-tracker as the sequence of consecutive fixations in *x,y* coordinates for that given task. To increase interpretability, we transform the scanpath into a sequence of AOIs that the user attended to, based on its sequence of fixations. Namely, we map each fixation in the scanpath into the AOI it falls into. For instance, Figure 11 represents a scanpath that is translated into the sequence "Ref(×5)-RBar(×2)-Lab," which indicates five consecutive fixations on the *Reference sentence* AOI (Ref), followed by two consecutive fixations on the *relevant bars* AOI (RBar), followed by a single fixation on the *Label* AOI. In the case that a fixation does not fall on any of the AOIs defined in Section 6, it is mapped to a "Dummy AOI" that captures all remaining areas of the MSNV. As a result, we obtain one scanpath per user and per task, with the average scanpath length of 177.4 fixations (SD = 99.9).

[7]Note that this is a very simplified scanpath for illustration purposes.

To extract representative patterns that occur during MSNV processing, we apply on the AOI sequences the differential pattern mining algorithm proposed by West et al. [75]. We do so because this algorithm has been extensively used to identify distinctive eye-tracking patterns from user scanpaths, and because of a previous work that showed that this algorithm was suitable for exploring the role of specific cognitive abilities on the users' gaze behaviors while processing stand-alone bar charts [66]. As a matter of fact, our analysis, described next, is inspired by Steichen et al. [66], which to the best of our knowledge is the only work that leveraged scanpaths to understand the influence of cognitive abilities in InfoVis tasks. Specifically, we mine all possible patterns of length 3 to length 10, as patterns shorter than three fixations were already considered in the previous analysis (i.e., fixations are patterns of length 1, and transitions between AOIs are patterns of length 2), and that patterns longer than 10 are very infrequent. We only consider patterns that appear at least in 40% of the scanpaths of tasks in either the control or the adaptive group, following the approach of Steichen et al. [66]. The rationale is that we aim at capturing changes in patterns that are representative of the MSNV processing behaviors, and that differences in infrequent patterns (appearing in less than 40% of the tasks) would not be informative in generating generalizable insights. We also do not consider patterns that include the "Dummy AOI," as here we are interested in understanding how the interventions influence the processing of the AOIs defined in Section 6.

As a result of this process, we obtain 28 unique patterns, which are reported in Table A in the appendix. To quantify how representative the patterns are during MSNV processing, we use two complementary metrics used in related work for pattern mining from eye-tracking scanpaths [32, 44, 66], described next:

- *Sequence Support*: The **Sequence Support (SS)** metric captures the proportion of scanpaths in which the user exhibits the pattern at least once, and is defined as the number of scanpaths that contain at least one instance of the pattern, divided by the total number of scanpaths. Because in our analysis each scanpath corresponds to a user task, this metric shows the prevalence of patterns across tasks.
- *Average Pattern Frequency*: The **Average Pattern Frequency (AFP)** metric indicates how frequently a pattern is exhibited in each scanpaths on average and is defined as the number of occurrences (including repetitions) of the pattern in all scanpaths divided by the total number of scanpaths. This metric complements SS by indicating how frequently a pattern recurs in the tasks.

7.2 Statistical Analysis

We evaluate differences in the users' scanpaths with and without adaptations using Mixed-Effect Models, similarly, as described in Section 6.2. For each of the 28 patterns of interest, we fit two models: one to evaluate differences in SS and one to evaluate differences in APF. For SS, we fit a GLMM for the binomial family using the *glmer* function in the lme4 package in R [4], with a binary value indicating whether the pattern is found or not in the scanpath of the task as the dependent variable, with *group* (adaptive, control) as the independent variable, *vis literacy* (low, medium, high) as the fixed effect, and *Participant-ID* and *MSNV-ID* as random effects. For APF, we fit an LMM using the lmerTest package in R [38] with the number of occurrences of the pattern in the scanpath of the task as the dependent variable, *group* as the independent variable, *vis literacy* as the fixed effect, and *Participant-ID* and *MSNV-ID* as random effects. As a result, we fit 28 patterns x 2 metrics = 56 models, which we accounted for by reporting results significant after applying the Benjamini–Hochberg procedure to control for the FDR [5].

Table 6. Patterns for Which a Significant Main Effect of Group Was Found

	Pattern	Metric	Main Effect of Group
Set 1	Ref(×2)-RBar	SS	More in adaptive $\chi^2(1) = 7.28$, $p = .0070$, $r = .27$
	Ref(×3)-RBar	SS	More in adaptive $\chi^2(1) = 8.43$, $p = .0037$, $r = .29$
	Ref(×4)-RBar	SS	More in adaptive $\chi^2(1) = 7.78$, $p = .0053$, $r = .28$
	Ref(×5)-RBar	SS	More in adaptive $\chi^2(1) = 9.45$, $p = .0021$, $r = .31$
	Ref(×6)-RBar	SS	More in adaptive $\chi^2(1) = 7.98$, $p = .0047$, $r = .29$
Set 2	RBar-Lab-RBar	APF	More in adaptive $\chi^2(1) = 8.49$, $p = .0036$, $r = .29$
	RBar(×2)-Lab	APF	More in adaptive $\chi^2(1) = 7.77$, $p = .0053$, $r = .28$
	Lab-RBar(×2)	APF	More in adaptive $\chi^2(1) = 7.17$, $p = .0073$, $r = .27$

Note: Successive fixations are indicated in parentheses (e.g., Ref(×6) means six consecutive fixations within the Ref AOI).

7.3 Results

The analysis of scanpaths yielded several main effects of group on different patterns for either SS or APF metrics. The patterns for which a main effect was found are reported in Table 6. The first column indicates the pattern of interest, the second column indicates the metric for which a significant effect was found, and the third column indicates the results of the statistical tests. No significant interaction effect with vis literacy was found for any of the patterns tested.

We divided the patterns in Table 6 into two sets based on their similarity. The first set, shown at the top of Table 6, captures significant main effects on SS for patterns involving transitions from reference sentences (Ref) to relevant bars (RBar). The second set, shown at the bottom of Table 6, captures significant main effects on APF for patterns involving transitions between relevant bars (RBar) and labels (Lab). As with the gaze metrics evaluated in Section 6, all of the main effects found indicate that users in the adaptive group present higher values of the discriminative patterns than users in the control group. We discuss these two sets of patterns next.

Pattern Set 1: Reference sentences (Ref) to relevant bars (RBar). This first set of patterns involves a series of two to six fixations on reference sentences before transitioning to a relevant bar. Because the effect for these patterns are quite similar, we solely show, in Figure 12, the effect for the first pattern in this set, namely "Ref(×2)-RBar." These effects all show that these patterns are substantially more frequent in the adaptive group as compared to the control group, which suggests that the adaptive interventions are successful at directing the users' attention toward the relevant bars when they read a reference sentence, as captured by the repeated fixations on the references. This is encouraging since dynamically guiding the user's attention to the appropriate bars in the charts when it is most relevant (i.e., when the user is attending to that piece of information in the text) is the main purpose of the adaptive interventions, as discussed in Section 3.2. These results are also consistent with the findings from the intervention processing analysis in Section 5, that most of the interventions were processed by the participants while active.

The fact that the main effects for these patterns are only found on SS is likely due to the fact that the MSNV snippets include few references (2.5 on average, cf. Section 3.2), and thus not so

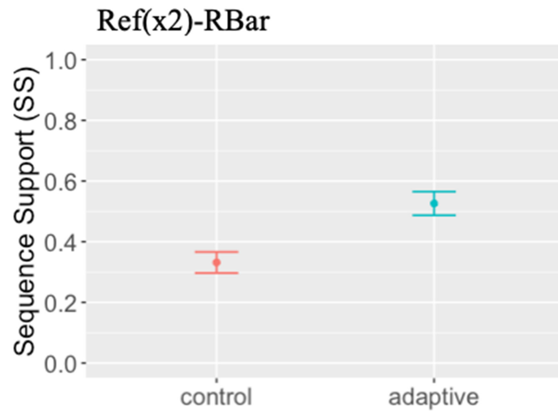


Fig. 12. Main effect for pattern set 1. Error bars show 95% confidence intervals.

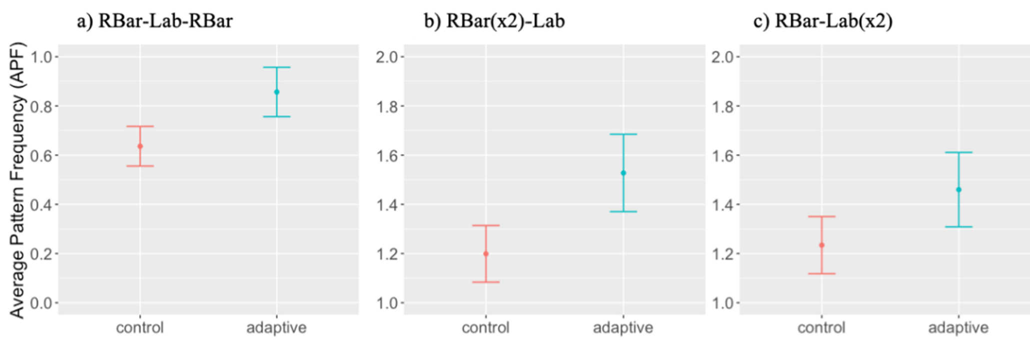


Fig. 13. Main effect for pattern set 2. Error bars show 95% confidence intervals.

many transitions from the references to the chart are needed within each task. The results for SS indicate that participants in the adaptive group exhibited these patterns in more tasks (up to about 55% of the tasks for the pattern in Figure 12) compared to the participants in the control group (33% of the tasks in Figure 12).

Interestingly, the number of transitions from Ref to R-Bar was already studied in the gaze metrics analysis in Section 6 but did not yield any significant main effect. Thus, the results for this set of patterns highlight the complementarity of the gaze metrics and scanpath analysis, as we needed to look at longer sequences to be able to capture this difference in behavior.

Pattern Set 2: Relevant Bars (RBar) to Labels (Lab), and vice versa. The second set of patterns involves different combinations of transition back and forth between relevant bars and labels, which are found to appear, on average, more times per task in the adaptive group compared to the control group (Figure 13). All of these main effects are found for APF but not for SS, suggesting that the participants in both the adaptive and the control group used these patterns in a similar amount of tasks, with the adaptive group reusing them more often within each task.

The fact that participants transitioned more frequently from the relevant bars to the labels in the adaptive group (Figure 13(b)) is consistent with the results from the gaze metrics analysis in Section 6.3, which revealed similar findings in terms of the number of transitions from Relevant Bars to Labels. In addition, here we found that this transition is contextualized within longer patterns (either preceded or followed by another fixation on the RBar AOI, see Figure 13(a) and (b)). We also found a significant effect for patterns comprising the opposite transition (Label to RBar, see Figure 13(a) through (c)), which was not found in the gaze metrics analysis. These opposite patterns are interesting because they provide additional indications that the adaptive interventions fostered higher integration of the relevant bars and the labels, which in turn may be indicative of better information processing on the visualizations.

As mentioned earlier, we found no interaction effect of group with vis literacy on any of the patterns, meaning that all participants in the adaptive group exhibited higher usage of the patterns found in Table 6 than the control group, regardless of their levels of vis literacy. Although integrating the text and the visualization is a known difficulty in processing MSNVs, it is still possible that these patterns are attributable to the improved performance of the low vis literacy users in the adaptive group. Namely, the first set of patterns may indicate that guiding the attention of the low vis literacy users to the relevant bars help them focus on the relevant information in the chart, while they read through the text. Similarly, the patterns in the second set could indicate that the interventions helped the low vis literacy users going back and forth from the relevant bars to the labels, thus integrating both. And although the interventions induced the same behaviors in the higher vis literacy users, it is possible that high vis literacy users did not benefit from them because they did not need such help, perhaps due to their already high vis literacy, although further analysis is needed to ascertain how these patterns relate to the performance of the high and low vis literacy users.

8 ANALYSIS ON USERS WHO WERE DISTRACTED BY THE INTERVENTIONS

In this section, we examine whether the self-reported levels of distraction caused by the interventions in the adaptive group may have influenced the eye movements of the participants. We especially want to ascertain whether the participants who reported some levels of distraction did not exhibit unwanted gaze behaviors, such as flickering toward the interventions. We do so by splitting users in two groups based on their ratings of the 7-point Likert scale item (ranging from 1 to 7) on distraction in the usability questionnaire (see Section 3.3): those who reported low levels of distraction (ratings lower than 4, $N = 21$) versus those who reported moderate to high levels of distraction (ratings greater or equal to 4, $N = 25$). We compare the AOI gaze metrics and scanpaths of these two groups, depending on the levels of vis literacy of the participants.

8.1 Statistical Analysis

We borrow methodologies for Sections 5 through 7 and run three set of statistical analyses on distraction:

1. We run a set of LMM over the metrics in Table 3, by selecting the metrics listed one at a time as the dependent variable, with *Distraction* (low, high) and *vis literacy* (low, medium, high) as the fixed effects, and *Participant-ID* and *MSNV-ID* as random effects.
2. We run a set of Mixed Models over the summative gaze metrics described in Section 6.1, by selecting each of the 30 metrics one at a time as the dependent variable, with *Distraction* (low, high) and *vis literacy* (low, medium, high) as the fixed effects, and *Participant-ID* and *MSNV-ID* as random effects.
3. We mine frequent patterns in the adaptive group (SS > .4) in the same way as in Section 7.1, leading to 26 patterns (see Table A in the appendix) and run a set of Mixed Models to evaluate AFP and SS differences. Analogously to Section 7.1, for SS we fit a GLMM for the binomial family with a binary value indicating whether the pattern is found or not in the scanpath of the task as the dependent variable, with *Distraction* (low, high) and *vis literacy*

Table 7. Gaze and Pattern Metrics for Which Significant Effect of Distraction Was Found

Measure	Main Effect of Distraction
Average pattern frequency	Less for distracted users
Ref(×3)	$\chi^2(1) = 12.87$, $p = .0003$, $r = .53$
Average pattern frequency	Less for distracted users
Ref(×4)	$\chi^2(1) = 12.36$, $p = .0004$, $r = .52$
Number of fixations	Less of distracted users
Ref	$\chi^2(1) = 9.31$, $p = .0023$, $r = .45$

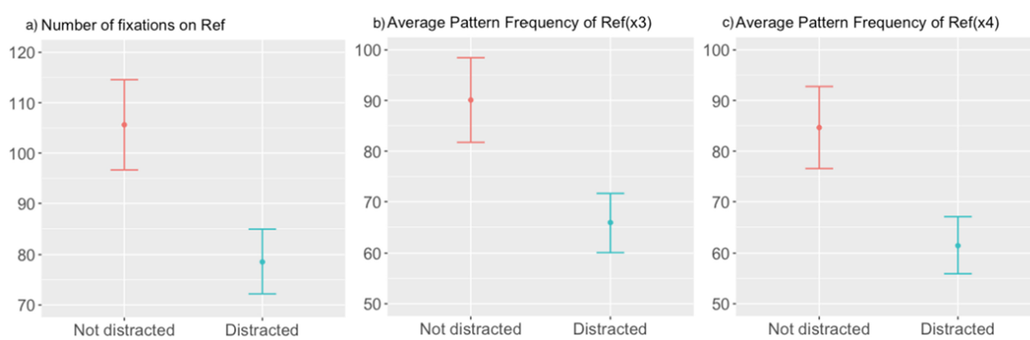


Fig. 14. Main effects found on Distraction. Error bars show 95% confidence intervals.

(low, medium, high) as the fixed effects, and *Participant-ID* and *MSNV-ID* as random effects. For APF, we fit an LMM with the number of occurrences of the pattern in the scanpath of the task as the dependent variable, *Distraction* (low, high) and *vis literacy* (low, medium, high) as the fixed effects, and *Participant-ID* and *MSNV-ID* as random effects.

8.2 Results

Table 7 presents the three main effects of distraction that were found across the three sets of analysis that we conducted, and the effects are shown in Figure 14. No significant interaction effects with vis literacy were found.

We found a significant effect of Distraction on the number of fixations on the Reference sentence AOI, indicating that distracted users do less fixations on this AOI than non-distracted ones (Figure 14, left). We also found a main effect of distraction on two patterns related to consecutive fixations on these references [Ref(×3) and Ref(×4)], showing that distracted users exhibit these patterns less than non-distracted users (Figure 14, middle and right). This indicates overall than distracted users process slightly less extensively the reference sentences, perhaps because their attention is pulled away from the text due to the interventions, hence their levels of reports distraction. However, non-distracted users seem to be able to dedicate more fixations to the reference sentences, possibly because they finish reading the references instead before processing an intervention. However, the analysis did not reveal any further significant differences that could help understand this behavior. In particular, we found no significant difference in how many interventions the distracted and non-detracted users process, nor in how long they process them. This suggests that the self-reported distraction overall has little impact on the MSNV gaze processing, which is also consistent with the fact that distraction was not found to impact comprehension and reading time in previous work [42]. We also did not find any interaction effect of vis literacy, which indicates that eye movements specific to distracted users are not characterized by levels of vis literacy.

9 DISCUSSION

We first discuss the implications that the presented eye-gaze analysis holds for evaluating adaptive visualizations (Section 9.1). We then discuss how our results inform the design of adaptive guidance in narrative visualizations, via personalization to the user's levels of vis literacy (Section 9.2). Last we discuss the main limitations of our work and how they can guide future work.

9.1 Eye-Tracking for Evaluating Adaptive Visualizations

Gaze-driven guidance to support visualization processing has recently attracted more attention from the InfoVis research community (e.g., [22, 42, 64]). These previous works, however, have not extensively evaluated how their adaptive guidance influences the user's gaze processing, as they mostly leveraged attention maps and fixation time over the adaptive component to evaluate [22, 64]. We extend these works by showing that fine-grained eye-tracking analysis can be valuable to understand how and why the target adaptive guidance may benefit specific groups of users in narrative visualizations. Specifically, we show that leveraging gaze metrics over different regions of the visualization, along with the entire user scanpath, is important as well to uncover the user's processing strategy. Scanpaths have also been extensively used in usability studies (e.g., [14, 17, 18]); however, to the best of our knowledge, we are the first to leverage scanpaths to evaluate the effectiveness of gaze-driven guidance in visualization tasks.

Our results show that highlighting one specific part of the MSNV seems to generate proactive processing of other important related components in the visualization, as well as increased transitions between these components. Thus, one should leverage processing metrics (scanpaths and transitions across AOIs) in addition to attention metrics (fixation within AOIs) to evaluate in depth the effectiveness of adaptive visualizations. Furthermore, we found that the scanpaths can better capture how the adaptive interventions guide the users' attention toward the highlighted components than summative gaze metrics do, as only our scanpath analysis revealed that the interventions generated increased transitions from relevant bars to the relevant bars, which was the main purpose of the dynamic interventions in the first place. This is especially interesting since these transitions captured by the scanpaths are the only one that involve processing of both modalities of the narrative visualizations (text and visualization), showing that scanpath may be very suitable to evaluate adaptive guidance in multimodal documents.

Our results suggest that the adaptive interventions promote specific gaze strategies to process the bar charts in the MSNVs, namely by transitioning back and forth from the relevant bars to the labels, as well as by transitioning from the labels to the legend, and then back to the relevant bars. These findings are quite interesting because these are the key components of a bar chart, and integrating these elements is crucial to fully understand the data. Interestingly, although users spent the same amount of time reading the MSNV across the control and adaptive groups, we found no salient gaze behaviors or scanpath patterns that the control group exhibits more than the adaptive group. This is likely because without the interventions, participants explore the MSNVs in more heterogeneous and unique ways and thus do not generate recurrent behaviors that could be captured in our analysis.

We also controlled for the participants' levels of self-reported distraction caused by the interventions, which to the best of our knowledge is novel, and found that distraction generates less fixations on the reference sentences. This suggest that users distracted by the interventions may process the references less extensively, possibly because their attention is pulled away too soon. However, we did not find any other results for distracted users, suggesting that overall, they process the MSNV in a similar way. In particular, we did not find any flickering or quick glancing that could have indicated unwarranted behaviors. This is consistent with the fact that self-reported distraction did not impact comprehension and reading time [39]. Nonetheless, future work could further explore the timing of the guidance as well as other aspects that could help minimize distracting factors for these users.

9.2 Implications for Personalization

In our previous work [39], we found that when we provide adaptive guidance, vis literacy significantly influences comprehension, namely the guidance only improved the comprehension of participants with low levels of vis literacy users. This suggests that there is to some extent a link between vis literacy, comprehension, and guidance. A main goal of this work (RQ2) was to explain this link further by examining if it can be captured in terms of salient gaze patterns that are intrinsic to low vis literacy users only. Our results show that providing guidance through adaptive interventions encourages several gaze behaviors, which can be leveraged to refine the adaptation mechanism via personalization to the user's levels of vis literacy. We elaborate in the following on the implications of these result, first for users with low vis literacy and second for participants with higher vis literacy.

Users with low vis literacy. These are the users who were found to significantly improve their comprehension of the MSNVs when receiving adaptive interventions in our earlier work [42], and our analysis identified gaze behaviors that can explain their improved performance. We found that, when provided with adaptive interventions, these users fixated more on the relevant bars, transitioned more between the relevant regions of the bar chart, and were more engaged in processing the legend (Sections 6.3 and 7.3). Although all users exhibited these behaviors, it is possible that users with low vis literacy benefited the most from them because they needed help the most due to their low literacy, particularly their lower ability to process the salient features of the chart [7]. Furthermore, users with low vis literacy transitioned more between the legend and the relevant bars when provided with adaptive interventions, suggesting that they better integrate these two key elements (Section 6.3). Altogether, these changes in gaze behavior may explain the increase in comprehension performance for these users, as mapping the relevant bars to their labels and to the information encoded in the legend is fundamental to make sense of bar charts. This may prove to be a challenging task for users with low vis literacy, and we provide insights into how the adaptive interventions help them in doing so.

Although our findings indicate that the current interventions are beneficial to low vis literacy users overall, it is still possible that some of the low vis literacy users do not perform the useful gaze behaviors that we have identified. Moving forward, it could be worthwhile to further personalize the guidance to these users, by encouraging them to perform these behaviors when they do not do it spontaneously, as captured by eye-tracking. We could do so by, for instance, highlighting the relevant labels and items in the legend, in addition to the current highlights.

Users with medium and high vis literacy. We found that the gaze behaviors of these users are overall influenced by the adaptive interventions, showing that they do notice and process the highlights as much as the low vis literacy users (Sections 5, 6.3, and 7.3). Importantly, these behaviors generated by the adaptation pertain to processing the relevant parts of the bar charts, which was the intended goal of the adaptation, meaning that the interventions, at least, do not generate unwarranted behaviors in these users. However, our previous work [42] showed that providing these adaptations to users with medium and high vis literacy does not help them better understand the visualizations. This means that the gaze behaviors we found in Sections 6.3 and 7.3 are not sufficient to influence the performance of these users. One possible explanation is that these users are already able to quickly identify the relevant information in the bar chart, due to their higher vis literacy, and they do not benefit as much from improved integration of the relevant bars with the labels and the legend. They may also not need to look at the relevant bars right away after reading a reference sentence, possibly because they prefer to do the mapping from references to relevant bars at their own timing. Based on these findings, we encourage future work to investigate whether these users might benefit from other forms of adaptive interventions more suited to their needs.

Altogether, our findings show that the effectiveness of adaptive interventions in narrative visualizations could be improved via personalization to the users' levels of vis literacy. To deliver such personalized interventions, one possible approach is to automatically infer the users' levels of vis literacy from their gaze behaviors, as done in the work of Conati et al. [12]. This would be especially suitable for narrative visualizations as our results indicate that users exhibit specific gaze behaviors that might accurately reveal their levels of vis literacy.

9.3 Limitations

There are two main limitations in our work that can drive future research:

First, our goal of this work with RQ2 was to examine what are the salient gaze patterns that are intrinsic to low vis literacy users only and could explain their improve comprehension as compared to high vis literacy users. We found one such interaction effect (see Table 5), as discussed in Section 9.2, and we argue that, together with the rest of main effects of group found (see Table 4 and Table 6), they can partly explain the link between vis literacy, comprehension, and guidance. However, this link could go beyond gaze behaviors, and there is a possibility that we are not able to fully capture it using analysis of eye movements alone. Further research is needed to validate whether comprehension accuracy at different levels of vis literacy is due to some other reasons that are not being captured by the analysis of users' eye movements.

Second, although eye-tracking technology is becoming more ubiquitous and accurate, there remain limitations related to its data quality and accuracy for certain users (representing ∼27% of users in our dataset, see Section 3.4), issues that are likely to be exacerbated outside the lab. These limitations are potentially problematic in gaze-based adaptive systems, given that an inaccurate gaze estimation could lead to unwanted interventions. Improving eye-tracking accuracy is outside the scope of our research; however, in future work we should further investigate mechanisms to better support users for which eye-tracking accuracy is low, such as the possibility to disable or modify the behavior of the interventions, minimizing the disadvantageous effects of inaccurate interventions.

10 CONCLUSION

In this article, we leveraged eye-tracking to evaluate the effectiveness of gaze-driven adaptive interventions in MSNVs, a widespread form of narrative visualizations in real-world sources. The evaluated interventions consist of dynamically highlighting the relevant datapoints (bars in a bar chart) when users read a reference sentence that describes them, as captured by an eye-tracker. Our previous work has shown that such adaptive interventions can improve comprehension of the MSNV, but only for users with low vis literacy. We extend this previous work by revealing the specific gaze behaviors that the adaptive interventions generate in users, depending on their levels of vis literacy. In particular, providing adaptive interventions lead to overall more fixations on the relevant datapoints, longer fixations on the legend, and more gaze transitions between the key components of the visualization (datapoints, labels, legend). These results provide encouraging evidence for this type of adaptive guidance, showing that this mechanism is effective not only in guiding the users' attention toward the relevant datapoints but also in facilitating processing of the relevant information in the MSNV. In addition, we found changes specific to users with low vis literacy, namely an increased amount of transitions from the relevant datapoints to the legend of the visualization.

This further explains how the adaptive interventions helped users with low vis literacy better contextualize and integrate the relevant datapoints with the rest of the components in the visualization, leading to improved comprehension. All in all, our evaluation sheds light on the underlying processing behaviors of users in adaptive narrative visualizations, driving the design of future adaptive guidance mechanisms, more personalized to the users' vis literacy.

APPENDIX A

Table A. Patterns Mined for the Analysis of Scanpaths

	Control		Adaptive	
Pattern	SS	APF	SS	APF
Lab(×3)	0.64	4.86	0.69	4.7
Lab(×2)-NRBar	0.44	0.91	0.44	0.94
Lab(×2)-RBar	0.56	1.39	0.61	1.53
Lab-NRBar(×2)	0.42	0.83	0.38	0.87
Lab-RBar(×2)	0.54	1.23	0.58	1.46
NRBar-Lab(×2)	0.46	1.01	0.47	1.05
NRBar(×2)-Lab	0.43	0.85	0.4	0.93
NRBar(×3)	0.42	2.24	0.38	1.82
Ref(×3)	0.98	74.3	1	76.93
Ref(×2) -RBar	0.33	0.52	0.53	1.09
RBar-Lab(×2)	0.56	1.39	0.66	1.72
RBar-Lab-RBar	0.36	0.64	0.44	0.86
RBar(×2)-Lab	0.55	1.2	0.61	1.53
RBar(×3)	0.6	3.33	0.61	3.76
Lab(×4)	0.49	3.15	0.51	2.96
Lab(×3)-RBar	0.43	0.83	0.43	0.8
Lab(×2)-RBar(×2)	0.39	0.68	0.42	0.77
Ref(×4)	0.97	69.42	0.99	72.07
Ref(×3)-RBar	0.3	0.45	0.5	1
RBar-Lab(×3)	0.43	0.8	0.48	0.89
RBar(×2)-Lab(×2)	0.39	0.64	0.44	0.81
RBar(×4)	0.42	1.98	0.43	2.19
Ref(×5)	0.97	65.14	0.99	67.89
Ref(×4)-RBar	0.28	0.4	0.47	0.88
Ref(×6)	0.97	61.3	0.98	64.15
Ref(×5)-RBar	0.25	0.35	0.44	0.81
Ref(×7)	0.96	57.85	0.98	60.75
Ref(×6)-RBar	0.22	0.31	0.42	0.73

Note: Patterns were of length 3–10 fixations and appeared in at least 40% of the tasks in the control or adaptive group.

APPENDIX B

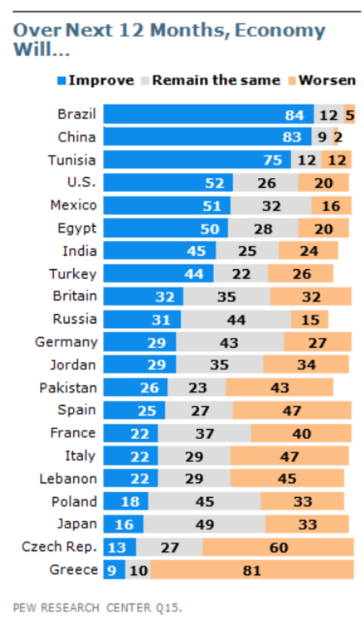
This appendix includes the following:

- A screenshot of all 14 MSNVs used in the user study reported in this article.
- The set of comprehension questions participants were asked about each MSNV. Note that the comprehension questions are asked separately right after reading the corresponding MSNV (i.e., the participants did not see both together).

Pervasive Gloom About the World Economy

July 12, 2012

In the wake of four years of economic turmoil around the world and political upheaval in a number of nations, very few people are satisfied with the way things are going in their country. In only six of the 21 nations surveyed do half or more of the population think national economic conditions will improve over the next 12 months. This includes very optimistic Brazilians (84%), Chinese (83%) and Tunisians (75%) and relatively optimistic Americans (52%), Mexicans (51%) and Egyptians (50%). In addition, a plurality of Indians (45%) and Turks (44%) see a better economy on the horizon. But in seven countries majorities or pluralities think economic conditions will worsen, including 81% of Greeks and 60% of Czechs.



Comprehension Questions:

☐ In which nation does the majority believe their economic situation will worsen?

- ○ Egyptian
- ○ Czechs

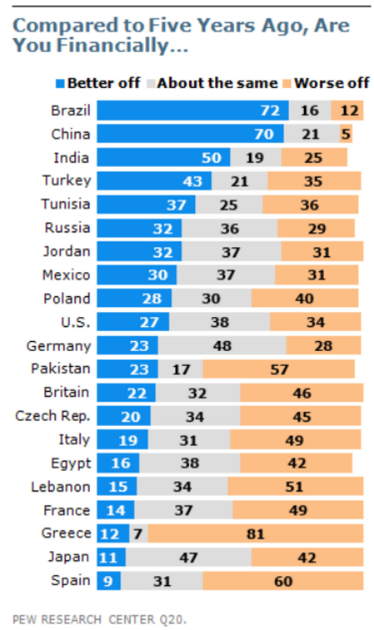
☐ The following is a suitable alternative title:

- ○ Economic turmoil around the world sees reversal with nations on the mend
- ○ Nations around the world have mixed feelings on economic situation

Pervasive Gloom About the World Economy

July 12, 2012

People may think their personal situation is better than economic conditions in their nation, but only in Brazil (72%) and China (70%) do large majorities think their families are better off than they were five years ago. On balance, Indians (50%) and Turks (43%) also say their situations have improved. However, majorities or pluralities in several nations say their financial situation has deteriorated. Solid majorities hold this view in Greece (81%), Spain (60%) and Pakistan (57%), as do at least four-in-ten in Lebanon, Italy, France, Britain, the Czech Republic, Japan, Egypt and Poland.



Comprehension Questions:

☐ In which nation does a plurality believe that their financial situation has deteriorated?

- ◯ France
- ◯ India

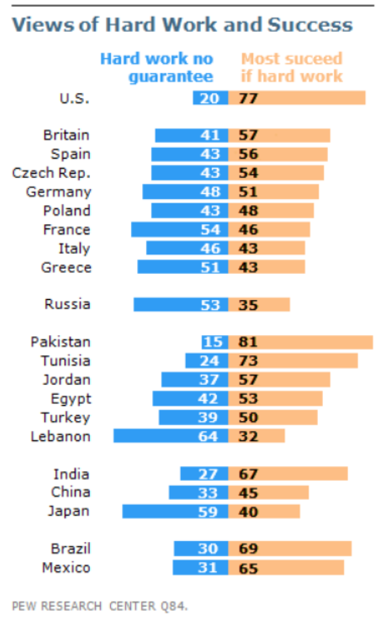
☐ The following is a suitable alternative title:

- ◯ Few families better off than five years ago
- ◯ Most people views their nation’s economic situation as worsening

Pervasive Gloom About the World Economy

July 12, 2012

The idea that hard work leads to material success is no longer, if it ever was, a uniquely Western value. Half or more in 13 of the 21 nations surveyed believe that most people can succeed if they are willing to work hard. This includes Pakistan (81%) and the U.S. (77%). It also includes Tunisia (73%), Brazil (69%), India (67%) and Mexico (65%).



Comprehension Questions:

☐ Which nation is among those that “Half or more in 13 of the 21 nations surveyed believe that most people can succeed if they are willing to work hard”?

- ◯ France
- ◯ India

☐ The following is a suitable alternative title:

- ◯ Widespread outlook on the world economy
- ◯ A survey of each nation’s views on their economic success

Percentage of people trusting various components that comprise the Financial Trust Index

October 30, 2012

"Trust in banks has returned to levels we've reported in the last year of the Index, bouncing back to 33 percent from 27 percent just three months ago. The low level of trust banks experienced last quarter was likely due to the effect of the JP Morgan scandal, demonstrating that very public cases of mismanagement can have short-run effects on trust," said Luigi Zingales, co-author of the Financial Trust Index and the Robert R. McCormack Professor of Entrepreneurship and Finance at the University of Chicago Booth School of Business.

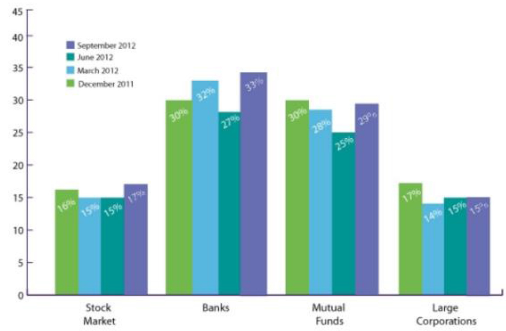


Figure 1: Trust

Comprehension Questions:

☐ According to the article: "bouncing back to 33 percent from 27 percent just three months ago," what month was three months ago?

- ◯ March 2012
- ◯ June 2012

☐ Trust in banks ________ since the levels reported three months ago.

- ◯ Increased
- ◯ Decreased

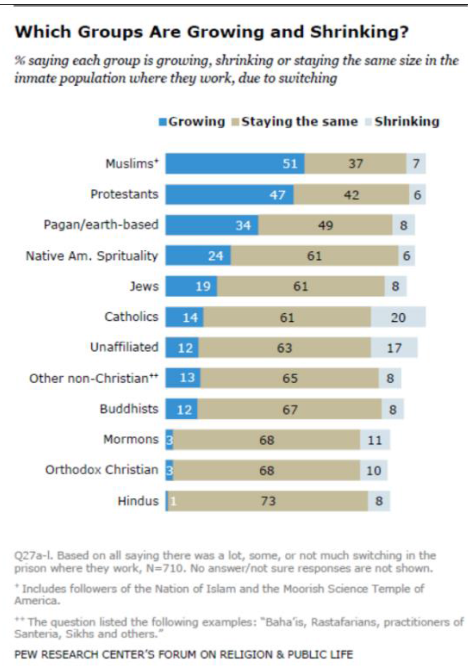
☐ The following is a suitable alternative title:

- ◯ Trust levels in the Financial Trust Index
- ◯ Trust levels in various components of the global stock market

Religion in Prisons – A 50-State Survey of Prison Chaplains

March 22, 2012

To get a sense of which religious groups are gaining the most converts, the Pew Forum survey asked chaplains to estimate whether the number of inmates in each of 12 religious groups is increasing, decreasing or staying at about the same level. Among chaplains who report that at least some switching occurs within the correctional facilities where they work, about half (51%) report that Muslims are growing in number, and 47% say the same about Protestant Christians. A sizable minority of chaplains answering this question also say that followers of pagan or earth-based religions are growing (34%). For nine of the 12 religious groups considered, however, a solid majority (61% or more) of chaplains answering the question report that the size of each group is stable. And for several religious groups, the chaplains are as likely, or even more likely, to report shrinkage as to report growth. For example, one-in-five chaplains answering this question (20%) say that the number of practicing Catholics behind bars is shrinking due to switching, while 14% say the ranks of Catholics are growing. Similarly, 17% say that the number of inmates with no religious preference is shrinking, while 12% say the ranks of the unaffiliated are growing. And about one-in-ten chaplains report a decline in Mormons and Orthodox Christians due to switching, while only 3% say those religious minorities are growing behind bars.



Comprehension Questions:

☐ Which religious group is among those that "a solid majority (61% or more) of chaplains answering the question report that the size of each group is stable."?

◯ Protestants

◯ Buddhists

☐ 17% of chaplains say that the number of inmates with no religious preference is ________.

◯ Shrinking

◯ Growing

☐ The following is a suitable alternative title:

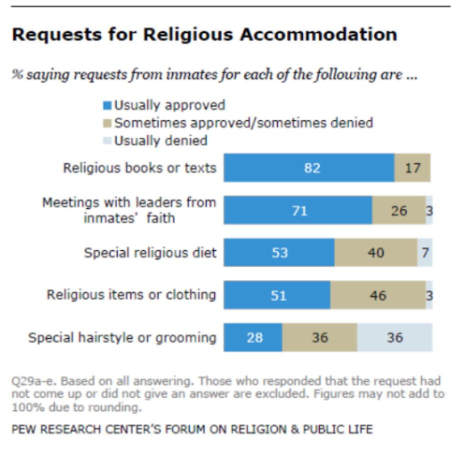
◯ Which religious groups are growing and shrinking?

◯ Practicing behind bars: facing off with other religious groups

Religion in Prisons – A 50-State Survey of Prison Chaplains

March 22, 2012

The Pew Forum survey included several questions designed to probe the kinds of requests that inmates make for accommodation of their religious beliefs and practices, as well as the frequency with which they are granted. An overwhelming majority of chaplains who responded to these questions say that inmates' requests for religious texts (82%) and for meetings with spiritual leaders of their faith (71%) are usually approved. And about half of chaplains say that requests for a special religious diet (53%) or for permission to have sacred items or religious clothing such as crucifixes, eagle feathers and turbans (51%) also are usually granted.



Comprehension Questions:

☐ Select the religious item requested in prisons that was mentioned in the article:

◯ Bibles

◯ Turbans

☐ Requests for special religious diets in prison are usually ________.

◯ Approved

◯ Not approved

☐ The following is a suitable alternative title:

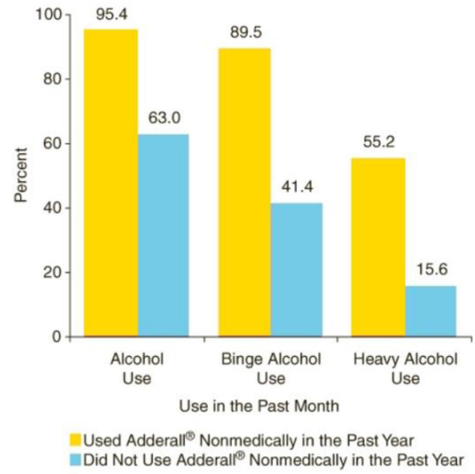
◯ Religious requests from inmates are running rampant in prison

◯ Prison chaplains provide feedback on religious accommodations in prison

Nonmedical Use of Adderall®, by College Enrollment Status

April 7, 2009

Among full-time college students aged 18 to 22, those who used Aderall® nonmedically in the past year were more than 1.5 times as likely as their counterparts to have used alcohol in the past month (95.4 vs. 63.0 percent), more than twice as likely to have been binge alcohol users (89.5 vs. 41.4 percent), and more than 3 times as likely to have been heavy alcohol users (55.2 vs. 15.6 percent) (see Figure). Similar patterns were observed for underage full-time college students (i.e., those aged 18 to 20) who used Adderall® nonmedically in the past year and for nonmedical Adderall® users of legal drinking age compared with their counterparts who had not used it nonmedically (data not shown).



Source: 2006 and 2007 SAMHSA National Surveys on Drug Use and Health (NSDUHs).

Comprehension Questions:

☐ What were students who used Adderall nonmedically in the past more than three times as likely to have done?

◌ Heavy Alcohol Use

◌ Binge Alcohol Use

☐ Students who used Adderall nonmedically in the past were more than _______ times as likely to have been alcohol users.

◌ 1.5

◌ 2.5

◌ 3.5

☐ The following is a suitable alternative title:

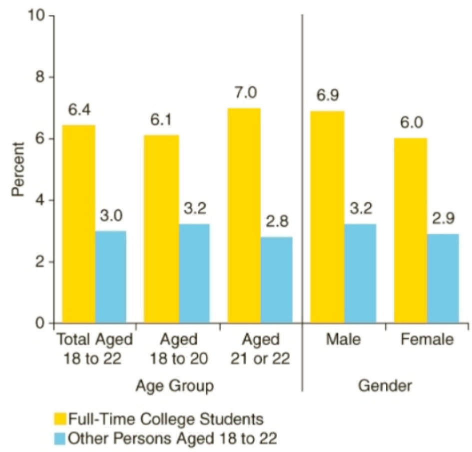
◌ Universities concerned over excessive Adderall use

◌ Results from national surveys on drug use and health

Nonmedical Use of Adderall®, by College Enrollment Status

April 7, 2009

Full-time college students aged 18 to 22 were twice as likely as their counterparts who were not full-time college students to have used Adderall® nonmedically in the past year (6.4 vs. 3.0 percent) (Figure 1). This pattern was found for both males and females and for persons aged 18 to 20 as well as for those 21 or 22 years old.



Source: 2006 and 2007 SAMHSA National Surveys on Drug Use and Health (NSDUHs).

Comprehension Questions:

☐ Full-time college students aged 21–22 were _______ as likely to have used Adderall nonmedically.

- ◯ Two times
- ◯ Four times
- ◯ One and a half times

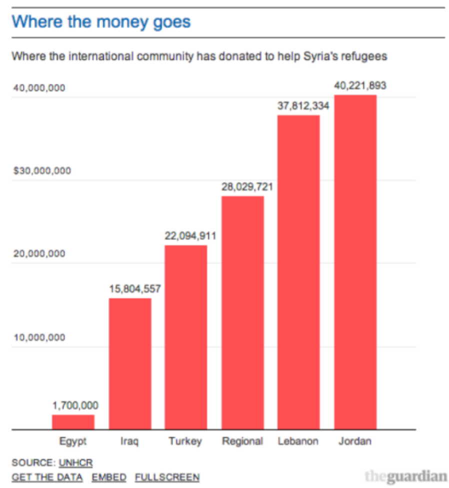
☐ The following is a suitable alternative title:

- ◯ College students abusing Adderall during exam season
- ◯ Nonmedical use of Adderall in students aged 18–22

Syrian refugees: how many are there and where are they?

March 6, 2013

The humanitarian fallout of the conflict in Syria reaches new proportions as the number of estimated refugees reaches one million. Who's Helping? Some contributions are made on a regional basis, but many donors prefer to contribute to efforts in a specific country. In line with the distribution of the refugees themselves, most funds are funnelled towards Jordan (28%), followed by Lebanon (26%), Turkey (15%) and Iraq (11%).



Comprehension Questions:

☐ Turkish funding is _____ than Lebanese funding.

○ Greater

○ Lesser

☐ The following is a suitable alternative title:

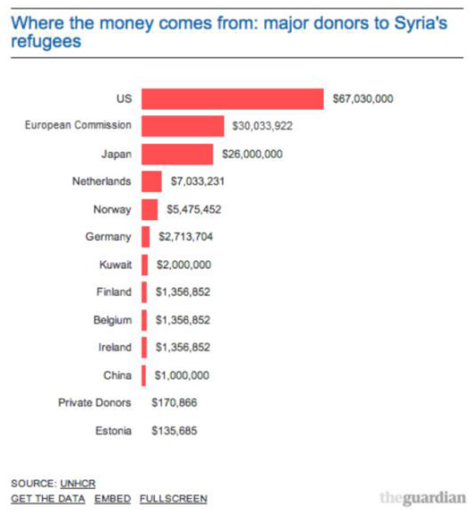
○ Major contributors supporting Syrian refugees

○ Syrian refugees supported in various countries

Syrian refugees: how many are there and where are they?

March 6, 2013

The humanitarian fallout of the conflict in Syria reaches new proportions as the number of estimated refugees reaches one million. Who's Helping? So far, over $145 million has been pledged to alleviate the crisis in 2013 - less than a third of the $494 million which the UN says is needed. Contributing 46%, the United States is by far the biggest donor, followed by the European Commission (21%) and Japan (18%).



Comprehension Questions:

☐ Of the “$145 million pledged to alleviate the crisis in 2013”, the largest donor is the.:

○ United Nations

○ United States

☐ 4. The following is a suitable alternative title:

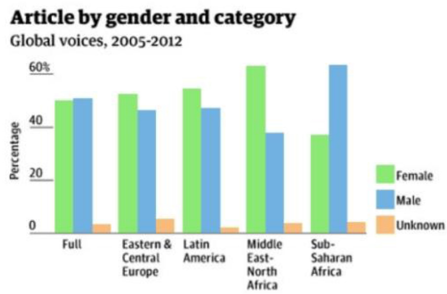
○ Major donors to Syrian refugees

○ Cost of Syrian war escalating

Gender diversity at Global Voices

March 8, 2013

Volunteer sites like Wikipedia often have trouble attracting women contributors, and women Wikipedians tend toward only some kinds of stories. That's not true at Global Voices, where women's writing is strong across all regions. Even in Sub-Saharan Africa, where the majority of posts are written by men, women are writing nearly 40% of the posts. In contrast, only 33% of the Guardian's news articles and 19% of the Telegraph's news articles include women as writers.



Comprehension Questions:

☐ In what part of the world are the majority of posts written by men?

- ○ Sub-Sarahan Africa
- ○ Latin America

☐ In Sub-Sarahan Africa, women are writing __________ than men.

- ○ More
- ○ Less

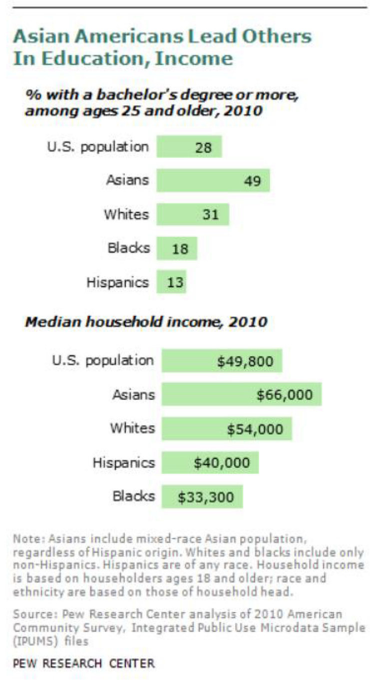
☐ The following is a suitable alternative title:

- ○ Women authors are leading the pack at Wikipedia
- ● Women authors at Global Voices strong across the world

The Rise of Asian Americans

June 19, 2012

Asian Americans trace their roots to any of dozens of countries in the Far East, Southeast Asia and the Indian subcontinent. Each country of origin subgroup has its own unique history, culture, language, religious beliefs, economic and demographic traits, social and political values, and pathways into America. But despite often sizable subgroup differences, Asian Americans are distinctive as a whole, especially when compared with all U.S. adults, whom they exceed not just in the share with a college degree (49% vs. 28%), but also exceed in median annual household income ($66,000 versus $49,800).



Comprehension Questions:

☐ Asian American median household income is ________ than that of all U.S. adults.

○ Greater

○ Lesser

☐ The following is a suitable alternative title:

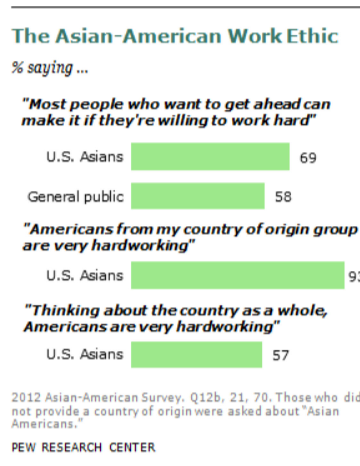
○ Asian Americans concerned over economic conditions

○ Asian Americans are distinctive as a whole

The Rise of Asian Americans

June 19, 2012

Asian Americans have a pervasive belief in the rewards of hard work. Nearly seven-in-ten (69%) say people can get ahead if they are willing to work hard, a view shared by a somewhat smaller share of the American public as a whole (58%). And fully 93% of Asian Americans describe members of their country of origin group as "very hardworking"; just 57% say the same about Americans as a whole.



Comprehension Questions:

☐ 57% of which group thinks that as a whole, Americans are "very hard-working"?

- (●) Asian Americans
- ( ) General public

☐ The following is a suitable alternative title:

- ( ) Asian American population increasing in U.S.
- (●) Examining the Asian American work ethic

REFERENCES

[1] Daniel Archambault, Helen C. Purchase, and Bruno Pinaud. 2010. The readability of path-preserving clusterings of graphs. *Computer Graphics Forum* 29, 3 (2010), 1173–1182. DOI: https://doi.org/10.1111/j.1467-8659.2009.01683.x

[2] P. Ayres and G. Cierniak. 2012. Split-attention effect. In *Encyclopedia of the Sciences of Learning*, Norbet M. Seel (Ed.). Springer, 3172–3175.

[3] Oswald Barral, Sébastien Lallé, and Cristina Conati. 2020. Understanding the effectiveness of adaptive guidance for narrative visualization: A gaze-based analysis. In *Proceedings of the 25th International Conference on Intelligent User Interfaces (IUI '20)*. ACM, New York, NY, 1–9. DOI: https://doi.org/10.1145/3377325.3377517

[4] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2014. Fitting linear mixed-effects models using lme4. arXiv:1406.5823.

[5] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 57, 1 (1995), 289–300.

[6] Katy Börner, Adam Maltese, Russell Nelson Balliet, and Joe Heimlich. 2016. Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization* 15, 3 (2016), 198–213. DOI: https://doi.org/10.1177/1473871615594652

[7] J. Boy, R. A. Rensink, E. Bertini, and J.-D. Fekete. 2014. A principled way of assessing visualization literacy. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1963–1972. DOI: https://doi.org/10.1109/TVCG.2014.2346984

[8] Stephen Boyd Davis, Olivia Vane, and Florian Kräutli. 2016. Using data visualisation to tell stories about collections. In *Proceedings of EVA London*.

[9] Andreas Buja, John Alan McDonald, John Michalak, and Werner Stuetzle. 1991. Interactive data visualization using focusing and linking. In *Proceedings of IEEE Visualization 1991*. 156–163.

[10] Michael Burch, Natalia Konevtsova, Julian Heinrich, Markus Hoeferlin, and Daniel Weiskopf. 2011. Evaluation of traditional, orthogonal, and radial tree diagrams by an eye tracking study. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2440–2448. DOI: https://doi.org/10.1109/tvcg.2011.193

[11] Nora Castner, Enkelejda Kasneci, Thomas Kübler, Katharina Scheiter, Juliane Richter, Thérése Eder, Fabian Hüttig, and Constanze Keutel. 2018. Scanpath comparison in medical image reading skills of dental students: Distinguishing stages of expertise development. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research and Applications.* 1–9.

[12] Cristina Conati, Sébastien Lallé, Md. Abed Rahman, and Dereck Toker. 2017. Further results on predicting cognitive abilities for adaptive visualizations. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence.* 1568–1574. DOI: https://doi.org/10.24963/ijcai.2017/217

[13] Francesco Di Nocera, Michela Terenzi, and Marco Camilli. 2006. Another look at scanpath: Distance to nearest neighbour as a measure of mental workload. In *Developments in Human Factors in Transportation, Design, and Evaluation*, D. de Waard, K. A. Brookhuis, and A. Toffetti (Eds.). Shaker Publishing, 295–303.

[14] Andrew T. Duchowski, Jason Driver, Sheriff Jolaoso, William Tan, Beverly N. Ramey, and Ami Robbins. 2010. Scanpath comparison revisited. In *Proceedings of the 2010 Symposium on Eye-Tracking Research and Applications (ETRA'10).* ACM, New York, NY, 219–226. DOI: https://doi.org/10.1145/1743666.1743719

[15] Wolfgang Einhäuser, Merrielle Spain, and Pietro Perona. 2008. Objects predict fixations better than early saliency. *Journal of Vision* 8, 14 (2008), 18.

[16] Sukru Eraslan, Victoria Yaneva, Yeliz Yesilada, and Simon Harper. 2019. Web users with autism: Eye tracking evidence for differences. *Behaviour & Information Technology* 38, 7 (2019), 678–700.

[17] Sukru Eraslan and Yeliz Yesilada. 2015. Patterns in eyetracking scanpaths and the affecting factors. *Journal of Web Engineering* 14, 5-6 (2015), 363–385.

[18] Sukru Eraslan, Yeliz Yesilada, and Simon Harper. 2016. Eye tracking scanpath analysis techniques on web pages: A survey, evaluation and comparison. *Journal of Eye Movement Research* 9, 1 (2016), 1–19.

[19] Leah Findlater and Joanna McGrenere. 2004. A comparison of static, adaptive, and adaptable menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'04).* ACM, New York, NY, 89–96. DOI: https://doi.org/10.1145/985692.985704

[20] Mary E. Frame, Rik Warren, and Anna M. Maresca. 2019. Scanpath comparisons for complex visual search in a naturalistic environment. *Behavior Research Methods* 51, 3 (2019), 1454–1470. DOI: https://doi.org/10.3758/s13428-018-1154-0

[21] Mohammad Ghoniem, J.-D. Fekete, and Philippe Castagliola. 2004. A comparison of the readability of graphs using node-link and matrix-based representations. In *Proceedings of the IEEE Symposium on Information Visualization.* IEEE, Los Alamitos, CA, 17–24. DOI: https://doi.org/10.1109/INFVIS.2004.1

[22] Fabian Göbel, Peter Kiefer, Ioannis Giannopoulos, Andrew T. Duchowski, and Martin Raubal. 2018. Improving map reading with gaze-adaptive legends. In *Proceedings of the ACM Symposium on Eye Tracking Research and Applications (ETRA'18).* ACM, New York, NY, Article 29, 9 pages. DOI: https://doi.org/10.1145/3204493.3204544

[23] Tamara van Gog. 2014. The signaling (or cueing) principle in multimedia learning. In *The Cambridge Handbook of Multimedia Learning* (2nd ed.), Richard Mayer (Ed.). Cambridge University Press, Cambridge, MA, 263–278. DOI: https://doi.org/10.1017/CBO9781139547369.014

[24] Joseph Goldberg and Jonathan Helfman. 2011. Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. *Information Visualization* 10, 3 (2011), 182–195. DOI: https://doi.org/10.1177/1473871611406623

[25] Lewis R. Goldberg. 1999. A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality Psychology in Europe* 7, 1 (1999), 7–28.

[26] Joseph A. Harsh, Molly Campillo, Caylin Murray, Christina Myers, John Nguyen, and Adam V. Maltese. 2019. "Seeing" data like an expert: An eye-tracking study using graphical data representations. *Life Sciences Education* 18, 3 (2019), Article 32. DOI: https://doi.org/10.1187/cbe.18-06-0102

[27] Nathalie Henry, Anastasia Bezerianos, and Jean-Daniel Fekete. 2008. Improving the readability of clustered social networks using node duplication. *IEEE Transactions on Visualization and Computer Graphics* 14, 6 (2008), 1317–1324. DOI: https://doi.org/10.1109/TVCG.2008.141

[28] Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost van de Weijer. 2015. *Eye Tracking: A Comprehensive Guide to Methods and Measures.* Oxford University Press.

[29] Anthony Jameson and Krzysztof Z. Gajos. 2012. Systems that adapt to their users. In *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications.* CRC Press, Boca Raton, FL, 431–456.

[30] Eija Kaasinen. 1999. Usability challenges in agent based services. In *Intelligence in Services and Networks Paving the Way for an Open Service Market*, Lecture Notes in Computer Science, Vol. 1597, Springer, 131–142. DOI: https://doi.org/10.1007/3-540-48888-X_14

[31] Slava Kalyuga. 2007. Enhancing instructional efficiency of interactive e-learning environments: A cognitive load perspective. *Educational Psychology Review* 19, 3 (2007), 387–399. DOI: https://doi.org/10.1007/s10648-007-9051-6

[32] John S. Kinnebrew and Gautam Biswas. 2012. Identifying learning behaviors by contextualizing differential sequence mining with action features and performance evolution. In *Proceedings of the 5th International Conference on Educational Data Mining.* 57–64.

[33] Ha-Kyung Kong, Wenjie Zhu, Zhicheng Liu, and Karrie Karahalios. 2019. Understanding visual cues in visualizations accompanied by audio narrations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, Article 50, 13 pages.
[34] Nicholas Kong, Marti A. Hearst, and Maneesh Agrawala. 2014. Extracting references between text and charts via crowdsourcing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Los Alamitos, CA, 31–40. DOI: https://doi.org/10.1145/2556288.2557241
[35] Philipp Koytek, Charles Perin, Jo Vermeulen, Elisabeth André, and Sheelagh Carpendale. 2017. MyBrush: Brushing and linking with personal agency. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2017), 605–615.
[36] Bill Kules and Robert Capra. 2012. Influence of training and stage of search on gaze behavior in a library catalog faceted search interface. *Journal of the American Society for Information Science and Technology* 63, 1 (2012), 114–138. DOI: https://doi.org/10.1002/asi.21647
[37] Jelmer Kuurstra. 2015. *Individual Differences in Human-Computer Interaction: A Review of Empirical Studies*. University of Twente.
[38] Alexandra Kuznetsova, Per B. Brockhoff, and Rune Haubo Bojesen Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82, 13 (2017), 1–26.
[39] Sébastien Lallé, Cristina Conati, and Roger Azevedo. 2018. Prediction of student achievement goals and emotion valence during interaction with pedagogical agents. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems*. 1222–1231.
[40] Sébastien Lallé, Cristina Conati, and Giuseppe Carenini. 2017. Impact of individual differences on user experience with a visualization interface for public engagement. In *Proceedings of the 2nd International Workshop on Human Aspects in Adaptive and Personalized Interactive Environments*. ACM, New York, NY, 247–252. DOI: https://doi.org/10.1145/3099023.3099055
[41] Sebastien Lallé, Cristina Conati, and Dereck Toker. 2019. A gaze-based experimenter platform for designing and evaluating adaptive interventions in information visualizations. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research and Applications*. ACM, New York, NY, 60. DOI: https://doi.org/10.1145/3314111.3322502
[42] Sébastien Lallé, Dereck Toker, and Cristina Conati. 2019. Gaze-driven adaptive interventions for magazine-style narrative visualizations. *IEEE Transactions on Visualization and Computer Graphics* 27, 6 (2019), 2491–2952. DOI: https://doi.org/10.1109/TVCG.2019.2958540
[43] Shahid Latif, Kaidie Su, and Fabian Beck. *Authoring Combined Textual and Visual Descriptions of Graph Data*. Eurographics Association.
[44] Yu Li, Carla Allen, and Chi-Ren Shyu. 2019. Quantifying and understanding the differences in visual activities with contrast subsequences. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research and Applications*. ACM, New York, NY, Article 42.
[45] Adam V. Maltese, Joseph A. Harsh, and Dubravka Svetina. 2015. Data visualization literacy: Investigating data interpretation along the novice—Expert continuum. *Journal of College Science Teaching* 45, 1 (2015), 84–90.
[46] Oludamilare Matthews, Sukru Eraslan, Victoria Yaneva, Alan Davies, Yeliz Yesilada, Markel Vigo, and Simon Harper. 2019. Combining trending scan paths with arousal to model visual behaviour on the web: A case study of neurotypical people vs people with autism. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation, and Personalization*. 86–94.
[47] Alejandra Meneses, José-Pablo Escobar, and Soledad Véliz. 2018. The effects of multimodal texts on science reading comprehension in Chilean fifth-graders: Text scaffolding and comprehension skills. *International Journal of Science Education* 40, 18 (2018), 2226–2244. DOI: https://doi.org/10.1080/09500693.2018.1527472
[48] Ronald Metoyer, Qiyu Zhi, Bart Janczuk, and Walter Scheirer. 2018. Coupling story to visualization: Using textual analysis as a bridge between data and interpretation. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces*. ACM, New York, NY, 503–507. DOI: https://doi.org/10.1145/3172944.3173007
[49] Tamara Munzner. 2014. *Visualization Analysis and Design*. CRC Press, Boca Raton, FL.
[50] Unaizah Obaidellah, Tanja Blascheck, Drew T. Guarnera, and Jonathan Maletic. 2020. A fine-grained assessment on novice programmers' gaze patterns on pseudocode problems. In *ACM Symposium on Eye Tracking Research and Applications*. ACM, New York, NY, 1–5. DOI: https://doi.org/10.1145/3379156.3391982
[51] Heather L. O'Brien, Rebecca Dickinson, and Nicole Askin. 2017. A scoping review of individual differences in information seeking behavior and retrieval research between 2000 and 2015. *Library & Information Science Research* 39, 3 (2017), 244–254.
[52] Yasmina Okan, Mirta Galesic, and Rocio Garcia-Retamero. 2016. How people with low and high graph literacy process health graphs: Evidence from eye-tracking. *Journal of Behavioral Decision Making* 29, 2-3 (2016), 271–294. DOI: https://doi.org/10.1002/bdm.1891
[53] Kristien Ooms, Philippe De Maeyer, and Veerle Fack. 2014. Study of the attentive behavior of novice and expert map users using eye tracking. *Cartography and Geographic Information Science* 41, 1 (2014), 37–54. DOI: https://doi.org/10.1080/15230406.2013.860255

[54] Alvitta Ottley. 2016. *Toward Personalized Visualizations*. Tufts University.
[55] Erol Ozcelik, Ismahan Arslan-Ari, and Kursat Cagiltay. 2010. Why does signaling enhance multimedia learning? Evidence from eye movements. *Computers in Human Behavior* 26, 1 (2010), 110–117. DOI: https://doi.org/10.1016/j.chb.2009.09.001
[56] Matthew D. Plumlee and Colin Ware. 2006. Zooming versus multiple window interfaces: Cognitive costs of visual comparisons. *ACM Transactions on Computer-Human Interaction* 13, 2 (2006), 179–209. DOI: https://doi.org/10.1145/1165734.1165736
[57] Alex Poole and Linden J. Ball. 2006. Eye tracking in HCI and usability research. In *Encyclopedia of Human Computer Interaction*. IGI, 211–219.
[58] George E. Raptis, Christina Katsini, Marios Belk, Christos Fidas, George Samaras, and Nikolaos Avouris. 2017. Using eye gaze data and visual activities to infer human cognitive styles: Method and feasibility studies. In *Proceedings of the 25th Conference on User Modeling, Adaptation, and Personalization*. 164–173.
[59] K. Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124, 3 (1998), 372–422. DOI: https://doi.org/10.1037/0033-2909.124.3.372
[60] Jonathan C. Roberts. 2007. State of the art: Coordinated and multiple views in exploratory visualization. In *Proceedings of the 5th International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV'07)*. IEEE, Los Alamitos, CA, 61–71.
[61] Darrell S. Rudmann, George W. McConkie, and Xianjun Sam Zheng. 2003. Eyetracking in cognitive state detection for HCI. In *Proceedings of the 5th International Conference on Multimodal Interfaces*. ACM, New York, NY, 159–163.
[62] E. Segel and J. Heer. 2010. Narrative visualization: Relling stories with data. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 1139–1148. DOI: https://doi.org/10.1109/TVCG.2010.179
[63] Frank Serafini. 2012. Reading multimodal texts in the 21st century. *Research in the Schools* 19, 1 (2012), 26–32.
[64] Nelson Silva, Tobias Schreck, Eduardo Veas, Vedran Sabol, Eva Eggeling, and Dieter W. Fellner. 2018. Leveraging eye-gaze and time-series features to predict user interests and build a recommendation model for visual analysis. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research and Applications*. ACM, New York, NY, Article 13, 9 pages.
[65] Ben Steichen, Michael M. A. Wu, Dereck Toker, Cristina Conati, and Giuseppe Carenini. 2014. Te,Te,Hi,Hi: Eye gaze sequence analysis for informing user-adaptive information visualizations. In *Proceedings of the 22nd International Conference on User Modeling, Adaptation, and Personalization*. 183–194. DOI: https://doi.org/10.1007/978-3-319-08786-3_16
[66] Ben Steichen, Michael M. A. Wu, Dereck Toker, Cristina Conati, and Giuseppe Carenini. 2014. Te,Te,Hi,Hi: Eye gaze sequence analysis for informing user-adaptive information visualizations. In *Proceedings of the 22nd International Conference on User Modeling, Adaptation, and Personalization*. 183–194. DOI: https://doi.org/10.1007/978-3-319-08786-3_16
[67] Markus Steinberger, Manuela Waldner, Marc Streit, Alexander Lex, and Dieter Schmalstieg. 2011. Context-preserving visual links. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2249–2258. DOI: https://doi.org/10.1109/TVCG.2011.183
[68] Robert H. Tai, John F. Loehr, and Frederick J. Brigham. 2006. An exploration of the use of eye-gaze tracking to study problem-solving on standardized science assessments. *International Journal of Research & Method in Education* 29, 2 (2006), 185–208.
[69] Rong Tang and Yeseul Song. 2018. Cognitive styles and eye movement patterns: An empirical investigation into user interactions with interface elements and visualisation objects of a scientific information system. *Information Research* 23, 2 (2018).
[70] Dereck Toker and Cristina Conati. 2014. Eye tracking to understand user differences in visualization processing with highlighting interventions. In *Proceedings of the 22nd International Conference on User Modeling, Adaptation, and Personalization*. 219–230. DOI: https://doi.org/10.1007/978-3-319-08786-3_19
[71] Dereck Toker, Cristina Conati, and Giuseppe Carenini. 2019. Gaze analysis of user characteristics in magazine style narrative visualizations. *User Modeling and User-Adapted Interaction* 29 (2019), 977–1011.
[72] Dereck Toker, Cristina Conati, Giuseppe Carenini, and Mona Haraty. 2012. Towards adaptive information visualization: On the influence of user characteristics. In *Proceedings of the 20th International Conference on User Modeling, Adaptation, and Personalization*. 274–285. DOI: https://doi.org/10.1007/978-3-642-31454-4_23
[73] Dereck Toker, Cristina Conati, Ben Steichen, and Giuseppe Carenini. 2013. Individual user characteristics and information visualization: Connecting the dots through eye tracking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 295–304. DOI: https://doi.org/10.1145/2470654.2470696
[74] Tamara Van Gog. 2014. The signaling (or cueing) principle in multimedia learning. In *The Cambridge Handbook of Multimedia Learning* (2nd ed.). Cambridge University Press, New York, NY, 263–278. DOI: https://doi.org/10.1017/CBO9781139547369.014

[75] Julia M. West, Anne R. Haake, Evelyn P. Rozanski, and Keith S. Karn. 2006. eyePatterns: Software for identifying patterns and similarities across fixation sequences. In *Proceedings of the 2006 Symposium on Eye Tracking Research and Applications (ETRA'06).* ACM, New York, NY, 149–154. DOI: https://doi.org/10.1145/1117309.1117360

[76] Daricia Wilkinson, Moses Namara, Karla Badillo-Urquiola, Pamela J. Wisniewski, Bart P. Knijnenburg, Xinru Page, Eran Toch, and Jen Romano-Bergstrom. 2018. Moving beyond a "one-size fits all" exploring individual differences in privacy. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–8.

[77] Qiyu Zhi, Alvitta Ottley, and Ronald Metoyer. 2019. Linking and layout: Exploring the integration of text and visualization in storytelling. *Computer Graphics Forum* 38, 3 (2019), 675–685. DOI: https://doi.org/10.1111/cgf.13719

[78] Mary C. Dyson and Mark Haselgrove. 2001. The influence of reading speed and line length on the effectiveness of reading from screen. *International Journal of Human-Computer Studies* 54, 4 (2001), 585–612. https://doi.org/10.1006/ijhc.2001.0458

Received July 2020; revised December 2020; accepted January 2021