SCM
Studies in Communication and Media

FULL PAPER

Doubters are more convincing than advocates

The impact of user comments and ratings on credibility perceptions of false news stories on social media

Zweifler sind überzeugender als Fürsprecher

Der Einfluss von Nutzerkommentaren und Ratings auf die Glaubwürdigkeitswahrnehmung von falschen Nachrichten-Storys auf sozialen Medien

Jan P. Kluck, Leonie Schaewitz & Nicole C. Krämer

Jan P. Kluck (M.Sc.), Social Psychology: Media & Communication at the Department of Computer Science and Applied Cognitive Science, University Duisburg-Essen, Forsthausweg 2, 47057 Duisburg, Germany; Contact: jan.kluck(at)uni-due.de

Leonie Schaewitz (Dr.), Educational Psychology, Institute of Educational Research, Ruhr University Bochum, Universitätsstraße 150, 44780 Bochum, Germany; Contact: leonie.schaewitz(at)rub.de. ORCID: https://orcid.org/0000-0002-4339-501X

Nicole C. Krämer (Prof. Dr.), Social Psychology: Media & Communication at the Department of Computer Science and Applied Cognitive Science, University Duisburg-Essen, Forsthausweg 2, 47057 Duisburg, Germany; Contact: nicole.kraemer(at)uni-due.de

FULL PAPER

Doubters are more convincing than advocates
The impact of user comments and ratings on credibility perceptions of false news stories on social media

Zweifler sind überzeugender als Fürsprecher
Der Einfluss von Nutzerkommentaren und Ratings auf die Glaubwürdigkeitswahrnehmung von falschen Nachrichten-Storys auf sozialen Medien

Jan P. Kluck, Leonie Schaewitz & Nicole C. Krämer

Abstract: False information on social media poses a crucial threat to our society, and calls for interventions to combat this problem are becoming louder. Users themselves may have the potential to diminish the impact of misleading information. In an online experiment with a 3 x 3 between-subjects design (credibility evaluation in user comments: positive vs. negative vs. none) x (numerical credibility rating: positive vs. negative vs. none), we tested the influence of bandwagon cues on the impact of a false news post on Facebook (N = 240). Contrary to prevalent assumptions regarding heuristic information processing, numerical credibility ratings had no influence on participants' credibility appraisals and intended sharing behavior. However, negative user comments diminished the believability of false news. Moreover, participants' willingness to share the news post publicly and privately was indirectly reduced by the effect of negative user comments on perceived news credibility.

Keywords: bandwagon cues, credibility, online news, false information, negativity bias

Zusammenfassung: Falsche Informationen auf sozialen Medien stellen eine ernsthafte Bedrohung für unsere Gesellschaft dar und Forderungen nach Interventionen zur Bekämpfung dieses Problems werden lauter. Nutzerinnen und Nutzer selbst könnten das Potential haben, den Einfluss falscher Informationen abzuschwächen. Mit einem 3 x 3 Between-Subjects-Design (Glaubwürdigkeitsbewertungen in Nutzerkommentaren: positiv vs. negativ vs. keine) x (numerische Glaubwürdigkeitsbewertungen: positiv vs. negativ vs. keine) haben wir in einem Online-Experiment getestet, welchen Einfluss Nutzerrepräsentationen auf die Wirkung eines falschen Nachrichtenposts auf Facebook haben (N = 240). Entgegen weitverbreiteter Vermutungen hinsichtlich heuristischer Informationsverarbeitung haben numerische Glaubwürdigkeitsbewertungen keinen Einfluss auf die Glaubwürdigkeitseinschätzung der Teilnehmenden und ihrer Intentionen, den Nachrichtenpost zu teilen. Allerdings mindern negative Nutzerkommentare die Glaubwürdigkeit der Falschnachrichten, Darüber hinaus wird die Bereitschaft der Teilnehmenden, den Nachrichtenpost öffentlich und privat zu teilen, indirekt durch den Effekt von negativen Nutzerkommentaren auf die wahrgenommene Glaubwürdigkeit reduziert.

Schlagwörter: Bandwagon Cues, Glaubwürdigkeit, Online-Nachrichten, falsche Informationen, Negativity Bias

Acknowledgements: This research was funded by the German Federal Ministry of Education and Research in the program IT Security ("autonomous and safe in the digitized world," grant number 16KIS0773) and was supported by the Digital Society research program funded by the Ministry of Culture and Science of the German State of North Rhine-Westphalia.

1. Introduction

Given the popularity of social media as a source of information and news, it is also assumed to serve as a fruitful channel for misleading content (Tandoc et al., 2017). When people consider false information to be believable, they are more likely to interact with it by commenting on it or sharing it (Kim & Dennis, 2017). In so doing, they help the information to spread quickly on social media. False information in the form of news seems to be particularly concerning because it often mimics real news organizations, rendering credibility appraisals much more difficult (Tandoc et al., 2017). Unfortunately, current approaches to counter false information on social media (e.g., fact-checking sites) are limited in their effectiveness (Lazer et al., 2018). Therefore, we need to examine new types of interventions.

One major challenge for users is that they are confronted with a huge amount of information on online platforms but do not have sufficient mental resources to process all content thoroughly (Pentina & Tarafdar, 2014). Instead, people are prone to draw on heuristic cues in order to make credibility appraisals in the social web (e.g., Sundar, 2008). One such heuristic that Internet users commonly use is based on social impact, according to which users consider other individuals' reactions when evaluating information on social media (Flanagin, 2017; Sundar, 2008). However, so far, research investigating the extent to which explicit credibility evaluations by others influence the perception of false information online is scarce. Since users are able to react to dubious information in real time, it is worth exploring whether explicit user credibility evaluations affect the believability of false information on social media. Therefore, the current study examines how user comments that explicitly express credibility evaluations, as well as explicit numerical credibility ratings, affect individuals' credibility perceptions and their willingness to share content.

2. False information, false news, and interventions

False information can take different forms. The most important differentiation is between *misinformation* and *disinformation*. Whereas misinformation is regarded as false and misleading information, disinformation is described as "misinformation with an attitude," due to its intentional and purposeful nature (Fetzer, 2004, p. 231). Thus, while both forms can be objectively identified as false, the question of whether false information is regarded as misinformation or disinformation depends on the sender's motivation. From the recipient's perspective, however, it seems less important whether false information is created deliberately or unintentionally. In

the case of uncertainty, the recipient uses credibility cues to make judgments about the believability of new information, irrespective of whether it is correct information, misinformation, or disinformation (Karlova & Fisher, 2013).

False information that mimics news posts on social media is currently drawing a great deal of scholarly attention. The concern is that such information has particular potential to mislead citizens and consequently to bias people's political decisions (e.g., Allcott & Gentzkow, 2017). In this context, Tandoc et al. (2017) argued that the definition of news has to be reconsidered given that even a short post on Twitter or Facebook can be considered as a piece of news regardless of who created it. Since several incidents of false information in the form of news (which we term *false news* in the following) have occurred around the globe, leading to political and societal repercussions, officials and institutions have begun to take this issue very seriously (Tandoc et al., 2017), and the call for effective interventions is becoming louder. Against this background, fact-checking interventions provided by independent organizations have been regarded as a promising counteragent to false information on the Internet. Indeed, it was shown that fact-checking interventions on social media can lower individuals' agreement with false political information that is congruent with their attitudes (Hameleers & van der Meer, 2019). However, fact-checking approaches have also been found to be limited in their effectiveness. For example, Pennycook, Cannon, and Rand (2017) found that labels indicating that stories were contested by fact-checkers were unable to diminish the perceived accuracy of false news on social media. A further problem with fact-checking is that information on social media seems to have a short durability, meaning that news posts have gone "through their cycle of life" before organizations have been able to check the content (Kim & Dennis, 2017, p. 3). Therefore, interventions are needed that can effectively counter false information quickly after publication. In this vein, Bode and Vraga (2018) found that social corrections (i.e., posting a related corrective story) can be effective in reducing health-related misperceptions on social media – even as effective as the same correction provided by an algorithm. Since the distribution of misleading content seems to be strongly promoted by users' interactional behavior, it is helpful to explore how users are also able to reduce the spread of false information. User-based interventions have an advantage over fact-checking approaches in that users can intervene in real time. Moreover, it is well established that individuals' aggregated judgments can be very accurate. Davis-Stober, Budescu, Dana, and Broomell (2014) demonstrated that the accuracy of such aggregated evaluations is highly robust. Thus, explicit credibility judgments of users are a promising approach to counter false information on social media. Since there is a paucity of research investigating the extent to which explicit credibility evaluations by others influence the perception of and interactional behavior with false information, this paper contributes to the current literature by addressing this research gap.

3. Credibility appraisals

When considering the extent to which information is believable, the concept of credibility plays a crucial role. In communication and social psychology, credibility appraisals are understood less as an objective attribution process, but rather as

based on recipients' subjective judgments of believability (e.g., Flanagin & Metzger, 2008; Winter & Krämer, 2014).

To explain how people make credibility appraisals in the complex online environment, Metzger (2007) proposed a dual processing model of web credibility based on the *Elaboration Likelihood Model* (ELM; Petty & Cacioppo, 1986) and the *Heuristic-Systematic Model* (HSM; Chaiken, 1987). In line with these models, Metzger's (2007) model assumes that people's motivation and ability to evaluate a message determine how elaborately they assess information credibility on the web. Motivation is based on factors such as a person's need to find accurate information or a special interest in a particular issue, while ability refers to the individual's knowledge about how to accurately evaluate online information and his/her general cognitive resources to do so. When people have little motivation to find credible information on the Internet, they either rely on heuristic cues or fail to make credibility evaluations at all. By contrast, people who have high motivation to access credible information probably take a more effortful, systematic route of evaluation. With regard to information on social media, it can be assumed that people are not always motivated to put their full effort into credibility evaluations, as they are less inclined to actively seek news on social media but are often exposed incidentally to news on these platforms (Gil de Zúñiga, Weeks, & Ardèvol-Abreu, 2017). Moreover, individuals are often looking for an overview of information rather than pursuing in-depth knowledge (Costera Meijer & Groot Kormelink, 2015).

In addition, people are frequently confronted with a vast amount of information and seem to feel too overloaded to deal with it (Holton & Chyi, 2012). Given this, in order to cope with cognitive restraints when processing the vast amount of information in the online environment, it is presumed that people balance the costs and benefits of information gain (Pentina & Tarafdar, 2014).

This prevalent tendency can be explained by the more general framework of the *Limited Capacity Model of Mediated Message Processing* (LC3MP; Lang, 2000). According to this model, information processing includes controlled as well as automatic sub-processes, which people cannot perform optimally, either because they choose to use fewer resources than needed for the processing task, or because they do not have sufficient mental capability to fulfill the requirements. Since people are often confronted with a vast amount of information on social media, it can be assumed that they take a heuristic route to evaluate information credibility in order to balance costs and benefits. In line with this, users seldom process information in isolation on social media, but rather process it in combination with reactions of other users as an essential part of the social web. Therefore, Flanagin (2017) emphasized the role of social influence when making credibility appraisals based on heuristics – especially in situations of ambiguity.

4. User representations as cognitive heuristics

Based on focus group interviews with participants covering a broad range of levels of internet expertise and sociodemographic characteristics, Metzger and colleagues (2010) reported that heuristics based on *social confirmation* are very

powerful when making credibility appraisals. The endorsement heuristic can be considered as the core heuristic of social confirmation, as it "suggests that people are inclined to perceive information and sources as credible if others do so also, without much scrutiny of the site content or source itself" (Metzger et al., 2010, p. 427). Sundar (2008) termed group-based endorsements as the *bandwagon heuristic*, which he assumes to be a very vigorous cognitive shortcut when making credibility appraisals, especially in the social media environment, because the designs of these platforms make bandwagon cues in the form of user reactions highly salient. Since user comments and numerical user representations are the most common types of bandwagon cues on social media (Walther & Jang, 2012), the current work focuses on these user representations.

User comments. In the context of online news, user comments are particularly relevant, because they can influence the perception of online news and the subsequent process of public communication. For instance, with regard to perceptions of journalistic quality of online news, researchers have found that comments which challenge the quality of an article lead to lower perceptions of journalistic quality compared to comments that praise the quality of an article (Dohle, 2018; Kümpel & Springer, 2016). Likewise, comments which disparage the quality of a news story on Twitter were found to decrease the message credibility of the story compared to positive comments on the quality of the news (Waddell, 2018).

However, although Dohle (2018) also found that negative user comments reduced participants' quality assessments of journalistic products compared to positively valenced comments, no difference was found between negative comments and no comments. Against this background, other scholars demonstrated that positive and negative comments do not simply have opposing effects. For instance, it was shown that uncivil user comments lowered the perceived quality of news articles from known and unknown media outlets (Prochazka, Weber, & Schweiger, 2018). Yet, when the same articles were accompanied by civil comments, quality evaluations did not improve; rather, the articles were perceived even more poorly compared to articles without comments. A study by Winter, Brückner, and Krämer (2015) demonstrated that comments opposing the slant of the article diminished the persuasive effect of the article, but that comments supporting the article's position did not increase persuasive effects. The finding that negative comments have a stronger effect than positive comments relative to a control condition (e.g., no or neutral comments) can be explained by the *negativity bias* (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001), which assumes that people process negative stimuli more thoroughly than positive stimuli because negative impressions are quicker to form. So far, there is no research investigating whether explicit credibility evaluations by other social media users influence the perceived credibility of false information. It is likely that people rely on comments from others to make credibility judgments (see also Metzger et al., 2010). Following the assumptions of the negativity bias, we hypothesize:

Hypothesis 1 (H1): Comments which express concerns regarding the credibility of false information on social media diminish recipients' perceived credibility of the information compared to no comments.

It remains unclear which effects positive user comments have relative to the absence of comments. Since there is no evidence that positively valenced user comments can enhance credibility perceptions of news compared to no comments, we ask:

Research Question 1 (RQ1): Do comments which endorse the credibility of false information on social media enhance recipients' perceived credibility of the information compared to no comments?

Numerical user representations. Drawing on the bandwagon heuristic, Sundar (2008) highlighted the role of aggregated popularity cues in the online environment, as such cues greatly simplify the representation of user opinions and preferences. Indeed, by varying the reputation of the source, the valence of community ratings, and the number of views, Winter and Krämer (2014) demonstrated that community ratings affected the perceived credibility of an online news text – but only in combination with high view counts. A study by Li and Sakamoto (2014) showed that people adopted the truthfulness rating of the majority of other users with regard to health-related statements. Furthermore, it was demonstrated that people rely to the same extent on user ratings as they do on expert evaluations (Jucks & Thon, 2017) or algorithms (Bode & Vraga, 2015) when validating information in the social web. It was also found that source credibility ratings by experts influence participants' credibility assessment of news content on Facebook in the direction given by the rating (Kim & Dennis, 2017). Numerical credibility ratings might be beneficial in combating false information as they enable people to assess information in real time with little effort. Moreover, numerical credibility ratings can represent credibility assessments of a larger group of users and are therefore more reliable than single user statements (Davis-Stober et al., 2014). Since people seem to rely to the same extent on user ratings as they do on expert evaluations when validating online information (Flanagin & Metzger, 2013; Jucks & Thon, 2017) and are likely to follow credibility appraisals of the majority (Li & Sakamoto, 2014), we propose the following hypothesis:

Hypothesis 2 (H2): A numerical credibility rating of social media content influences recipients' perceived credibility of false information in the direction that is indicated by the rating.

Since numerical user representations on social media are often accompanied by user comments, the relative and potentially intertwined influence of both has to be considered. Previous studies indicate that the influence of user comments might be dependent on numerical user representations (Neubaum, Rösner, Ganster, Hambach, & Krämer, 2018; Walther, Liang, Ganster, Wohn, & Emington, 2012). Neubaum et al. (2018), for example, found a disordinal interaction between comment valence and numerical user representations (i.e., likes and shares) insofar as the effect of comment valence only emerged when likes and shares were high. However, using a more explicit numerical user representation (i.e., helpfulness rating), Walther et al. (2012) found a more directional interaction between numerical and textual user reactions when they had the same valence. Therefore, it can be assumed that an explicit user rating of the credibility of news content on social media has similar effects:

Hypothesis 3 (H3): User credibility ratings have a stronger effect on recipients' perceived credibility of false information when accompanied by user comments which indicate credibility appraisals in the same direction.

So far, we have focused on external factors that influence people's credibility perceptions. However, the way in which people process information is based not only external factors but also on individual dispositions. Individuals' cognitive preferences, i.e., *thinking styles*, are particularly relevant in this regard because they determine how people approach information processing.

5. Thinking styles

It has already been demonstrated that *need for cognition* (NFC) is a strong predictor of young people's general credibility concerns regarding online information and their ability to evaluate such information correctly (Metzger, Flanagin, Markov, Grossman, & Bulger, 2015). NFC has been defined as "people's tendency to engage in and enjoy effortful cognitive endeavors" (Cacioppo, Petty, Feinstein, & Jarvis, 1996, p. 197). It is assumed that individuals with a high NFC tend to reflect on and think about all available information. In contrast, individuals with a low NFC are assumed to process information solely based on social comparison, cognitive heuristics, or other individuals (e.g., experts; Cacioppo et al., 1996). In this vein, it has been shown that people with a high NFC are more likely to take community ratings into account rather than user comments when estimating a general opinion climate (Lee & Jang, 2010) or when community ratings are relevant for completing a related task (Winter & Krämer, 2014). On the contrary, people with a low NFC seem to rely equally on the different forms of user representations when estimating others' opinions (Lee & Jang, 2010) or to rely only on user comments when executing a related task (Winter & Krämer, 2014). Moreover, Lee and Jang (2010) showed that when shaping their personal opinions about a societal topic, people with a low NFC are only affected by user comments and not by numerical user representations.

Lee and Jang (2010) ascribed this pattern of findings to the *exemplification theory*. According to this theory, individuals have an attentional preference for concrete and vivid exemplary descriptions of an event when making judgments about it, because it is cognitively less effortful to process single-view descriptions than to process statistical summaries (Zillmann, 1999). Therefore, it can be assumed that individuals with a high NFC are more likely to consider all available cues that might be relevant for credibility appraisals. Individuals with a low NFC, on the other hand, are considerably less motivated to take numerical representations into account and are thus likely to rely only on user comments. Therefore, the following hypothesis is derived:

Hypothesis 4 (H4): The influence of a numerical credibility rating on recipients' credibility evaluations of false information is greater for people with a high need for cognition than for those with a low need for cognition.

Unlike NFC, *faith in intuition* (FII) is conceptualized as the tendency to unconsciously evaluate information based on instincts, feelings and immediate impressions Epstein, Pacini, Denes-Raj, and Heier (1996). Danziger, Moran, and Rafaely

(2006) demonstrated that people with high FII evaluate information based on the ease with which informational factors come to mind. Such findings indicate that people with high FII only perceive parts of messages and are prone to rely on immediate impressions. In this regard, Metzger et al. (2015) also found that young people with high FII were more likely to trust online information in general. This suggests that such people tend to focus on the information itself and rely on their own immediate impression when making credibility appraisals. Therefore, the following hypotheses are derived:

Hypothesis 5a/b (H5a/b): The influence of (a) a numerical credibility rating and (b) user comments on recipients' credibility evaluations of false information is greater for people with low faith in intuition than for those with high faith in intuition.

6. Sharing information

On social media, large amounts of false information are widely spread, because it is shared and forwarded by individual users and not by automated robots (Vosoughi, Roy, & Aral, 2018). Hence, in order to diminish the dissemination of false information, it seems vital to understand when and why individuals react to social media content in this way. Kim and Dennis (2017) found that participants were more likely to read, comment on, and share a news story on Facebook when they believed that the story was true. Moreover, the authors found that the perceived believability fully mediated the effects of source ratings by a third party on users' actions. Against this background, Costera Meijer and Groot Kormelink (2015) noted that sharing news on Facebook plays an important role in personal impression management and serves a communicative function.

To date, there are few studies focusing on the impact of aggregated user reactions on people's tendency to share informational content on social media. One exemplary study in this field by Li and Sakamoto (2014) showed that participants not only made credibility appraisals based on the collective opinion of other social media users, but also followed the sharing intentions of the crowd – no matter whether the statement was true, false, or debatable. Thus, it seems that individuals take user reactions into account when sharing content in order to decide what will make them look good and what might have negative consequences for their reputation. Accordingly, people would avoid sharing information that is evaluated negatively by the crowd. If people decide to share information based on credibility indications of other users, it is likely that this effect is indirectly affected by their own credibility judgments of the content. Consequently, in line with H1, RQ1 and H2, we expect and ask:

Hypothesis 6 (H6): Compared to no comments, user comments indicating poor credibility of false information diminish individuals' willingness to share the information through the indirect pathway of perceived credibility.

Research Question 2 (RQ2): Do comments endorsing the credibility of false information increase individuals' willingness to share the information through the indirect pathway of perceived credibility?

Hypothesis 7 (H7): A numerical credibility rating influences recipients' willingness to share false information in the direction that is indicated by the rating through the indirect pathway of perceived credibility.

7. Method

The present study employed a 3 (user comments: positive vs. negative vs. no comments) x 3 (numerical credibility rating: positive vs. negative vs. no rating) between-subjects design to investigate the influence of user representations on individuals' perceived credibility of false news and their willingness to share it. In an online survey, participants were exposed to a mockup Facebook news post containing false information, which was fabricated for the purpose of the study, and corresponding user reactions (comments and ratings). The respective mockup was varied according to the experimental conditions. The local ethical review board approved the study procedure. The study was conducted in January 2018.

7.1 Sample

Most of the participants were recruited via Facebook groups and surveycircle.com, an online platform for scientific surveys. As an incentive to take part in the study, participants were offered the chance to enter a lottery to win gift cards from an online retail store. A total of 246 participants completed the survey, of which six datasets were excluded because the stimulus material was viewed for less than 5 seconds. Thus, the final sample consisted of 240 individuals (170 females, 69 males, 1 did not specify gender), aged between 18 and 68 years (M = 27.58, SD = 8.13). The majority of participants were students (60.4%) and employees (24.6%), and most had a high school diploma or higher (88.4%). Participants also rated their social media habits with respect to different behaviors (from 1 = never to 7 = very often for all scales). Most of the respondents stated that they used social media (M = 5.55, SD = 1.56). Moreover, the majority of the sample indicated reading news via social media on a regular basis (M = 4.44, SD = 1.77). However, sharing content publicly on social media was less habitual among participants (M = 2.28, SD = 1.36), and their general commenting behavior was also quite low (M = 2.77, SD = 1.60). This also applied for the more specific commenting behavior regarding news on social media (M = 1.85, SD = 1.30).

7.2 Procedure

Participants received a short briefing at the beginning of the survey, informing them that they would take part in a study about online news. Next, they were randomly assigned to one of the experimental conditions and exposed to a mockup news post which contained false information. In the conditions with numerical credibility rating, participants received additional information stating that Facebook had introduced the use of star ratings for different contents (e.g., credibility ratings for news or quality ratings for products). It was explained that these ratings depict the average value of other users' evaluations and that a higher number

of blue stars indicates a better appraisal. After reading the news post, participants answered the questions regarding their credibility assessment and sharing intentions (the dependent measures). Subsequently, they were exposed to a retraction and were informed that the presented news post had been fictitious. Following this, they answered some further questions concerning several control variables, the perception of the source, NFC, and FII as well as person-related characteristics such as sociodemographic information (gender, age, level of education). In addition, the questionnaire also assessed participants' social media self-efficacy, their willingness to interact with the post (e.g., to comment on or read the post), and their recall accuracy of the post's content. However, these latter variables were not relevant for the hypotheses and were therefore not considered in the present analysis. On the final page of the survey, a debriefing was provided. On average, the full procedure took 13 minutes.

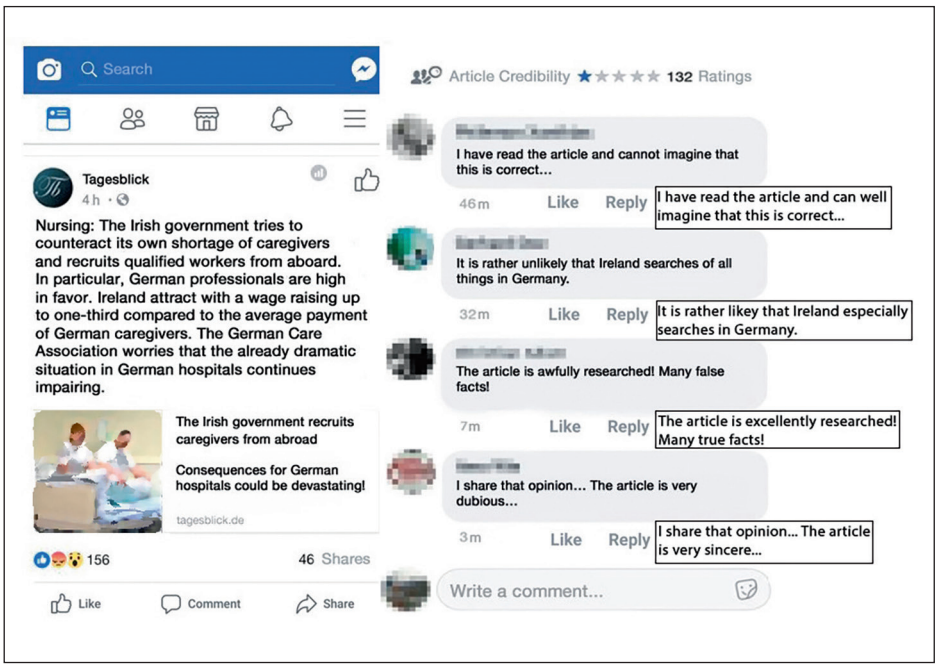
7.3 Stimulus material

Since Facebook has been assumed to benefit the propagation of false information in the form of news to a particularly high degree (e.g., Allcott & Gentzkow, 2017), we tested our assumptions within this setting. To choose appropriate stimulus material for the study, an online pretest (N = 61) was conducted. To find a topic for false news and a name of an online news outlet which evoke uncertainty regarding credibility judgments, the pretest participants were exposed to five fictitious news stories and were asked to rate the false articles' credibility, the topic relevance, and to indicate their prior knowledge regarding the topic of the articles. Moreover, they were asked to rate the trustworthiness and familiarity of eleven news outlet names. For the main study, a topic with a medium level of credibility, high level of topic relevance, and low level of prior knowledge was selected. The chosen fictitious news outlet's name was rated with a medium level of trustworthiness and low levels of familiarity by the pretest participants.

The stimulus material for the main study was constructed based on the results of our pretest: The mockup news post claimed that the Irish government was trying to entice nurses away from Germany. The fictitious news outlet of the post was named "Tagesblick" ("Daily View") and the story was written in the contemporary form of a Facebook news post. The number of likes and shares was kept constant across all conditions.

For the variation of the numerical credibility rating, a star rating was placed underneath the news post, which was introduced as a new feature for user reactions. It was labeled "article credibility" and participants were either exposed to a negative credibility rating (one blue star) or a positive credibility rating (four blue stars). The number of ratings was kept constant at 132. In the control groups, no star rating was presented. Moreover, depending on the experimental condition, participants were either exposed to four user comments that indicated trust in the content of the news article (positive valence) or to four user comments that expressed doubts in the content (negative valence). In the control groups, the content of the four comments was blurred. The comments were designed such that they explicitly expressed trust or mistrust. The stimulus material is pictured in Figure 1.

Figure 1. Visualization of the stimulus material showing the article on the left side and the related user reactions on the right side. Only negative user reactions are displayed in the Facebook design; positive user comments are displayed in the additional frames. Stimulus material was translated from German to English for the purpose of illustration. The symbolic picture was blurred for publication.



7.4 Measures

Credibility. Participants rated the mockup Facebook news post according to the following eight pairs of adjectives, which were adapted from literature on information credibility (e.g., Appelman & Sundar, 2016; Meyer, 1988; van der Kaa & Krahmer, 2014): "unreliable – reliable", "dishonest – honest", "inaccurate – accurate", "imprecise – precise", "based on false facts – based on facts", "doubtful – credible", "unconvincing – convincing", "dubious – serious", and "not trustworthy – trustworthy" (on a semantic differential from 1 to 7). The internal consistency was high (Cronbach's $\alpha = 0.94$).

Willingness to share. To measure participants' willingness to share the mockup Facebook news post, two single items were used. One item asked how likely participants would be to share the post privately via social media (e.g., via Facebook Messenger or WhatsApp) and the other asked how likely they would be to share the post publicly on social media (e.g., via Facebook newsfeed or on Twitter). Both items were measured on a 7-point Likert scale (from 1 = *highly unlikely* to 7 = *highly likely*). We distinguished between these two forms of sharing because the varying number of potential recipients might have a crucial impact on the intentions to share the false news (Oeldorf-Hirsch & Sundar, 2015).

The Rational-Experiential Inventory (REI). A German version of the REI (Keller, Bohner, & Erb, 2000) was used to operationalize need for cognition (NFC) and faith in intuition (FII). The 14-item NFC scale measures the degree of enjoyment of and engagement in cognitive activities. Example items include: "The notion of thinking abstractly is not appealing to me." or "Thinking is not my idea of fun." The 15-item FII scale measures the degree of people's confidence in their initial impressions and feelings as a basis for subsequent behavior. Example items include: "I am a very intuitive person" or "I am quick to form impressions about people." All items were rated on a 7-point Likert scale (from 1 = *does not apply at all* to 7 = *definitely applies*). For both scales, the internal consistency was high (Cronbach's $\alpha = 0.88$ for each scale).

Source perception and control variables. In addition to the key dependent measures, some potentially influencing variables were assessed. One single item asked about participants' perceived familiarity with the news source (scale from 1 = *not familiar at all* to 7 = *very familiar*). To check whether the source credibility was also affected by our manipulations, a second item asked how trustworthy participants considered the source to be (scale from 1 = *not trustworthy at all* to 7 = *very trustworthy*). The general attitude towards the Irish government was assessed on a 7-point Likert scale (from 1 = *very negative* to 7 = *very positive*). To assess participants' perceived relevance of the topic of the mockup news post to their own lives, we asked: "How important is the issue 'nursing in Germany' for you personally?" Participants answered on a 7-point Likert scale (from 1 = *very unimportant* to 7 = *very important*). Moreover, we assessed how high respondents estimate their prior knowledge regarding the article's topic to be (from 1 = *very low* to 7 = *very high*).

Manipulation check. As a manipulation check, in the comments condition, participants were asked: "When you think of the comments below the message, how did the commenters evaluate the corresponding message?" Similarly, in the rating conditions, participants were asked: "When you think of the star ratings regarding the article's credibility, how was the message rated on the credibility scale by other Facebook users?" Both questions were measured on a 7-point Likert scale (from 1 = *not credible at all* to 7 = *very credible*). Participants perceived the credibility rating by other users in the negative rating condition to be lower ($M = 2.69$, $SD = 1.78$) than participants in the positive rating condition ($M = 4.98$, $SD = 1.41$). This effect was significant ($t(158) = -9.01$, $p < .001$, $d = -1.43$). Likewise, participants in the negative comment condition perceived the valence of the comments regarding the message credibility as more negative ($M = 1.66$, $SD = 0.93$) than participants in the positive comment condition ($M = 5.60$, $SD = 1.55$). Again, this effect was significant ($t(158) = -19.43$, $p < .001$, $d = -3.08$).

8. Results

All analyses were conducted using IBM SPSS 25 and IBM SPSS Amos 25. To test our conceptual model, a path analysis with manifest variables and maximum likelihood estimation was conducted. Indirect effects were tested using bias-corrected bootstrapping with 5,000 resamples (95% confidence interval).

8.1 Preliminary analysis

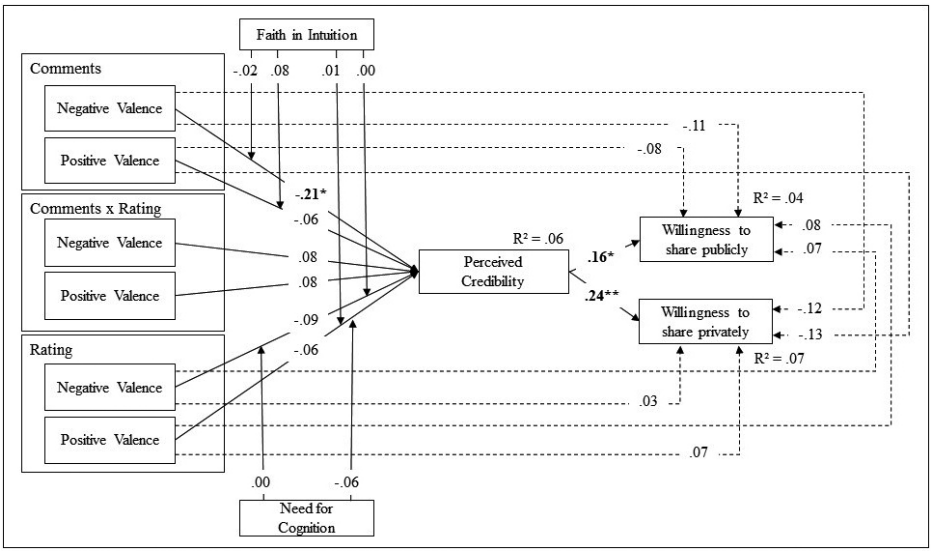
To test whether the control variables as well as perceived source credibility were constant among the experimental conditions, a two-way multivariate analysis of variance (MANOVA) was conducted with the between-subject variables including type of numerical rating as well as type of user comment as independent variables. No significant overall effect emerged regarding the different types of numerical ratings (negative vs. positive vs. none; *Wilks'* $\lambda = 0.97$, $F(10, 454) = 0.47$, $p = .801$, $\eta^2 = .013$). The means of perceived source credibility ($F(2, 231) = 0.30$, $p = .739$, $\eta^2 = .003$), perceived source familiarity ($F(2, 231) = 0.65$, $p = .525$, $\eta^2 = .006$), perceived personal relevance ($F(2, 231) = 0.08$, $p = .925$, $\eta^2 = .001$), prior knowledge ($F(2, 231) = 0.63$, $p = .532$, $\eta^2 = .005$), and attitudes towards the Irish government ($F(2, 231) = 1.28$, $p = .281$, $\eta^2 = .011$) did not differ between the numerical rating conditions. Likewise, there was no main effect for the presence of different types of user comments (negative vs. positive vs. none; *Wilks'* $\lambda = 0.96$, $F(10, 454) = 0.93$, $p = .508$, $\eta^2 = .020$). No differences emerged between the user comments conditions regarding the dependent variables perceived source credibility ($F(2, 231) = 1.13$, $p = .324$, $\eta^2 = .010$), perceived source familiarity ($F(2, 231) = 2.08$, $p = .128$, $\eta^2 = .018$), perceived personal relevance ($F(2, 231) = 1.04$, $p = .354$, $\eta^2 = .009$), prior knowledge ($F(2, 231) = 0.79$, $p = .455$, $\eta^2 = .007$), and attitudes towards the Irish government ($F(2, 231) = 1.29$, $p = .278$, $\eta^2 = .011$). Moreover, no interaction effects between type of numerical rating and type of user comments emerged (*Wilks'* $\lambda = 0.96$, $F(20, 753.82) = 0.45$, $p = .983$, $\eta^2 = .010$).

8.2 Main analysis

As the independent variables were multicategorical, indicator dummy coding as described by Hayes and Montoya (2017) was employed. The indicator method codes the groups in which a case arises as "1" and all other cases as "0." The influence of the treatment variables was tested relative to the corresponding control group.

The fit of the model was assessed by indices and cut-off criteria recommended by Byrne (2010) and Hu and Bentler (1999): $\chi 2$ should be non-significant, the Tucker-Lewis Index (TLI) and comparative fit index (CFI) should be above .95, the root mean square error of approximation (RMSEA) should lie below .06, and the standardized root mean squared residuals (SRMR) should not exceed .05. The indices suggest a good fit: χ^2 (62) = 64.379, $p = .393$, $\chi^2/df = 1.04$, CFI = 1.00 TLI = 1.00, RMSEA = .01 (90% CI = [.00, .04]), SRMR = .04. The model is visualized in Figure 2.

Figure 2. Path model (*N* = 240). Standardized effect sizes are reported.



*p < .05; **p < .01

H1 assumed that comments indicating negative credibility would reduce participants' perceived credibility of the mockup Facebook news article, and H2 assumed that a numerical credibility rating system would influence participants' credibility perception in the direction of its valence. Regarding the interplay of numerical credibility ratings and user comments, H3 stated that credibility ratings would have a stronger effect on participants' perceived credibility when accompanied by user comments in the same direction. Supporting H1, comments with a negative valence negatively affected participants' perceived article credibility relative to no comments (β = –.21, p = .015). To answer RQ2, we also tested the effect of positive comments on participants' perceived credibility. The data showed that comments with a positive valence did not influence the perceived article credibility relative to the control condition (β = –.06, p = .463).

Contrary to the assumptions of H2, participants' perceived article credibility was not affected either by negative credibility ratings (β = –.09, p = .376) or by positive credibility ratings (β = –.06, p = .496) relative to no credibility ratings. Therefore, H2 had to be rejected. Concerning the interaction between comments and ratings, no significant interaction effect was found. The perceived article credibility was not affected when comments and ratings had the same valence (positive: β = .08, p = .446; negative: β = .08, p = .349). Thus, H3 had to be rejected.

In the next step, the moderation hypotheses (H4, H5a/b) were addressed. The moderation variables NFC and FII were mean-centered and multiplied by the dummy-coded independent variables to obtain the interaction terms. H4 assumed that the influence of a numerical credibility rating on participants' perceived cred-

ibility of false information would be stronger for people with a high NFC than for those with a low NFC, and H5a/b expected FII to moderate the relationship between the independent variables and perceived article credibility. Contrary to H4, the interaction of NFC with negative credibility ratings (β = –.00, p = .958) and with positive credibility ratings (β = –.06, p = .459) did not significantly affect participants' perceived article credibility. Concerning H5a/b, none of the interactions between FII and the independent variables had a significant impact on participants' credibility evaluations (negative comments x FII: β = –.02, p = .808, positive comments x FII: β = .08, p = .462, negative ratings x FII: β = .00, p = .982, and positive ratings x FII: β = .01, p = .901). Thus, H5a/b had to be rejected. Overall, 6% of the variance in perceived credibility was explained by the independent variables.

H6 assumed that participants would be less likely to share false information when it is accompanied by user comments with a negative valence, through the indirect pathway of perceived credibility. RQ2 asked whether positive comments increase the willingness to share the article through the indirect pathway of perceived credibility. H7 proposed that a numerical credibility rating would influence recipients' willingness to share false information in the direction indicated by the rating through the indirect pathway of perceived credibility. The data revealed that perceived credibility significantly affected participants' willingness to share the content publicly (β = .16, p = .044) and privately (β = .24, p = .002). In contrast, negative comments had no direct effect on participants' public (β = –.11, p = .116) or private (β = –.12, p = .120) sharing intentions (compared to the control group). However, negative user comments had an indirect negative effect on the willingness to share the article publicly (β = –.03, p = .028, 95% CI [-0.09, -0.003]) and an indirect negative effect on the willingness to share the article privately (β = –.05, p = .009, 95% CI [-0.11, -0.01]) via perceived credibility. Positive user comments did not directly affect participants' willingness to share the article publicly (β = –.08, p = .346) or to share the article privately (β = –.13, p = .100), and also did not indirectly affect the public (β = –.01, p = .304, 95% CI [-0.6, 0.01]) or private (β = –.02, p = .386, 95% CI [-0.07, 0.03]) sharing intentions via perceived credibility.

Negative credibility ratings did not directly affect the intention to share the article publicly (β = .07, p = .314) or privately (β = .03, p = .674) and did not indirectly affect participants' sharing intentions via perceived credibility (sharing publicly: β = –.01, p = .224, 95% CI [-0.06, 0.01] and sharing privately: β = –.02, p = .288, 95% CI [-0.07, 0.02]). Likewise, positive ratings did not directly influence participants' willingness to share the post publicly (β = .08, p = .273) or privately (β = .07, p = .329), and also did not indirectly affect the willingness to share the article publicly (β = –.01, p = .323, 95% CI [-0.05, 0.01]) or privately (β = –.01, p = .399, 95% CI [-0.06, 0.03]). Thus, H6 was supported, whereas H7 had to be rejected. The results suggest that negative user comments indirectly affected the willingness to share the mockup news article (publicly and privately) via the perceived credibility. Overall, 4% of the variance in the public sharing intention and 7% of the variance in the private sharing intention was explained by the model.

9. Discussion

Since false information on social media has become an urgent problem, the aim of the current study was to examine whether textual and numerical user representations can be useful counteragents against false information on social media. By manipulating bandwagon cues related to a mockup Facebook news post containing false information, the study showed that explicit credibility evaluations can reduce the believability of the news post and indirectly reduce the willingness to share it. However, this effect was only found for user comments concerning the credibility of the news post, while numerical credibility ratings did not affect participants' perceived credibility. In the following, the current findings are discussed against the background of the theoretical assumptions.

The first set of hypotheses addressed participants' perception of credibility regarding the content of a false news post. In line with the assumption of H1, comments doubting the credibility of the article reduced participants' credibility appraisals relative to the appraisals of participants were not exposed to comments. In contrast, comments with a positive valence did not enhance the credibility perceptions (RQ1). As suggested in the theoretical section, this pattern of findings can be explained by the negativity bias (Baumeister et al., 2001): Comments contesting the news content might be processed more thoroughly because they arouse more attention and because negative impressions are quicker to form than positive ones. However, it might also be that people consider trusting potentially false information to be more detrimental than distrusting potentially valid information. In this case, user comments which express concerns regarding the credibility of an article might be more heavily weighted by the recipient than positive user comments. In contrast, when others express endorsement, people might not be sufficiently convinced and rather take other cues into account. Moreover, the literature indicates that people consider negative user statements as more honest than positive ones. User comments which are entirely positive might be particularly likely to evoke reactance in the recipient and render the related message rather non-credible (Metzger et al., 2010).

Contrary to our assumptions, participants did not rely on numerical user ratings as a more representative bandwagon cue (H2). Although we assumed that such credibility ratings would be more unambiguous than likes and shares when assessing other people's appraisals, neither positive credibility ratings nor negative credibility ratings affected credibility perceptions relative to no ratings. There are several reasons why our participants might have been less prone to attend to the numerical rating. First of all, in line with exemplification theory (Zillmann, 1999), people might show an attentional preference for concrete comments when making judgments about social media content, as opposed to aggregated user feedback, because such comments provide vivid exemplars which can be processed with less effort. Following the assumptions of the LC3MP (Lang, 2000), people save mental resources by disregarding numerical user presentations, because they have already obtained a satisfactory outcome by reading the related comments. Notably, the manipulation check indicated that the numerical credibility rating did not attract participants' attention in the same way as did user comments.

Nevertheless, other studies found that aggregated user representations do influence Internet users' attitudes towards products and product reviewers (e.g., Walther et al., 2012), purchasing decisions (Metzger et al., 2010), and credibility judgments of online information (Li & Sakamoto, 2014; Winter & Krämer, 2014). This being said, it appears that quantitative information (e.g., user ratings) is predominant over exemplar information (e.g., comments) only when it has greater diagnostic relevance and when the exemplar information is comparatively uninformative (Zillmann & Brosius, 2012, p. 86). Neither of these criteria apply to the current study, in which the employed comments were rather brief and did not contain any arguments. People may process comments differently when they are more complex (e.g., when they provide additional information) and may also rely more on numerical user representations when they are accompanied by such comments. Moreover, in the current study, the numerical rating was integrated as a new feature in the otherwise familiar Facebook environment. As such, participants might have been skeptical of or inattentive to this new feature, because evaluating new features costs a certain amount of cognitive resources (Al-Debei, Al-Lozi, & Papazafeiropoulou, 2013).

We also expected to find an interaction effect between numerical credibility ratings and user comments insofar as the effect of user ratings would be stronger when accompanied by user comments indicating credibility ratings in the same direction (H3). However, we were unable to demonstrate this interaction effect. It is conceivable that this was due to a lack of effect of user ratings on the credibility perceptions, as participants may have relied rather on the more familiar user representations (i.e., likes and shares) as numerical bandwagon cues. As the number of likes and shares was kept constant across the experimental conditions in the present study, it was not possible to analyze the effects of these cues.

Furthermore, contrary to the assumption that people with a high NFC would be more likely to take numerical credibility ratings into account than those with a low NFC (H4), the data showed no significant interaction between user ratings and NFC. This may be due to a ceiling effect, because participants' NFC was generally high across the experimental conditions. Since participants in the current study were exposed to a comprehensive preview text, it might be suggested that those participants with a high NFC focused more on the content of the message than on user representations. Accordingly, participants might have considered the informational content itself to be the most valuable information rather than user representations (Cacioppo et al., 1996).

It was also assumed that people with high FII would be less likely to take other users' reactions into account when making credibility evaluations (H5a/b). Again, no such moderation effects were found. Since the participants generally showed a high NFC and the two thinking styles can co-occur (Epstein et al., 1996), it is possible that participants with high FII revised their initial impression of the article once they cognitively reflected on their intuitive impressions upon looking at the negative comments (see also Alós-Ferrer & Hügelschäfer, 2016).

The next set of hypotheses addressed participants' willingness to share the mockup news post (H6, H7). In line with our assumptions, the results revealed that negative user comments negatively affected participants' willingness to share

the mockup news article publicly and privately through the indirect pathway of perceived credibility (H6), but not on a direct path. With regard to RQ2, positive user comments did not affect participants' public or private sharing intentions via credibility perceptions. However, the results demonstrate that perceived credibility is an important predictor of individuals' public and private sharing intentions. As Costera Meijer and Groot Kormelink (2015) explain, most people are afraid of negative responses from others when sharing informational content on social media. This tendency might be reinforced when people have doubts about the credibility of contents, because they worry about their reputation when sharing false or doubtful news. However, it is also possible that people think that doubtful news is not worth sharing with others, or they simply do not want to misinform others. In contrast to negative user comments, numerical credibility ratings did not influence participants' sharing intentions – either directly or indirectly (H7). Since credibility ratings did not affect the mediating variable perceived credibility, it is unsurprising that negative ratings did not indirectly affect participants' willingness to share the content.

9.1 Limitations and future research

The current study has some limitations that should be acknowledged when interpreting the results. First, participants were only exposed to one false news post, which concerned only one topic. This might decrease the generalizability of the findings and might have diminished the ecological validity. Thus, future investigations should expose participants to different pieces of false information in order to increase the generalizability of the results. Second, the manipulated user comments were consistently either in a positive or in a negative direction. Future studies should integrate balanced user comments into the manipulation of the stimulus material. Moreover, future work should examine the effect of user representations for attitude-congruent and attitude-incongruent topics, because people's stance on an issue might affect their credibility judgments of false information. Third, due to the cross-sectional nature of the present study, participants were only exposed to the user representations once during the experiment. Hence, the participants might have been too unfamiliar with the credibility rating to make proper use of it. Fourth, since the source of information is also an important heuristic for people's credibility evaluations (e.g., Sundar, 2008), we need to acknowledge that the news source in our study was entirely fictitious. Future studies should examine how bandwagon cues affect credibility evaluations of false information when it is presented as stemming from a known "alternative" news source as compared to a well-known serious news source, because this might elicit different effects.

Finally, the sample consisted of mainly young, female, and highly educated individuals. Subsequent research should examine the influence of user reactions on other users' credibility appraisals of information by drawing on a more heterogeneous sample. Moreover, as our sample size was relatively small, the analyses had limited statistical power, which might have affected the significance in our path model.

9.2 Practical implications

Although we expected user-generated numerical credibility ratings of online news articles to be successful in counteracting false information on social media, we were unable to demonstrate that people rely on these ratings when making credibility appraisals. This suggests that numerical credibility ratings do not appear to provide an advantage over current interventions (e.g., fact-checkers). Therefore, researchers and social media providers should not overestimate the effectiveness of such interventions. However, although not necessarily accurate, comments expressing concerns about the veracity of a news post influence people's credibility evaluations of false news, whereas comments expressing trust in the story do not affect such evaluations. This finding can be viewed from two different perspectives: From the perspective of false information, our results can be regarded as beneficial, as we found that comments expressing credibility evaluations only diminished the believability of a false news story when they were negatively valenced, but did not enhance the believability when they were positively valenced. From the perspective of serious news, however, this finding can also be considered as detrimental, as comments which shed doubt on information posted by serious news outlets might also cause harm by threatening the perception of the outlets' trustworthiness. Thus, on the one hand, it would make sense to encourage social media users to comment on their credibility evaluations when reading doubtful information from a suspicious source, thus enhancing other users' skepticism and reducing the potential impact of false information. On the other hand, news outlets should be aware that comments addressing the believability of their stories might undermine their credibility. Therefore, further research should investigate how to utilize comments to their full potential in order to diminish the impact of false news without being harmful for real news (e.g., by users demanding arguments or external sources for assertions).

9.3 Conclusion

Overall, in the current study, numerical bandwagon cues largely failed to affect participants' credibility appraisals regarding the message of a false news post and are consequently not a promising counteragent against false information on social media. Yet, negative user comments had a negative effect on people's credibility appraisals in the social media environment and, moreover, indirectly reduced the likelihood of sharing the false news content. Our study shows that people do not perceive false news on social media in isolation; on the contrary, user comments that explicitly express disbelief in a story can support decisions regarding what is true and what is not true. This finding should be considered when designing future interventions to counter false information on social media. However, researchers and serious news outlets should be aware that user comments expressing doubts about the veracity of a news item also have the potential to reduce the perceived credibility of real news.

References

Al-Debei, M. M., Al-Lozi, E., & Papazafeiropoulou, A. (2013). Why people keep coming back to Facebook: Explaining and predicting continuance participation from an extended theory of planned behaviour perspective. *Decision Support Systems*, *55*(1), 43–54. https://doi.org/10.1016/j.dss.2012.12.032

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Alós-Ferrer, C., & Hügelschäfer, S. (2016). Faith in intuition and cognitive reflection. *Journal of Behavioral and Experimental Economics*, *64*, 61–70. https://doi.org/10.1016/j.socec.2015.10.006

Appelman, A., & Sundar, S. S. (2016). Measuring message credibility: Construction and validation of an exclusive scale. *Journalism & Mass Communication Quarterly*, *93*(1), 59–79. https://doi.org/10.1177/1077699015606057

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370. https://doi.org/10.1037//1089-2680.5.4.323

Bode, L., & Vraga, E. K. (2015). In related news, that was wrong: The correction of misinformation through related stories functionality in social media. *Journal of Communication*, *65*(4), 619–638. https://doi.org/10.1111/jcom.12166

Bode, L., & Vraga, E. K. (2018). See something, say something: Correction of global health misinformation on social media. *Health Communication*, *33*(9), 1131–1140. https://doi.org/10.1080/10410236.2017.1331312

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming, second edition*. New York, NY, USA: Routledge Taylor & Francis Group. https://doi.org/10.4324/9780203805534

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, *119*(2), 197–253. https://doi.org/10.1037/0033-2909.119.2.197

Chaiken, S. (1987). The heuristic model of persuasion. In M. P. Zanna, J. Olson, & C. P. Herman (Eds.), *Social influence: The Ontario symposium* (pp. 3–39). Hillsdale, NJ: Lawrence Erlbaum Associates.

Costera Meijer, I., & Groot Kormelink, T. (2015). Checking, sharing, clicking and linking. *Digital Journalism*, *3*(5), 664–679. https://doi.org/10.1080/21670811.2014.937149

Danziger, S., Moran, S., & Rafaely, V. (2006). The influence of ease of retrieval on judgment as a function of attention to subjective experience. *Journal of Consumer Psychology*, *16*(2), 191–195. https://doi.org/10.1207/s15327663jcp1602_9

Davis-Stober, C. P., Budescu, D. V, Dana, J., & Broomell, S. B. (2014). When is a crowd wise? *Decision*, *1*(2), 79–101. https://doi.org/10.1037/dec0000004

Dohle, M. (2018). Recipients' assessment of journalistic quality: Do online user comments or the actual journalistic quality matter? *Digital Journalism*, *6*(5), 563–582. https://doi.org/10.1080/21670811.2017.1388748

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive–experiential and analytical-rational thinking styles. *Journal of Personality and Social Psychology*, *71*(2), 390–405. https://doi.org/10.1037/0022-3514.71.2.390

Fetzer, J. H. (2004). Information: Does it have to be true? *Minds and Machines*, *14*(2), 223–229. https://doi.org/10.1023/B:MIND.0000021682.61365.56

Flanagin, A. J. (2017). Online social influence and the convergence of mass and interpersonal communication. *Human Communication Research*, *43*(4), 450–463. https://doi.org/10.1111/hcre.12116

Flanagin, A. J., & Metzger, M. J. (2008). Digital media and youth: Unparalleled opportunity and unprecedented responsibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 5–27). Cambridge, MA, USA: MIT Press. https://doi.org/10.1162/dmal.9780262562324.005

Flanagin, A. J., & Metzger, M. J. (2013). Trusting expert- versus user-generated ratings online: The role of information volume, valence, and consumer characteristics. *Computers in Human Behavior*, *29*(4), 1626–1634. https://doi.org/10.1016/j.chb.2013.02.001

Gil de Zúñiga, H., Weeks, B., & Ardèvol-Abreu, A. (2017). Effects of the news-finds-me perception in communication: Social media use implications for news seeking and learning about politics. *Journal of Computer-Mediated Communication*, *22*(3), 105–123. https://doi.org/10.1111/jcc4.12185

Hameleers, M., & van der Meer, T. G. L. A. (2019). Misinformation and polarization in a high-choice media environment: How effective are political fact-checkers? *Communication Research*. https://doi.org/10.1177/0093650218819671

Hayes, A. F., & Montoya, A. K. (2017). A tutorial on testing, visualizing, and probing an interaction involving a multicategorical variable in linear regression analysis. *Communication Methods and Measures*, *11*(1), 1–30. https://doi.org/10.1080/19312458.2016.1271116

Holton, A. E., & Chyi, H. I. (2012). News and the overloaded consumer: Factors influencing information overload among news consumers. *Cyberpsychology, Behavior, and Social Networking*, *15*(11), 619–624. https://doi.org/10.1089/cyber.2011.0610

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jucks, R., & Thon, F. M. (2017). Better to have many opinions than one from an expert? Social validation by one trustworthy source versus the masses in online health forums. *Computers in Human Behavior*, *70*, 375–381. https://doi.org/10.1016/j.chb.2017.01.019

Karlova, N. A., & Fisher, K. E. (2013). A social diffusion model of misinformation and disinformation for understanding human information behaviour. *Information Research*, *18*(1), 1–17. Retrieved from http://www.informationr.net/ir/18-1/paper573.html

Keller, J., Bohner, G., & Erb, H.-P. (2000). Intuitive und heuristische Verarbeitung – verschiedene Prozesse? Präsentation einer deutschen Fassung des 'Rational–Experiential Inventory' sowie neuer Selbstberichtskalen zur Heuristiknutzung [Intuitive and heuristic judgment – different processes? Presentation of a German version of the Rational–Experiential Inventory and of new self-report scales of heuristic use]. *Zeitschrift Für Sozialpsychologie*, *31*(2), 87–101. https://doi.org/10.1024//0044-3514.31.2.87

Kim, A., & Dennis, A. R. (2017). Says who? How news presentation format influences perceived believability and the engagement level of social media users. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2987866

Kümpel, A. S., & Springer, N. (2016). Commenting quality: Effects of user comments on perceptions of journalistic quality. *Studies in Communication and Media*, *5*(3), 353–366. https://doi.org/10.5771/2192-4007-2016-3-353

Lang, A. (2000). The limited capacity model of mediated message processing. *Journal of Communication*, *50*(1), 46–70. https://doi.org/10.1111/j.1460-2466.2000.tb02833.x

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., … Zittrain, J. L. (2018). The science of fake news. *Science*, *359*(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

Lee, E.-J., & Jang, Y. J. (2010). What do others' reactions to news on internet portal sites tell us? Effects of presentation format and readers' need for cognition on reality perception. *Communication Research*, *37*(6), 825–846. https://doi.org/10.1177/0093650210376189

Li, H., & Sakamoto, Y. (2014). Social impacts in social media: An examination of perceived truthfulness and sharing of information. *Computers in Human Behavior*, *41*, 278–287. https://doi.org/10.1016/j.chb.2014.08.009

Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology*, *58*(13), 2078–2091. https://doi.org/10.1002/asi.20672

Metzger, M. J., Flanagin, A. J., Markov, A., Grossman, R., & Bulger, M. (2015). Believing the unbelievable: Understanding young people's information literacy beliefs and practices in the United States. *Journal of Children and Media*, *9*(3), 325–348. https://doi.org/10.1080/17482798.2015.1056817

Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication*, *60*(3), 413–439. https://doi.org/10.1111/j.1460-2466.2010.01488.x

Meyer, P. (1988). Defining and measuring credibility of newspapers: Developing an index. *Journalism Quarterly*, *65*(3), 567–574. https://doi.org/10.1177/107769908806500301

Neubaum, G., Rösner, L., Ganster, T., Hambach, K., & Krämer, N. C. (2018). United in the name of justice: How conformity processes in social media may influence online vigilantism. *Psychology of Popular Media Culture*, *7*(2), 185–199. https://doi.org/10.1037/ppm0000112

Oeldorf-Hirsch, A., & Sundar, S. S. (2015). Posting, commenting, and tagging: Effects of sharing news stories on Facebook. *Computers in Human Behavior*, *44*, 240–249. https://doi.org/10.1016/j.chb.2014.11.024

Pennycook, G., Cannon, T. D., & Rand, D. G. (2017). Prior exposure increases perceived accuracy of fake news. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2958246

Pentina, I., & Tarafdar, M. (2014). From "information" to "knowing": Exploring the role of social media in contemporary news consumption. *Computers in Human Behavior*, *35*, 211–223. https://doi.org/10.1016/j.chb.2014.02.045

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 19, pp. 123–205). New York, NY, USA: Academic Press. https://doi.org/10.1016/S0065-2601(08)60214-2

Prochazka, F., Weber, P., & Schweiger, W. (2018). Effects of civility and reasoning in user comments on perceived journalistic quality. *Journalism Studies*, *19*(1), 62–78. https://doi.org/10.1080/1461670X.2016.1161497

Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 73–100). Cambridge, MA, USA: The MIT Press. https://doi.org/10.1162/dmal.9780262562324.073

Tandoc, E. C., Ling, R., Westlund, O., Duffy, A., Goh, D., & Zheng Wei, L. (2017). Audiences' acts of authentication in the age of fake news: A conceptual framework. *New Media & Society*. https://doi.org/10.1177/1461444817731756

van der Kaa, H., & Krahmer, E. (2014). Journalist versus news consumer: The perceived credibility of machine written news. In *Proceedings of the Computation+Journalism conference*. New York, NY, USA. Retrieved from https://pure.uvt.nl/portal/files/4314960/c

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Waddell, T. F. (2018). What does the crowd think? How online comments and popularity metrics affect news credibility and issue importance. *New Media and Society*, *20*(8), 3068–3083. https://doi.org/10.1177/1461444817742905

Walther, J. B., & Jang, J. (2012). Communication processes in participatory websites. *Journal of Computer-Mediated Communication*, *18*(1), 2–15. https://doi.org/10.1111/j.1083-6101.2012.01592.x

Walther, J. B., Liang, Y., Ganster, T., Wohn, D. Y., & Emington, J. (2012). Online reviews, helpfulness ratings, and consumer attitudes: An extension of congruity theory to multiple sources in Web 2.0. *Journal of Computer-Mediated Communication*, *18*(1), 97–112. https://doi.org/10.1111/j.1083-6101.2012.01595.x

Winter, S., Brückner, C., & Krämer, N. C. (2015). They came, they liked, they commented: Social influence on Facebook news channels. *Cyberpsychology, Behavior, and Social Networking*, *18*(8), 431–436. https://doi.org/10.1089/cyber.2015.0005

Winter, S., & Krämer, N. C. (2014). A question of credibility – Effects of source cues and recommendations on information selection on news sites and blogs. *Communications*, *39*(4), 435–456. https://doi.org/10.1515/commun-2014-0020

Zillmann, D. (1999). Exemplification theory: Judging the whole by some of its parts. *Media Psychology*, *1*(1), 69–94. https://doi.org/10.1207/s1532785xmep0101_5

Zillmann, D., & Brosius, H.-B. (2012). *Exemplification in communication: The influence of case reports on the perception of issues*. Mahwah, NJ: Lawrence Erlbaum Associates.