

What prompts users to click on news headlines? Evidence from unobtrusive data analysis

Evidence from
unobtrusive
data analysis

49

Tingting Jiang

*School of Information Management, Wuhan University, Wuhan, China and
Center for Studies of Information Resources,
Wuhan University, Wuhan, China, and*

Qian Guo, Shunchang Chen and Jiaqi Yang

School of Information Management, Wuhan University, Wuhan, China

Received 17 April 2019
Revised 27 July 2019
Accepted 16 September 2019

Abstract

Purpose – The headlines of online news are created carefully to influence audience news selection today. The purpose of this paper is to investigate the relationships between news headline presentation and users' clicking behavior.

Design/methodology/approach – Two types of unobtrusive data were collected and analyzed jointly for this purpose. A two-month server log file containing 39,990,200 clickstream records was obtained from an institutional news site. A clickstream data analysis was conducted at the footprint and movement levels, which extracted 98,016 clicks received by 7,120 headlines ever displayed on the homepage. Meanwhile, the presentation of these headlines was characterized from seven dimensions, i.e. position, format, text length, use of numbers, use of punctuation marks, recency and popularity, based on the layout and content crawled from the homepage.

Findings – This study identified a series of presentation characteristics that prompted users to click on the headlines, including placing them in the central T-shaped zones, using images, increasing text length properly for greater clarity, using visually distinctive punctuation marks, and providing recency and popularity indicators.

Originality/value – The findings have valuable implications for news providers in attracting clicks to their headlines. Also, the successful application of nonreactive methods has significant implications for future user studies in both information science and journalism.

Keywords Clicking behaviour, Clickstream data analysis, Crawled data, News headline presentation, News selection, Unobtrusive methods

Paper type Research paper

Introduction

The shift of news consumption to online sources has been taking place globally. According to the Pew Research Center, 93 percent of American adults read news online at least occasionally, either via mobile or desktop devices (Katerina and Elisa, 2018). Readers turn most frequently to social media and news organizations' websites and applications for news (www.journalism.org/fact-sheet/digital-news/). In China, the size of online news consumers has reached 663m, and 95 percent of them are mobile users (www.cnnic.net.cn/hlwzfzyj/hlwxzbg/hlwtjbg/201808/P020180820630889299840.pdf). An important fact found by the Chinese Online News Market Survey 2016 is that many people tended to read merely the news headlines or the news in the first screen and barely involved themselves in in-depth reading. This could be primarily attributed to the news explosion on the internet (www.cnnic.net.cn/hlwzfzyj/hlwxzbg/mtbg/201701/P020170112309068736023.pdf).

Nowadays, news organizations worldwide produce online news continuously, which is complemented by social media that have greatly accelerated news dissemination and consumption. Most online news consumers feel overloaded with the amount of news available



Aslib Journal of Information
Management
Vol. 72 No. 1, 2020
pp. 49-66

© Emerald Publishing Limited
2050-3806
DOI 10.1108/AJIM-04-2019-0097

This research has been made possible through the financial support of the National Natural Science Foundation of China under Grants Nos 71774125 and 71420107026.

(Holton and Chyi, 2012). News selection is a forced move for them to take in order to cope with information overload (Edgerly *et al.*, 2014). Unlike print newspapers, the display of online news features an entering page that lists headlines hyperlinked to news article pages. The presentation of headlines on the entering page has a substantial impact on what will be read and what will be ignored (Knobloch-Westerwick, Sharma, Hansen and Alter, 2005).

As a result, the major role of an online news headline is no longer a succinct and accurate summary of the corresponding news story. Instead, it should at least be able to arouse users' curiosity about the news story, and sometimes it needs to entice them to click and open the news articles (Arapakis *et al.*, 2014). If a headline loses the competition for users' clicks, the news article will stay ignored despite its high quality (Kuiken *et al.*, 2017). News providers are thus striving to increase headline salience in virtue of visual cues, including both headline components (e.g. use of numbers and/or questions) and peripheral indicators (e.g. recency and/or popularity indicators) (Xu, 2013).

A number of studies have been conducted on the influence of various headline characteristics on news consumers' attention, perception and behavior, which will be reviewed in detail in the next section. Existing findings derived mainly from experiments, lab-based or field, that intimated real-world news consumption scenarios. Instead, this study made creative use of the clickstream data generated from users interacting with online news in the real world, coupled with the page layout and content crawled from a working news site. The goal is to reveal the relationships between headline presentation and users' news selection behavior. The following research questions are addressed:

- RQ1. What are the characteristics of online news headline presentation in terms of position, format, text length, use of numbers, use of punctuation marks, currency and popularity, respectively?
- RQ2. Do the differences in each characteristic dimension engender different news selection behavior? If yes, what are the differences?

Literature review

News selection

News selection in early studies actually referred to how journalists selected news to report (Welbers *et al.*, 2016). In contrast, how the audience selected news to read remained under-researched until the explosion of online news. Audience news selection can be made based on content or form (Knobloch-Westerwick, Sharma, Hansen and Alter, 2005). Content-based news selection is driven by user motivations and goals (Edgerly *et al.*, 2014). The personal utility of news content basically determines news selectivity (Knobloch-Westerwick, Dillman Carpentier, Blumhoff and Nickel, 2005). Due to information overload, however, highly routinized news consumption has been replaced gradually by stumbling upon news online. News acquisition is increasingly associated with users' low involvement and low expectation (Yadamsuren and Erdelez, 2016). This has given rise to form-based news selection that emphasizes news display features, such as headline typeface size and accompanying illustrations. Also, collaborative filtering recommendations, including implicit (e.g. times read) and explicit (e.g. ratings) indicators, have been introduced to affect news selection (Knobloch-Westerwick, Sharma, Hansen and Alter, 2005).

Selection is a process consisting of three levels: selective attention determines to which stimuli in the environment our perceptual system is addressed; selective perception determines how information adapted to our cognitive system will be processed and stored; and selective retention determines which of the prior stored information our cognitive system will recollect (Donsbach, 2004). Form-based news selection has been investigated at the first two levels. Bucher and Schumacher (2006) compared the attention patterns in the

reception of print and online news in an eye-tracking study. They found that visual cues and information hierarchies guided the attention process as stimuli for an active, intention-driven selection process. Go *et al.* (2014) explored the effects of interface cues on users' perceptions of online news. Their findings indicated that news from an expert source had higher perceived quality and that news from an expert source and recommended by more people had higher perceived credibility.

Selective retention, the highest level of selection, can be revealed by such observable audience behavior as clicking, recommending and commenting, etc. (Wendelin *et al.*, 2015). It is most relevant to this study which focused on users clicking on headlines, the most straightforward behavioral manifestation, in form-based news selection.

Semiotics and news headline techniques

Semiotics is a philosophical approach to interpret messages in terms of signs (Holsanova *et al.*, 2006). A sign can be a word, image, or sound. A semiotic process divides into three phases which are perception, manipulation, and consummation. That is, a person becomes aware of a sign, interpret it, and respond to it in sequence (Morris, 1964). Since semiotics can be used to generate and convey meaning and thus influence people's response, it has wide application in the real world, especially in marketing. Recent related studies have explored the interplay of verbal and visual elements in semiotics and found that their congruence or coordination would enhance the effectiveness of communication (Adegoju and Ademilokun, 2015; Yu and Song, 2017). The prevalence of visual resources on the internet has given rise to visual semiotics which has been found to positively affect users' attitudes, intention, and/or behavior (Wanick *et al.*, 2018; Valentini *et al.*, 2018).

When it comes to increasing the attractiveness of a headline, news providers mainly depend on verbal techniques, including simplification, sensationalism, selectivity, negativity, forward-reference, using questions, quotes, numbers, buzzwords, punchline, and metaphors and metonymies, embedding surprising, controversial, and gossip elements, controlling headline length and word length, and so forth (Tenenboim and Cohen, 2013; Lai and Farbroth, 2014; Blom and Hansen, 2015; Kuiken *et al.*, 2017; Shie, 2011; Rubin *et al.*, 2016; Rony *et al.*, 2017). Some of these techniques have been criticized as "clickbait," which aim to attract attention but actually distract attention as constant clicking into new articles will prevent in-depth reading (Chakraborty *et al.*, 2016).

Despite the variety of attractive headlines, there still lacks sufficient evidence concerning whether they actually encourage readership, and even conflicting empirical results have been obtained from different studies. Safran (2013) investigated five types of news headlines, i.e., normal, question, how to, number, and reader-addressing, and found that number headlines resonated the most, especially for females, while question headlines resonating the least. In a clickbait detection study, Chakraborty *et al.* (2016) found that clickbait headlines were longer than conventional non-clickbait headlines, taking the form of well-formed English sentences with both content and function words. A focused study on question headlines indicated that question headlines were significantly more effective than declarative ones, particularly those with self-referencing cues (Lai and Farbroth, 2014). However, Kuiken *et al.* (2017) found that the absence of a question would lead to better headline performance measured in terms of click-through rate and the use of numbers and headline length had no significant influence on performance.

Online information presentation and users' attention, perception and behavior

Besides news headlines, researchers have investigated widely how the presentation of various types of online information, such as search results, product information, advertisements, and social media content, influences users' attention, perception and behavior. This study

identified seven major factors from a comprehensive review of existing related studies in information science, journalism, marketing and linguistics, etc.

Position. When users are viewing webpages, there is a dominant F-shaped pattern of eye movements, i.e., eyes first scanning the top of the page and then downwards (Nielsen, 2006). This pattern also applies to search result pages. The click-through rate decreases rapidly with the rank of the result item. The first item could receive three times more clicks than the second one (Penna and Quaresma, 2015). Advertisers have a strong tendency to secure the top positions for their sponsored results (Jansen *et al.*, 2014). The positioning of banner ads is an important predictor of users' response in terms of ad processing, attention, click and attitude (Broeck *et al.*, 2018).

Format. The "dual coding" theory states that humans encode verbal and visual information with two separate mental systems (Yang *et al.*, 2017). The superiority of images over words has been identified for product presentation in online shopping. Product images might help attract attention, reduce uncertainty, increase credibility and thus encourage purchase decisions (Blanco *et al.*, 2010; Xu *et al.*, 2015). The positive effect of images could be further enhanced by increasing image size or adding motion (Park *et al.*, 2005; Song and Kim, 2012). Visual elements have also been proved more effective than textual ones in attracting attention to ads (Kuisma *et al.*, 2010).

Title length. The title is an important component of a post shared on social media. Users tend to spend more time scanning titles than reading posts. Longer titles might increase users' reading load, thus ineffective in attracting views or replies to posts (Wagner *et al.*, 2012; Ma and Chen, 2017). On Amazon, each customer review has a title, and title length was negatively related to readership (Salehan and Dan, 2016). An interesting finding for psychology articles is that amusing titles were usually shorter, and shorter titles had more citations (Subotic and Mukherjee, 2014). There also exist various forms of short text without a title on the Web. It has been revealed that tweets shorter than 100 characters had a higher engagement rate, and ultra-short 40-character Facebook posts received much more likes and comments (Lee, 2014).

Use of numbers. The use of numbers is frequently seen in the health domain. Numbers are often associated with higher scientific credibility as they can engender more accurate perceptions than probability phrases and graphical displays, and such accuracy can be verified. Numeric information could facilitate informed decision making by reducing uncertainty and resolving conflicts (West *et al.*, 2013). Statistical evidence had a stronger influence on beliefs and attitude than narrative evidence (Zebregs *et al.*, 2015). In online shopping or advertisement, the price of a product is a typical type of number. Providing pricing information would decrease the click-through rate for banner ads but increase that for search engine ads and group-buying products (Chtourou *et al.*, 2002; Atkinson *et al.*, 2014; Liu *et al.*, 2015).

Use of punctuation marks. A paralinguistic feature, punctuation is used to indicate pause, tone and term attribution and functions in written languages. Appropriate punctuation marks enable sentences to be read smoothly while inappropriate ones may puzzle readers and lower their interest (Huang and Yang, 2010). Unambiguous sentences with internal punctuation would be read faster than without (Hirotnani *et al.*, 2006). It has been found that academic authors used various punctuation marks in their article titles, with colon, hyphen and coma the most frequently used (Fumani *et al.*, 2015). Noguti (2016) considered punctuation when examining the titles of posts on social media but identified no significant influence of punctuation marks on users' interaction with the posts.

Recency. Recency or timeliness refers to how recent the information is. It is an important aspect of information quality and has an impact on users' assessment of the credibility

of information (Metzger, 2007). News is time-sensitive information, and recency is a dominant criterion for selecting online news and judging its trustworthiness (Choi and Stvilia, 2015). On social news aggregators, when the credibility of the source was low, more recent news received more attention; and concerning high-credibility sources, more recent news had higher perceived newsworthiness (Xu, 2013). Further evidence has suggested that faster updates on social media would lead to higher source credibility as mediated by cognitive elaboration (Westerman *et al.*, 2014). Also, the recency of electronic word-of-mouth messages has been found to positively influence customers' perception of their usefulness which in turn predicts purchase intention (Cheung, 2014).

Popularity. It is natural for individuals to follow others' choices and attitudes in the face of an overwhelming amount of information rather than relying on their own judgment (Wang and Lin, 2011). User-generated content on social media, e.g., micro-posts, will receive likes, reposts and comments and replies, which together indicate popularity. Higher popularity is often associated with greater perceived usefulness and stronger preferences and thus increases the intention of interaction and possibility of actual interaction (Chang *et al.*, 2015; Chin *et al.*, 2015). The popularity of a product in online shopping, i.e., having a large number of people liking or purchasing it, would positively affect customers' trust, perception of product quality and value as well as purchase intention and behavior (Mou and Shin, 2017). Ksiazek *et al.* (2016) found that more popular videos with larger numbers of views, favorites and ratings would attract more comments while less popular ones more replies to comments.

Clickstream data analysis

As social sciences are attaching more and more importance to nonreactive research, researchers have made wide use of trace data to explore humans' online behavior. The value of trace data exists in its unobtrusiveness. That is, data are automatically accumulated as a result of real-world activities (Jansen, 2009). Clickstream data is a typical type of trace data generated on Web servers when users visit websites or applications for their own purposes. It captures all the clicks or page requests made by users in sequence, from entering to leaving a site. Each clickstream record basically informs us which user performs which type of action on which page at what time (Sen *et al.*, 2006). Clickstream data analysis was first employed by E-commerce researchers to perform customer clustering and to characterize and model customers' visiting and purchasing behavior. It has also been widely applied in user behavior studies in other contexts, such as social media, social commerce, online courses and so on.

A general framework for analyzing clickstream data has taken shape, which consists of three levels, i.e., footprint, movement and pathway (Chi *et al.*, 2017). When a user visits a site, each click causes a movement, the changing of location from one page to another, and leaves a footprint, a mark showing the user's presence on a page. The click series during that visit, i.e., chaining all the movements in a chronological order, engenders a pathway, indicating the process in which the user interacts with the site. In particular, the footprint left on a destination page where intended information is provided or intended task is enabled is a "core footprint," and the movement that leads the user to the destination page is a "pivotal movement." The footprint level analysis aims to provide a basic understanding of traffic distribution within a site or subsite. The movement level analysis concerns whether and how users' goals are achieved. The pathway-level analysis focuses on characterizing and visualizing individual users' or their aggregated paths of site navigation. This framework has been successfully applied in the studies of users' information-seeking behavior in social library systems (Jiang, 2014) and academic library OPAC systems (Jiang *et al.*, 2017).

Methodology

Data collection

This study obtained a 4.23-gigabyte server log file from an institutional news site that affiliates to a renowned university in China. The log file contains a total of 39,990,200 clickstream records generated between March 1, 2017 and April 30, 2017. It was reduced to six basic fields, i.e. User-IP (client IP address, e.g. "202.114.65.*"), Date (date on which request is made, e.g. "28/Mar/2017"), Time (time when request is made, e.g. "08:00:10"), Method (type of client to server request, e.g. "GET"), URL (URL of the resource requested, e.g. "/info/1002/40929.htm") and Status (HTTP status code returned by server, e.g. "200"), which were useful for analyzing users' news selection behavior.

During the two-month time period, a Web crawler tool was meanwhile used to scrape the homepage of the news site once a day at 23:00, for the purpose of capturing the layout of the homepage and its daily update of content. The crawled data helped the researchers characterize how news headlines were presented on the homepage. Each news article had a distinct URL which could be used to identify in the log file all the records generated as a result of users opening it by clicking on the headlines. The clickstream data and crawled data were processed separately and then linked together via URLs in data analysis.

Clickstream data preparation

Data cleaning. The four primary steps of data cleaning include: sorting all the fields in sequence and deleting the errors that appear on the top of, bottom of, or grouped in each sorted field; deleting unsuccessful request records whose status codes (Status field) do not belong to the 200 class, e.g. 400 (not found) and 500 (internal server error); deleting data submission records whose request methods (Method field) are "POST"; and deleting redundant records whose URLs (URL field) end with ".jpg," ".css" and ".js," etc. (requests for pictures, styles, scripts and other resources).

Data parsing. Each session is composed of all the records deriving from one visit to the site, and a user may have more than one session for visiting the site multiple times during the two months. So different users were identified with their IP addresses (User-IP field); for the same user, if the time interval between any two records exceeded 30 min, they would be divided into different sessions. Given that some visits recorded in the log file might be attributed to search engine spiders, a cut-off of 101 records was adopted to determine non-human sessions: if a session contained 101 or more records, it was assumed to represent a non-human visit and thus excluded. As a result of data cleaning and parsing, approximately 10 percent ($n = 3,987,030$) of the records remained, which involved 839,685 sessions belonging to 284,685 distinct users.

Data coding. The final preparation was to translate the clickstream data into meaningful page requests. A coding system was created based on the simple information architecture of the news site which is composed of four major types of pages: the homepage (H) devotes a whole page to display hundreds of news headlines divided into different blocks; a news category page (C) provides a list of news headlines under a particular category, e.g. people, academic and alumni, etc.; a search result page (S) returns a list of news headlines in response to a user query; and a news article page (A) shows the full article of a news story and 8 related news headlines below the article. As in Table I, the type of page was identified with the specific string in its URL, and each page type was assigned a particular code. A code was appended to each record in the log file according to the URL field, which was completed by a Python program.

Crawled data preparation

As mentioned above, the crawled data mainly depicted the presentation of news headlines on the homepage of the institutional news site. Taking into account the site's own interface

and information design, the researchers decided to identify the characteristics of headline presentation from seven dimensions: a headline's position on the homepage and its format, the length of a headline (count of characters), the use of numbers (Arabic numbers) and the use of punctuation marks (Chinese punctuation marks) in a headline, and the recency (time of publishing) and popularity (frequency of accessing) of the corresponding news story. Table II shows the categories under these dimensions, but the basis for categorization will be explicated in the next section.

Evidence from
unobtrusive
data analysis

Results

General usage of the institutional news site

Since users' navigation patterns were not the focus of this study, a clickstream data analysis was performed only at the footprint and movement levels. A footprint is left on a page upon it being viewed. During the two months, a total of 3,987,030 footprints were created within the institutional news site, with an average of 4.75 footprints per session or visit. The distribution of these footprints among the four major page types can be found in Figure 1. News article pages (A) attracted 72.97 percent of the footprints, suggesting that users visited the site mainly for reading news stories. News category pages (C, 15.17 percent), the homepage (H, 11.75 percent), and search result pages (S, 0.11 percent) explained different proportions of the footprints. When browsing news categories, users were most interested in such categories as Important (2.02 percent), Media (1.13 percent), Academic (0.88 percent), General (0.77 percent) and Headline (0.73 percent).

Page type	Code	URL string
Homepage	H	/index.htm OR /titlelist.htm
News category page	C	/ttxw.htm, /wdyw.htm, /zhxw.htm, /mtwd.htm, /ljrjw.htm, /ztbd.htm, /wdsp.htm, /kydt.htm, /bfxy.htm, /djpx.htm, /gjil.htm, /xyzs.htm, /ljlt.htm, /xsgc.htm, /ljyx.htm, /ljfk.htm, /xwrx.htm, OR /lgxd.htm (each string represents a news category)
Search result page	S	/search.jsp?
News article page	A	/info/xxxx/xxxxx.htm ("x" represents an Arabic number)

Table I.
Clickstream data
coding system

Dimension	Category	Description
Position	Zones 1–8	Refer to Figure 5 for the segmentation of zones
Format	Text-only	Textual title without image
	Mixed	Text and image displayed side by side
	Image-dominated	Image with small caption in the carousel
Text length	Short	3–10 characters
	Medium	11–18 characters
	Long	19–26 characters
Use of numbers	Inclusion	Contains at least one Arabic number
	Exclusion	Contains no Arabic number
Use of punctuation marks	Inclusion	Contains at least one of the punctuation marks “【】,”“(),” and “《》”
	Exclusion	Contains no punctuation mark
Recency	High	Article published within 7 days
	Low	Article published more than 7 days ago
Popularity	High	Article read more than 1,000 times
	Low	Article read 0–1,000 times

Table II.
A framework for
characterizing the
presentation of
news headlines

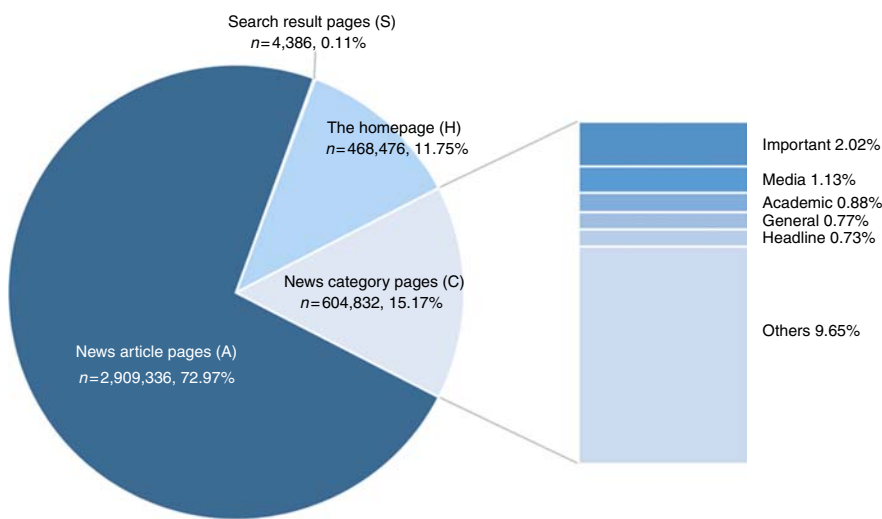


Figure 1.
Distribution of
footprints among
different page types

A physical footprint left on the ground can be deep or shallow. This study defined footprint depth as the time spent viewing a page, an indicator of the user's involvement in interacting with the page. The mean footprint depths differed greatly among the four types of pages, as shown in Figure 2. On average, users spent a little more than 2 min on a news article, which is reasonable given that most articles on the site were 500 to 800 Chinese characters in length. However, they stayed about four times longer on search result pages than on news category pages, although the former displayed much less information than the latter, i.e., 10 and 25 headlines on each page, respectively. Unexpectedly, the homepage occupied users only for an extremely short duration (3.80 s) despite the abundance of news headlines aggregated on it.

The footprint level analysis then concentrated on core footprints. As news article pages provide the news stories users intend to read, they are the destination pages in this site. Core footprints derived from users opening news article pages and were identifiable in the log file with the records appended the "A" code. Overall, the 2,909,336 core footprints involved 179,484 distinct users and 18,303 distinct news articles. Among the 284,685 users who visited the site during the two months, 63.05 percent of them accessed at least one news article.

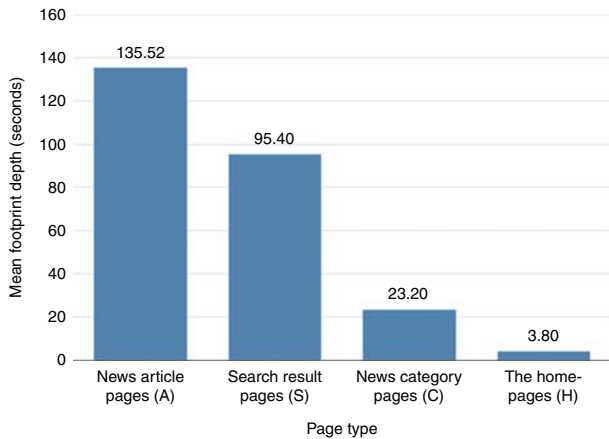


Figure 2.
Mean footprint depth
of each page type (i.e.
time spent on page
measured in second)

As the site has accumulated roughly 27,000 news articles (estimated by adding the numbers in all news categories), around two-thirds of them were accessed at least once. Figures 3(a) and 3(b) are the log-log scale plots of the distributions of core footprints among users and news articles, respectively. Each distribution fits a power law for being linear. In other words, there existed a few active users accessing many news articles and a few popular news articles being accessed many times, while low-activity users and low-popularity news articles were in the majority.

The analysis of core footprints was followed by that of pivotal movements which engendered core footprints, i.e., the movements leading users to news article pages. The four types of pages in the site all give access to news articles via news headlines, as described above. A Python program was used to capture the footprint preceding each core footprint in the log file and identify its type. The results were visualized in a Sankey diagram using the visual analytics tool BDP (<https://me.bdp.cn/home.html>) (Figure 4). The widths of the flow branches are proportional to the frequencies of different pivotal movements. Surprisingly, 70.58 percent of the core footprints originated from external sources. That is, the users arrived at the news article pages by following direct links, e.g., links shared on social media. However, the four types of pages still played their roles in helping users find news articles of interest.

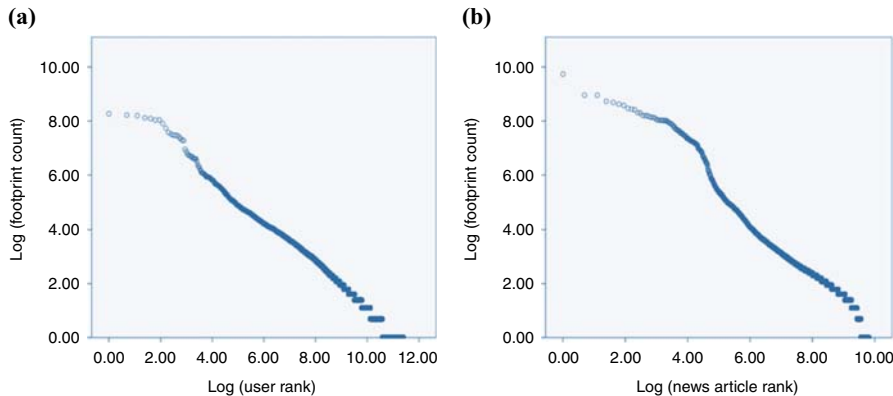


Figure 3. Distribution of core footprints among (a) users and (b) news articles

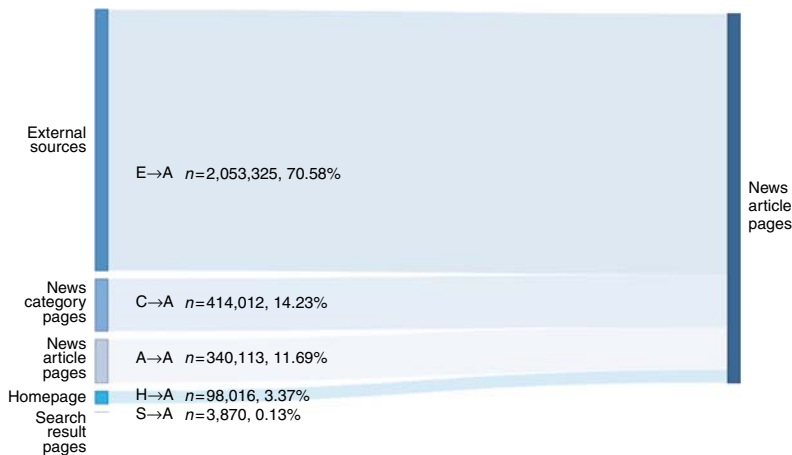


Figure 4. Sankey diagram of the core footprint analysis results

News category pages (14.23 percent) contributed a little more views of news articles than news article pages (11.69 percent), while the homepage only 3.37 percent and search result pages (0.13 percent) almost ignorable.

Relationships between news headline presentation and users' clicking behavior

The following analysis concentrated on the news headline selection on the homepage. The selection of headlines on other three types of pages would be more or less affected by their relatedness to the specific category, search query or news story, thus falling into the scope of content-based selection. Instead, the homepage mixes a large number of headlines on various topics and is more likely to afford form-based selection. The homepage of the institutional news site demonstrates a traditional layout with a global navigation bar on the top and news headlines enclosed into blocks below it. These blocks correspond to different news categories. Newly published headlines will edge out older ones from the blocks. The most vibrant blocks (e.g. Important and Media) will be updated daily, while some others (e.g. People and Alumni) weekly or even monthly. According to the crawled data, the total numbers of news headlines on the homepage varied slightly every day, ranging from 115 to 120. There was an overlap between the headlines on different days. Nevertheless, even if a headline stayed on the homepage for days, its presentation might keep changing. For example, its position would shift down gradually; its format might be reduced from mixing text and image to text only, and its recency would decrease, and popularity might increase.

Therefore, this study treated every headline on a single day as a distinct headline. The number of clicks on each distinct headline was extracted from the clickstream data. All the distinct headlines that ever appeared on the homepage during the two months added up to 7,120. They attracted 98,016 clicks in total. The following analysis examined the relationships between the presentation of news headlines and users' clicking behavior. The Kruskal-Wallis H test was adopted for the three dimensions with three or more categories, i.e., position, format and length, while the Mann-Whitney U test for the use of numbers, use of punctuation marks, recency and popularity which each has two categories.

Position. For the convenience of analysis, the blocks on the homepage were further divided into eight zones. As seen in Figure 5, the homepage was partitioned into three screens vertically. The first screen was what users saw when they initially opened the homepage, and the second and last screens would be revealed in sequence if they scrolled down. The horizontal partition was based on the intrinsic boundaries in the layout. Zones 1 and 2 were situated, respectively, at the left and right side of the first screen. The second screen comprised Zones 3, 4 and 5, and the last screen Zones 6, 7 and 8. This study found that these zones attracted significantly different numbers of clicks ($H=2,462.933$, $p=0.000$). According to *post hoc* comparisons, significant differences existed in all pairs of the zones except for Zones 3 and 5, Zones 3 and 6, and Zones 5 and 6. Users' clicking was the most likely to occur in the first screen, including Zones 1 (mean rank = 6,574.62) and 2 (mean rank = 5,736.10) and the middle of the second screen, i.e., Zone 4 (mean rank = 4,374.58).

Format. Although purely textual news headlines dominate the homepage ($n=6,680$), there are also a handful of images scattering in different zones. A conspicuous carousel occupies Zone 2, displaying five images rotationally and each with a line of small text on the top. In some of the news categories, such as people and special, images are provided side by side with certain headlines, probably for adding visual richness. Both image-dominated headlines ($n=328$) and mixed ones ($n=112$) were in the minority. Significant differences in clicking behavior were found among the headlines in different formats ($H=414.553$, $p=0.000$). Image-dominated headlines received more clicks than both mixed ($U=1,636.223$, $p=0.000$) and text-only ones ($U=2,297.615$, $p=0.000$), and mixed headlines more than text-only ones ($U=661.392$, $p=0.002$).



Figure 5.
The eight zones on
the homepage of the
institutional news site

Text length. As extracted from the crawled data, the lengths of the news headlines published on the homepage during the two months ranged between 3 and 26 characters. This study defined three groups of headlines of different lengths with tertiles: short headlines (3–10 characters), medium headlines (11–18 characters) and long headlines (19–26 characters). Most of the headlines fell into the medium group ($n = 4,448$), while short ($n = 1,656$) and long headlines ($n = 1,016$) were less frequently seen. The three groups of headlines aroused significantly different numbers of clicks ($H = 436.307$, $p = 0.000$). Long headlines were clicked on more frequently than both medium ($U = 1,120.347$, $p = 0.000$) and short ones ($U = 1,672.290$, $p = 0.000$). Also, medium headlines were clicked on more frequently than short ones ($U = 551.943$, $p = 0.000$).

Use of numbers or punctuation marks. The headlines in the institutional news site were basically written in Chinese characters, though Arabic numbers and punctuation marks also used at times. Arabic numbers only appeared in 18.43 percent of the headlines ($n = 1,312$) mainly as years, with a few rankings, percentages, etc. There exists no significant difference in users' clicking behavior between the headlines, including and those excluding Arabic numbers ($Z = -6.771$, $p = 0.742$). The most common punctuation marks in the headlines included “【】” (square brackets), “()” (round brackets) and “《》” (guillemets) from the

Chinese punctuation system. The square brackets ($n = 1,664$) were often used to enclose news sources or categories at the beginning of headlines, while the round brackets ($n = 520$) supplementary content at the end of headlines. The guillemets ($n = 160$) helped indicate the titles of books, papers, journals and other documents. Significant results were obtained for the square brackets ($Z = -9.025$, $p = 0.000$) and round brackets ($Z = -8.864$, $p = 0.000$), but not for the guillemets ($Z = -2.475$, $p = 0.130$). More clicks were engendered with the inclusion of square brackets (mean rank: inclusion 3,950.87 > exclusion 3,441.44) or round brackets (mean rank: inclusion 4,314.90 > exclusion 3,501.06) in the headlines.

Recency or popularity. Two of the blocks, i.e., important and media, additionally provided recency and/or popularity indicators after each news headline. The important block displayed both the date of publishing and the number of reads, and the media block only the former. All the distinct headlines that were ever enclosed in both blocks during the two months added up to 888, with 488 in important and 400 in media, respectively. This study determined two levels of news recency based on the mean recency (around seven days): high (published within seven days) and low (published more than seven days ago). The mean popularity (around 1,000 reads) was used to distinguish two levels of news popularity: high (read 1,000 times and more) and low (read less than 1,000 times). Highly recent headlines ($n = 698$, 78.60 percent) and highly popular ones ($n = 312$, 63.93 percent) were in the majority. The two groups of headlines of different recency ($Z = -15.366$, $p = 0.000$) demonstrated a significant difference in the number of clicks, so did those of different popularity ($Z = -17.889$, $p = 0.000$). Users were more likely to click on high-recency headlines (mean rank: high 513.50 > low 191.00) or high-popularity ones (mean rank: high 396.50 > low 158.76).

Discussion

News selection on the institutional news site

The research context of this study is an institutional news site which reports exclusively news events happening to the university and mainly targets its faculty, students, administration, and alumni, etc. According to the clickstream data analysis, 72.97 percent of all the footprints were left on news article pages. It can be inferred that this institutional news site was much like the electronic version of a print newspaper which mainly supported traditional linear news consumption, i.e., article by article. Visiting the site had become a daily routine to keep informed about current events in the university, mainly for working and/or educational purposes.

Given that 70.58 percent of these news article pages were opened by following external links, the site should pay attention to the underuse of its own navigation system. In particular, the role of within-site search tools was almost ignorable, as indicated by the rarity of footprints on search result pages (0.11 percent) and the infrequency of accessing news articles via search (0.13 percent). The movement level analysis indicates that 25.92 percent of the news articles were accessed from news category pages and news article pages. Users could discover older news when they browsed by category or followed related recommendations. Because two-thirds of all the news articles on the site were accessed during the two months, users did have an interest in obsolete news.

This study was the most interested in the homepage of the site. It was responsible for displaying the most recent and popular news headlines, however playing an insignificant part in directing users to news articles (3.37 percent). The average 3.8-s stay on the homepage also prevented users from spotting many news headlines. Such results contradict to a certain degree the previous finding that online news consumers preferred serendipitous exposure to headlines “pushed” onto them on news portals or aggregators (Pentina and Tarafdar, 2014). A possible explanation for the unattractiveness of this homepage is the text-heavy “link-rich” design made the page crowded and visually unappealing (Spencer, 2010).

Implications for presenting news headlines

Given the extremely short stay on the homepage of the institutional news site, the clicks on a headline were more likely to be driven by its external presentation than users' considered judgment. Among the seven characteristic dimensions of headline presentation examined in this study, except for the use of Arabic numbers and the use of guillemets, all others showed significant differences in clicking behavior, based on the joint analysis of the clickstream data and crawled data.

News headlines placed in conspicuous positions or presented with images were found to be clicked on more frequently. A little different from the F-shaped pattern for general webpage browsing (Nielsen, 2006), a T-shaped pattern was detected for this homepage where the first screen and the middle of the second screen attracted the most clicks. Zone 4 stood out probably for containing the Media block which reposted the university news reported on mainstream media. In regard to format, the neglect of text-only headlines implies users' preference for images. It is because the perception of images involves an automatic, parallel, fast and effortless process while written words a voluntary, serial, slow and effortful process (Flores *et al.*, 2014). The "dual coding" theory (Yang *et al.*, 2017) also applies to the presentation of news headlines.

Although shorter titles have been found more effective for customer reviews (Salehan and Dan, 2016) and research articles (Subotic and Mukherjee, 2014), the users of this institutional news site were less interested in shorter headlines. It could be more difficult to provide necessary information for users to predict what the news stories are about with fewer words. Whether a headline contained numbers had no impact on users' clicking, despite the role of numbers in increasing credibility, as mentioned above. This may be explained by the fact that most numbers appeared as years which were not deemed an informative element. As for punctuation marks, only the inclusion of square brackets or round brackets was associated with more clicks. The former ("【】") are boldfaced symbols that are much thicker and heavier than the surrounding text; and the latter ("（ ）") are made up of two curves which easily stand out from straight-line Chinese characters. In contrast, guillemets ("《》") exhibit less eye-catching shapes. They were probably deemed an organic component of a headline for used exclusively to enclose titles of books and other publications.

When confronted with a tremendous amount of information, people are apt to avoid extra efforts and rely more on peripheral cues to make a selection. Recency and popularity indicators are widely seen on news sites. These indicators are useful cues that help users evaluate information quality, especially when they are involved in casual scanning (Xu, 2013). As found in this study, more recent or more popular news headlines attracted more clicks. Inherently important to news, recency determines its value by affecting users' trust (Westerman *et al.*, 2014). The pursuit of popular news has its psychological root. It reflects the "bandwagon effect," which suggests that individuals tend to believe something when many other people also believe it (Lee and Sundar, 2013).

The above findings have valuable implications for news providers in attracting clicks to their headlines. Nevertheless, the researchers have a concern about the potential risks of misusing news headline presentation techniques. Clickbaits have been created by constructing extremely long headlines (Chakraborty *et al.*, 2016) or overusing certain punctuation marks (Noguti, 2016). Some news providers have even made use of computer-written news articles to keep a high updating frequency to allure users into addictive news reading (Van der Kaa and Krahmer, 2014). Visual prominence might be added to selected news deliberately with superiority in position and/or format. Responsible news providers should strive every effort to achieve authenticity and objectivity, both of which contribute to news trustworthiness (Gunter *et al.*, 2009). Therefore, it is unwise to mix normal news with clickbaits, advertisements, or duplicates, which will soon cause users' disgust and expel them instead. The "bandwagon effect" needs to be controlled appropriately as it may lock

users in information cocoons (Zuiderveen *et al.*, 2016). If popular news contains bias, the negative effects of the bias can be strengthened as the news gets more popular. This is detrimental to the diversification of knowledge, opinions and interests.

The pros and cons of unobtrusive data

This study introduced two types of unobtrusive data, i.e., the clickstream data from the Web server and the data crawled from the homepage. The former was generated as a result of users visiting the institutional news site, while the latter reflected the intrinsic characteristics of headline presentation. The collection of both types of data was preceded by and thus had no interference with the interaction between users and the site. In this sense, nonreactive methods are superior to experiments that lack naturalness as well as self-report methods that rely on the participant to provide data. Besides, the clickstream data involved in this study was incomparable in size. The original 39,990,200 clickstream records were processed and then analyzed to extract 98,016 clicks on the headlines. The results were obtained based on the population rather than a sample during the two months. As March and April belong to the spring semester, the researchers plan to request a long-term (e.g. one-year) logfile from the site to see whether users' behavior would change over time.

Although ubiquitous on Web servers, clickstream data are still underused in the studies of information behavior. It can be attributed to some inherent limitations of such data. First and foremost, academic researchers have no easy access to clickstream data. Websites are often very cautious about releasing their server logs for fear of offending users' privacy. The generalizability of the findings of this study could be enhanced if clickstream data were collected from mainstream news sites. Second, clickstream data analysis is useful in revealing users' behavioral patterns while leaving what causes their behavior uninterpreted. Complementary methods, e.g., surveys and interviews, are needed to explore why users click on a link and what they actually do after opening a page (Jansen, 2009).

Conclusions

Online news consumers have increasingly depended on form-based news selection to cope with information overload. This study investigated the relationships between news headline presentation and users' news selection behavior based on unobtrusive data. It was found that the differences in headline position, format, length, use of punctuation marks (square brackets or round brackets), recency, and popularity, aroused significantly different numbers of clicks on the headlines, respectively.

The theoretical contribution of this study is twofold. On the one hand, it provided a comprehensive overview of the "form" of online news headlines by taking into account seven characteristic dimensions of presentation in the same study. These dimensions constitute a framework of the formal considerations for evaluating headlines. On the other hand, this study filled the gap in audience news selection research by probing into the highest level of selection, i.e., selective retention, while former studies confined to selective attention and perception. Clicking is the most straightforward behavioral manifestation of news selection. It will be interesting to explore how behavior is affected by attention and perception in the process of news selection.

This study also presents practical implications for news providers. Headlines are entries to news stories on the internet. A series of presentation techniques were obtained for creating effective online news headlines, including placing headlines in the central T-shaped zones, using images, increasing text length properly for greater clarity, using visually distinctive punctuation marks, and providing recency and popularity indicators. Taking advantage of these techniques in a responsible fashion, news providers can encourage news consumption while alleviating news consumers' information overload, which is key to the vitality of the news media industry.

Furthermore, the acquirement of the large-scale logfile from the institutional news site enabled the researchers to demonstrate a new way of examining users' interaction with news headlines. There is no denying that this study is also limited by the context-specific data, and the findings may not be applicable to other populations. However, the advocacy of nonreactive methods has significant implications for future user studies in both information science and journalism.

References

- Adegoju, A. and Ademilokun, M. (2015), "Verbal and visual signifiers of advertising shares offers in Nigeria's 2005 bank recapitalisation", *Discourse & Communication*, Vol. 9 No. 5, pp. 519-533.
- Arapakis, I., Lalmas, M., Cambazoglu, B.B., Marcos, M.C. and Jose, J.M. (2014), "User engagement in online news: under the scope of sentiment, interest, affect, and gaze", *Journal of the Association for Information Science & Technology*, Vol. 65 No. 10, pp. 1988-2005.
- Atkinson, G., Driesener, C. and Corkindale, D. (2014), "Search engine advertisement design effects on click-through rates", *Journal of Interactive Advertising*, Vol. 14 No. 1, pp. 24-30.
- Blanco, C.F., Sarasa, R.G. and Sanclemente, C.O. (2010), "Effects of visual and textual information in online product presentations: looking for the best combination in website design", *European Journal of Information Systems*, Vol. 19 No. 6, pp. 668-686.
- Blom, J.N. and Hansen, K.R. (2015), "Click bait: forward-reference as lure in online news headlines", *Journal of Pragmatics*, Vol. 76, pp. 87-100.
- Broeck, E.V.D., Poels, K. and Walrave, M. (2018), "An experimental study on the effect of ad placement, product involvement and motives on Facebook ad avoidance", *Telematics and Informatics*, Vol. 35 No. 2, pp. 470-479.
- Bucher, H.J. and Schumacher, P. (2006), "The relevance of attention for selecting news content. An eye-tracking study on attention patterns in the reception of print and online media", *Communications*, Vol. 31 No. 3, pp. 347-368.
- Chakraborty, A., Paranjape, B., Kakarla, S. and Ganguly, N. (2016), "Stop clickbait: detecting and preventing clickbaits in online news media", paper presented at the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, San Francisco, CA.
- Chang, Y.T., Yu, H. and Lu, H.P. (2015), "Persuasive messages, popularity cohesion, and message diffusion in social media marketing", *Journal of Business Research*, Vol. 68 No. 4, pp. 777-782.
- Cheung, R. (2014), "The influence of electronic word-of-mouth on information adoption in online customer communities", *Global Economic Review*, Vol. 43 No. 1, pp. 42-57.
- Chi, Y., Jiang, T., He, D. and Meng, R. (2017), "Towards an integrated clickstream data analysis framework for understanding web users' information behavior", paper presented at the iConference 2017 Proceedings, Wuhan.
- Chin, C.Y., Lu, H.P. and Wu, C.M. (2015), "Facebook users' motivation for clicking the 'Like' button", *Social Behavior & Personality An International Journal*, Vol. 43 No. 4, pp. 579-592.
- Choi, W. and Stvilia, B. (2015), "Web credibility assessment: conceptualization, operationalization, variability, and models", *Journal of the Association for Information Science & Technology*, Vol. 66 No. 12, pp. 2399-2414.
- Chtourou, M.S., Chandon, D.J.L. and Zollinger, M. (2002), "Effect of price information and promotion on click-through rates for internet banners", *Journal of Euromarketing*, Vol. 11 No. 2, pp. 23-40.
- Donsbach, W. (2004), "Psychology of news decisions: factors behind journalists' professional behavior", *Journalism*, Vol. 5 No. 2, pp. 131-157.
- Edgerly, S., Vraga, E.K., McLaughlin, B., Alvarez, G., Yang, J. and Kim, Y.M. (2014), "Navigational structures and information selection goals: a closer look at online selectivity", *Journal of Broadcasting & Electronic Media*, Vol. 58 No. 4, pp. 542-561.

- Flores, W., Chen, J.C.V. and Ross, W.H. (2014), "The effect of variations in banner ad, type of product, website context, and language of advertising on Internet users' attitudes", *Computers in Human Behavior*, Vol. 31 No. 1, pp. 37-47.
- Fumani, M.R.F.Q., Goltaji, M. and Parto, P. (2015), "The impact of title length and punctuation marks on article citations", *Annals of Library & Information Studies*, Vol. 62 No. 3, pp. 126-132.
- Go, E., Jung, E.H. and Wu, M. (2014), "The effects of source cues on online news perception", *Computers in Human Behavior*, Vol. 38, pp. 358-367.
- Gunter, B., Campbell, V., Touri, M. and Gibson, R. (2009), "Blogs, news and credibility", *Aslib Proceedings*, Vol. 61 No. 2, pp. 185-204.
- Hirotnani, M., Frazier, L. and Rayner, K. (2006), "Punctuation and intonation effects on clause and sentence wrap-up: evidence from eye movements", *Journal of Memory & Language*, Vol. 54 No. 3, pp. 425-443.
- Holsanova, J., Rahm, H. and Holmqvist, K. (2006), "Entry points and reading paths on newspaper spreads: comparing a semiotic analysis with eye-tracking measurements", *Visual Communication*, Vol. 5 No. 1, pp. 65-93.
- Holton, A.E. and Chyi, H.I. (2012), "News and the overloaded consumer: factors influencing information overload among news consumers", *Cyberpsychology, Behavior, and Social Networking*, Vol. 15 No. 11, pp. 619-624.
- Huang, Y.K. and Yang, W.I. (2010), "A study of internet book reviews and borrowing intention", *Library Review*, Vol. 59 No. 7, pp. 512-521.
- Jansen, B.J. (2009), *Understanding User-Web Interactions via Web Analytics*, Morgan and Claypool Publishers, San Rafael, CA.
- Jansen, B.J., Liu, Z. and Simon, Z. (2014), "The effect of ad rank on the performance of keyword advertising campaigns", *Journal of the American Society for Information Science & Technology*, Vol. 64 No. 10, pp. 2115-2132.
- Jiang, T. (2014), "A clickstream data analysis of users' information seeking modes in social tagging systems", paper presented at the iConference 2014 Proceedings, Berlin.
- Jiang, T., Chen, S. and Gao, H. (2017), "Visualizing users information seeking pathways on OPAC of academic library", *Journal of Academic Library*, Vol. 35 No. 1, pp. 63-71.
- Katerina, E.M. and Elisa, S. (2018), "News use across social media platforms 2018", available at: www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/ (accessed March 12, 2019).
- Knobloch-Westerwick, S., Dillman Carpentier, F., Blumhoff, A. and Nickel, N. (2005), "Informational utility effects on selective exposure to good and bad news: a cross-cultural investigation", *Journalism and Mass Communication Quarterly*, Vol. 82 No. 1, pp. 181-195.
- Knobloch-Westerwick, S., Sharma, N., Hansen, D.L. and Alter, S. (2005), "Impact of popularity indications on readers' selective exposure to online news", *Journal of Broadcasting & Electronic Media*, Vol. 49 No. 3, pp. 296-313.
- Ksiazek, T.B., Peer, L. and Lessard, K. (2016), "User engagement with online news: conceptualizing interactivity and exploring the relationship between online news videos and user comments", *New Media & Society*, Vol. 18 No. 3, pp. 261-270.
- Kuiken, J., Schuth, A., Spitters, M. and Marx, M. (2017), "Effective headlines of newspaper articles in a digital environment", *Digital Journalism*, Vol. 5 No. 10, pp. 1300-1314.
- Kuisma, J., Simola, J., Uusitalo, L. and Öörni, A. (2010), "The effects of animation and format on the perception and memory of online advertising", *Journal of Interactive Marketing*, Vol. 24 No. 4, pp. 269-282.
- Lai, L. and Farbroth, A. (2014), "What makes you click? The effect of question headlines on readership in computer-mediated communication", *Social Influence*, Vol. 9 No. 4, pp. 289-299.
- Lee, J.Y. and Sundar, S.S. (2013), "To tweet or to retweet? That is the question for health professionals on twitter", *Health Communication*, Vol. 28 No. 5, pp. 509-524.

- Lee, K. (2014), "The proven ideal length of every tweet, Facebook post, and headline online: fast company", available at: www.cubanxgiants.com/berry/250/documents/twitter_lengths.pdf (accessed March 12, 2019).
- Liu, Y., Yuan, P., Liu, W. and Li, X. (2015), "What drives click-through rates of tourism product advertisements on group buying websites?", *Procedia Computer Science*, Vol. 55, pp. 221-230.
- Ma, Y. and Chen, J. (2017), "Effects of platform and content attributes on information dissemination on we media: a case study on WeChat platform", paper presented at the Wuhan International Conference on e-Business, Wuhan.
- Metzger, M.J. (2007), "Making sense of credibility on the web: models for evaluating online information and recommendations for future research", *Journal of the Association for Information Science & Technology*, Vol. 58 No. 13, pp. 2078-2091.
- Morris, C.W. (1964), *Signification and Significance*, Mass The MIT Press, Cambridge.
- Mou, J. and Shin, D. (2017), "Effects of social popularity and time scarcity on online consumer behavior regarding smart healthcare products: an eye-tracking approach", *Computers in Human Behavior*, Vol. 78, pp. 74-89.
- Nielsen, J. (2006), "F-Shaped pattern for reading web content", available at: www.nngroup.com/articles/f-shaped-pattern-reading-web-content/ (accessed March 12, 2019).
- Noguti, V. (2016), "Post language and user engagement in online content communities", *European Journal of Marketing*, Vol. 50 Nos 5/6, pp. 695-723.
- Park, J., Lennon, S.J. and Stoel, L. (2005), "On-line product presentation: effects on mood, perceived risk, and purchase intention", *Psychology & Marketing*, Vol. 22 No. 9, pp. 695-719.
- Penna, L. and Quaresma, M. (2015), "How we perceive search engines", paper presented at the International Conference of Design, User Experience, and Usability, Los Angeles, CA.
- Pentina, I. and Tarafdar, M. (2014), "From 'information' to 'knowing': exploring the role of social media in contemporary news consumption", *Computers in Human Behavior*, Vol. 35, pp. 211-223.
- Rony, M.M.U., Hassan, N. and Yousuf, M. (2017), "Diving deep into clickbaits: who use them to what extents in which topics with what effects?", paper presented at the IEEE/ACM International Conference on Advances in Social Networks Analysis & Mining, July 31-August 03, 2017, Sydney.
- Rubin, V., Conroy, N., Chen, Y. and Cornwell, S. (2016), "Fake news or truth? Using satirical cues to detect potentially misleading news", paper presented at the Proceedings of the second workshop on computational approaches to deception detection, June 17, 2016, San Diego, California.
- Safran, N. (2013), "Data insights into the headlines readers click", available at: www.immagic.com/eLibrary/ARCHIVES/GENERAL/BNILE_US/B130717S.pdf (accessed March 10, 2019).
- Salehan, M. and Dan, J.K. (2016), "Predicting the performance of online consumer reviews: a sentiment mining approach to big data analytics", *Decision Support Systems*, Vol. 81 No. C, pp. 30-40.
- Sen, A., Dacin, P.A. and Pattichis, C. (2006), "Current trends in web data analysis", *Communications of the ACM*, Vol. 49 No. 11, pp. 85-91.
- Shie, J.S. (2011), "Metaphors and metonymies in New York Times and Times supplement news headlines", *Journal of Pragmatics*, Vol. 43 No. 5, pp. 1318-1334.
- Song, S. and Kim, M. (2012), "Does more mean better? An examination of visual product presentation in e-retailing", *Journal of Electronic Commerce Research*, Vol. 13 No. 4, pp. 345-355.
- Spencer, D. (2010), *A Practical Guide to Information Architecture*, Five Simple Steps, Penarth.
- Subotic, S. and Mukherjee, B. (2014), "Short and amusing: the relationship between title characteristics, downloads, and citations in psychology articles", *Journal of Information Science*, Vol. 40 No. 1, pp. 115-124.
- Tenenboim, O. and Cohen, A.A. (2013), "What prompts users to click and comment: a longitudinal study of online news", *Journalism*, Vol. 16 No. 2, pp. 198-217.
- Valentini, C., Romenti, S., Murtarelli, G. and Pizzetti, M. (2018), "Digital visual engagement: influencing purchase intentions on Instagram", *Journal of Communication Management*, Vol. 22 No. 4, pp. 362-381.

- Van der Kaa, H.A.J. and Krahmer, E.J. (2014), "Journalist versus news consumer: the perceived credibility of machine written news", *British Medical Journal*, Vol. 2 No. 5147, p. 305.
- Wagner, C., Rowe, M., Strohmaier, M. and Alani, H. (2012), "What catches your attention? An empirical study of attention patterns in community forums", paper presented at the International Conference on Weblogs and Social Media, Dublin.
- Wang, S.M. and Lin, C.C. (2011), "The effect of social influence on the bloggers' usage intention", *Online Information Review*, Vol. 35 No. 1, pp. 50-65.
- Wanick, V., Stallwood, J., Ranchhod, A. and Wills, G. (2018), "Can visual familiarity influence attitudes towards brands? An exploratory study of advergame design and cross-cultural consumer behaviour", *Entertainment Computing*, Vol. 27, pp. 194-208.
- Welbers, K., Van Atteveldt, W., Kleinnijenhuis, J., Ruigrok, N. and Schaper, J. (2016), "News selection criteria in the digital age: professional norms versus online audience metrics", *Journalism*, Vol. 17 No. 8, pp. 1037-1053.
- Wendelin, M., Engelmann, I. and Neubarth, J. (2015), "User rankings and journalistic news selection: comparing news values and topics", *Journalism Studies*, Vol. 18 No. 2, pp. 135-153.
- West, S.L., Squiers, L.B., McCormack, L., Southwell, B.G., Brouwer, E.S., Ashok, M. and Sullivan, H.W. (2013), "Communicating quantitative risks and benefits in promotional prescription drug labeling or print advertising", *Pharmacoepidemiology & Drug Safety*, Vol. 22 No. 5, pp. 447-458.
- Westerman, D., Spence, P.R. and Heide, B.V.D. (2014), "Social media as information source: recency of updates and credibility of information", *Journal of Computer-Mediated Communication*, Vol. 19 No. 2, pp. 171-183.
- Xu, P., Chen, L. and Santhanam, R. (2015), "Will video be the next generation of e-commerce product reviews? Presentation format and the role of product type", Vol. 73, pp. 85-96.
- Xu, Q. (2013), "Social recommendation, source credibility and recency: effects of news cues in a social bookmarking website", *Journalism & Mass Communication Quarterly*, Vol. 90 No. 4, pp. 757-775.
- Yadamsuren, B. and Erdelez, S. (2016), "Incidental exposure to online news", in Marchionini, G. (Ed.), *Synthesis Lectures on Information Concepts, Retrieval, and Services*, Morgan and Claypool Publishers, San Rafael, California.
- Yang, S.B., Hlee, S., Lee, J. and Koo, C. (2017), "An empirical examination of online restaurant reviews on Yelp.com: a dual coding theory perspective", *International Journal of Contemporary Hospitality Management*, Vol. 29 No. 2, pp. 817-839.
- Yu, H. and Song, Z. (2017), "Picture-text congruence in translation: images of the Zen master on book covers and in verbal texts", *Social Semiotics*, Vol. 27 No. 5, pp. 604-623.
- Zebregs, S., Van, d.P.B., Neijens, P. and De, G.A. (2015), "The differential impact of statistical and narrative evidence on beliefs, attitude, and intention: a meta-analysis", *Health Communication*, Vol. 30 No. 3, pp. 1-8.
- Zuiderveen, B., Frederik, J., Trilling, D., Moeller, J., Bodó, B., De Vreese, C.H. and Helberger, N. (2016), "Should we worry about filter bubbles?", *Internet Policy Review*, Vol. 5 No. 1, pp. 1-16.

Further reading

- Jiang, T., Yu, C. and Jia, W. (2014), "Exploring users' within-site navigation behavior: a case study based on clickstream data", *Chinese Journal of Library & Information Science*, Vol. 7 No. 4, pp. 63-76.

Corresponding author

Tingting Jiang can be contacted at: tij@whu.edu.cn

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgroupublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.