

Extended Abstract: Don't Take the Bait: Users' Engagement with Clickbait and Its Effect on Editorial Considerations

The content overflow characterizing the information environment on social network sites (SNS) demands that news publishers compete for audience attention. As part of this competition, several tactics have been adopted by news editors (Molyneux & Coddington, 2020). Prominent among them is the *clickbait* headline (Bazaco, 2019; Beleslin et al., 2017; Mukherjee et al., 2019; Orosa, 2017). According to the Oxford English Dictionary, clickbait headlines are “(On the Internet) content whose main purpose is to attract attention and encourage visitors to click on a link to a particular web page,” (“Clickbait” in Oxford English Dictionary, n.d), usually by using catchy, exaggerated, exciting, and sensational headlines, in order to maintain and exploit users’ curiosity gap (Thiel, 2018).

There is evidence that clickbait’s use of attention-grabbing headlines does successfully entice visitors to click on links (Zhang et al., 2019). However, this practice is coming under increasing criticism from journalists, scholars and users. Journalists and scholars claim that clickbait headlines lead to the “death of journalism” (Dvorkin, 2016), as their only aim is to attract audiences towards vacuous texts, which has a negative effect on media quality and on public information norms and standards (Bazaco, 2019; Cable & Mottershead, 2018; Thurman & Myllylahti, 2009). Users equally dislike clickbait and find it deceptive (Beleslin et al., 2017; Chakraborty et al., 2016; Scacco & Muddiman, 2016). Furthermore, clickbait affects the public image and credibility of outlets who use them (Chakraborty et al., 2016; Molyneux & Coddington, 2020; Vultee et al., 2020). Thus, one could argue that publishers face a dilemma: clickbait seems as an effective tactic in the short run, but with questionable aftereffects in the long run (Zhang et al., 2019).

No research hitherto has tried to untangle the mutual longitudinal influence between audience engagement and editorial choices. Yet, limited findings suggest that users’ engagement with clickbait is indeed a complex phenomenon. In particular, Zhang et al. (2019) found contradicting patterns when

observing reactions to specific post and when observing traffic to websites. At the post level, clickbait gained attention. In contrast, an inverted U-shaped relationship emerges between the amount of clickbait created by a publisher and the traffic to the focal publisher. Zhang et al. (2019) speculated that the contradiction between users' actual behavior and perception could be attributed to a learning process: over time, users learn to identify and then refrain from engaging with clickbait.

Therefore, we first ask:

R1: Is the engagement of Facebook users with “clickbait” headlines lower than with non-clickbait headlines?

And, following Zhang et al. (2019) we hypothesize:

H1: Over the tested period, Facebook users' engagement with “clickbait” headlines will decrease.

Lastly, we argue that clickbait could serve as a case study through which the influence of audience engagement on publishers could be better understood. Can changes and adaptations be identified in both users and publishers? Do interactivity and learning processes occur? To this aim we examine publishers' behavior over time. If, as has been suggested by previous research (Blanchett-Neheli, 2018; Lischka, 2018; Tandoc, 2014; Tsuruel et al., 2019), publishers react to audience behavior as monitored by engagement matrices – user perceptions of clickbait, and their sanctions on publishers who use them – this may lead to changes in publishers' practices.

Therefore, due to both the increase in user dissatisfaction with clickbait and the importance news editors place on their audience's response to and engagement with their news content, we expect that:

H2: Over the tested period, news outlets' use of clickbait headlines on Facebook will decrease.

H3: We expect to find cross-lagged relations between users' (decreasing) engagement with clickbait headlines and the (decreasing) use of clickbait headlines by news outlets, such as: the decrease in engagement will be followed by a decrease in news outlets' use.

Method

Our analysis is based on a sample of posts from 35 English-speaking news outlets (see Table 1). Facebook posts, links to news articles, and related social media information were retrieved by data collection company NewsWhip, which monitors the social media activity of more than 50,000 publications worldwide (Kilgo et al., 2018; Mourao & Robertson, 2019).

For the present study, we only included Facebook posts which we could connect with their respective articles on the news outlets' websites: overall, approximately 1,048,575 post-article combinations, posted by news outlets over approximately three years (2017-2019).

In order to detect the clickbait in our data we followed the recent approach which relies on automatic tools to detect clickbait headlines (e.g. Biyani et al., 2016; Elyashar et al., 2017). We adopted a clickbait detection algorithm which was submitted as part of the Clickbait Challenge 2017 (Potthast et al., 2018). The algorithm was further validated using four human coders, ensuring high validity of the automatic processing step (for further elaborations see the appendix).

Audience engagement was measured based on numbers of Likes, comments, and shares for each post, as captured by Newswhip data.

In addition, our analysis controlled for time of day in which the post was posted, and the characteristics of the publisher (TV news, digitally-born website-only media outlets, legacy news media with both website and print, radio, news aggregators).

Preliminary Results

Our preliminary results show that the level of engagement with clickbait posts is lower across all measures (likes, comments, and shares) and that it decreases over time. However, the differences over time, although significant, are rather small. Clickbaits are only responsible for a 5% change.

In addition, the results show a decrease in the probability of a headline being clickbait over time.

Lastly, we expected to find a cross-lagged relation between users' (decreasing) engagement with clickbait headlines and the (decreasing) use of clickbait headlines by news outlets. Here we found very robust results: engagement with clickbait headlines predicted the percentage of clickbait published in the following month.

Table 1*List of Outlets in Our Data*

Outlet	First Day	Last Day	Total Days	Total Posts	Primary Medium	Format (for Newspapers)	DB \ Legacy	Owner -ship	Aggrega -tor	News agency	Based country
dailymail	01/01/2017	08/12/2019	979	97027	Newspaper	Tabloid	legacy	Private	No	No	UK
metrouk	01/01/2017	07/12/2019	947	49116	Newspaper	Tabloid	Legacy	Private	No	No	UK
Theindependent-online	01/01/2017	07/12/2019	947	46897	Newspaper	Tabloid	Legacy	Private	No	No	UK
nytimes	01/01/2017	07/12/2019	947	46806	Newspaper	Broadsheet	Legacy	Private	No	No	USA
theguardian	01/01/2017	07/12/2019	945	45971	newspaper	Broadsheet	legacy	private	No	No	UK
telegraph.co.uk	01/01/2017	07/12/2019	947	45342	Newspaper	Broadsheet	Legacy	Private	No	No	UK
wsj	01/01/2017	05/12/2019	885	44968	Newspaper	Broadsheet	Legacy	Private	No	No	USA
irishtimes	01/01/2017	08/12/2019	979	43060	Newspaper	Broadsheet	legacy	Private	No	No	Ireland
irishtimes	01/01/2017	08/12/2019	978	39468	Newspaper	Broadsheet	Legacy	Private	No	No	Ireland
ctvnews	01/01/2017	08/12/2019	979	38116	TV		Legacy	Private	No	No	Canada
breakingnews	01/01/2017	08/12/2019	978	37247	Website		DB	Private	Yes	No	Ireland
nationalpost	01/01/2017	08/12/2019	969	34266	Newspaper	Broadsheet	Legacy	Private	No	No	Canada
cbcnews	01/01/2017	08/12/2019	952	31005	TV + Radio		Legacy	Public	No	No	Canada
cbsnews	01/01/2017	07/12/2019	566	23317	TV + Radio		legacy	Private	No	No	USA
theglobeandmail	01/01/2017	08/12/2019	970	22324	Newspaper	Broadsheet	Legacy	Private	No	No	Canada UK +
skynews	01/01/2017	07/12/2019	940	21659	TV		Legacy	Private	No	No	Ireland
cnn	01/01/2017	05/12/2019	662	20303	TV		Legacy	Private	No	No	USA
buzzfeed	01/01/2017	08/12/2019	963	18352	Website		DB	Private	Yes	No	USA
bbc	01/01/2017	08/12/2019	964	15838	TV		Legacy	Public	No	No	UK
mashable	01/01/2017	07/12/2019	712	15042	Website		DB	Private	No	No	USA
vice	01/01/2017	08/12/2019	962	11179	Newspaper	Tabloid	Legacy	Private	No	No	Canada
newstalkfm	01/01/2017	08/12/2019	972	10989	Radio		Legacy	Private	No	No	Ireland
itvnews	01/01/2017	08/12/2019	841	9614	TV		Legacy	Private	No	No	UK

apnews	03/01/2017	07/12/2019	903	5538			Legacy	Private	No	Yes	USA
businessinsider	01/01/2017	05/12/2019	877	5297	Website		DB	Private	No	No	USA
thecanaryuk	04/01/2018	08/12/2019	507	4004	Website		DB	Private	No	No	UK
upi	10/01/2018	07/12/2019	316	3903			Legacy	Private	No	Yes	USA
todayfm	03/01/2017	08/12/2019	905	3094	Radio		Legacy	Private	No	No	Ireland
reuters	01/01/2017	05/12/2019	585	1655			legacy	Private	No	Yes	UK
yahoonews	01/01/2017	07/12/2019	463	1501	Website		DB	Private	Yes	No	USA
newsweek	02/01/2017	18/07/2019	350	889	Newspaper	Tabloid	Legacy	Private	No	No	USA
buzzfeednews	09/01/2018	08/12/2019	272	676	Website		DB	Private	Yes	No	USA
huffpost	05/01/2017	04/12/2019	270	469	Website		DB	Private	No	No	USA
vox	08/01/2017	09/12/2019	178	268	Website		DB	Private	No	No	USA

Appendix

Clickbait is a complex concept thus complicated to detect. Scholars take various approaches, while trying to detect clickbaits. Most of the tools are *headline* based trying to decide whether a headline is clickbait or not using the headline itself, by detecting the textual features of the headline, such as words and sentences length (Biyani et al., 2016; Chakraborty et al., 2016; Elyashar et al., 2017; Papadopoulou et al., 2017), the use in internet slang and emoticons (Chakraborty et al., 2016; Papadopoulou et al., 2017), the use in exclamation and question marks (Biyani et al., 2016; Chakraborty et al., 2016; Papadopoulou et al., 2017), the text sentiment (Biyani et al., 2016, Papadopoulou et al., 2017), and many more textual features. Some of the tools use this approach combined with the various visual features (Papadopoulou et al., 2017; Kumar et al., 2018).

An alternative approach is based on the claim that clickbait headlines are intentionally overpromising and create expectations that the article does not fulfill (Kumar et al., 2018). Thus, the tools developed according to this approach *compare the headline and the article* trying to decide whether the headline is clickbait (Kumar et al., 2018; Dong et al., 2019). Another approach is based on *users' behavior*, using engagement metrics. Facebook clickbait detecting model is an example to this approach, by using click-to-share ratio and time spent on article as indicators whether the story is a clickbait and remove it (Chakraborty et al., 2016).

In order to accelerate the development and improvement of clickbait detection automatic tools Potthast and his colleagues (2018) established the clickbait challenge 2017 encouraging scholars to address this issue. The challenge dataset contained a large corpus of tweets (38,517), published by 27 English language news outlets in the period from December 1, 2016 through April 30, 2017. Each tweet was graded by five human annotators in a four-point scale from “not” via “slightly” and “considerably” to “heavily” clickbaiting (for further information about the clickbait challenge, the challenge corpus and the “clickbaitiness” grading method please see Potthast et al., 2018). The

participant teams had to develop a regression technology that rates the level of “clickbaitiness” of a post. The teams’ technology performance was primarily evaluated by its mean squared error with respect to the mean judgment of the annotators.

The best-performing approaches submitted to the challenge are available and employed to process the headlines. Thus, our clickbait detection builds on the state of the art of detecting clickbait automatically. Although none of the best-performing approaches have reached perfection as of yet, the detection performance (mean squared error of 0.024) of the currently best-performing one (goldfish) is sufficient for practical purposes. Furthermore, the graded scale of clickbait scoring, as well as an analysis of the confidence values returned by the machine learning algorithms allows for selecting subsets of headlines that are clickbait with high confidence, thus ensuring a high validity of the automatic processing step.

Although the winning algorithm was validated and tested, as part of the challenge, we carried out further validation using four human coders for a sample of 100 posts. First, coders were asked to evaluate the level of clickbaitiness for each headline according to the four-point scale implemented in the challenge. We created a new variable: for each headline we averaged all human coders evaluations, and then correlated it against the standardized clickbait score the algorithm yielded, with satisfying results (Person’s $r=.721$, $p<0.01$). Second, taking a more conservative approach we wished to single-out the point in which human coders reached a consensus that the headline is indeed a clickbait and identify the cutting-point for the algorithmic standardized score, that corresponds with it. Thus, we asked coders to perform a dichotomous evaluation, judging whether a headline is a clickbait. We then compared the mean standardized score according to the number of coders that identify the headline as clickbait (ranging from 0=none, 4=all coders). Based on the results we set our cutting-point on 1.5.

References

- Bazaco, A. (2019). Clickbait as a strategy of viral journalism: conceptualisation and methods.
- Beleslin, I., Njegovan, B. R., & Vukadinović, M. S. (2017, October). Clickbait titles: Risky formula for attracting readers and advertisers. In *XVII International Scientific Conference on Industrial Systems (IS'17) Novi Sad, Serbia* (pp. 364-369).
- Biyani, P., Tsioutsoulouklis, K., & Blackmer, J. (2016, February). " 8 amazing secrets for getting more clicks": detecting clickbaits in news streams using article informality. In *Thirtieth AAAI conference on artificial intelligence*.
- Blanchett Neheli, N. (2018). News by numbers: The evolution of analytics in journalism. *Digital Journalism*, 6(8), 1041-1051.
- Cable, J., & Mottershead, G. (2018). 'Can I click it? Yes you can': Football journalism, Twitter, and clickbait. *Ethical Space*, 15(1/2).
- Chakraborty, A., Paranjape, B., Kakarla, S., & Ganguly, N. (2016, August). Stop clickbait: Detecting and preventing clickbaits in online news media. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 9-16). IEEE.
- 'Clickbait.' (n.d.) In *Oxford Dictionaries*. Retrieved on November 5, 2020, from: <http://www.oxforddictionaries.com/definition/english/clickbait>
- Dong, M., Yao, L., Wang, X., Benatallah, B., & Huang, C. (2019, April). Similarity-aware deep attentive model for clickbait detection. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 56-69). Springer, Cham.
- Dvorkin, F. (27/04/2016). Column: Why click-bait will be the death of journalism, in *pbs.org*. retrieved on January 2, 2020, from: <https://www.pbs.org/newshour/economy/what-you-dont-know-about-click-bait-journalism-could-kill-you>.
- Elyashar, A., Bendahan, J., & Puzis, R. (2017). Detecting Clickbait in Online Social Media: You Won't Believe How We Did It. *arXiv preprint arXiv:1710.06699*.

- Harcup, T. (2016). Asking the readers: Audience research into alternative journalism. *Journalism Practice*, 10(6), 680-696.
- Hurst, N. (2016). *To clickbait or not to clickbait? an examination of clickbait headline effects on source credibility* (Doctoral dissertation, University of Missouri--Columbia).
- Kilgo, D. K., Harlow, S., García-Perdomo, V., & Salaverría, R. (2018). A new sensation? An international exploration of sensationalism and social media recommendations in online news publications. *Journalism*, 19(11), 1497-1516.
- Kumar, V., Khattar, D., Gairola, S., Kumar Lal, Y., & Varma, V. (2018, June). Identifying clickbait: A multi-strategy approach using neural networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval* (pp. 1225-1228).
- Lischka, J. A. (2018). Logics in social media news making: How social media editors marry the Facebook logic with journalistic standards. *Journalism*, 1464884918788472.
- Molyneux, L., & Coddington, M. (2020). Aggregation, clickbait and their effect on perceptions of journalistic credibility and quality. *Journalism Practice*, 14(4), 429-446.
- Mourão, R. R., & Robertson, C. T. (2019). Fake news as discursive integration: An analysis of sites that publish false, misleading, hyperpartisan and sensational information. *Journalism Studies*, 20(14), 2077-2095.
- Mukherjee, P., Dutta, S., & Mani, D. (2019). The Antecedents and Rebroadcast Consequences of Clickbait. *IIM Bangalore Research Paper*, (592).
- Orosa, B. G. (2017). Use of clickbait in the online news media of the 28 EU member countries.
- Papadopoulou, O., Zampoglou, M., Papadopoulos, S., & Kompatsiaris, I. (2017). A two-level classification approach for detecting clickbait posts using text-based features. *arXiv preprint arXiv:1710.08528*.
- Pengnate, S. F., Young, A., & Chen, J. (2018). NEW DETAILS EMERGE! Revealing the Effects of Clickbait Headlines on User Responses.

- Potthast, M., Gollub, T., Hagen, M., & Stein, B. (2018). The clickbait challenge 2017: Towards a regression model for clickbait strength. *arXiv preprint arXiv:1812.10847*.
- Scacco, J. M., & Muddiman, A. (2019). The curiosity effect: information seeking in the contemporary news environment. *New media and society* 22(3), 429-448.
- Tandoc Jr, E. C. (2014). Journalism is twerking? How web analytics is changing the process of gatekeeping. *New media & society*, 16(4), 559-575.
- Thiel, K. (2018). *Avoiding Clickbait*. Cavendish Square Publishing, LLC.
- Thurman, N., & Myllylahti, M. (2009). Taking the paper out of news: A case study of Taloussanommat, Europe's first online-only newspaper. *Journalism Studies*, 10(5), 691-708.
- Tsuriel, K., Dvir Gvirsman, S., Ziv, L., Afriat-Aviv, H., & Ivan, L. (2019). Servant of two masters: How social media editors balance between mass media logic and social media logic. *Journalism*, 1464884919849417.
- Vultee, F., Burgess, G. S., Frazier, D., & Mesmer, K. (2020). Here's What to Know About Clickbait: Effects of Image, Headline and Editing on Audience Attitudes. *Journalism Practice*, 1-18.
- Zhang, W., Jiang, Q., & Peng, C. H. (2019). Unfolding the Clickbait: A Siren's Call in the Attention Economy. In *The 27th European Conference on Information Systems (ECIS) 2019 European Conference on Information Systems*. Association for Information Systems. AIS Electronic Library (AISeL).