Gaze-Driven Adaptive Interventions for Magazine-Style Narrative Visualizations

Sébastien Lallé, Dereck Toker, and Cristina Conati

Abstract—In this article, we investigate the value of gaze-driven adaptive interventions to support the processing of textual documents with embedded visualizations, i.e., Magazine Style Narrative Visualizations (MSNVs). These interventions are provided dynamically by highlighting relevant data points in the visualization when the user reads related sentences in the MSNV text, as detected by an eye-tracker. We conducted a user study during which participants read a set of MSNVs with our interventions, and compared their performance and experience with participants who received no interventions. Our work extends previous findings by showing that dynamic, gaze-driven interventions can be delivered based on reading behaviors in MSNVs, a widespread form of documents that have never been considered for gaze-driven adaptation so far. Next, we found that the interventions significantly improved the performance of users with low levels of visualization literacy, i.e., those users who need help the most due to their lower ability to process and understand data visualizations. However, high literacy users were not impacted by the interventions, providing initial evidence that gaze-driven interventions can be further improved by personalizing them to the levels of visualization literacy of their users.

Index Terms—Narrative visualizations, gaze-driven adaptation, personalization, highlighting, eye-tracking, user characteristics

1 Introduction

Visualizations are typically designed following a one-size-fits-all approach, meaning that they do not take into account individual differences in their users. There is, however, mounting evidence that user characteristics such as cognitive abilities and personality traits, can significantly influence user experience during information visualization (InfoVis) tasks, even with well-designed, thoroughly evaluated visualizations, e.g., [1], [2], [3], [4]. These findings have prompted researchers to study *user-adaptive information visualizations*, i.e., visualizations that can recognize the specific needs and abilities of their users, and adapt various aspects of the visualization accordingly.

The first examples of adaptive visualizations leveraged user actions with an interactive visualization system to detect evidence that the user is not working well with the current visualization, and suggest a suitable alternative, e.g., [5], [6]. More recently, researchers have been investigating eye-tracking data as a source of information to predict user needs and drive adaptation. Eye-tracking is especially suitable for delivering adaptation in visualization because it can directly capture visual processes that are fundamental for working with a visualization. Eye-tracking data has also the advantage of being available in both interactive and non-interactive visualizations.

Existing research has shown that eye-tracking data can be used to predict in real-time several long-term user characteristics and short-term states known to influence visualization effectiveness, such as users' perceptual abilities [7], [8], [9], interest [10], confusion [11] and cognitive load [12]. Whereas these predictions of user long-term characteristics and short-term states have yet to be used for adaptation, there is initial evidence on the effectiveness of adaptation simply based on detecting specific user gaze behaviors, i.e., *gaze-driven adaptation*. Specifically, Göbel *et al.* [13] and Bektas *et al.* [14] have shown that gaze-driven adaptation can facilitate the processing of maps.

We contribute to this body of research by investigating gaze-driven adaptation as a means to support the processing of visualizations embedded in narrative text, known as *Magazine-Style Narrative Visualization* (*MSNV* for short) [15]. We focus on MSNVs featuring bar charts, one of the most ubiquitous visualizations found in MSNV documents such as newspapers, scientific articles, blogs, textbooks [16]. We also investigate the potential value of long-term user characteristics to further personalize the delivery of gaze-driven adaptation in MSNVs.

As it is often the case for multimodal documents, processing MSNVs can be challenging due to the need to split attention between two information sources, with a possible increase in cognitive load and a negative impact on comprehension [17]. This challenge can be exacerbated in MSNVs as there are often multiple sentences in the text, called *references*, that solicit attention to different aspects of the accompanying visualization (see example in Fig. 1). In particular, identifying which data points in the visualizations correspond to each reference is a well-known difficulty in MSNVs [18], [20], [21], [22], [23].

Based on these results, in this paper we design and evaluate a form of cuing (i.e., adding visual prompts that guide user attention) that aims to facilitate MSNV processing by highlighting the relevant data points in the visualization

• *The authors are with the Department of Computer Science, University of British Columbia, 2366 Main Mall, Vancouver, BC V6T 1Z4, Canada. E-mail: {lalles, dtoker, conati}@cs.ubc.ca.*

Manuscript received 29 Apr. 2019; revised 17 Nov. 2019; accepted 23 Nov. 2019. Date of publication 9 Dec. 2019; date of current version 29 Apr. 2021. (Corresponding author: Sébastien Lallé.) Recommended for acceptance by B. Lee. Digital Object Identifier no. 10.1109/TVCG.2019.2958540

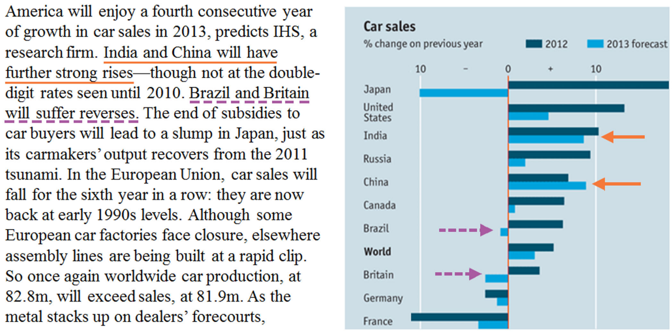

Fig. 1. An example of MSNV document with multiple references, with the first two underlined by us for easier identification. Arrows identify the different data points the underlined sentences refer to in the MSNV bar graph. Source: The Economist - Dec 22, 2012.

when a user reads a reference in the text, as detected by an eye-tracking device. We also conducted an exploratory analysis to understand whether the effectiveness of this gaze-driven adaptation depends on user characteristics previously shown by Toker *et al.* [21] to exacerbate difficulties in processing bar-chart-based MSNV. We do so to ascertain if the gaze-driven adaptation should be further personalized to any of these user characteristics. Thus, the research questions we examine in this paper are as follows:

Does gaze-driven highlighting of relevant parts of an MSNV graph improve user performance and subjective experience with bar-chart-based MSNVs, compared to users who received no such highlighting? Do the results depend on user characteristics that were previously found to influence MSNV processing?

To answer these questions, we conducted a user study during which participants read a set of 14 MSNV with our proposed gaze-driven highlighting interventions, and we compared their outcomes with those of users who underwent the same task without interventions in the study described in Toker *et al.* [21].

Our results show that the proposed highlighting interventions specifically helped users with low levels of *visualization literacy (vis literacy* for short*)*, i.e., users with a lower ability to process and understand data visualizations [22]. This finding is significant because it shows that the adaptive interventions benefited those users who needed help the most. Our results also show that users with high vis literacy were not impacted by the interventions, but their performance still had room for improvement. This suggests to further examine adaptive interventions personalized to be helpful for high vis literacy users.

Our work contributes to existing research in two ways. First, it broadens the evidence on the value of eye-tracking for adaptive visualizations. Prior to our work, gaze-driven adaptation has only been used to support the processing of map-based visualizations [13], [14]. Our results extend these findings both by looking at a different visualization type (bar charts) and by focusing on MSNV, a widespread context of usage for visualizations that has not yet been investigated for gaze-driven adaptation.

A second contribution is that our results are the first to show that *long-term user characteristics* can influence the effectiveness of gaze-driven adaptation in visualization, calling for further research on how these interventions can be personalized to these long-term characteristics.

2 Related Work

2.1 Need for Adaptation in InfoVis

Short-term states such as *cognitive overload* [23] and *confusion* [24] can directly reveal when the user is struggling while performing a task with a visualization, signalling that real-time support addressing the specific user difficulties would be beneficial.

There is extensive evidence that *long-term user characteristics* (e.g., cognitive abilities, personality traits) significantly influence visualization processing, in a way that warrants providing support adapted to these user differences. For instance, low levels of the cognitive abilities *perceptual speed* and *visual working memory* (WM) have been linked to lower performance in simple analytic tasks with bar-chart-based visualizations [1], [25]. Low levels of *spatial abilities* have been linked to worse performance in map reading tasks [26] and probabilistic reasoning tasks performed with icon array visualizations [27]. Low *vis literacy* was found to hinder user experience during decision-making tasks supported by maps and deviation charts [4], as well as during processing network visualizations in science museums [28]. Users with low levels of *need for cognition*, a personality trait, obtained worse performance than their counterparts in low-level analytical tasks with colored boxes [29].

Several long-term user characteristics have also been linked to user performance during static (i.e., non-adaptive) MSNV processing. Toker *et al.* [21] tested the influence of nine user characteristics on user performance when completing the task of reading and answering comprehension questions about static bar-chart-based MSNVs extracted from real-world sources. Results showed that users with low levels of *reading proficiency* and *verbal WM* were significantly slower in task completion than users with high levels of these abilities. Users with low levels of *vis literacy*, *need for cognition* and *verbal IQ* were significantly less accurate on comprehension questions than their counterparts. Off-line analysis of users' eye-tracking data showed that these worse performances were due in part to difficulties in identifying referenced data points in the MSNV visualizations. In this paper, we leverage the same set of static MSNVs used in Toker *et al.* [21], but embed them in a platform that delivers gaze-driven interventions that dynamically highlight referenced data points. We evaluate their effectiveness and if/how it is modulated by the five user characteristics that were found to impact performance in Toker *et al.* [21], with results linking the intervention effectiveness to the user's levels of vis literacy.

2.2 Eye-Tracking for User Adaptation

Existing research has investigated the potential of eye-tracking to support adaptation by detecting in real-time relevant user gaze patterns, *short-term* states, and *long-term* user characteristics.

Gaze-driven adaptation, which reacts to specific user gaze patterns, has been investigated in several domains. For instance, gaze-driven prompts were used to refocus student attention back to the screen when they look away while interacting with educational software, with positive results on student learning [30]. Shirazi *et al.* [31] adapted online advertisements based on what information users look at in e-commerce webpages. In InfoVis, gaze-driven adaptation has been studied to support map reading [13], [14]. Göbel *et al.* [13] dynamically placed the legend of a map next to where the user was looking and highlighted in the legend the symbols that lied in the area of gaze location, with positive results for processing time and user satisfaction. Bektas *et al.* [14] deemphasized the parts of a map that were outside the user's focus of attention, but this adaptation did not benefit users in simple visual search tasks.

Research has shown that eye-tracking can reveal more about the users than where they look. In particular, eye-tracking data has been used to predict user short-term states such as boredom [32] and mind wandering [33] in educational settings, as well as interest [10], confusion [11] and cognitive load [12] during visualization tasks. In the context of visualization processing, there has been results on leveraging eye-tracking to predict long-term cognitive abilities relevant for adaptation, such as perceptual speed, verbal WM, visual WM, visual scanning [7], [8], [9]. These results showed that eye-tracking can reveal rich information about users. However, thus far only the work by D'Mello *et al.* [33] on predicting mind wandering while studying educational text has been used to drive adaptation, in this case prompting the user to refocus their attention.

2.3 Cuing for InfoVis and Multimodal Documents

There has been a recent interest in studying cuing, i.e., adding visual prompts that guide user attention, to support visual analytics tasks, see Collins *et al.* [34] for an overview. So far these cues have been typically predefined and provided upfront to everyone, e.g., [35], [36], [37], which may not suit the needs of all users. Other research showed that cuing can facilitate the processing of multimodal instructional material consisting of text and diagrams or pictures (but not visualizations), see Van Gog [19] for an overview. In particular, color coding to match parts of the text and graphics was found to increase comprehension [38], [39]. This color coding was provided either upfront [38] or at the user's request when clicking on a specific paragraph [39].

Other work [18], [40], [41], [42], [43], [44] investigated cuing for supporting the processing of MSNV documents. Specifically, Steinberger *et al.* [18] studied colored lines that were drawn upfront over the document to link words in the text to the corresponding information in the visualization. The cues were evaluated in a task requiring users to seek specific information (targets) in MSNVs. The targets were predefined so that the linking could be provided upfront, and the number of targets was limited to avoid clutter. Results showed that this form of cuing reduced search time. However, providing all cues upfront is hard to scale to MSNVs with a large number of references, as it is often the case in real-world documents (e.g., Pew Research documents on public policy can include up to 30 references [20]), because many cues can visually clutter the document and create overlaps, thus diminishing the effectiveness and readability of the visualization. Zhi *et al.* [40] offered visual cues at the user's initiative, by highlighting relevant data points in the visualization when selecting a reference in the text, and vice versa. Results showed that the highlighting improved user-perceived satisfaction and amount of interaction with an MSNV, but not their performance on comprehension and recall tasks compared to not receiving the highlighting. Metoyer *et al.* [43] proposed a similar approach to Zhi *et al.* [40] for sports narratives with visualizations, albeit with no evaluation. Latif *et al.* [42] proposed an approach to ease the implementation of visual cues in MSNVs triggered by hovering the mouse over references, whereas we show how to design and implement cues triggered by gaze. Adar *et al.* [44] proposed a system to personalize visualizations and texts based on a user's location and mouse clicks, with no end-user evaluation. Badam *et al.* [41] investigated interactive MSNVs in which specified visualizations were displayed at the user's demand when the user clicked on a reference in the text. They found that this approach improved user performance compared to documents with no visualization at all. We extend [18], [40], [41], [42], [43], [44] in two ways: first, we use eye-tracking to trigger the interventions, thus offering an effortless way to trigger the visual cues. Second, we investigate the influence of user characteristics on the effectiveness of the interventions, as it is possible that the tested visual cues do not suit all users depending on their needs and preferences.

The form of gaze-driven cuing we adopt in this paper was originally proposed at a workshop [45], albeit without implementation nor evaluation, and is further inspired by a study that compared different ways of highlighting relevant bars in stand-alone bar charts, i.e., not embedded in MSNVs [46]. We rely on this work because it explicitly investigated cuing displayed during the task as opposed to provided upfront, as done in [18], [35], [36], [37], or upon specific user requests, as done in [40], [41]. The study tested various forms of highlighting interventions, including thickening the border of the relevant bars and deemphasizing non-relevant bars. These interventions were tested in a series of fictitious, low-level analytical tasks. For each task, the user would see a bar graph and a question asking to retrieve and compare specific data points, with the bars corresponding to these data points being highlighted all at once after a predefined time delay. This delay was added to mimic the effects of receiving cuing dynamically during bat chart processing and ascertain if it could be annoying or distracting. Results showed that users performed significantly better when receiving the interventions, compared to receiving none, providing preliminary evidence on the potential of cuing for bar chart processing. We extend this work by studying a more realistic and challenging usage context of bar charts, namely MSNVs, and by using eye-tracking to trigger the interventions in a timely manner based on reading behaviors. We also link interventions effectiveness to the levels of visualization literacy of the user, thus highlighting the need for personalization in MSNVs.

3 Adaptive MSNV

In this section, we first describe the MSNVs that we leverage to evaluate our proposed gaze-driven highlighting interventions. Then, we describe the design of these interventions and their implementation for the target MSNVs.

3.1 MSNV Dataset

We used a set of 14 static bar-chart-based MSNVs that Toker *et al.* [21] derived from an existing dataset of MSNVs extracted from real-world sources, e.g., Pew Research, The Guardian, The Economist [20]. To keep the study complexity

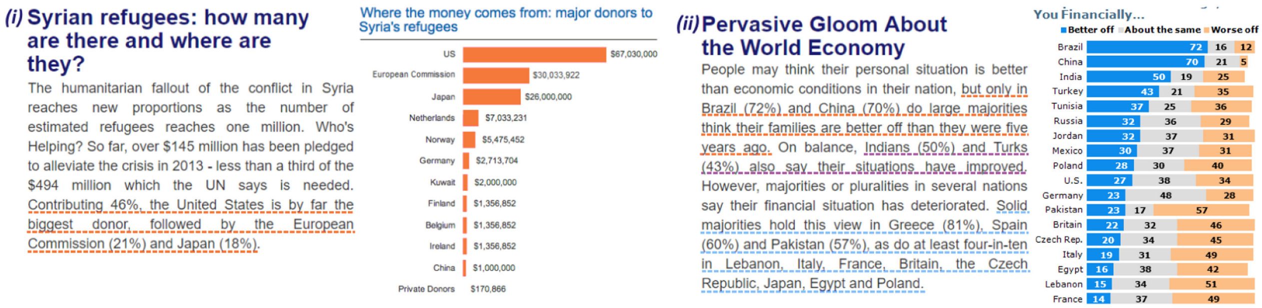

Fig. 2. Two MSNVs with different levels of complexity: *(i)* the one on the left with one reference (dashed underlines, added to this figure for clarity) and a simple bar chart, *(ii)* one on the right with three references and a stacked bar charts showing many more data points.

manageable, these MSNVs feature only bar charts, one of the most commonly used visualizations in real-world documents [16]. The references in this dataset had been previously identified via a rigorous coding process, indicating which data points in each visualization correspond to each sentence(s), as detailed in Kong *et al.* [20]. Each MSNV in this dataset consisted of "snippets" of larger source documents whereby each snippet included a self-contained excerpt from the original text and one accompanying bar chart. We use this format in order 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 bar chart types (4 simple, 6 stacked, 4 grouped), length (measured in terms of the number of words and references) and the number of referenced data points. Fig. 2 shows two MSNVs with different complexity, whereas Table 1 shows summary statistics on the composition of the 14 MSNVs. The selection process is fully described in Toker *et al.* [21].

3.2 Design of the Gaze-Driven Highlighting

Our proposed gaze-driven interventions dynamically highlight those bars in an MSNV chart corresponding to the reference the user is reading, as detected by eye-tracking.[1] The rationale for this form of guidance is to drive users' attention to the appropriate data in the charts when it is most relevant, i.e., when the user is attending to that piece of information in the text.

Designing these highlighting interventions entails several challenges related to determining the main properties of the interventions, namely: *(i)* what type of highlighting to use; *(ii)* when exactly to trigger the intervention during the reading of a reference; *(iii)* whether interventions should be incrementally added to the bar chart as references are read, or whether previous interventions should be removed so that only one is active at any given time.

Testing values for all these properties in a formal study is not feasible as it would generate too many study conditions. Instead, we conducted dedicated pilot studies to identify a suitable value for each of these three properties, based on the feedback provided by the pilot participants. In these pilots, the participants read the same set of 14 MSNVs and completed the same task as in the main study (described in Section 4.2), with various versions of the adaptive interventions based on the property values we wanted to test. Participants were then interviewed to elicit their preferences and feedback.

Highlighting Type. For this property, we chose to pilot test the two types of highlighting found to be most effective at supporting bar chart processing [46] (see related work): thickening the border of the relevant bars, and desaturating non-relevant bars (see Fig. 3 for examples).

Five out of 6 pilot users reported that desaturating bars was too disruptive because it removes context by making it difficult to see the desaturated bars even if they wanted to. Based on this feedback we retained *thickening* of the border of the bars for the study. The borders are always thickened in black so as to use a neutral color that has no other visual encoding in the bar charts of the MSNVs, as done in Carenini *et al.* [46]. To ensure that the black outlines are noticeable for all the bars in the dataset, we adjusted the brightness and saturation of all bar colors so that their contrast ratio with the black outline is consistent across MSNVs. We ensured that the colors remain well visible and distinguishable.

Intervention Timing. Because we want to trigger interventions when a user is reading references in the text, we use an eye-tracker to track the user's fixations (gaze maintained at one point of the screen) over these references. An intervention for a reference is then triggered whenever a *sufficient* number of fixations on the related sentences have been detected. An open question, however, is when exactly to trigger the intervention during the reading of a reference, e.g., at the start of the sentence, when the sentence has been fully read, somewhere in between.

For this study, we chose to test the option of triggering the interventions when the user has read more than half of the sentence, so that they have sufficient context to

TABLE 1
Summary Statistics for the Properties of the MSNVs

MSNV Property	Min	Max	Mdn	Mean	SD
Total number of words in narrative text	43	228	75	90.8	49.7
Total number of sentences in narrative text	2	14	4	4.9	3.0
Total number of references in narrative text	1	7	3	2.8	1.8
Total number of data points in visualization	4	63	14	22.1	19.7
Total number of referenced data points	2	24	6	10.1	7.8

1. A video of the interventions is available in the documentation folder at *http://github.com/ATUAV/ATUAV_Experimenter_Platform.*

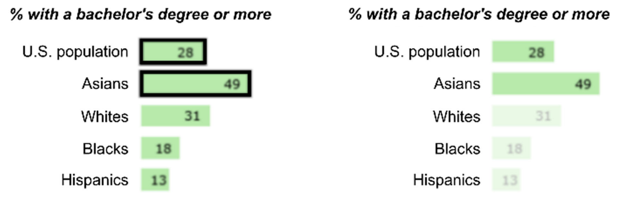

Fig. 3. Sample highlighting of the top two bars, via thickening of their borders (left), and desaturating the other bars (right).

process the relevant bars in the chart. The opposite approach is to trigger the intervention as soon as the user starts reading the sentence, but we deemed this to be potentially too distracting, and prone to error as the interventions might be triggered by a few spurious or slightly inaccurate fixations. The challenge here is that the number of fixations needed to process a reference can depend on its length, difficulty or phrasing, as well as on the reading speed of each user. We leveraged the eye-tracking data collected in Toker *et al.* [21], which used the same dataset as we use here, to compute the average number of fixations users spent on each reference in our target MSNVs. This average ranges from 8 for the shortest reference to 45 for the longest one (overall $\mathrm{mean} = 24, \mathrm{st.dev.} = 10$).

Next, we needed to define which percentage of this average number of fixations per reference should be considered as sufficient for having processed the reference and for triggering the intervention. A high percentage is risky because it is prone to triggering interventions too late, such as when the user has finished reading the reference and has moved to subsequent text. This is especially true for fast readers or readers who skim through the text, as they would generate fewer than average fixations on a sentence.

We chose to first pilot a trigger threshold of 60 percent, i.e., for each reference, the corresponding intervention is triggered when 60 percent of the average fixations required to read it are detected. Five out of 6 pilot users reported that the interventions seemed to appear too late (e.g., "*It feels like there is a delay*"). We also noticed that 3 of these users each triggered less than 70 percent of the interventions, either because they were reading fast and thus did not generate enough fixations over the reference, or because our 60 percent threshold was generally too conservative. Based on this feedback we lowered the trigger threshold to 40 percent, to better ensure that the interventions would be triggered, even by faster readers. Although for some users this threshold might trigger interventions when they are just partway through reading a sentence, for our purposes, it is important to make sure that as many interventions as possible are delivered, given the objective to investigate the effectiveness of these interventions compared to not receiving them. We tested this 40 percent threshold with two additional pilot users, who triggered most interventions and found their timing to be suitable, and thus we retained this threshold for the study.

Intervention Removal. Because most MSNVs contain multiple references, we had to determine whether previously triggered interventions should be left active or removed upon the delivery of a new one. Leaving all interventions active is shown in Fig. 4A, in which four references (underlined) were read, of which the current one is at the bottom, and all the corresponding bars are all highlighted. This approach facilitates going back to the previous references, which can be useful if the user forgot some information, or wants to compare data points across references. A possible drawback is that having too many highlighted bars might become overwhelming, and might make it difficult to discern the bars related to the current reference read (e.g., the 'Catholics' bars in Fig. 4 A, corresponding to the reference at the end of the text), especially in documents with many references.

An alternative strategy is to remove all previous interventions, as shown in Fig. 4B, where only the 'Catholics' bars (i.e., the latest intervention) are left highlighted. With this approach, the user can easily focus on the most current intervention but cannot refer to previous ones anymore.

Pilot testing both strategies with 6 users revealed mixed feelings, with no clear winner. Four users reported that removing the previous highlighting was unhelpful because they could not remember what bars they already processed. Two of them said they often had to go back and re-read the text due to that. Three pilot users liked having all the interventions after reading the entire text because it provides "*a good visual summary of the salient information described in the MSNV*". However, five out of 6 users reported difficulties in distinguishing between the previously highlighted bars and the recent ones with this strategy.

To account for this feedback we implemented a third strategy designed to leverage the pros of the previous two without the drawbacks. This strategy involves keeping previous highlights but desaturating the thickening so that the black outline becomes gray (see Fig. 4C), thus distinguishing the most recent highlighting from the previous ones. We pilot tested this strategy with two additional users, who provided very positive feedback. Thus, we retained it for the main study.

3.3 Platform to Test Adaptive MSNV

The gaze-driven interventions evaluated in this paper are generated by a platform (briefly described in this section) we have developed to deliver and evaluate interventions in user-adaptive visualizations. The platform,[2] described in Lallé *et al.* [47], consists of the following components.

A *back end* component processes eye-tracking data in real-time by establishing a connection with the eye-tracker, fetching the raw data at the eye-tracker's sampling rate, and processing it to extract fixations and other higher-level information that can be used to trigger adaptation. Fixations are detected by a real-time implementation of the popular ID-T algorithm [48], which identifies fixations as groups of consecutive gaze samples with low dispersion and sufficient duration. We set the maximum dispersion to 35 pixels and the minimum duration to 100ms, as we found that these values approximate the fixations detected offline by Tobii's proprietary algorithm on the gaze data collected in Toker *et al.* [21]. Fixations are captured both over the entire screen and within pre-specified Areas of Interest (AOIs). For this study, the AOIs are the individual reference sentences in each MSNV in our dataset, and the platform tracks user fixations over these AOIs.[3]

2. Link to code and documentation: *http://github.com/ATUAV/ATUAV_Experimenter_Platform/tree/master/*

3. We used a JS script (included in the platform on Github) to automatically extract the AOIs coordinates from the documents.

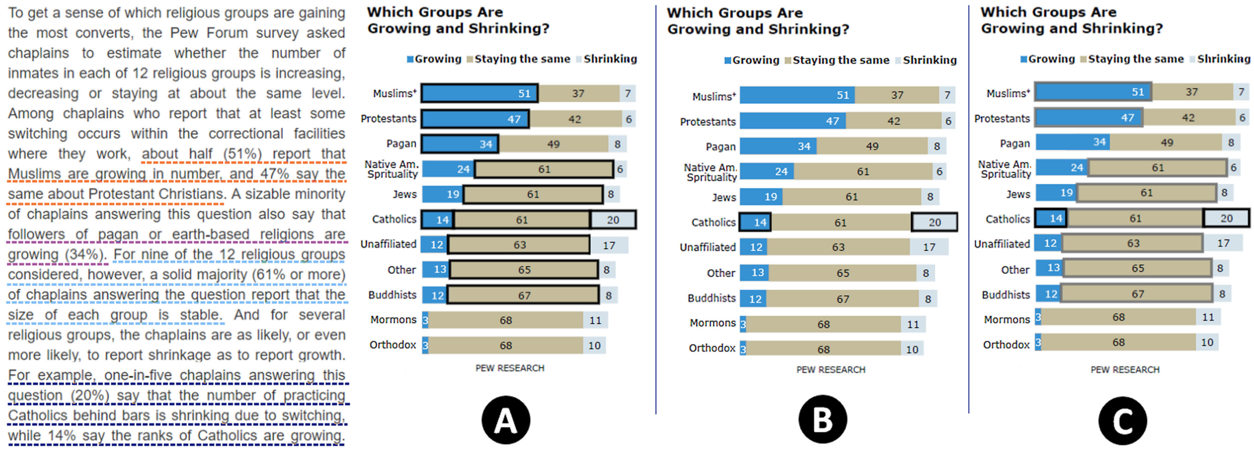

Fig. 4. Example removal strategies. The 4 references underlined in the text have been read, with the one at the bottom being the current one active. (A) keep all highlighting; (B) remove previous highlighting; (C) desaturate previous highlighting in grey.

A *middle end* component manages the delivery of adaptation by evaluating a set of *adaptation rules* over the live-stream of eye-tracking features generated by the back end. For this MSNV study, there is one general rule that initiates an intervention when the number of detected fixations over a given reference exceeds its triggering threshold (see Section 3.2). The middle end also provides functionalities to evaluate the adaptation via user studies, e.g., data logging, task randomization, and the display of questionnaires and tests as needed.

A *client side* component displays the target visualization (in our case the MSNVs) and the adaptive interventions upon notification from the middle end. Interventions are displayed in the visualizations via JavaScript callbacks, using the D3 and Angular JavaScript libraries.

4 User Study

To evaluate the gaze-driven adaptive highlighting for MSNVs described in the previous section, we used a *between-subject design*, where we compare the performance of a group of users who read MSNV with the highlighting interventions (*adaptive group)* and a *control group* that reads the same MSNV with no highlighting. The data for the control group comes from the study reported in Toker *et al.* [21], referred to as *control study* from now on. The data for the adaptive group comes from the study we describe in the rest of this section (*adaptive study* from now on). The two studies use the exact same task and procedure, fully described in [21] and summarized in Section 4.2.

4.1 Participants

A total of 119 subjects were recruited in the studies via advertising at our campus and on Craigslist and were paid \$35 for participating. The control study [21] included 56 subjects (32 female), with age from 19 to 69 (M = 28, SD = 11). For the adaptive study, we recruited 63 participants (34 female), with age from 18 to 59 (M = 25, SD = 8). In both studies, about 60 percent of the participants were university students, and the others were from a variety of backgrounds (e.g., retail manager, restaurant server, artist, nurse, retired).

4.2 Study Procedure

The study procedure involves a single session lasting at most 90 minutes. The session starts with the participant undergoing calibration with the eye-tracker, a Tobii T-120 camera-based remote eye-tracker embedded in a display of 1280 x 1024 pixels, with a sampling rate of 120 Hertz. Next, participants are given the task of reading an MSNV document on the computer screen, and signal when they are done by clicking 'next'. At this point, they see a screen with a set of questions that elicit their opinion of the document and test their comprehension (see Section 4.4). Participants perform this same task for the 14 MSNVs described in Section 3.1, taking on average 20 minutes. The ordering of the 14 MSNVs is randomized for each participant. Participants are not given a time limit to read the MSNVs, nor training on the interventions, to mimic how they might be used in realistic settings. To ensure that participants dedicated adequate effort to the task, a \$50 bonus was promised for the three participants with the best performance, evaluated in terms of both speed and accuracy. In the adaptive study, participants also filled out a postquestionnaire to rate the usefulness and ease of use of the adaptation (Section 4.4). They were also briefly interviewed to discuss their ratings.

Standard tests from psychology were used to assess, for each participant, a battery of user characteristics (described in the next section) that might influence how MSNVs are processed and how participants react to adaptive interventions. To reduce fatigue, some of these tests, which are computer-based and do not require an invigilator, were given to participants to do at home prior to the experiment. The rest, which was either paper-based or required specialized software not available remotely, were administered during the study session. See Toker *et al.* [21] for more details on the test administration procedure, which was kept identical in the adaptive study.

4.3 User Characteristics

Nine user characteristics were measured in both the control and the adaptive study, to keep the same study procedure. As we discussed in the related work, only five out of these nine were found to influence user performance when processing MSNVs in Toker *et al.* [21], i.e., low levels of these

TABLE 2
Set of User Characteristics and Tests, and Summary Statistics for the Scores of the Participants

User Char.	Definition	Instrument	Score range	Mean score (SD)
Vis Literacy	Ability to use a visualization to translate questions specified in the data domain into visual queries in the visual domain, and to interpret visual patterns in the visual domain as properties in the data domain [24].	*Visualization Literacy 101 – Bar Chart Test* [24]	*−2 ; 2*	*.35 (±.6)*
Verbal WM	Quantity of verbal information (e.g., words) that can be temporarily maintained and manipulated in working memory [43].	*OSPAN (Operation-word span)* [44]	*0 ; 6*	*4.9 (±.9)*
Reading Proficiency	Vocabulary size and reading comprehension ability in English [45].	*X_Lex Vocabulary Test* [45]	*0 ; 100*	*85 (±10)*
Verbal IQ	Overall verbal intellectual abilities that measures acquired knowledge, verbal reasoning, and attention to verbal materials [46].	*North American Adult Reading Test* [46]	*75 ; 125*	*102 (±9)*
Need for Cognition	Extent to which individuals are inclined towards effortful cognitive activities [47].	*Need for Cognition Scale* [47]	*−36 ; 36*	*6.5 (±10)*

characteristics were linked to worse performance. Because of their link to performance, we focus on these five for our analysis of how user characteristics might influence the effectiveness of adaptive interventions for MSNV.

These five user characteristics are defined in Table 2, along with the standard tests from psychology we used to measure them, and the mean scores collected over participants. The first characteristic (vis literacy) relates to how well users can process visualization; the next three (verbal WM, reading proficiency, and verbal IQ) relate to the ability to process textual elements; lastly, need for cognition is a personality trait defining how much users like effortful cognitive activities. A Kendall rank correlation test shows that these five measures are not correlated ($\tau < .30$).

4.4 Dependent Measures

Two sets of dependent measures used in the adaptive study were the same as those collected in Toker *et al.* [21] and relate to *task performance* and *user experience*. In addition, in the adaptive study we collected measures related to the user *perception* of the interventions.

TABLE 3
Questionnaire on the Perception of the Interventions

Component	Statements
Usefulness	1. I found the interventions useful. 2. I found the interventions useful to understand the document. 3. I found the interventions useful to focus on the relevant information. 4. I found the interventions distracted. 5. I found the interventions confusing.
Ease of use	6. I found that the timing of the intervention was right. 7. I found the interventions easy to notice. 8. I found the intervention well-integrated into the document.
Satisfaction	9. I was satisfied with the intervention. 10. I would use the interventions in my daily life.

i) *Task performance.* This set comprises of *time on task* and *accuracy* in processing each MSNV. Time on task is the time elapsed between when a participant is shown an MSNV, and when they signal that they are done reading it by hitting "next". Accuracy is the ratio of correct answers to the comprehension questions that appear after pressing "next". These were adapted from Dyson *et al.* [49] and include two *recognition questions* asking to recall specific information from the MSNV, and a *title question* asking to select a suitable alternative title for the MSNV as a way to test overall comprehension.

ii) *User experience.* This set comprises two subjective measures about the perceived *ease-of-understanding* and *interest* of the document, assessed on a 5-point Likert scale.

iii) *Perception of the interventions.* This set comprises 10 measures related to user-perceived *usefulness*, *ease-of-use*, and *satisfaction* with the interventions. These measures were collected via a web questionnaire displayed after a participant read all 14 MSNVs, in which the participant rated the 10 statements listed in Table 3 on a 7-point Likert scale ranging from "Strongly disagree" to "Strongly agree". Statements 1-5 ask about the perceived usefulness of the interventions, including possible negative aspects, namely if they were distracting or confusing. Statements 6-8 ask about aspects related to the ease-of-use of the interventions. The last statements (9-10) gauge satisfaction with the interventions. This questionnaire was derived from the Usefulness, Satisfaction, and Ease of use (USE) questionnaire [50], a popular instrument to evaluate the usability of user interface features.

5 Analysis of the Triggering Mechanism

To make sure that the intervention triggering mechanism discussed in Section 3.2 generated enough interventions to answer our research question, we examined the percentage of interventions each participant triggered, given the maximum number of 35 available. Fig. 5 reports these percentages. It should be noted that we discarded 5 users because

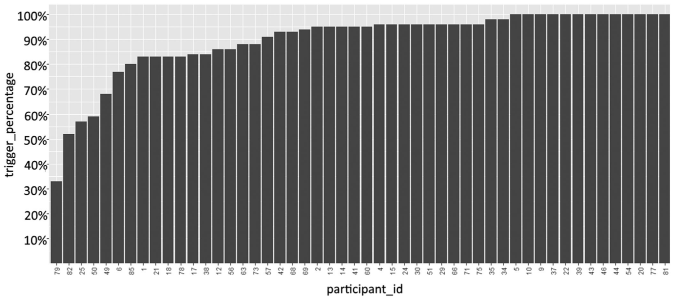

Fig. 5. Percentage of triggered interventions per participants.

of too many invalid gaze samples (as reported by the eye-tracker), leaving 58 users for this and subsequent analyses.

On average, these 58 participants triggered 81 percent of the interventions (SD = 19%), and Fig. 5 shows that a large majority triggered most of the interventions. Specifically, 46 users (about 79 percent) triggered more than 75 percent of the interventions. The fact that not all interventions were triggered is to be expected, as it is a normal reading behavior to skim or even skip some sentences at times, for example when a text is not of interest to the participant.

For the 13 participants who triggered less than 75 percent of the interventions, we investigated whether this was due to problems with eye-tracking, or to them being fast/skim readers. To do so, we looked at the attention map for each of these users, as the attention maps can reveal tracking issues if they are misaligned with the sentences in the text and the bar chart. We also checked the proportion of gaze samples marked as invalid by the Tobii eye-tracker.

The attention map of 8 of these 13 participants showed a misalignment (see example on Fig. 6) and a high rate of invalid gaze samples. This technical issue likely interfered with the intervention delivery mechanism, because accurate tracking of which reference sentence the user is reading is necessary to trigger the right intervention at the right time (see Section 3.3). Therefore, we discarded these users from further analysis.

There was no obvious issue with the attention maps and invalid gaze samples for the 5 remaining participants who triggered less than 75 percent of interventions, and their low trigger percentage is likely due to a high reading speed or a tendency to skim the text. Thus, we retained these users, remaining with 50 participants who, on average, triggered 89 percent of the interventions ($\mathrm{st.\,dev.} = 14\%$). We consider the proportion of interventions triggered by these 50 users to be adequate to proceed with our analysis. However, because there is some variance in their trigger percentage, we will discuss if/how it affects outcomes for the adaptive group in the results section.

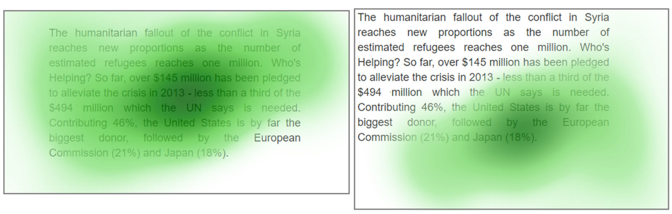

Fig. 6. Attention map aligned with the text (left) vs misaligned (right).

TABLE 4
Summary Statistics of Dependent Measures

Measure	Control	Adaptive
ACCURACY (%)	71.9 (±30)	74.4 (±31)
TIME-ON-TASK (SECS)	56.3 (±32)	60.1 (±33)
INTEREST	3.37 (±1)	3.31 (±1)
EASE-OF-UNDERSTANDING	4.00 (±1.2)	4.05 (±1.2)

6 Analysis and Results

Section 6.1 presents the analysis and results related to our research questions, namely whether the gaze-driven interventions improve user performance and experience with MSNVs, depending on the tested user characteristics. The next sections present further results within the adaptive group, namely on how participants perceived the interventions (Section 6.2) and were influenced by the number of interventions received (Section 6.3).

6.1 Comparison of Control and Adaptive Groups

We compare the performance and experience of the participants in the control and adaptive groups while accounting for the possible influence of the 5 user characteristics presented in Section 4.3 (UC from now on). Recall that we measure performance in terms of *accuracy* and *time on task*, and user experience in terms of perceived *ease-of-understanding* and *interest* of the MSNV (see Section 4.4). Statistics for these dependent measures are shown in Table 4.

We ran four mixed models, one per dependent measure, with group (adaptive, control) and the five UC as the independent variables. Participant ID and MSNV ID were added as random effects in the mixed models, to account for variability across the participants and the documents, respectively. Each mixed model was fitted with a bidirectional stepwise algorithm for model selection based on AIC, using the *lmerTest* package in R. To account for family-wise error, resulting *p*-values from all four mixed models are adjusted using the Benjamin–Hochberg procedure to control for the false discovery rate (FDR) [51]. We report statistical significance at the .05 level, as well as effect sizes as high for $r > .5$, medium for $r > .3$, and low otherwise.

Results. Because we are interested in the impact of having or not having adaptive interventions, we focus on results pertaining to effects involving groups. We found no significant main effect of group on the dependent variables, indicating that these interventions are not helpful for all users, as it is also suggested by the high variance in performance among participants (see st. dev. in Table 4).

There is a significant interaction effect of *vis literacy* with *group* on *accuracy* ($F_{1,97} = 12.71$, $p = .0006$, $r = 0.34$), indicating that the effectiveness of the interventions depends on the user's levels of this ability. To interpret this interaction effect, we divide all participants into three bins that include an equal number of participants, based on a 3-way split of their vis literacy scores (Low, Medium, High). As a result, the High bin includes participants with a *vis literacy* score of 1 and above, the Low bin participants with a score of -0.43 and under, and the Medium bin everyone else. This approach is

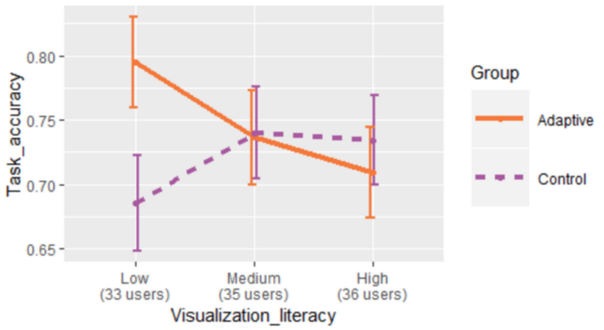

Fig. 7. Interaction effect of vis literacy with group on accuracy. Error bars show 95 percent confidence intervals.

used because the distribution of the vis literacy scores is asymmetrical with not established breakpoints to form the groups.[4] Fig. 7 shows the interaction effect between *vis literacy* with *group* on *accuracy* using this three-way split. Post-hoc FDR adjusted pairwise comparisons show two significant effects in this interaction:

i) Users with *low levels of vis literacy* were *more accurate* in the adaptive group than in the control group ($p = .0009$, $r = .33$), with a substantial boost in accuracy of nearly 11 percent on average (see Fig. 7, left). It should be noted that the mixed models revealed no significant interaction effect of vis literacy and group on time on task ($p = .19$, $r = .15$), indicating that the increased accuracy of the low vis literacy users with the adaptive interventions did not come at the expense of longer time on task.
ii) In the adaptive condition, *low vis literacy* users were significantly ($p = .006$, $r = .28$) *more accurate* than *high vis literacy* users (see the plain orange line in Fig. 7), with an improvement in accuracy of about 9 percent. As for these high vis literacy users, there is no significant difference in their accuracy ($p = 0.567$, $r = 0.08$) and time on task ($p = 0.954$, $r = 0.04$) among the control and adaptive group.

6.2 Perception of the Adaptive Interventions

We analyze the perception of the interventions in the adaptive group via the ratings that these users provided for the 10 statements related to usefulness, ease-of-use, and satisfaction, described in Section 4.3. Because the ratings for some of these statements were highly correlated, we selected only the four most distinct measures, namely those related to perceiving the interventions as *useful*, *delivered timely*, *confusing* and *distracting*. Table 5 provides summary statistics for these measures. Overall, the participants' ratings were positive for *useful*, *timing* and *confusion*, with modes of respectively 5 ("somewhat useful"), 6 ("right timing",) and 2.5 (between "not confusing" and "somewhat not confusing"). However, the interventions were found to be "somewhat distracting" (mode of 5).

4. An alternative is to create groups with an equal range of vis literacy scores (e.g., -2 to -.67 is Low, .67 to 2 is high, in-between is medium), which in our dataset yields less balanced groups, but does not change the statistical results of the post-hoc analysis (effect sizes are even slightly larger).

TABLE 5
Statistics on Perception of the Interventions

Measure	Mean	Mode	Sd	Min	Max
Useful	4.51	5	1.54	1	7
timely delivery	4.78	6	1.57	1	7
Confusing	3.07	2.5	1.66	1	7
Distracting	4.25	5	1.83	1	7

6.3 Impact of Percentage of Interventions Received

As discussed in Section 5, there were differences in the percentage of interventions triggered by each participant in the adaptive group ($\text{mean} = 89\%$, $\text{st. dev.} = 14\%$, $\text{min} = 31\%$, $\text{max} = 100\%$). We investigate whether these differences influence user performance, experience, and perception of the interventions. To facilitate the analysis, we discretized the number of received interventions into two bins (High and Low) via a median split. Users in the Low bin received 77 percent of the interventions on average ($\text{st. dev.} = 17\%$, $\text{min} = 31\%$, $\text{max} = 88\%$), and users in the High bin 97 percent on average ($\text{st. dev.} = 2\%$, $\text{min} = 91\%$, $\text{max} = 100\%$).

For each of the four measures of performance and experience (Table 4), we run a mixed model with the percentage of interventions triggered (Low or High) as the factor, and participant ID and MSNV ID as random effects. For each of the four measures of intervention perception in Table 5, we run a Kruskal-Wallis test with the percentage of interventions triggered as the independent. All results are adjusted using the Benjamin–Hochberg procedure.

We found significant main effects of percentage of interventions triggered on: *(i)* time-on-task ($F_{1,171} = 74.97$, $p < .0001$, $r = .55$); *(ii)* distraction ($x^2_{(1)} = 6.77$, $p < .009$, $r = .35$); *(iii)* confusion ($x^2_{(1)} = 5.53$, $p < .019$, $r = .38$). The directionality of these effects indicates that participants who triggered fewer interventions had shorter task times, but reported more confusion and more distraction.

7 Discussion

We discuss our results in more details and provide insights for designing and evaluating gaze-driven cues meant to guide the user's attention in narrative visualizations.

7.1 Value of Gaze-Driven Adaptation

Our results show that the gaze-driven interventions can improve comprehension of MSNVs, depending on the user's levels of vis literacy. This is noteworthy because so far, previous work only studied gaze-driven adaptation in maps with no narrative text [13], thus we broaden the evidence on the value of eye-tracking for adaptive visualizations. Furthermore, our findings are important because MSNVs constitute a widespread context of usage of information visualizations, in fact, they are the most common form of narrative visualizations found in real-world media (e.g., press, blogs, scientific reports) [15]. However, MSNVs are known to be challenging due to the need to integrate the two modalities. While there has been an interest in alleviating this difficulty with cues that guide understanding of the narrative, e.g., [15], [34], so far such cues have been provided upfront all at once [18], which does not scale to longer documents with multiple, possibly overlapping references. Cues were also displayed at the user initiative [40], however without improving comprehension, possibly because not everybody can effectively use and process on-demand cues. We contribute to this previous work by showing that gaze-driven adaptive cues can effectively support MSNV processing.

7.2 Role of User Characteristics

Our results are the first to show that long-term user characteristics (in our case *vis literacy*) can influence the value of adaptive visualizations. So far, previous studies have linked user characteristics to user performance in visualizations, e.g., [21], [25], [26], [27], [28], [29], [52], but not to the effectiveness of adaptive interventions as we do. Thus, our results provide strong rationales for leveraging user characteristics for the evaluation of other adaptive visualizations and call for studying personalization of the interventions.

The fact that users with low levels of vis literacy benefitted from the interventions is especially noteworthy because previous works have confirmed that having low vis literacy creates a disadvantage when working with visualizations, e.g., [4], [21], [28], [53]. Our finding provides promising initial evidence that gaze-driven interventions can help alleviate such disadvantage in the context of MSNV processing, so much so that low vis literary users outperformed (in terms of accuracy) the high vis literacy users in the presence of the interventions, without taking longer time on task. While high vis literacy users in our study were neither helped nor harmed by the interventions, they did not achieve a ceiling effect in accuracy, as shown by the fact that they were outperformed by low vis literacy users in the adaptive group. Based on these findings, it is worthwhile to explore other forms of guidance that suit better the specific needs of these users, starting with experimenting with other forms of highlights proposed for multimodal documents, such as visual links [18].

7.3 Perception of the Interventions

Participants overall found the interventions to be useful, delivered at the right timing, and not confusing, suggesting that the highlights generated a feeling of support to the task, which matches the finding that some users benefitted from the interventions. The low levels of confusion are especially noteworthy because participants received no training with the interventions prior to the study tasks, thus they could have misunderstood their meaning or the reasons for their appearance. Our results suggest that this was not a major factor and that the interventions were intuitive enough for most participants. It is possible, however, that introducing the users to the interventions beforehand might further improve the user experience.

Participants also found the interventions to be somewhat distracting, which is to be expected given that the interventions are provided dynamically during reading. Nonetheless, the levels of distraction remain moderate on average, and participants still reported that they found the interventions to be useful despite this distraction. Furthermore, the distraction did not escalate into confusion. Still, moving forward it will be important to study the specific reasons for distraction and ways to mitigate them.

7.4 Design Implications

Our study indicates that cumulatively adding thick borders to the relevant bars as users read the corresponding references in the text is suitable to support MSNV processing. Pilot tests revealed, however, that deemphasizing non-relevant bars was found to be too intrusive, which contradicts Carenini *et al.* [46] who found no difference in terms of user experience among border thickening and deemphasizing in low-level analytic tasks with stand-alone bar charts. This suggests that the effectiveness of different highlighting strategies depends in part on the target task and visualization, with deemphasizing or transparency being not suitable for MSNVs.

Our results indicate that the delivery timing of the interventions was suitable, since participants reported the timing of the interventions to be right, and were able to trigger about 90 percent of them on average. This also suggests that there were not many instances of interventions wrongly triggered because of spurious fixations detected on a reference, although a detailed analysis of the eye-tracking logs is needed to ascertain this.

Our results show that the amount of triggered interventions influences the user experience. Specifically, a high number of triggered interventions correlates with higher times on task, suggesting that receiving interventions slow users down. However, as discussed in Section 6.1, we found no difference in time on task between the adaptive group and the control group ($p = .19$, $r = .15$), which contradicts this hypothesis. Another possible explanation is that users who have a shorter time on task were fast readers who, because of their faster reading speed, happened to trigger fewer interventions. Additionally, users who received fewer interventions were significantly more distracted and more confused. The higher distraction might be due to the fact that interventions were triggered more erratically. The higher confusion might be due to the fact users expected some important information to be highlighted in the bar charts, whereas it was not always the case since they received less highlighting overall. This suggests that improving the consistency of the delivery mechanism could further increase the effectiveness of the adaptation.

8 Conclusion and Future Work

This paper proposes and evaluates a novel mechanism that provides gaze-driven highlighting interventions to help users process MSNVs, i.e., visualizations embedded in narrative text. Specifically, we leverage eye-tracking to detect in real time when the user is reading a sentence describing specified data points in the visualization, so as to dynamically highlight these data points in the visualization. We compared the performance and subjective experience of participants who received the gaze-driven highlighting against a control group who received none, and we also examined if users' performance and experience were influenced by long-term user characteristics known to play a significant role during MSNV processing. Our results show that the interventions were overall well perceived by the users and they benefitted specifically to users with low levels of vis literacy, who have lower ability in processing and understanding data visualizations. Overall, our work provides new insights on the value of gaze-driven adaptive guidance in visualization tasks, a topic that is receiving increasing interest in InfoVis [34]. This research is a first step towards designing personalized gaze-driven support for MSNV processing. As such, it provides a proof of concept for the potential of this guidance but also has limitations to be addressed in future work, as we discuss next.

The documents used in the study were excerpts from real-world MSNVs, usually quite shorter than the original documents, and we do not know how our results would generalize to these. We argue that the type of gaze-driven guidance we investigate should be even more helpful in longer, more challenging documents. However, it is possible that some of our findings on which intervention properties worked well (e.g., for the type of highlighting, removal strategies) might have to be adjusted. We plan to address this point by running a new study focusing on testing the proposed interventions with full-length MSNV. We also plan to investigate whether and how properties of the MSNVs (e.g., length, type of bar chart) influence user performance and the effectiveness of the interventions.

Although users overall appreciated the interventions, they found them to be somewhat distracting (although this distraction did not appear to hinder performance or user experience). We plan to examine how to reduce distraction and further improve the other scores of user perception, starting with the analysis of the study post-interviews to identify specific feedback that we can use to improve the intervention's design and delivery.

Although participants were able to trigger most interventions, our mechanism is quite sensitive to reading speed, but currently is not calibrated to it, causing some users to not trigger all the available interventions. We plan to investigate how to include this calibration in a way that is realistic for real-world settings so that the timing of the interventions can be personalized to users' reading speed.

There are several other steps of future work on our agenda, which will be facilitated by the software platform we have devised to support the implementation and the evaluation of gaze-based adaptive visualizations. For instance, we will experiment with adding cuing that guides the user's attention from the visualization back to the relevant reference in the text (i.e., the relevant references). We will examine ways to support high vis literacy users with alternative forms of adaptation that better suit their needs. We will compare our highlighting interventions with other types of dynamic cuing, i.e., a gaze-driven version of the static links presented in Steinberger *et al.* [18]. We also plan to investigate gaze-driven adaptation in MSNVs with different visualizations than bar charts. Altogether, thanks to this platform, we aim to better understand the value of gaze-driven adaptation across visualizations and tasks.

Acknowledgments

This work was supported by NSERC Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] D. Toker, C. Conati, G. Carenini, and Mona Haraty, "Towards adaptive information visualization: On the influence of user characteristics," in *Proc. 20th Int. Conf. User Modeling Adaptation Personalization*, 2012, pp. 274–285.

[2] A. Ottley, H. Yang, and R. Chang, "Personality as a predictor of user strategy: How locus of control affects search strategies on tree visualizations," in *Proc. 33rd Annu. ACM Conf. Human Factors Comput. Syst.*, 2015, pp. 3251–54.

[3] C. Ziemkiewicz, A. Ottley, R. Crouser, K. Chauncey, S. Su, and R. Chang, "Understanding visualization by understanding individual users," *IEEE Comput. Graphics Appl.*, vol. 32, no. 6, pp. 88–94, Nov./Dec. 2012.

[4] S. Lallé, C. Conati, and G. Carenini, "Impact of individual differences on user experience with a visualization interface for public engagement," in *Proc. 25th Conf. User Modeling Adaptation Personalization*, 2017, pp. 247–252.

[5] D. Gotz and Z. Wen, "Behavior-driven visualization recommendation," in *Proc. 14th Int. Conf. Intell. User Interfaces*, 2009, pp. 315–324.

[6] M. Mouine and G. Lapalme, "Using clustering to personalize visualization," in *Proc. 16th Int. Conf. Inf. Vis.*, 2012, pp. 258–263.

[7] B. Steichen, G. Carenini, and C. Conati, "User-adaptive information visualization: Using eye gaze data to infer visualization tasks and user cognitive abilities," in *Proc. Int. Conf. Intell. User Interfaces*, 2013, pp. 317–328.

[8] M. J. Gingerich and C. Conati, "Constructing models of user and task characteristics from eye gaze data for user-adaptive information highlighting," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 1728–1734.

[9] C. Conati, S. Lallé, M. A. Rahman, and D. Toker, "Further results on predicting cognitive abilities for adaptive visualizations," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, 2017, pp. 1568–1574.

[10] Y.-M. Jang, R. Mallipeddi, and M. Lee, "Identification of human implicit visual search intention based on eye movement and pupillary analysis," *User Model. User-Adapt. Interact.*, vol. 24, no. 4, pp. 315–344, Oct. 2014.

[11] S. Lallé, C. Conati, and G. Carenini, "Predicting confusion in information visualization from eye tracking and interaction data," in *Proc. 25th Int. Joint Conf. Artif. Intell.*, 2016, pp. 2529–2535.

[12] S. T. Iqbal, P. D. Adamczyk, X. S. Zheng, and B. Bailey, "Towards an index of opportunity: Understanding changes in mental workload during task execution," in *Proc. ACM Conf. Human Factors Comput. Syst.*, 2005, pp. 311–320.

[13] F. Göbel, P. Kiefer, I. Giannopoulos, A. T. Duchowski, and M. Raubal, "Improving map reading with gaze-adaptive legends," in *Proc. ACM Symp. Eye Tracking Res. Appl.*, 2018, pp. 29:1–29:9.

[14] K. Bektas, A. Cöltekin, J. Krüger, and A. T. Duchowski, "A testbed combining visual perception models for geographic gaze contingent displays," in *Proc. Eurographics Conf. Vis.*, 2015, pp. 67–71.

[15] E. Segel and J. Heer, "Narrative visualization: Telling stories with data," *IEEE Trans. Vis. Comput. Graphics*, vol. 16, no. 6, pp. 1139–1148, Nov. 2010.

[16] T. Munzner, *Visualization Analysis and Design*. Boca Raton, FL, USA: CRC Press, Taylor & Francis Group, 2015.

[17] P. Ayres and G. Cierniak, "Split-attention effect," in *Encyclopedia of the Sciences of Learning*, Berlin, Germany: Springer, 2012, pp. 3172–3175.

[18] M. Steinberger, M. Waldner, M. Streit, A. Lex, and D. Schmalstieg, "Context-preserving visual links," *IEEE Trans. Vis. Comput. Graphics*, vol. 17, no. 12, pp. 2249–2258, Dec. 2011.

[19] T. van Gog, "The signaling (or cueing) principle in multimedia learning," in *The Cambridge Handbook of Multimedia Learning*, Cambridge, U.K.: Cambridge Univ. Press, 2014, pp. 263–278.

[20] N. Kong, M. Hearst, and M. Agrawala, "Extracting references between text and charts via crowdsourcing," in *Proc. ACM Conf. Human Factors Comput. Syst.*, 2014, pp. 31–40.

[21] D. Toker, C. Conati, and G. Carenini, "Gaze analysis of user characteristics in magazine style narrative visualizations," *User Model. User-Adapted Interact.*, vol. 29, pp. 977–1011, 2019.

[22] J. Boy, R. A. Rensink, E. Bertini, and J.-D. Fekete, "A principled way of assessing visualization literacy," *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 12, pp. 1963–1972, Dec. 2014.

[23] E. W. Anderson, K. C. Potter, L. E. Matzen, J. F. Shepherd, G. A. Preston, and C. T. Silva, "A user study of visualization effectiveness using EEG and cognitive load," in *Proc. 13th Eurographics Conf. Vis.*, 2011, pp. 791–800.

[24] S. Lee, S.-H. Kim, Y.-H. Hung, H. Lam, Y. Kang, and J. Yi, "How do people make sense of unfamiliar visualizations?: A grounded model of novice's information visualization sensemaking," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 1, pp. 499–508, Jan. 2016.

[25] C. Conati, G. Carenini, E. Hoque, B. Steichen, and D. Toker, "Evaluating the impact of user characteristics and different layouts on an interactive visualization for decision making," *Comput. Graphics Forum*, vol. 33, pp. 371–380, 2014.

[26] B. Allen, "Individual differences and the conundrums of user-centered design: Two experiments," *J. Am. Soc. Inf. Sci.*, vol. 51, no. 6, pp. 508–520, 2000.
[27] A. Ottley *et al.*, "Improving Bayesian reasoning: The effects of phrasing, visualization, and spatial ability," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 1, pp. 529–538, Jan. 2016.
[28] K. Börner, A. Maltese, R. N. Balliet, and J. Heimlich, "Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors," *Inf. Vis.*, vol. 15, no. 3, pp. 198–213, Jul. 2016.
[29] C. Conati and H. Maclaren, "Exploring the role of individual differences in information visualization," in *Proc. Working Conf. Advanced Visual Interfaces*, 2008, pp. 199–206.
[30] S. D'Mello, A. Olney, C. Williams, and P. Hays, "Gaze tutor: A gaze-reactive intelligent tutoring system," *Int. J. Hum.-Comput. Stud.*, vol. 70, no. 5, pp. 377–398, May 2012.
[31] F. Alt, A. S. Shirazi, A. Schmidt, and J. Mennenöh, "Increasing the user's attention on the web: using implicit interaction based on gaze behavior to tailor content," in *Proc. 7th Nordic Conf. Human-Comput. Interact.*, 2012, pp. 544–553.
[32] N. Jaques, C. Conati, J. M. Harley, and R. Azevedo, "Predicting affect from gaze data during interaction with an intelligent tutoring system," in *Proc. 12th Int. Conf. Intell. Tutoring Syst.*, 2014, pp. 29–38.
[33] S. D'Mello, C. Mills, R. Bixler, and N. Bosch, "Zone out no more: Mitigating mind wandering during computerized reading," in *Proc. 10th Conf. Educ. Data Mining*, 2017, pp. 8–15.
[34] C. Collins *et al.*, "Guidance in the human–machine analytics process," *Vis. Inform.*, vol. 2, no. 3, pp. 166–180, Sep. 2018.
[35] J. Boy, L. Eveillard, F. Detienne, and J. Fekete, "Suggested interactivity: Seeking perceived affordances for information visualization," *IEEE Trans. Vis. Comput. Graphic*, vol. 22, no. 1, pp. 639–648, Jan. 2015.
[36] P. Ruchikachorn and K. Mueller, "Learning visualizations by analogy: Promoting visual literacy through visualization morphing," *IEEE Trans. Vis. Comput. Graphics*, vol. 21, no. 9, pp. 1028–1044, Sep. 2015.
[37] J. Hullman, N. Diakopoulos, and E. Adar, "Contextifier: Automatic generation of annotated stock visualizations," in *Proc. ACM Conf. Human Factors Comput. Syst.*, 2013, pp. 2707–2716.
[38] E. Ozcelik, I. Arslan-Ari, and K. Cagiltay, "Why does signaling enhance multimedia learning? Evidence from eye movements," *Comput. Hum. Behav.*, vol. 26, no. 1, pp. 110–117, Jan. 2010.
[39] S. Kalyuga, "Enhancing instructional efficiency of interactive E-learning environments: A cognitive load perspective," *Educ. Psychol. Rev.*, vol. 19, no. 3, pp. 387–399, Aug. 2007.
[40] Q. Zhi, A. Ottley, and R. Metoyer, "Linking and layout: Exploring the integration of text and visualization in storytelling," *Comput. Graphics Forum*, vol. 38, pp. 675–685, 2019.
[41] S. K. Badam, Z. Liu, and N. Elmqvist, "Elastic documents: Coupling text and tables through contextual visualizations for enhanced document reading," *IEEE Trans. Vis. Comput. Graphics*, vol. 25, no. 1, pp. 661–671, 2018.
[42] S. Latif, K. Su, and F. Beck, "Authoring combined textual and visual descriptions of graph data," in *Proc. Eurographics Conf. Vis.*, 2019, pp. 115–119.
[43] R. Metoyer, Q. Zhi, B. Janczuk, and W. Scheirer, "Coupling story to visualization: Using textual analysis as a bridge between data and interpretation," in *Proc. ACM Int. Conf. Intell. User Interfaces*, 2018, pp. 503–507.
[44] E. Adar, C. Gearig, A. Balasubramanian, and J. Hullman, "PersaLog: Personalization of news article content," in *Proc. ACM Conf. Human Factors Comput. Syst.*, 2017, pp. 3188–3200.
[45] G. Carenini, C. Conati, E. Hoque, and B. Steichen, "User task adaptation in multimedia presentations," in *Proc. 1st Int. Workshop User-Adaptive Inf. Vis.*, 2013, pp. 1–4.
[46] G. Carenini, C. Conati, E. Hoque, B. Steichen, and D. Toker, and J. Enns, "Highlighting interventions and user differences: Informing adaptive information visualization support," in *Proc. ACM Conf. Human Factors Comput. Syst.*, 2014, pp. 1835–1844.
[47] S. Lallé, C. Conati, and D. Toker, "A gaze-based experimenter platform for designing and evaluating adaptive interventions in information visualizations," in *Proc. ACM Symp. Eye Tracking Res. Appl.*, 2019, pp. 60–64.
[48] D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in *Proc. Symp. Eye Tracking Res. Appl.*, 2000, pp. 71–78.
[49] M. C. Dyson and M. Haselgrove, "The influence of reading speed and line length on the effectiveness of reading from screen," *Int. J. Human-Comput. Studies*, vol. 54, no. 4, pp. 585–612, Apr. 2001.
[50] A. M. Lund, "Measuring usability with the USE questionnaire12," *Usability Interface*, vol. 8, no. 2, pp. 3–6, 2001.
[51] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Royal Statist. Soc. Series B Methodol.*, vol. 57, no. 1, pp. 289–300, 1995.
[52] C. Conati, G. Carenini, D. Toker, and S. Lallé, "Towards user-adaptive information visualization," in *Proc. 29th Conf. Artif. Intell.*, 2015, pp. 4100–4106.
[53] N. Kodagoda, B. L. W. Wong, C. Rooney, and N. Khan, "Interactive visualization for low literacy users: From lessons learnt to design," in *Proc. ACM Conf. Human Factors Comput. Syst.*, 2012, pp. 1159–1168.

S. Lallé, received the PhD degree in computer science from the University of Grenoble. He is a research associate at the University of British Columbia (UBC). His research interest include user modeling and adaptation, aI in education, and machine learning. He has won a Best Paper Award at IVA'16 (Intel. Virtual Agents) and was runner-up for Best Student Paper at AIED'13 and TICE'12. He serves on the Board of Distinguished Reviewer of *ACM Trans. Interact. Intell. Syst.*

D. Toker, received the PhD degree from the University of British Columbia (UBC), in 2019. His areas of interests include user modeling, information visualization, eye-tracking, and machine learning. His publications in these fields have won the Best Paper Award at UMAP'14 (User Modeling, Adaptation and Personalization) and have been nominated for the Best Paper IUI'14.

C. Conati, is a professor of Computer Science at UBC. Her areas of research interest include user-adaptive interaction, user modeling, intelligent tutoring systems, and affective computing. She has more than 100 peer-reviewed publications in these fields, and her research has received awards from a variety of venues, including UMUAI (2002), IUI (2007), UMAP (2013, 2014), TiiS (2014), and IVA 2016. She is an associate editor for *UMUAI*, *ACM TiiS*, *IEEE Trans. Affect. Comp.*, and *IJAIED*, and serves on the executive committee of AAAI.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/csdl.