Readers' Perceptions of Personalized News Articles

Yixue Wang
yixue.wang@u.northwestern.edu
Northwestern University
Evanston, Illinois

Nicholas Diakopoulos
nad@northwestern.edu
Northwestern University
Evanston, Illinois

Abstract

The idea of article-level personalization is still distinct and vastly under-researched, and few studies explore readers' perceptions of personalized news articles. This research examines how readers perceive personalized content compared to non-personalized content via a study performed with Amazon Mechanical Turk participants. Our preliminary analysis shows that there is not much difference in readers' perceptions between personalized and non-personalized articles, and that readers perceive personalized and non-personalized articles slightly differently in subtle connection with their self-reported rating of the importance level of the news topic (i.e. gun violence). We also identified several avenues for future studies based on readers' feedback and the limitations of the study.

Keywords

News Personalization, News Automation, News Perception

1 Introduction

Modern online news platforms may use various data points about their audience, such as their location, to personalize the news reading experiences. Thurman and Schifferes [9] defined news personalization as "a form of user-to-system interactivity that uses a set of technological features to adapt the content, delivery, and arrangement of a communication to individual users' explicitly registered and/or implicitly determined preferences." News platforms believe that personalization can "surface content you may like, keep you up-to-date with topics you're interested in, and ultimately help you better understand what's happening in the world" [1] and therefore better help "manage information overload." [2]

Previous research helps explain why news organizations are interested in tailoring content to their users. The relationship between levels of customization and attitude toward a web portal are mediated by perceived relevance, interactivity, involvement, community, and novelty, and higher levels of customization will lead to higher revisits and engender more positive attitudes [3]. Gathering information from readers' personal characteristics could make news more relevant to readers, which is the paramount driver of news consumption [7], encouraging readers to return. News organizations can collect and generate user models for personalization via explicit feedback, including asking users about their preferences (explicit direct) and using users' feedback (explicit semidirect), and via implicit feedback, including monitoring users and deducing their preferences [6, 8].

There are two distinct forms of personalization [9]: active personalization through explicit feedback, and passive personalization through implicit feedback. Much of the news media's interest to date has revolved around the possibilities of implicitly personalized recommender systems for content, rather than actually personalizing the writing of an individual piece of content [3]. Some previous studies have, however, begun to provide guidelines on how to design personalized content, such as anticipating potential bias, providing reader control, and signaling personalization [1].

Studies have also revealed how readers perceive automated content via experiments and surveys. Automated text produced by an NLG system, Valtteri, received lower ratings than that produced by journalists, but the overall ratings were satisfactory based on credibility, likability, quality and representativeness [5]. Both U.S. and Chinese readers think automated reports have high quality when the reports are published in online media and tend to trust those reports as much as those produced by journalists [12]. In a different study, Wölker and Powel [11] found that European newsreaders don't perceive human-created, automated, and combined content credibility differently. On the contrary, another study shows that news attributed to a machine is perceived as less credible than news reported by human journalists [10]. For news organizations whose news was more trusted, automated news bots enhanced perceived objectivity [4].

Many studies have covered perceptions of automated content and the reasons behind personalized content, but there are few studies that seek to understand how readers perceive personalized news articles, which motivates our study. This study examines how readers perceive personalized content compared to non-personalized content. Specifically, how do

[1] https://help.nytimes.com/hc/en-us/articles/360003965994-Personalization

[2] https://niemanreports.org/articles/the-power-of-personalization/

[3] https://www.cjr.org/tow_center/journalism-newsfeeds-ai-artificial-intelligence.php

newsreaders perceive personalized news content's credibility, likeability, quality, and representativeness? In the following section, we introduce our study procedure and our analysis of the data. Our results indicate there is not much difference in readers' perceptions between personalized and non-personalized articles. And readers perceive personalized and non-personalized articles slightly differently in subtle connection with their self-reported importance level of the news topic (i.e. gun violence). We also propose several takeaways and limitations from this study for researchers and journalists to better design personalized news stories based on our results.

2 Study Design and Procedure

In this pilot study, we dynamically personalize a news story about gun violence in the U.S. and present that to participants recruited through Amazon Mechanical Turk (AMT) to get their feedback. We used Arria NLG Studio to personalize content by participants' location and gender information. The personalization process was based on a prototype developed at the Northwestern Knight Lab [4].

Participants were first asked for consent by reading through a consent form. Then, they were asked to provide their demographic information, including location and gender information (e.g. male, female, non-binary), and their self-reported interest level, knowledge level and importance level of gun-violence related news articles on a five-point Likert scale (1 being low and 5 being high), along with their news reading frequency and the news consumption platforms they typically use. After answering the questions about demographics and reading habits, they were randomly placed in a condition where either they read an article that was personalized for them (experimental condition) or not personalized for them (control condition) through a randomization function. Article personalization was implemented using the gender and location information provided in the initial questionnaire by adapting the inclusion and focus of information. One example of a personalized article as well as the non-personalized article are shown in Table 1. The survey recorded the time they spent on the article so we can understand whether personalized content promotes more engagement. We also used the time per word on the page of the news article to decide whether the result is valid or not. If participants spent too much or little time on the article page, we determine their answers are not valid and thus remove them. To implement this we used Isolation Forest in the scikit-learn package to remove any outliers based on the time they spent per word.

After reading the article, the survey asked questions around Sundar's four main receivers' criteria for the perception of news on a five-point Likert scale: 1) credibility (i.e., biased, fair, objective), 2) likability (i.e., boring, enjoyable, interesting, lively, pleasing), 3) quality (i.e., clear, coherent, comprehensive, concise, well-written), 4) representativeness (i.e., important, relevant, timely). The survey also asked participants to describe what they liked and didn't like about the articles so we could understand how to better design personalized articles in future studies.

Our study sample consisted of AMT users with an approval rate greater than 98% and the number of HITs approved greater than 1000. All of our study participants were English speakers and adults living in the U.S. since our news stories focus on issues of gun violence in the U.S. We determined the payment of USD 2.50 based on the average work time (1168s), and we estimated the sample size to be bigger than 164 by using power analysis to calculate the average absolute value of effect sizes (cohen's d = 0.40, power = 0.80, α = 0.05, Wilcoxon-Mann-Whitney tests) between controls and experiments based on an initial 20-participant pilot test. Our survey design received Institutional Review Board approval in May 2019.

In total, we collected data from 300 participants on Oct 9, 2019. 237 participants (88 female and 2 non-binary) provided valid answers after removing people who spent too much or too little time on the article page. On average, people with valid answers spent 940.4 seconds on the survey. 125 participants read the personalized articles, and the rest saw the non-personalized version.

3 Results

In Figure 1, we note that **non-personalized** articles have more ratings of 4 or 5 for *well-written*, *relevant*, *pleasing*, *important*, *fair*, *concise*, *coherent*, *clarity* , while **personalized** articles have more ratings of 4 or 5 for *timely*, *objectivity*, *lively*, *enjoyable*, *comprehensive*, *boring*, and *biased*.

We calculate *exciting* and *unbiased* using 6 minus *boring* and *biased* respectively in order to construct the four main factors proposed (i.e., credibility, likability, quality, and representativeness) by Sundar. Cronbach's α can show the internal consistency within a set of measurements. In our results, we see that all the αs from the four main factors are greater than 0.7, suggesting the metrics we measured in the survey preserving an acceptable internal consistency to represent the four main factors. We then construct the composite indicators using Benefit of the Doubt approach (BoD) in R *Compind* package, with all questionnaire items considered as outputs and a dummy input equal to one for all observations [2]. We report the four composite indicators built by BoD in Table 2.

We don't observe significant differences based on typical thresholds of significance between personalized articles and non-personalized articles in the data collected. However, for *concise* and *comprehensive*, there is some indication that users distinguish the length difference between the two versions,

[4]https://nuknightlab.github.io/studio-personal-story/

Table 1: Examples of personalized and non-personalized articles in the survey

Gun violence in the U.S. is more prevalent than in any other developed country, but Americans don't experience the symptoms of this epidemic equally. The U.S. gun death rate in 2017 hit a 20-year high after 39,773 people died from firearms, according to the Centers for Disease Control and Prevention. Nearly 12 people per 100,000 die from firearms each year — higher than the number who die in car accidents, once the leading cause of fatal injury. The U.S. has the world's highest rate of gun ownership by civilians, with 88 guns for every 100 people. This problem may be closer to home than you think. A recent gun-related crime in Illinois occurred on October 7, 2019 in Chicago. 2 people were injured. Firearm mortality in Illinois is 12.1 per 100,000 people. This is the 30th highest in the nation. (*Personalized based on IL*) One-third of all gun deaths are homicides. The U.S. gun homicide rate is 25 times that of other high-income countries, and these deaths are heavily concentrated in urban areas. More than 50 percent of murders occur in just 2 percent of the nation's 3,142 counties, according to the Crime Prevention Research Center. More than 50 percent of U.S. counties had zero murders in 2014. In New Orleans, Detroit and Baltimore, the homicide rate is up to 10 times higher than the national average, with almost 40 murders for every 100,000 people. (*Personalized based on urban*) Approximately 93 percent of suspects and nearly 82 percent of victims are male, while slightly more than 6 percent of suspects and 18 percent of victims are female. (*Personalized based on urban and male*) Although gun homicides tend to draw more attention from readers and viewers, suicides may be the bigger problem. Two-thirds of gun-related deaths are suicides, and suicide by firearm accounts for almost half of all suicides in the U.S. In 2017, homicide firearm deaths totaled 14,542, while suicides by firearm totaled 23,846. For men, firearm suicide rates largely increase with age, peaking in the 65-and-older group. For women, firearm suicide rates are highest in the 45 to 60 age range. Men represent 86 percent of firearm suicide victims, and are more than six times more likely than women to die by firearm suicide, according to the CDC. (*Personalized based on male*) ***Personalized article example*** *(Location: IL and urban area; Gender: male)* Average word count: 295	Gun violence in the U.S. is more prevalent than in any other developed country, but Americans don't experience the symptoms of this epidemic equally. The U.S. gun death rate in 2017 hit a 20-year high after 39,773 people died from firearms, according to the Centers for Disease Control and Prevention. Nearly 12 people per 100,000 die from firearms each year — higher than the number who die in car accidents, once the leading cause of fatal injury. The U.S. has the world's highest rate of gun ownership by civilians, with 88 guns for every 100 people. One-third of all gun deaths are homicides. The U.S. gun homicide rate is 25 times that of other high-income countries, and these deaths are heavily concentrated in urban areas. More than 50 percent of murders occur in just 2 percent of the nation's 3,142 counties, according to the Crime Prevention Research Center. More than 50 percent of U.S. counties had zero murders in 2014. In New Orleans, Detroit and Baltimore, the homicide rate is up to 10 times higher than the national average, with almost 40 murders for every 100,000 people. Gun Homicides get far more attention than suicides in the press, but the latter may be the bigger problem. Two-thirds of the gun-related deaths are suicides, and suicide by firearm accounts for almost half of all suicides in the U.S. In 2017, homicide firearm deaths totaled 14,542, while firearm suicides totaled 23,846. ***Non-personalized article*** word count: 202

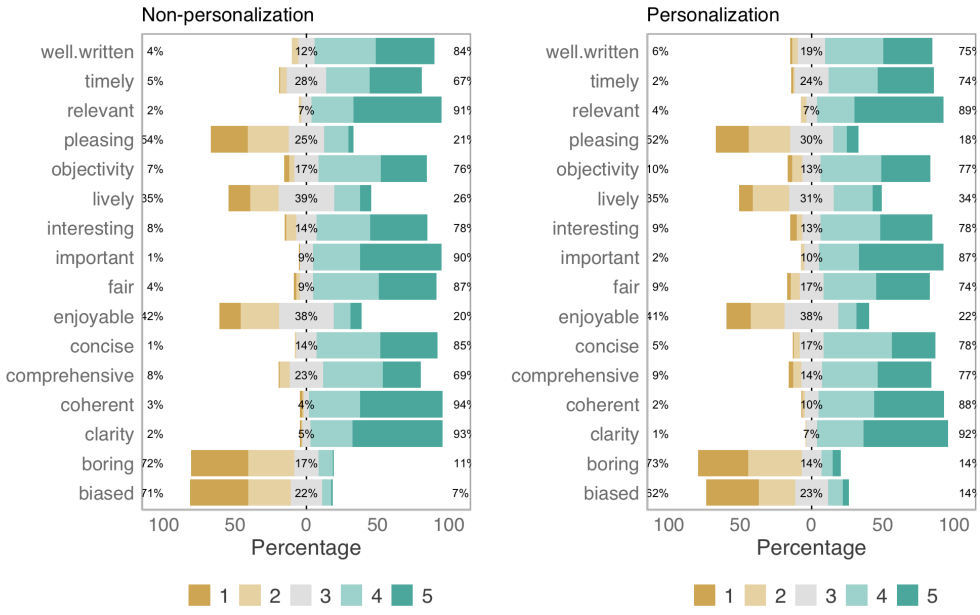

Figure 1: Ratings for personalized and non-personalized news stories

and perhaps future work, or a higher powered study could elaborate on this statistical trend (see Table 2).

Regression Models

In order to understand how users' knowledge, importance, and interest levels in gun violence impact on readers' perceptions of personalized articles, we build four regression models on the four composite indicators calculated (Table 2).

Figure 2 presents the coefficients and the robust standard errors of the four scaled regression models of the four composite indicators on credibility, likability, quality, and representativeness. From Figure 2, we note that personalized articles appear to have a slight negative impact on readers' perception of quality. Within a similar scope, male readers tend to perceive personalized articles as slightly lower quality, and longer articles have a slight positive impact on the perception of quality when controlling for other factors.

Figure 3 shows a potential interaction impact between the importance level and personalization on the perceptions of likability and representativeness. Even though we observe a slight positive impact on the two composite indicators, the boost from the personalized articles is not much beyond the non-personalized articles even at the highest level of importance. Possible reasons could be that our personalized articles are longer and thus difficult for readers to comprehend in a short time.

Table 2: Ratings of the evaluations of the main factors between non-personalized articles and personalized articles

	Non-Personalized Mean (Median)	**Personalized** Mean (Median)	**p-value** (Mann–Whitney U test)
Quality			
Composite indicators	0.95 (1)	0.94 (1)	0.543
		Cronbach's α= 0.775	
well-written	4.20 (4)	4.02 (4)	0.131
concise	4.24 (4)	4.03 (4)	**0.068**
comprehensive	3.87 (4)	4.02 (4)	**0.085**
coherent	4.47 (5)	4.34 (4)	0.106
clarity	4.54 (5)	4.50 (5)	0.526
Likability			
Composite indicators	0.88 (1)	0.86 (1)	0.307
		Cronbach's α= 0.799	
pleasing	2.44 (2)	2.50 (2)	0.736
lively	2.84 (3)	2.95 (3)	0.433
enjoyable	2.71 (3)	2.73 (3)	0.868
interesting	4.09 (4)	4.01 (4)	0.700
exciting	4.00 (4)	3.89 (4)	0.516
Credibility			
Composite indicators	0.89 (1)	0.86 (1)	0.373
		Cronbach's α= 0.815	
objectivity	3.97 (4)	3.97 (4)	0.819
fair	4.20 (4)	4.00 (4)	0.175
unbiased	4.04 (4)	3.08 (4)	0.185
Representativeness			
Composite indicators	0.94 (1)	0.94 (1)	0.465
		Cronbach's α= 0.703	
important	4.46 (5)	4.44 (5)	0.943
relevant	4.51 (5)	4.47 (5)	0.963
timely	3.97 (4)	4.09 (4)	0.362
time per word	0.445s (0.396s)	0.412s (0.380s)	0.587

4 Discussions

In this paper, Our preliminary analysis shows that: (1) In general, there is not much difference in readers' perceptions between personalized and non-personalized articles; (2) Readers perceive personalized and non-personalized articles slightly differently, in subtle connection with their self-reported rating of the importance level of the news topic (i.e., gun violence).

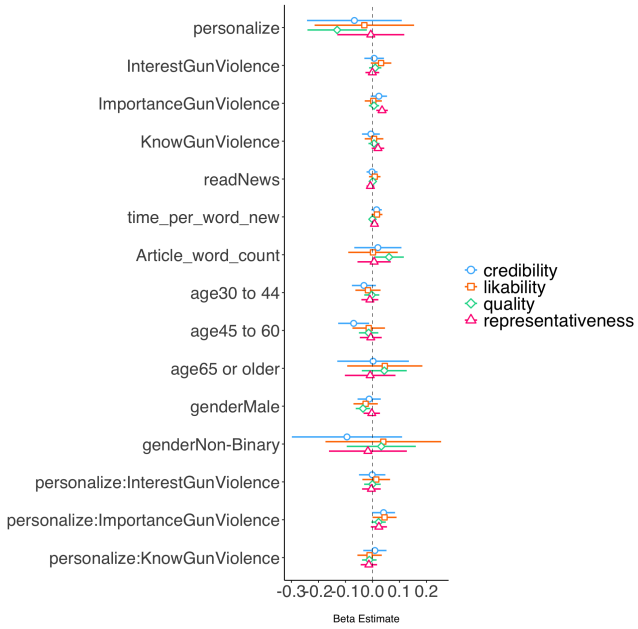

Figure 2: Coefficients and 95% CIs of the four scaled regression models of the composite indicators on credibility, likability, quality, and representativeness

This finding could help news organizations better design their used of personalized articles, especially when providing personalized articles to their readers. Our results suggest that it may be better to personalize the articles for readers who think the topic is important, rather than readers who like the topic or know the topic.

There are also some interesting insights from participants' feedback on the stories. People mentioned that the articles could be better if (1) adding personal stories and expert/citizen quotes, (2) visualizing the numbers in graphs because of too many numbers in the article, (3) starting with a clear introduction to explain the viewpoints and structure of the articles, and (4) changing the dry writing style. Some of them said that they didn't like the topic and there was no solution to the gun violence mentioned in the articles, and some people thought the articles were "clearly biased" and "one sided only showing fatality rates from gun violence and doesn't give you the other side of how guns could prevent violence."

On the other hand, people also mentioned that they liked that the articles contained facts and numbers linked to other sources, were straightforward and concise with clear numbers and statistical information, used numbers rather than opinions to convey, and were non-biased because they mention that suicide as well as other gun crimes are not evenly distributed.

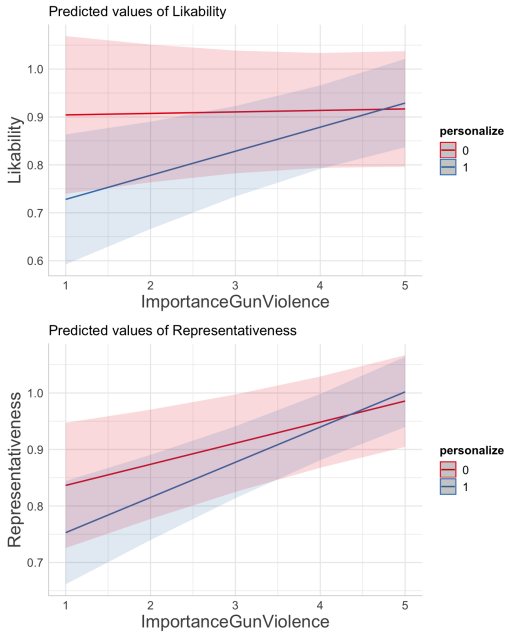

Figure 3: Interaction impact from the importance level and the personalized variable on (a) likability and (b) representativeness

It turns out that some readers realized the articles were personalized even without any hints in the survey, but their reactions to the personalization were different. Some didn't care if the gun-related accident mentioned in the personalized article happened close to them: they found it "irrelevant to the point" or even thought it used their "IP address to target" them. Some liked it because it was "relevant for their state" and added a "personalized regional touch" to the article "with recent and semi local statistics." Because of the variance in reception, future work may find it interesting to study the impact of labeling personalized articles as such.

Limitations

There are some limitations of this pilot study that could inform better study designs in the future. The first one is the controversial topic (i.e., gun-violence) that evokes personal emotions. Participants might have provided their ratings based on their positions on gun violence, rather than their actual perceptions of the articles. In future work we hope to expand this study to more neutral topics in order to see how the results may vary.

Another limitation is that our study was conducted on AMT, which doesn't provide the most organic reading experience and thus our results could be biased due to the monetary incentive over the interest incentive. Future studies need to consider how to make the reading and rating experience more organic, and may therefore benefit from more naturalistic field deployments.

Last but not least, by including extra personalized content, the personalized articles might "jump around to different subjects rapidly" as one participant mentioned, which could make the personalized articles hard to understand. In our future work, we plan to make sure that the personalized and non-personalized articles follow the exact same structure to make sure the two versions' length are the same.

5 Acknowledgments

We thank Nneoma Oradiesgwu, Alexis Shanes, Hayden Udelson and Daniel Zhou for participating in the news personalization prototype design at Northwestern University Knight Lab and Hayden Udelson and Daniel Zhou for participating in the survey design on AMT using Javascript and HTML.

References

[1] Eytan Adar, Carolyn Gearig, Ayshwarya Balasubramanian, and Jessica Hullman. 2017. PersaLog: Personalization of news article content. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3188–3200.

[2] Kristof De Witte and Nicky Rogge. 2009. Accounting for exogenous influences in a benevolent performance evaluation of teachers. *Available at SSRN 1462690* (2009).

[3] Sriram Kalyanaraman and S Shyam Sundar. 2006. The psychological appeal of personalized content in web portals: does customization affect attitudes and behavior? *Journal of Communication* 56, 1 (2006), 110–132.

[4] Bingjie Liu and Lewen Wei. 2019. Machine Authorship In Situ: Effect of news organization and news genre on news credibility. *Digital Journalism* 7, 5 (2019), 635–657.

[5] Magnus Melin, Asta Bäck, Caj Södergård, Myriam D Munezero, Leo J Leppänen, and Hannu Toivonen. 2018. No Landslide for the Human Journalist-An Empirical Study of Computer-Generated Election News in Finland. *IEEE Access* 6 (2018), 43356–43367.

[6] Miquel Montaner, Beatriz López, and Josep Lluís De La Rosa. 2003. A taxonomy of recommender agents on the internet. *Artificial intelligence review* 19, 4 (2003), 285–330.

[7] Kim Christian Schrøder. 2019. What do news readers really want to read about? How relevance works for news audiences. (2019).

[8] Michal Sela, Talia Lavie, Ohad Inbar, Ilit Oppenheim, and Joachim Meyer. 2015. Personalizing news content: An experimental study. *Journal of the Association for Information Science and Technology* 66, 1 (2015), 1–12.

[9] Neil Thurman and Steve Schifferes. 2012. The future of personalization at news websites: Lessons from a longitudinal study. *Journalism Studies* 13, 5-6 (2012), 775–790.

[10] T Franklin Waddell. 2018. A robot wrote this? How perceived machine authorship affects news credibility. *Digital journalism* 6, 2 (2018), 236–255.

[11] Anja Wölker and Thomas E Powell. 2018. Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism. *Journalism* (2018), 1464884918757072.

[12] Yue Zheng, Bu Zhong, and Fan Yang. 2018. When algorithms meet journalism: The user perception to automated news in a cross-cultural context. *Computers in Human Behavior* 86 (2018), 266–275.