

## The power of related articles – improving fake news detection on social media platforms

Henner Gimpel, Sebastian Heger, Julia Kasper, Ricarda Schäfer

### Angaben zur Veröffentlichung / Publication details:

Gimpel, Henner, Sebastian Heger, Julia Kasper, and Ricarda Schäfer. 2020. "The power of related articles – improving fake news detection on social media platforms." In *Proceedings of the 53rd Hawaii International Conference on System Sciences (HICSS)*, Maui, Hawaii, USA, January 7-10, 2020, edited by Tung Bui, 6063–72. Honolulu, HI: University of Hawai'i at Manoa. <https://doi.org/10.24251/hicss.2020.743>.

## The Power of Related Articles – Improving Fake News Detection on Social Media Platforms

Henner Gimpel

FIM Research Center, University of Augsburg  
Project Group Business & Information Systems  
Engineering of the Fraunhofer FIT  
henner.gimpel@fim-rc.de

Julia Kasper

University of Augsburg  
julia@kasper-team.de

Sebastian Heger

FIM Research Center, University of Augsburg  
Project Group Business & Information Systems  
Engineering of the Fraunhofer FIT  
sebastian.heger@fim-rc.de

Ricarda Schäfer

FIM Research Center, University of Augsburg  
ricarda.schaefer@fim-rc.de

### Abstract

*Social media is increasingly used as a platform for news consumption, but it has also become a breeding ground for fake news. This serious threat poses significant challenges to social media providers, society, and science. Several studies have investigated automated approaches to fighting fake news, but little has been done to improve fake news detection on the users' side. A simple but promising approach could be to broaden users' knowledge and thus the perceptual process in order to improve detection behavior. This study evaluates the impact of a digital nudging approach, which aims to fight fake news through the help of related articles. 322 participants took part in an online experiment simulating the Facebook Newsfeed. In addition to a control group, three treatment groups were exposed to different combinations of related articles. Results indicate that the presence of controversial related articles has a positive influence on the detection of fake news.*

### 1. Introduction

In the past decade, social media has revolutionized the way people interact and consume information. A 2016 survey from the Pew Research Centre showed that 62% of US adults consumed news through social media, as compared to only 49% in 2012 [6]. By 2018, this number had reached 69% [25].

Yet, in addition to numerous benefits, the ease with which news can now be distributed and accessed poses serious challenges. Social media makes it easy for users to create and share content with the public [15]. As a result, news through social media is not

necessarily reliable compared to news from traditional sources, as rigorous control is harder to implement. This lack of control enables the spread of so-called fake news [30], a term used to describe news “that are intentionally and verifiably false” [1]. In many cases, fake news are purposefully used to influence and manipulate the audience [21]. Common targets are politics and financial markets, where fake news are used to discredit politicians or affect the financial value of stocks and options [21].

It is immensely important to ensure that fake news (by both private users as well as organizations, e.g. Breitbart) do not cause harmful manipulation. This importance is further increased through the discovery that fake news diffuse much faster and on deeper levels than true stories in an online context [37]. Therefore, detecting such news must be a key priority.

Multiple approaches can be used to detect fake news, including algorithms, data mining, or other automated IT-centered approaches [7, 28, 29]. Yet, to date, no technical solution has been able to fully control the problem. Another research stream, however, takes an alternative approach: Rather than relying on technological solutions, it examines factors, which may influence the credibility perception among the audience itself. Fact checking, source credibility ratings, and the specific design and format of news sources have all been identified as possible means of fake news detection [2, 16]. However, previous research has raised concerns about the effectiveness of these approaches [18, 26]. Assessing the source has been found to be an important and helpful factor for deciding whether a news post is true or false, with known sources often creating more trust in the presented news [2, 16]. However, anyone can publish news in social media and therefore large amounts of

content is published by unknown sources. Consequently, it is all the more important to analyze possibilities to improve fake news detection in situations where the source is unknown [15]. Further, Lazer et al. address the necessity of multidisciplinary approaches when attempting to detect fake news [18]. These include structural changes as well as empowering individuals.

To empower individuals in a way that supports fake news detection, one needs to understand how human perception, processing, and sense-making of information and content work. One of the influential factors in the perception process is knowledge [11], which enables information perception and thus plays a key role in whether one classifies information as true or false.

Research has also shown that the increase in a person's prior knowledge leads to improved sense-making in news consumption and better handling of large amounts of information [27]. In the specific context of fake news detection, a person's knowledge base has shown to be one of the core resources used for assessing the truthfulness of news [9]. Hence, improving a user's knowledge base on a given topic in order to increase analytical thinking in social media holds promising potential in the fight against fake news [4].

When assessing possibilities to reach an enhancement in user behavior, there is currently much discussion about the approach of digital nudging [34, 39]. As Sunstein observes, the nudge "disclosure" can improve a user's knowledge. In social media, such a digital nudge could be implemented in the form of related articles, which address the same subject as the main article but do not necessarily take the same point of view [32]. This may help provide users with more information on a particular topic and, hence, empowers them to better detect fake news through knowledge improvement.

Therefore, the purpose of our research is to investigate whether related articles improve fake news detection by the users in social media environments. We investigate three possibilities on how to provide related articles by conducting an online experiment simulating the Facebook Newsfeed. Our results suggest that related articles have a significant effect on improved fake news detection. We achieved the best results when providing a mix of controversial related articles.

Section 2 of this paper provides the theoretical background on fake news, digital nudging and the derivation of the research hypotheses. Section 3 expands on the experimental method used to address our hypotheses. In Section 4, we present the results of the experiment and in Section 5, we conclude by

discussing results, highlighting the theoretical and practical implications and pointing out limitations of our research.

## **2. Theoretical Background and Hypotheses Development**

### **2.1 Fake News**

In contrast to fake news, truthful news are those stories that cover content that is verifiable through data, facts, and research. Even though the notion of "true news stories" is widely used, one still has to consider that journalism of this kind will always include some extent of a personal footprint, such as through the author's selection of topics covered [13].

Fake news stories are created for a variety of purposes, including financial and political gains [3]. The phenomenon is not new, but it is of growing importance given its increased reach via social media and the consequent magnitude of its detrimental effects [1].

Authors of fake news would like to intentionally mislead people, encouraging them to believe false information. They also change the way people interpret and respond to real news [30]. This leads to an increasing state of widespread distrust and confusion about real news [6]. Therefore, the detection of fake news has become an important task for social media platforms, who must attempt to remove fake news or at least mitigate the negative effects [29].

Recent studies have already examined automated methods which attempt to identify fake news among the almost endless stream of posts on social media platforms [7, 28, 29]. But due to its complex structure, the detection of fake news is difficult to automate, and it is uncertain whether full automation will ever be possible [6, 29]. As the largest online social network, Facebook is particularly affected by fake news stories and has been working towards reducing the amount of fakes on its site since 2015 [23]. The platform works with third party fact-checking organizations who manually check suspicious information and, if necessary, flag articles as doubtful.

### **2.2 News Consumption and Fake News Detection in Social Media**

To improve a user's ability to detect fake news, it is important to understand how users consume news in social media. According to Pentina and Tarafdar, the way news is consumed via social media has changed drastically compared to consumption via more conventional media [27]. They theorize that sense-

making of news in social media environments and the consequential formation of knowledge relies on information overload strategies. Two specific aspects characterize news consumption [27]:

First, the screening news stimuli explains which channels, sources, and content are considered by an individual. Second, sense-making relies on processing and interpreting information from news, which involves the interpretation of the meaning of the obtained news and its transformation into knowledge.

Social media is a news provider, which offers diversity of opinion and social legitimacy, but simultaneously, also offers space for subjective opinions and fake news stories [27]. This poses a severe challenge due to the fact that humans are fairly ineffective at recognizing deception [7]. This ineffectiveness is caused by humans' frequent inability to determine whether information presented to them is true or false [29]. Fake news detection is the correct decision of an individual that information is false. However, in the context of decision-making, natural human deficits resulting from cognitive and behavioral biases often lead to erroneous assessments [12].

Throughout the cognitive processing and interpreting of news information, news consumers evaluate trustworthiness and reliability [27]. Therefore, individuals access a variety of sense-making strategies, like source reliability or comparing news content to their own knowledge to evaluate the credibility of news [9].

However, prior exposure to false knowledge related to a news statement also increases believability of fake news. That is, due to the "illusory truth effect" individuals tend to misinterpret fake news as true when they were exposed to the false knowledge beforehand [26]. From a positive point of view, it also indicates that improving users' knowledge while consuming news in social media environments improves their ability to correctly decide which news are fake.

## 2.3 Related Articles

Researchers have shown growing interest in ways to influence decision-making actively in situations strongly affected by biases. This type of influencing behavior is often referred to as soft or asymmetric paternalism or nudging [19]. Thaler and Sunstein define a nudge as "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives." [34]. Nudges target biases and attempt to overcome them by conscientiously altering the "choice architecture" (design of possible choices) presented to the respective target group.

The goal is to foster decision-making contexts which promote behavior that is beneficial both to individuals and to society [33]. Nudges should not hinder freedom of choice and should aim to make life simpler, safer or easier for people to navigate, while remaining transparent and open rather than hidden [32]. Where the nudging philosophy has been transferred into the digital environment, it is referred to as "Digital Nudging" [39]. Common environments for digital nudging are social media platforms [39] such as Facebook, Twitter or Instagram. Social media providers are choice architects who choose the content and format of information presented to their users. Nudges could therefore be used to improve fake news detection in social media environments and become a countermeasure against fake news.

Sunstein presents a list of ten important "nudges", one of which is disclosure [32]. Disclosure describes the adding of supplementary information to a specific topic or situation such as nutritional details on the packaging of food. In the context of digital disclosure, it is explained that "more detailed and fuller disclosure might be made available online for those who are interested in it" [32]. Yet, simplicity remains essential, which means that information presented must both be comprehensible and accessible.

Since 2017, Facebook has been working with a feature called related articles, which shows users a range of articles under an original post to help them make decisions that are more informed. This feature is mainly used in the U.S., although there have been plans for a future roll-out in Germany, France, and the Netherlands [8]. From a theoretical perspective, related articles are a type of disclosure nudge.

Disclosure has already been used successfully in the context of social media, e.g. in the field of privacy protection [41]. We hypothesize that disclosure is a promising nudging approach to improve a user's knowledge, and can consequently stimulate analytical thinking. This is because users have access to more information and can better reflect the news, leading to an improvement in fake news detection [4].

Looking at news consumption in social media, news can be either true or false (actual state) and for both situations the user can further perceive the news post as true or false (perceived state). The results of this classification task can be described with the help of specificity. Specificity describes the proportion of actual fake news that is detected to be false by the users. We summarize this within our first hypothesis.

**H1: The presence of related articles improves fake news detection in terms of improved specificity.**

## 2.4 Design of Related Articles

Using related articles, there are various ways to implement a disclosure nudge. Since fake news are published to spread false information, the easiest way to recognize them is by knowing the truth [29]. Hence, when knowing all facts on a given topic, detecting fake news is simple.

From this point of view, users may better detect fake news as such while articles telling the truth are available to them. Following a modified form of the “illusory truth effect” [26] exposure of true statements may enable users to detect fake news even though they do not know the truth. Therefore, a disclosure nudge in the form of related articles that solely tell the truth ought to increase fake news detection.

Thus, we assume that related articles that are all true achieve the best results in fake news detection, measured by the highest specificity.

**H2a: Related articles that tell the truth lead to an increase in specificity in fake news detection compared to having no related articles.**

Finding verifiably true related articles is a difficult and costly task for the platform provider and it likely is impossible to fully automate [29].

An alternative is controversial news, which is easier to implement. Current technical approaches exist that build on detecting controversies in news [36]. Controversial news in social media represent those topics that find supporters for conflicting sides of an argument and are debated heatedly [10]. A popular example for such a controversy would be believing and disbelieving that climate change is real. From a user’s perspective, previous studies have already demonstrated that controversial discussions with supporting and opposing arguments (e.g. pro and contra the existence of climate change) help to improve people’s analytical thinking. Results suggest that people presented with point/counterpoint arguments are less susceptible to biases [40]. Controversial news, in addition, encourage users to actively think about a certain topic, which has previously been shown by analyzing online search behavior [38]. Thus, controversial related articles may increase active, open-minded and analytical thinking which is expected to be promising in the fight against fake news [4].

Controversy can be implemented in different ways when designing related articles. Firstly, in the sense of controversial related articles that show different viewpoints on a given topic. Herein, the collection of related articles itself includes controversy, with some articles agreeing and some disagreeing with the main article. Research has shown the potential of offering balanced viewpoints on a topic to counter cognitive biases and thus hinting at the potential of mixed,

controversial related articles in the fight against fake news [5]. Hence, we hypothesize that a mix of controversial related articles also increase fake news detection in terms of higher specificity compared to users who do not receive controversial related articles.

**H2b: A mixed of controversial related articles lead to an increase in specificity in fake news detection compared to having no related articles.**

Secondly, controversy can be achieved through using related articles that oppose the main article in the strongest possible way. “Considering the opposite” is a strategy to overcome cognitive biases that has been thoroughly discussed in research in psychology [20]. The strategy consists of directly pointing people to the opposite perspective on a given topic or question. Consequentially, we hypothesize that designing controversial related articles in a way that they strongly oppose the main article and thus hint at the opposite alternative increases fake news detection through an increase in specificity as compared to users who do not receive controversial related articles.

**H2c: Opposing related articles lead to an increase in specificity in fake news detection compared to having no related articles.**

Note that H2a, H2b, and H2c suggest that three mutually exclusive versions of related articles all are effective. There is no line of theoretical argumentation unequivocally suggesting which of the three designs of related articles performs best. In case multiple hypotheses are empirically supported, it is an empirical matter to identify which of the designs performs best.

Besides improving specificity, it is important not to aggravate sensitivity. Sensitivity describes the proportion of actual true news that is detected to be true news by the user. We do not hypothesize any effect but we include sensitivity in our analysis and test for possible aggravation.

## 3. Experiment Design and Procedures

To test our hypotheses, we conducted an online experiment that assesses the impact of related articles on fake news detection. The experiment included presenting an interactive newsfeed similar to that of Facebook, which the participants could scroll through and interact with to achieve a scenario that is as realistic as possible. Facebook was used as a template because it is the most used social media platform and therefore provides a real-life, natural, and known setting [24]. All parts of the experiment were in German language.

The experiment consisted of six stages: (1) introduction, (2) questions on demographic factors, (3) introductory newsfeed, (4) natural interaction run,

(5) questions on fake news detection, and (6) debriefing. Figure A in the appendix summarizes the experiment procedure.

Throughout the first four stages the context of fake news was not explained to avoid biases in the participants' behavior. Rather, the context mentioned was that of general online behavior in social media. All participants were asked demographic questions to help us develop a clear picture of who completed the experiment. Questions focused on age, gender, education, current profession, and intensity of social media usage. In order to make sure all participants were accustomed to the functionality of the feed, an introductory newsfeed was presented using a single news post, in which all functions (like, share, comment, report, open articles) had to be tested.

Afterwards, all participants went through a natural interaction run with the actual news feed, which consisted of news posts such as those published by newspapers or news services. The participants received six news posts about current events (politics, environment, and celebrities) of which three were truthful news and three were fake news. Each of the three topic categories was covered via one truthful and one fake news post in order to avoid a bias caused by topic selection. Further, all six articles are real archived news articles that were published online. For each article - and also all related articles - we have conducted a review of fact-checking websites (e.g. snopes.org) to ensure that it is either true or false news. As the news source strongly impacts the perception of credibility, we chose six articles from rather unknown sources [16]. In addition, we also avoided to include other criteria that would allow participants to clearly identify a fake news article as such (e.g. spelling mistakes). We did this to minimize other external effects in order to better observe the change in specificity caused by related articles. The order of the posts varied, but all participants were presented with the six identical news stories. Table A in the appendix gives an overview over all main and related article headlines.

All participants were randomly assigned to one of four treatments:

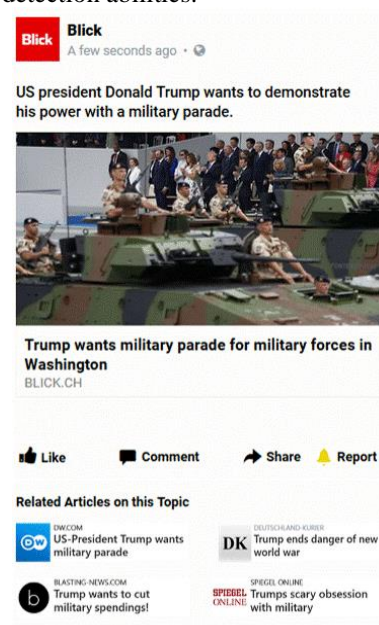
- **Treatment 1 (T1) - Control Group:** Six main articles are shown. No related articles are shown.
- **Treatment 2 (T2) - True:** Six main articles are shown. Four true related articles are shown under each main article, regardless of whether the main article is true or fake.
- **Treatment 3 (T3) - Mixed:** Six main articles are shown. Two fake and two true related articles are shown under each main article, regardless of whether the main article is true or fake.

- **Treatment 4 (T4) - Opposing:** Six main articles are shown. For each main article, four articles with opposing content (in relation to the main article) are shown. Consequently, four true related articles are shown for each fake main post, and four fake related articles are shown for each true main post.

Our control group T1 allows for comparing the general effect of related articles vs. no related articles (H1). The treatment groups T2-T4 are used to assess H2a, H2b and H2c. T2 includes the highest degree of truthfulness in the related articles, with all related articles showing the truth (H2a). While T3 offers the highest controversy within the related articles (2 true, 2 fake – H2b), T4 portrays the largest possible controversy between the main article and the related articles (H2c). One must consider that in the case of a fake main article, the related articles in T2 and T4 are identical (four true related articles in both cases). For all four treatments, the newsfeed allowed us to track the activities (liking, sharing, commenting, reporting, opening articles) executed during the experiment, which enabled us to evaluate the social media usage behavior of the participants.

Figure 1 shows an example of a real news post, including related articles. Clicking on either the main or the related article opened the full text of the article.

After the newsfeed interaction, all participants were shown the original six news posts (without related articles) and were asked to explicitly state for each post whether they regard it as a fake or true article. This step provided the basis for evaluating the individual fake news detection abilities.



**Figure 1:** True news post including related articles

Finally, the experiment ended with a debriefing session in which the participants were informed about the fact that we had manipulated different features of the posts and that these do not necessarily correspond to the real world.

Participants for the experiment were recruited via multiple channels. We used Facebook as the main source for recruiting and linked to our experiment in various groups, mainly targeting students and young professionals. Among others, the targeted groups included sports groups, student associations, university groups and the authors' social network. Other sources of participants were messenger providers such as WhatsApp, or email lists in a university and work context. The choice of recruiting channels was selected because the majority of Facebook users in Germany are aged between 25 and 34 [35]. Thus, we ensured that participants are used to social media environments and are in fact part of the target audience mainly affected by fake news in social media.

## 4. Results

### 4.1 Characteristics of the Sample

In total, 322 people completed our experiment, with 146 female and 176 male participants. 311 participants are between 18 and 35 years old, and 11 are older than 35. Regarding education, 221 participants have at least one academic degree. Analyzing the participants' use of social media, we found that 44.7% use social media daily, and 92.9% use Facebook. Consequently, we conclude that participants are familiar with the use of social media such as Facebook.

77 participants were randomly assigned to treatment T1 by the assignment algorithm of the survey provider, 77 to T2, 93 to T3 and 75 to T4. To ensure comparability between treatments, we analyze their homogeneity. For discrete variables we use chi-squared-tests and for continuous variables ANOVA. The test statistics show that participants do not differ significantly between the treatments in terms of gender, age, social media usage or education. Therefore, the groups can be assumed to be comparable.

68.0% of the participants actively interacted with the newsfeed, meaning they either shared, commented, liked, opened, or reported articles. A range of 1 to 34 interactions per active user was recorded, with a mean of 3.4 and a median of 2.0. Previous studies of social media usage report a 52% activity rate among users [22]. Based on this figure, participants in our study were slightly more active than the average user.

Participants shared between 0 and 6 posts, with a mean of 0.2 and a median of 0, and commented on 0 to

6 posts, with a mean of 0.7 and a median of 0. They liked 0 to 6 posts, with a mean of 0.4 and a median of 0. Participants who had the related articles feature implemented (T2-T4) opened related articles between 0 and 19 times, with a mean of 0.8 and a median of 0. On average, participants in T2, T3, and T4 used the news feed for a longer period of time. Participants assigned to treatment T1 interacted with the newsfeed for a mean of 80 seconds and a median of 62 seconds, T2 for a mean of 110 seconds and a median of 66 seconds, T3 for a mean of 92 seconds and a median of 77 seconds and T4 for a mean of 109 seconds and a median of 77 seconds.

Overall, participants reported an average of 0.7 articles with a median of 0 and a spread from 0 to 6. 64.6% of the participants did not report a single article.

For all interactions mentioned, we test whether there are significant differences between the treatments. Interactions (like, share, comment, report, or open related articles) do not differ significantly between treatments, excluding T1 for related articles as they were not available to this treatment.

### 4.2 Fake News Detection Behavior

After interacting with the news feed, in stage 5 all participants were asked to explicitly state whether they considered each of the six posts to contain fake or true news. The above-mentioned task represents a classification task consisting of the true state of the article (true or fake) and the participant's perception of the article (true or fake). As an adequate and common tool for comparing different approaches to classification, the confusion matrices including true positives (TP), false negatives (FN), false positives (FP) and true negatives (TN) of all four treatments are constructed [31]. Table 1 shows an exemplary confusion matrix for participants in T1. The overall n of that matrix describes the number of participants in T1 \* 6 (every participant was shown 6 news post for classification).

The same information can be extracted for all treatments, considering six classifications for each participant in each treatment group. To generally assess the difference between showing and not showing related articles, the confusion matrices of T2-T4 can be aggregated. Table 2 shows the confusion matrix of all groups in an aggregated manner.

**Table 1:** Confusion matrix of T1

	Perceived True	Perceived Fake
Actual True	182 (TP)	49 (FN)
Actual Fake	37 (FP)	194 (TN)

From these responses, specificity ( $\frac{TN}{TN+FP}$ ) and sensitivity ( $\frac{TP}{TP+FN}$ ) values are calculated for all four groups as a performance indicator. The respective results are presented in Table 3. For both, specificity and sensitivity higher values are preferable, as they indicate better classifications results.

Observing the performance results, Table 3 shows an increase in specificity, when comparing the aggregated results of all groups being shown related articles (T2-T4) with T1. Looking at differences between individual treatments, especially between T1 as compared to T2 and T3, T3 (mixed) exhibits the highest specificity, followed by T2. T4 shows the smallest increase in specificity when comparing the different alternatives of designing related articles. Between all groups that show related articles in the experiment, T3 with its high levels of controversy among related articles performs best in respect to both, specificity and sensitivity.

**Table 2:** Classification results of all groups

	T1: Control	T2: True	T3: Mixed	T4: Opposing
TP	182	169	218	177
FN	49	56	61	54
FP	37	26	25	36
TN	194	199	254	195

**Table 3:** Performance metrics

	Specificity	Sensitivity
T1: Control	83.98%	78.79%
T2: True	88.44%	75.11%
T3: Mixed	91.04%	78.14%
T4: Opposing	84.42%	76.62%
T2-T4 (aggregated)	88.16%	76.73%

In a next step, the significance of the observed performance differences is assessed. The tests are performed to evaluate whether the increase in specificity and sensitivity are likely caused by the treatment. As three fake and three true articles were shown to each participant, there exist only four levels for both accuracy metrics (0 to 3 out of 3 articles classified correctly) for each individual participant. We therefore use chi-squared-tests to determine, whether significant differences exist between the treatment groups.

As H1 regards the general effectiveness of related articles, we perform a chi-squared-test on the specificity and sensitivity values between T1 and T2-T4. While there are no significant differences in treatment sensitivity ( $p=0.632$ ), specificity is significantly lower (significance level 10%) in the control group as compared to treatments T2-T4 ( $p=0.063$ ). Thus, we find support for H1. There is a positive effect, as people being shown related articles perform significantly better at detecting fake news. Simultaneously, the decrease in sensitivity observed from no related articles to related articles (78.79% to 76.73%) is not significant, so we do not find an effect of related articles deteriorating one's ability to detect true news posts as such.

To evaluate the impact of the degree of truthfulness and controversy in the news posts, we conduct pairwise comparisons between all treatment groups using chi-squared tests. The aim is to test H2a, H2b and H2c. Table 4 shows the resulting p-values of all pairwise comparisons for specificity and sensitivity. The upper right triangle in Table 4 shows the results for specificity, the lower left triangle those for sensitivity.

**Table 4:** P-values of pairwise comparisons for specificity (upper right triangle) and sensitivity (lower left triangle, grey)

	T1: Control	T2: True	T3: Mixed	T4: Opposing
T1	-	0.178	0.008	0.178
T2	0.543	-	0.449	0.306
T3	0.934	0.748	-	0.029
T4	0.571	0.939	0.885	-

Looking at all significant values (significance level 5%) in Table 4, we conclude that participants in T2 did not perform significantly better than participants in the control group (T1). Therefore, we can reject H2a. True related articles do not necessarily lead to improved fake news detection.

Beyond that, the results indicate that specificity levels are significantly higher in T3 as compared to T1 and T4. Participants in T3 achieved better results in detecting and classifying fake news. This supports H2b and thus, the positive influence of controversy within the related articles when trying to improve fake news detection abilities.

Furthermore, the slight increase in specificity from T1 to T4 appears not to be caused by the treatment, as the corresponding p-value is very high. Consequently, we can reject H2c and the assumption that related articles that strongly contradict the main article lead to an improvement in users' fake news detection abilities.

Sensitivity is not affected, as all pairwise comparisons yield high p-values. This indicates that while related articles improve fake news detection behavior under specific conditions, they seem not to deteriorate the subjects' ability to classify true news as such. The results presented above are also robust when equalizing the treatment group sizes (T3 n=76) through random sampling and then conducting the analyses as before.

## 5. Discussion & Conclusion

### 5.1 Contributions

This paper details an online experiment we conducted to test the influence of related articles on fake news detection in social media. The results indicate that related articles in general can improve fake news detection. More precisely, a mixture of controversial articles under a main article foster the ability to detect and classify fake news articles.

In response to our first hypothesis (H1), this study finds indication that related articles have a positive impact on fake news detection behavior in the context of social media. Related articles portray a digital nudge, and fall under the category of disclosure. The key characteristic of this nudge is that it supplies additional knowledge and triggers analytical thinking to those who read the news post [32]. Disclosure-type nudges have proven to be effective at improving peoples' behavior in several environments [14, 17]. Our study demonstrates that this positive effect also applies in the context of fake news detection.

However, in response to hypothesis H2a our results indicate that providing solely true related articles does not necessarily enable better fake news detection. Thus, we cannot confirm our assumption of exposure of truthful knowledge (see [26]) being an effective approach. Also, showing related articles that strongly contradict the main article (H2c) does not lead to a significant improvement in fake news detection.

In any case, automatically providing only related articles which are true seems almost impossible in practice [29]. For this reason we hypothesized that a mix of controversial related articles may be an effective approach (H2b). Our results indicate that a mix of controversial related articles enable users to better classify and detect fake news articles as they achieved significantly higher specificity than treatments T1 and T4. This is in line with previous studies, which demonstrated that people presented with point/counterpoint arguments are less susceptible to biases [40]. Thus, a mix of controversy may lead users to reflect more thoroughly on information, rather than

simply accepting the information as it is presented in the main article.

Moreover, controversial articles posted below true articles do not reduce users' ability to recognize and classify true articles, which is immensely important if the nudge is to have the desired effect.

In addition to these theoretical contributions, the results of our study also have practical implications. For social media platform operators, our results suggest a simple means of supporting their users in fake news detection. In particular, we provide valuable and specific insights as to how the feature might be implemented, as the use of only true articles in this type of feature would be complex and costly, if not impossible to achieve [7]. In contrast to that, the use of a mix of controversial articles (as in the mixed treatment T3) is a much more practical and feasible approach. Based on the results of our study, this is the most effective way to improve the accuracy of fake news detection. Further, the usage of related articles increased users' screen time in our study and thus, can be deemed compatible with common social media platform business models.

Overall, this paper demonstrates empirically that a specific form of related articles improves fake news detection in social media, while not compromising the users' ability to identify truthful news as such.

### 5.2 Limitations and Further Research

Our study has some limitations, which highlight the need for future research. First, the generalizability of our research needs to be validated in future research. This is because the age range of our participants was limited to relatively young Germans. Further tests with different age groups and nationalities should be carried out. However, since our survey targets social media, a young average age of participants fits the user group, as a recent study showed that the largest share of Facebook users in Germany is aged between 25 and 34 years old [35].

Further, there are various factors crucial to the detection of fake news [9]. Prior research has shown that the source of an article is an important factor for its perceived credibility [2]. Even though the impact of source credibility is purposely omitted from the scope of this paper, it should be further investigated in future research whether the observed effects hold even when the news come from known sources.

The participants' previous knowledge of the topics covered in the articles may have influenced detection behavior, and should also be a focus of future research. However, with random assignment to treatments this can hardly explain the observed treatment effects. In addition to that, not all selected related articles are of

the exact same relevance to the main article. The choice of related articles in T3 therefore potentially impacts the perceived helpfulness of this treatment.

Our study focuses on short-term behavior. Future attempts should investigate long-term behavior in order to provide a more holistic understanding of behavioral patterns. In this sense, it will also be important to investigate users' reactions to the related articles disclosure nudge once it has been in use for an extended period of time and is no longer a novel feature.

Lastly, although we have attempted to authentically recreate the original newsfeed, our study took place in an experimental environment, a fact that may have influenced the behavior of our participants. To address this issue, it would be insightful to gather data on real social media usage.

Overall, and despite these limitations, our study will help with creating a better environment in terms of improved detection of fake news.

Nonetheless, research needs to investigate the impact of other digital nudges and to evaluate their effectiveness, not only in terms of improving fake news detection but also in terms of reporting suspicious content. There are promising approaches (such as social norms) in this field, and one might potentially combine multiple nudges to combat fake news more effectively.

To conclude, our study indicates that a mix of controversial related articles improve users' ability to detect fake news. Related articles are a comparatively simple concept and are already present on some platforms. They may prove to be a valuable weapon in the fight against fake news. Our finding that a mix of controversial articles are effective in this context makes the use of related articles much more feasible, as this removes the need to ensure that all related articles are true or opposing. The use of digital nudges may significantly advance the fight against fake news and ideally, these will be combined with the development of automated, IT-centered solutions. In the long run, the problem of fake news will not be easy to solve, particularly with respect to political manipulation, which pose a central challenge for society. Consequently, a variety of approaches will be needed to prevent the spread of false and misleading information via social media in the future.

## References

- [1] Allcott, H. and M. Gentzkow, "Social Media and Fake News in the 2016 Election", *Journal of Economic Perspectives*, 31(2), 2017, pp. 211–236.
- [2] Amin, A., J. Zhang, H. Cramer, L. Hardman, and V. Evers, "The effects of source credibility ratings in a cultural

heritage information aggregator", in *Proceedings of the 3rd Workshop on Information Credibility on the Web*, K. Tanaka, Editor, Madrid, Spain. 2009. ACM: New York, NY.

- [3] Aral, S., "Truth, Disrupted.", *Harvard Business Review*, 2018, pp. 3–11.

[4] Bronstein, M.V., G. Pennycook, A. Bear, D.G. Rand, and T.D. Cannon, "Belief in Fake News is Associated with Delusionality, Dogmatism, Religious Fundamentalism, and Reduced Analytic Thinking", *Journal of Applied Research in Memory and Cognition*, 8(1), 2019, pp. 108–117.

[5] Caraban, A., E. Karapanos, D. Gonçalves, and P. Campos, "23 Ways to Nudge", in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, S. Brewster, G. Fitzpatrick, A. Cox, and V. Kostakos, Editors, the 2019 CHI Conference, Glasgow, Scotland Uk, 5/4/2019 - 5/9/2019. 2019. ACM Press: New York, New York, USA.

[6] Chatfield, A.T., C.G. Reddick, and K.P. Choi, "Online Media Use of False News to Frame the 2016 Trump Presidential Campaign", in *Proceedings of the 18th Annual International Conference on Digital Government Research*, C.C. Hinnant, Editor, Staten Island, NY, USA. 2017. ACM: New York, NY.

[7] Conroy, N.J., V.L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news", in *Proceedings of the Association for Information Science and Technology*, A. Grove, Editor. 2015. Association for Information Science and Technology: Silver Spring, Maryland.

[8] <https://www.forbes.com/sites/kathleenchaykowski/2017/08/03/facebook-expands-fight-against-fake-news-with-automatic-related-articles/#59fa46b31d2c>, accessed 6-9-2019.

[9] Flinham, M., C. Karner, K. Bachour, H. Creswick, N. Gupta, and S. Moran, "Falling for Fake News: Investigating the Consumption of News via Social Media", in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, R. Mandryk, M. Hancock, M. Perry, and A. Cox, Editors, Montreal QC, Canada. 2018. ACM Press: New York, USA.

[10] Garimella, K., G.D.F. Morales, A. Gionis, and M. Mathioudakis, "Quantifying Controversy on Social Media", *ACM Transactions on Social Computing*, 1(1), 2018, pp. 1–27.

[11] Goldstein, B.E., *Sensation and Perception*, 8th edn., Wadsworth Cengage Learning, Belmont, Calif., 2010.

[12] Hansen, P.G. and A.M. Jespersen, "Nudge and the Manipulation of Choice", *European Journal of Risk Regulation*, 4(1), 2013, pp. 3–28.

[13] Hermida, A., "TWEETS AND TRUTH", *Journalism Practice*, 6(5-6), 2012, pp. 659–668.

[14] Ho, D.E., "Fudging the Nudge: Information Disclosure and Restaurant Grading", *Yale Law Journal*, 122, 2012, pp. 574–688.

- [15] Kaplan, A.M. and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media", *Business Horizons*, 53(1), 2010, pp. 59–68.
- [16] Kim, A. and A.R. Dennis, "Says Who?: How News Presentation Format Influences Perceived Believability and the Engagement Level of Social Media Users", *MIS Quarterly*, Forthcoming, 2018.
- [17] Kroese, F.M., D.R. Marchiori, and D.T.D. de Ridder, "Nudging healthy food choices: a field experiment at the train station", *Journal of public health*, 38(2), 2016, 133–137.
- [18] Lazer, D.M.J., M.A. Baum, Y. Benkler, A.J. Berinsky, K.M. Greenhill, F. Menczer, M.J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, M. Schudson, S.A. Sloman, C.R. Sunstein, E.A. Thorson, D.J. Watts, and J.L. Zittrain, "The science of fake news", *Science* (New York, N.Y.), 359(6380), 2018, pp. 1094–1096.
- [19] Leonard, T.C., "Richard H. Thaler, Cass R. Sunstein, *Nudge: Improving decisions about health, wealth, and happiness*", *Constitutional Political Economy*, 19(4), 2008, pp. 356–360.
- [20] Lord, C.G., M.R. Lepper, and E. Preston, "Considering the opposite: A corrective strategy for social judgment", *Journal of Personality and Social Psychology*, 47(6), 1984, pp. 1231–1243.
- [21] Maasberg, M., E. Ayaburi, C. Liu, and Y. Au, "Exploring the Propagation of Fake Cyber News: An Experimental Approach", in *Proceedings of the 51st Hawaii International Conference on System Sciences*, T. Bui, Editor, Hawaii International Conference on System Sciences. 2018. Hawaii International Conference on System Sciences.
- [22] <https://www.statista.com/statistics/275788/share-of-facebook-user-activities/>, accessed 6-9-2019.
- [23] Mustafaraj, E. and P.T. Metaxas, "The Fake News Spreading Plague: Was it Preventable?", in *Proceedings of the 2017 ACM Web Science Conference*, P. Boldi, K. Kinder-Kurlanda, P. Fox, D. McGuinness, and L. Poirer, Editors, Troy, New York, USA. 2017. ACM Association for Computing Machinery: New York, NY, USA.
- [24] Newman, N., R. Fletcher, A. Kalogeropoulos, D.A.L. Levy, and R.K. Nielsen, *Digital News Report 2018*, 2018.
- [25] <http://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/>, accessed 6-9-2019.
- [26] Pennycook, G., T.D. Cannon, and D.G. Rand, "Prior exposure increases perceived accuracy of fake news", *Journal of Experimental Psychology: General*, 147(12), 2018, pp. 1865–1880.
- [27] Pentina, I. and M. Tarafdar, "From 'information' to 'knowing': Exploring the role of social media in contemporary news consumption", *Computers in Human Behavior*, 35, 2014, pp. 211–223.
- [28] Rubin, V.L., Y. Chen, and N.J. Conroy, "Deception detection for news: Three types of fakes", in *Proceedings of the Association for Information Science and Technology*, A. Grove, Editor. 2015. Association for Information Science and Technology: Silver Spring, Maryland.
- [29] Shu, K., A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake News Detection on Social Media: A Data Mining Perspective", *ACM SIGKDD Explorations Newsletter*, 19(1), 2017, pp. 22–36.
- [30] Shu, K., S. Wang, and H. Liu, "Understanding User Profiles on Social Media for Fake News Detection", in *IEEE 1st Conference on Multimedia Information Processing and Retrieval*, IEEE, Editor, Miami, FL. 2018. IEEE: Piscataway, NJ.
- [31] Sokolova, M. and G. Lapalme, "A systematic analysis of performance measures for classification tasks", *Information Processing & Management*, 45(4), 2009, pp. 427–437.
- [32] Sunstein, C.R., "Nudging: A Very Short Guide", *Journal of Consumer Policy*, 37(4), 2014, pp. 583–588.
- [33] Thaler, R.H. and C.R. Sunstein, "Libertarian Paternalism", *American Economic Review*, 93(2), 2003, pp. 175–179.
- [34] Thaler, R.H. and C.R. Sunstein, *Nudge: Improving Decisions About Health, Wealth, and Happiness: The Politics of Libertarian Paternalism*, Yale University Press, New Haven, 2008.
- [35] <https://www.statista.com/statistics/867622/number-of-facebook-users-in-germany-by-age-and-gender/>, accessed 9-21-2018.
- [36] Vicario, M.D., W. Quattrociochi, A. Scala, and F. Zollo, "Polarization and Fake News: Early Warning of Potential Misinformation Targets", *ACM Transactions on the Web*, 13(2), 2019, pp. 1–22.
- [37] Vosoughi, S., D. Roy, and S. Aral, "The spread of true and false news online", *Science* (New York, N.Y.), 359(6380), 2018, pp. 1146–1151.
- [38] Weeks, B.E., L.M. Friedenberg, B.G. Southwell, and J.S. Slater, "Behavioral consequences of conflict-oriented health news coverage: the 2009 mammography guideline controversy and online information seeking", *Health Communication*, 27(2), 2012, pp. 158–166.
- [39] Weinmann, M., C. Schneider, and J. Vom Brocke, "Digital Nudging", *Business & Information Systems Engineering*, 58(6), 2016, pp. 433–436.
- [40] Wiley, J., "A fair and balanced look at the news: What affects memory for controversial arguments?", *Journal of Memory and Language*, 53(1), 2005, pp. 95–109.
- [41] Yang, W., P.G. Leon, K. Scott, X. Chen, A. Acquisti, and L.F. Cranor, "Privacy nudges for social media: an exploratory Facebook study.", in *Proceedings of the 22nd International Conference on World Wide Web Companion*, D. Schwabe, Editor. 2013. International World Wide Web Conferences Steering Committee: Republic and Canton of Geneva.