MindMargin: An Article-Adjacent Commenting Platform

Daniel Haehn
School of Engineering and
Applied Sciences
Harvard University
Cambridge, MA USA
haehn@seas.harvard.edu

Sharon Zhou
Harvard College
Cambridge, MA USA
zhou12@college.harvard.edu

Krzysztof Z. Gajos
School of Engineering and
Applied Sciences
Harvard University
Cambridge, MA USA
kgajos@eecs.harvard.edu

Copyright is held by the author/owner(s).
CHI'14, April 26–May 1, 2014, Toronto, Canada.
Copyright © 2014 ACM ISBN/14/04...$15.00.
DOI string from ACM form confirmation

Abstract
Commenting systems are a popular means for facilitating conversation among readers on many websites. Reading and writing comments can increase active user engagement in information exchange, personal reflection, and lively discussion, among others. We explore how user engagement can be increased by proposing a new commenting system and interface, MindMargin. In contrast to traditional commenting systems where comments are featured below the article, MindMargin presents comments adjacent to the article. Users can post and navigate comments and replies on a horizontal infinite scroll. Comments are anchored by users to specific sections or phrases of the article. This system exposes users to a diverse and relevant array of opinions as they read.

Author Keywords
Comments; anchored; commenting systems; opinions.

ACM Classification Keywords
H.5.2. [Information Interfaces and Presentation (e.g. HCI)]: User Interfaces

Introduction
News pages, media sites, online shops, blogs and social networks support the ability to give content-related

feedback in the form of comments. Online commenting holds a promise that readers might share their views, provide additional relevant information and generally engage in lively, thoughtful discussion with others. Given increasing polarization of the public political discourse, and prevailing opinion that presenting people with diverse points of view helps people develop more nuanced views [4, 5, 7], the promise of deep and thoughtful discussion among the readers is particularly valuable. However, this promise is not fulfilled: in particular, comments are frequently irrelevant or of low quality [3, 6].

We hypothesize that low quality of many of the comments is only part of the challenge. Even high quality comments have less impact on helping other readers develop more nuanced opinions than what might be possible. In particular, we note that commenting systems on most sites are featured under, and separate from, the main content. However, there is compelling prior research suggesting that separating additional content and discussion from the main content [[***has the following ill effects]] [1, 2, 8, 10, 9, 11, 12].

An alternative, explored previously particularly in educational settings [1, 2, 8, 10], is to present comments on the margin of the primary text and visibly linked to the relevant parts of the text. Such anchored discussions improve conversation among readers of academic texts by making the context of the conversation much more clear [1, 2, 8]. A study that compared the effect of placements of the notes on students' comprehension of the text, found that students who saw notes presented adjacent to the main text displayed deeper understanding of the text than students who saw those same notes placed at the end of the text [10].

We build on these insights to investigate whether placing user comments side-by-side the main text can improve how much people consider diverse view points when forming their own opinion of the issue. To enable empirical investigation of this question, we have built MindMargin, a system that exposes the user to comments while reading the main text. The comments are anchored to relevant sections of the content and are placed on an infinite horizontal scroll.

We propose two hypotheses for MindMargin's effect on its users in comparison to the traditional vertical interface:

1. Users of MindMargin will develop more thoughtful and nuanced opinions of an article than users presented with those same comments using a traditional vertical interface. Specifically, we expect that users who interact with MindMargin will report more moderate stances on controversial issues than users of a traditional vertical interface.

2. Users of MindMargin will report a more positive subjective impression of the existing comments because MindMargin displays anchored comments that appear alongside relevant passages of the text making the relevance of the comments to the primary text more apparent.

The results of our online study show that participants had a significantly more positive impression of the comments related to a controversial article when those comments were presented using MindMargin compared to when those same comments were presented below the main article. Participants who used MindMargin also reported a more moderate stance on the controversial issue raised in the article than participants who interacted with the traditional vertical layout interface. Although this

difference was not yet significant, it provides sufficient evidence for us to pursue a larger-scale study.

Approach

We implemented two commenting systems. The first commenting system is MindMargin with anchored comments on a horizontal infinite scroll next to the reference medium (see Figure 1). The second commenting system is a traditional vertical interface (see Figure 2). The two prototypes consist of clean user interfaces to avoid design clutter and distraction.

MindMargin is split into two sides: The reference media on the left and and an adjacent commenting system on the right. The commenting system displays comments in a horizontal infinite scroll. Thus, an unrestricted amount of comments can be linked to the reference media. Navigation within the infinite scroll component can be performed via mousewheel interaction (either left/right or top/down scrolling with the same effect) or by adjustment of a slider on the bottom of the right split screen. While navigating through the infinite scroll, the reference medium remains fixed on the left. Similar to [11, 12], comments are anchored to the horizontal reference point of the media. We minimize disturbance by avoiding interactions with the primary text, as defined by [11], and using thin dotted lines for anchoring to the article's right edge. This design decision was motivated by MindMargin's inherently more visible and thus distracting interface than the traditional vertical system's.

If a comment has replies, a dropdown button appears on the comment's footer. Lighter in color, replies to comments appear vertically under their comment when the button is clicked. This arrangement optimizes horizontal real estate by reserving horizontal space for parent comments. Readers can upvote and downvote comments. Most recent comments and most popular comments are displayed directly next to the reference media for increased visibility.

Our implementation of the traditional vertical commenting system follows a vertically ordered design: The reference media appears first and on top of the commenting system that follows below. Navigation within the article as well as within the comments can be performed via top/down scrolling. The replies, upvoting, and downvoting function and are organized in the same manner as MindMargin.

Experiment

We performed a between-subjects online experiments with young adults. Participants were randomly assigned to one of two conditions: MindMargin or a vertical commenting interface.

Participants. 106 online participants landed on our page for our user study and evaluation, of which 46 proceeded to begin and complete the study (30 female). 19 participants were assigned to the Mind Margin condition and 27 to the vertical interface condition. Participants were recruited online through social media and college listservs. Participants were college students, aged 18 to 25, and 68% hailed from the local university. The self-reported reading frequency of online news among participants ranged from daily to almost never.

Experimental Conditions. The two conditions in our study were MindMargin and the traditional vertical interface, both seeded with existing comments from a relevant news article. The article was selected on the basis of its opinionated nature and its relevance both in recent news and to our anticipated participant pool. We chose an opinion piece from our university's undergraduate

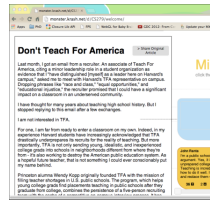

Figure 1: The Min
system with the ref
on the left and an a
commenting system

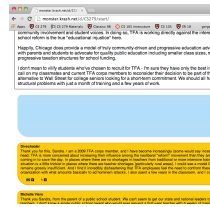

Figure 2: The trad
commenting system
vertically ordered de

publication, titled Don't Teach for America.

The article already had over fifty comments by affiliates and non-affiliates of the university alike, from which we selected the top 39 comments as ranked by Disqus, the existing commenting system on the publication's website, to be used in our study. The same comments were used in both conditions. In the traditional vertical interface, the comments appeared in the identical order as ranked in the original article. In MindMargin, we anchored them to the article based on textual references, specific phrases, quotes, and relevant content in each comment.

Users could make new comments by writing in the static new comment box above existing comments in the traditional interface and by clicking any part of the article to open a new comment box in MindMargin. In the MindMargin condition, we provided participants with simple, temporary instructions: "click the text to comment."

XXXX include that last paragraph? XXXX

Tasks. To ensure that our results would be informative for the design of real-world commenting systems, we designed the experimental tasks to focus participants' attention on the content of the article. The study design did not emphasize that the evaluation of the commenting system was the object of the study ([[XXX but was it disclosed in the consent form?]]

Participants were presented with an article and they were instructed to [[XXX ??? XXX]]. Once they completed reading the article, we asked them verification questions to ensure that they have indeed read and comprehended the article. Specifically we asked them about the overall stance of the article and, in a free-text response, for two pieces of supporting evidence used in the article. All 46 participants gave correct and thorough answers to these verification questions. Participants were then asked to complete a post-experiment questionnaire and were not permitted to refer back to the article once the questionnaire was administered. [[XXX state what questions were included in this questionnaire]]

To further incentivize focus on the content of the article and reflection on the issue it discussed, we used the following tagline to advertise the study: "Do you (really) think like a Harvard student?" and at the end of the study we provided them with feedback showing how their stance on the issue discussed in the article compared to those of other participants [[XXX other participants or Harvard students?]].

Procedure. Participants were given an initial questionnaire asking basic demographics and reading frequency. Before given the article, they were also asked either to provide a username or pseudonym, or to remain anonymous. During the reading of the article, participants were allotted 10 minutes. After 2 minutes, they were permitted to proceed to the questionnaire. The 2-minute delay was to ensure the reading of the article, and did not seem to prevent fast readers from moving too slowly, as the average reading time was 3 minutes 47 seconds.

In the follow-up questionnaire, reading verification questions were first posed. Participants were then asked their personal stance on the article, whether they liked the article, and whether they agreed with the article. They were also asked to self-report whether they read the comments in the article and to provide two adjectives that described either their reaction to, or a description of, the comments.

Design and Analysis. We use a between-subject factorial design w/ 2 factors 1st: condition 2nd factor: prior exposure to article and we used the following metrics/measures: ... extreme factor + and a number of subjective measures (adjectives==state that we coded these adjectives; discuss agreement across coders) explain what kinds of statistical tests we use

Results

Hypothesis 1

Our first hypothesis predicted that MindMargin would have an impact on individual opinions, prompting new readers to develop a stance on the issue and encouraging readers with existing views to consider alternate views. The reasoning behind this hypothesis was that MindMargin exposes readers to a greater number of opinions while reading the article. Consequently, readers think more independently, reconsidering their own views in light of others', and develop more thoughtful and nuanced opinions.

All participants were asked their stance on a Likert scale from Strongly For TFA to Strongly Against TFA. In our analysis, we excluded data from participants who reported to have not read the comments (there was statistically no difference in stance between prototypes Num dropped=XX and Nums for ea. avg stance=XX). We also excluded participants who did not toggle the Likert scale and answer this question (Num==XX). The data of the remaining participants (N=XX) was remapped for their "Stance Polarization," or the deviation of their stance from neutral, as $|50 - stance|$ (range 0-50).

Participants also reported on their familiarity with the article, which ranged from "read" if they had read or skimmed the article, to "seen" if they had seen the article in the past but did not read or skim it, to "none" if they had never encountered the article prior to this study. We then performed an analysis for both prototypes that included familiarity with the article as a covarying factor and the dependent variable SP, their "Stance Polarization."

We found that for users whose familiarity of the article was "none," there was no significant difference in the polarization of their stance between MindMargin and the traditional. For participants who were previously exposed to the issue in the article, and whose familiarity with the article was thus "seen" or "read," we did in fact observe a difference in stance polarization between the two prototypes (see Figure YYYY). Participants whose familiarity of the article was "seen" had a SP value of 7 with MindMargin and 17 with the traditional system. Participants whose familiarity of the article was "read" had a SP value of 17 with MindMargin and 27 with the traditional system.

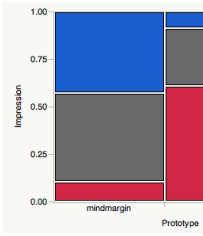

Figure 3: When us
MindMargin, the m
participants describ
comments as positiv

Hypothesis 2

Our second hypothesis predicted an overall increase in positive impressions on comments when using the MindMargin interface. We asked participants who read the comments to input two adjectives in free-text describing either their reaction to the comments or a description of the comments. We then asked four independent volunteers, blind to the experiment, to classify these adjectives using a four-bin classifier ("Positive," "Negative," "Neutral," and "Invalid"). They were told to classify the adjectives provided that they were in answer to the question, posed on the survey: "What did you think of the comments (from article X)¿'

We removed adjectives given two or more "Invalid" classifications as outliers, which only appeared in the

vertical condition. We found that adjectives without uniform encoding observed an uncontradictory mix of classifications: "Positive" and "Neutral" or "Neutral" and "Negative," but never "Positive" and "Negative". We were thus confident in using the resulting median encodings for the final classification of the specified adjectives.

Then, we compared the impressions between MindMargin and the traditional commenting system. We observed a drastic change of impressions when using MindMargin. As seen in figure 3, the greater majority of comment impressions in the traditional interface was negative (negative at 34.8%, neutral at 17.2%, positive 4.7%) while most of the comment impressions among users of MindMargin were classified either positive or neutral (negative at 4.7%, neutral at 20.3%, positive 18.8%).

Discussion

The lower SP values among MindMargin users reveal that participants with MindMargin who had prior exposure to the article reported less polarized views to reading the article for a second read or glance.

However, there are limitations to our observations, as the results are suggesting that our trends are not yet statistically significant. This could likely result from using a between-subjects methodology. To address this in subsequent research, we plan to significantly increase the participant pool and to employ a within-subjects methodology, in which we query participants for their individual stance prior to their reading of the article and observe the deltas in SP values afterward.

MindMargin users had a significantly more positive impression of the comments. Since all participants were exposed to identical comments, we conclude that MindMargin readers consider the comments more substantial and ultimately place greater trust and consideration into others' opinions.

In addition to our quantitative results, we would also like to acknowledge qualitative feedback from a MindMargin user that suggests actions he/she took beyond the scope of reading and commenting article: "This article showed me a new perspective on TFA, which after doing research, I have realized I agree with." No written feedback suggesting actions outside the scope of the article was received from participants with the traditional commenting system. While there is insufficient evidence to conclude that MindMargin motivated the user's pursuit of further research into the issue, it nevertheless indicates that this reader using MindMargin thought critically and independently about the issue.

Future Work

In the future, we plan to expand the participant pool to include participants of all ages and backgrounds. We would like to explore if MindMargin causes increased difficulty for readers to leave inflamed comments because MindMargin forces them to choose an appropriate anchoring place for their comments. Finally, we plan to pursue research on a commenting system like MindMargin, but for videos and music, that anchors comments to certain times or time-intervals within a given recording. Research into annotations on visual pieces, other than text, is also being considered.

References

[1] Brush, A. B., Bargeron, D., Grudin, J., Borning, A., and Gupta, A. Supporting interaction outside of class: Anchored discussions vs. discussion boards. In *In: Stahl, G. (Ed.): Proc. of CSCL 2002* (2002), 425–434.

[2] Guzdial, M., and Turns, J. Effective discussion through a

computer-mediated anchored forum. *Journal of the Learning Sciences 9*, (c) 2002 Inst. For Sci. Info (2000), 437–469+.

[3] Hsu, C.-F., Khabiri, E., and Caverlee, J. Ranking comments on the social web. In *Proceedings of the 2009 International Conference on Computational Science and Engineering - Volume 04*, CSE '09, IEEE Computer Society (Washington, DC, USA, 2009), 90–97.

[4] Kriplean, T., Morgan, J. T., Freelon, D., Borning, A., and Bennett, L. Considerit: Improving structured public deliberation. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, ACM (New York, NY, USA, 2011), 1831–1836.

[5] Munson, S. A., Lee, S. Y., and Resnick, P. Encouraging reading of diverse political viewpoints with a browser widget. In *ICWSM* (2013).

[6] O'Sullivan, P. B., and Flanagin, A. J. Reconceptualizing 'flaming' and other problematic messages. *New Media & Society 5*, 1 (Mar. 2003), 69–94.

[7] Park, S., Kang, S., Chung, S., and Song, J. Newscube: delivering multiple aspects of news to mitigate media bias. In *CHI*, D. R. O. Jr., R. B. Arthur, K. Hinckley, M. R. Morris, S. E. Hudson, and S. Greenberg, Eds., ACM (2009), 443–452.

[8] van der Pol, J., Admiraal, W., and Simons, P. R.-J. The affordance of anchored discussion for the collaborative processing of academic texts. *I. J. Computer-Supported Collaborative Learning 1*, 3 (2006), 339–357.

[9] Wagner, E. J., Liu, J., Birnbaum, L., and Forbus, K. D. Rich interfaces for reading news on the web. In *Proceedings of the 14th international conference on Intelligent user interfaces*, IUI '09, ACM (New York, NY, USA, 2009), 27–36.

[10] Wolfe, J. Annotations and the collaborative digital library: Effects of an aligned annotation interface on student argumentation and reading strategies. *I. J. Computer-Supported Collaborative Learning 3*, 2 (2008), 141–164.

[11] Zellweger, P., Regli, S. H., Mackinlay, J. D., and Chang, B.-W. The impact of fluid documents on reading and browsing: an observational study. In *CHI*, T. Turner and G. Szwillus, Eds., ACM (2000), 249–256.

[12] Zyto, S., Karger, D., Ackerman, M., and Mahajan, S. Successful classroom deployment of a social document annotation system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, ACM (New York, NY, USA, 2012), 1883–1892.