

## *Do You Really Know If It's True?* How Asking Users to Rate Stories Affects Belief in Fake News on Social Media

Patricia L. Moravec  
Kelley School of Business  
Indiana University  
[pmoravec@indiana.edu](mailto:pmoravec@indiana.edu)

Antino Kim  
Kelley School of Business  
Indiana University  
[antino@indiana.edu](mailto:antino@indiana.edu)

Alan R. Dennis  
Kelley School of Business  
Indiana University  
[ardennis@indiana.edu](mailto:ardennis@indiana.edu)

Randall K. Minas  
Shidler College of Business  
University of Hawai'i at Mānoa  
[rminas@hawaii.edu](mailto:rminas@hawaii.edu)

### Abstract

*The rise of “fake news” has become a major concern for social media platforms. In response, Facebook has proposed and tested the idea of users flagging and rating news articles and sources, much akin to how consumers rate products and services on the Internet. One obvious challenge with this crowdsourced rating approach is whether the users really know enough to rate news articles and sources. Perhaps, a side benefit of asking users to evaluate an article—and asking about their personal experience with the event described in the article—is making them realize that they do not know enough about the event to make an accurate judgment, thus pushing them to become more skeptical. We asked 68 social media users to assess the believability of 42 social media headlines. We found that, while users were generally more likely to believe articles that agreed with their point of view, asking users to rate pushed them to think more critically about the truthfulness of the articles. Moreover, once users had been asked to rate some articles, they remained critical of other articles as well, even without the rating prompt. Overall, our findings suggest that asking users to evaluate the truthfulness of articles may not only produce rating information that can be a useful reference at a later point in time but also have an immediate benefit of alerting users to think more critically about all articles they see.*

### 1. Introduction

Fake news on social media rose to global attention in 2016 during the US presidential election, where disinformation campaigns to influence the election results were widespread [2, 3, 37]. About 60% of adults get news from social media (primarily Facebook), and the proportion is increasing [15], suggesting that the problem will likely worsen [34]. In addition, past research shows that users are poor at assessing whether a news story on social media is real or fake [39] because confirmation bias leads them to believe articles that align with their *a priori* beliefs and disbelieve those that do not [25]. As a result, more fake news articles are shared on social media than real news [39], and fake news also spreads faster than real news [47].

The prevalence of fake news has not only shaken the public's trust in journalism but also stirred up criticism towards social media platforms for not taking more proactive actions to stop the spread of fake news [3]. In 2017, Facebook tried flagging fake news articles based on user reports—which were reviewed by third-party fact-checking organizations—but this proved to be ineffective and was later removed [31]. In early 2018, Facebook announced that it will adopt user ratings of news sources—an Internet staple in determinations of credibility, quality, and value [5]—following the route of eBay and Yelp, which have users rate businesses [10] or Amazon and BestBuy.com, which have users rate products. However, a unique challenge in rating news articles and sources is, do users know enough to rate them? When users rate a business, it is expected that they have actually used the business' products or services and thus have personal knowledge of it. Credible reviews of products are grounded in actual purchase and use of the product. How can users accurately rate news sources unless they have been personally involved with the news articles produced by a news source? Without personal knowledge of the events in a news article, users may lack the basis on which to assess whether the news source has done a good job in reporting on the events. Of course, the users may attempt to aggregate knowledge from additional news sources and references, but that level of knowledge is still one level below personal experience; they still have no direct personal knowledge of the validity of the report. Combine this lack of personal experience together with common biases and noise, user ratings may pose a serious challenge in producing useful information that could help other users assess the credibility of news sources.

Perhaps therein lies a potentially unintended positive consequence. Although there may be several hurdles we must cross before user ratings are able to help other users who would consume the rating information, could prompting users to evaluate the truthfulness of articles—and asking them about their personal knowledge of the events described in the articles—help those same users who are providing the ratings? Will being asked to deliberately form a judgment about the credibility of a news article induce users to think more carefully about whether a social media article is true or

fake? Some users may give little thought to rating and simply provide the first assessment that comes to mind. Other users may pause and realize that they actually have no direct way to assess the credibility of a news article—unless, of course, they have personally experienced the reported event or done extensive unbiased research—and this may cause them to be a bit more skeptical of the articles they see, both those they are asked to rate, as well as those they are not.

In this paper, we investigate whether asking users to rate the credibility of articles they see influences their assessment of those articles, and whether this spills over to influence how credible they perceive other articles to be on social media. Our results show that, while users are still susceptible to confirmation bias, prompting them to evaluate the truthfulness of the articles nudges them to think more critically. We also found support for a carry-over effect; once users are asked to rate some articles, they remain critical toward other articles as well even without the rating prompt. How Facebook can aggregate potentially biased user ratings and present the information in an effective way remains an open question. Still, our work suggests that prompting users to evaluate the articles themselves may be effective in helping users to think more critically about the articles they see on social media.

## 2. Prior Theory and Research

Fake news has been defined as “news articles that are intentionally and verifiably false and could mislead readers” [2]. Disinformation and fake news have long been a problem, but fake news on social media became an important societal issue when it was reported that Russian intelligence agencies used social media to propagate fake news in an attempt to influence the 2016 presidential election in the United States [2, 3]. In response to fake news, a number of fact-checking initiatives have been launched [16, 29], and fact checking has been shown to influence the perceived credibility of an author or an article [48], though the issue is fact-checking articles in a timely manner before any fake news can spread and cause damage. Fact-checking has traditionally been done by experts (e.g., PolitiFact), and Facebook’s fake news flag also leveraged expert fact-checking in addition to user reports.

An alternative solution to expert fact-checking is to have users rate the credibility or truthfulness of news articles and/or news sources. Just as eBay has users rate individual transactions that are aggregated into an overall seller rating, social media platforms could have users rate individual news articles which would then be aggregated to provide an overall news source rating. Alternately, users could be asked to rate the news source

directly, without consideration of a specific news article. These source ratings would then be applied to articles when they are first published, the same way that sellers on eBay and other e-commerce sites have their ratings applied to all new products they offer for sale. Research shows that source ratings influence the extent to which users believe social media stories [25] and influence online news consumption [4].

Prior research on fact checking and news source ratings has focused on the end result of fact checking and ratings (e.g., whether social media users’ beliefs about article credibility are influenced by fact-checking reports or ratings [48]). However, one unanswered question is whether—and how—asking a user to evaluate articles affects the user doing the assessment. Does the act of rating induce greater skepticism? Does it influence users to pause from undiscerningly going through articles and to view them with more critical eyes? We begin by considering how users process information on social media and the cognitive process that takes place when users are prompted to evaluate articles.

### 2.1. Information processing in social media

People use the Internet for many different purposes, such as accomplishing tasks or seeking hedonistic pleasure [49]. Most individuals use social media for hedonistic purposes [18], such as seeking entertainment or connecting with friends [23], rather than utilitarian purposes, such as completing work tasks. Individuals in a hedonistic mindset may be less likely to consider information critically than those with a utilitarian mindset, as their consumption is tied to enjoyment and pleasure [19].

Facebook strives to maximize user enjoyment; therefore, its algorithms are designed to identify and display content that matches the users’ preferences, so users are likely to see articles that align with their existing beliefs [44]. Such a process causes a decrease in the range of information that the users encounter, and, as a result, Facebook users often exist in information “bubbles”—also referred to as echo chambers [4]—that reinforce their beliefs and make them believe that others around the world are more like them [44]. Therefore, many articles that users encounter on social media are related to topics that users have previously viewed [44] and have already formed an opinion about.

Humans have two very different cognitive processes [24]. Many dual process models have been posed under a host of different names; see Evans [11] for an analysis. In this study, we adopt the commonly used terminology of Keith Stanovich [41] and Daniel Kahneman [24], who call these two distinct cognitive processes System 1 and System 2 cognitions. System 1 cognition is

automatic and runs continuously, involuntarily providing us with conclusions without our conscious thought [24]. System 1 cognition is our ‘fast thinking,’ where our simple heuristics produce perceptions and actions in less than a second [24]. System 1 is our intuitive decision making system [1]; when we have an intuition or a “gut reaction,” that is System 1 talking. The quick nature of System 1 is what enables us to do intuitive tasks without direct thought, such as walk, talk, recognize faces, and effortlessly retrieve certain facts from memory [24].

However, System 1 comes with certain drawbacks. When we process information using System 1, we only use the information immediately at hand with the vividness and saliency of available information driving our decisions, rather than a more nuanced and carefully considered model [6, 24]. The associative memory processing of System 1 is strongly influenced by framing [17], because it quickly searches for confirming evidence of the question posed. The questions “Is Pat friendly?” and “Is Pat unfriendly?” are fundamentally different questions because they trigger our System 1 to retrieve entirely different instances of Pat’s behavior [24]. As long as we can form the information into a coherent story—right or wrong—we are likely to follow our immediate System 1 response [24].

The second process is System 2, which involves a more effortful, deliberate cognition [24]. System 2, or “slow thinking,” takes much more time to arrive at a conclusion and is laborious [24, 27]. We have physiological symptoms that show the effort involved in the process: our pupils dilate, the heart rate changes, the blood pressure rises, and extra blood flows to different areas of the brain that are active [24]. Examples of System 2 processing are doing a standardized computation that requires holding numbers in working memory, monitoring the appropriateness of our behavior in tense social situations, comparing two products for value, and checking the validity of a complex argument [32].

Because System 2 cognition requires more effort than System 1 cognition, humans are predisposed to avoid System 2 unless there is a need for it [24, 42]. We usually adopt the perceptions and actions produced by System 1 unless we are motivated to invest effort, or our System 1 warns us that something is wrong its result through feelings of unease [24].

When users encounter information that aligns with their pre-existing opinions, their System 1 produces an instant confirmation that the information is sensible, and the users are inclined to believe it [9, 26, 33]. When users encounter information that is contrary to their opinions, their System 1 instantly produces a negative reaction because the information does not align with what System 1 knows to be true, resulting in cognitive

dissonance [13]. When an individual is presented with two contradictory facts, both of which are plausible (e.g., John is honest, but a story says he lied), he/she must resolve the inconsistency. This can be done either by concluding that the two facts are not contradictory (e.g., John lied, but he is still honest because lying is not related to honesty) or by accepting one and rejecting the other (e.g., John is honest, and thus, I do not believe he lied; or John lied, and thus, I do not believe he is honest) [13].

Resolving such a cognitive dissonance takes cognitive effort, and humans tend to be cognitive misers who resist invoking System 2 and expending effort [40]. This tendency is exacerbated when humans are in a hedonistic mindset [19]. Because rejecting the new information is simpler cognitively than reassessing one’s pre-existing opinions, most people accept their System 1 instant conclusion; they retain their existing opinion and discard the new information as being false [9, 26, 30]. This tendency to favor information that confirms one’s pre-existing opinions and ignore information that challenges them is called confirmation bias [9, 26, 33]. Thus, people are more likely to believe information that matches their pre-existing opinions (i.e., attitude homophily or alignment [2, 20]). Past research shows that confirmation bias has a significant effect on the belief in articles posted on social media [25].

In summary, we theorize that when a user views a news article on social media, their System 1 cognition will produce an instant assessment of the believability of the article based on its alignment with their prior opinions on the topic. The context of social media is an important one to study, as the influence of confirmation bias while in a hedonistic mindset has not been thoroughly studied, especially with the potential for confirmation bias on social media to grow, causing real, demonstrable changes to our political landscape [39, 45]. Social media users are usually in a hedonistic mindset, so they are unlikely to invest the effort needed to invoke System 2 cognition to override their System 1 results. The net effect is a belief that is heavily influenced by prior opinions. Therefore:

**H1:** *Pre-existing opinions on a topic directly influence the extent to which a news headline is perceived to be believable.*

## 2.2. User rating of news articles

Users are commonly asked to rate products and services on the Internet. Though some ratings are fake, our focus in this paper is on ratings produced in good faith. When we ask users to rate products or services, we expect them to have used the product or service. When users are

prompted to rate, they consider their own experiences and draw upon those experiences to produce the rating.

Now, consider ratings of news articles. Unless the user was actually involved in the events described in a social media article (either as a participant or a witness), the user has no personal experiences to draw upon. Unless the user witnessed the event reported in the article, he/she is incapable of providing a direct assessment of its truthfulness because he/she has no personal knowledge of the facts. Therefore, except in very rare circumstances where the user was a participant or witness of the events in the article, users will be unable to provide a rating based on personal knowledge of events, thus leading to divergent outcomes between individuals in assessing credibility. Now, some users may not have personal experience but are informed on the topic and have triangulated information from various sources. Others may accept the information based solely on their System 1 conclusion on the information. In the case of fake news articles that are complete fabrications, there are no actual events that anyone could have witnessed! The onus is then on individual user to decide to “look into” the issue, accept the article as fact, or reject the article.

We theorize that when users are asked to rate the truthfulness of an article based on their personal experience, they will recognize that they have no personal knowledge on which to base a rating. Since not all users will recognize this without an appropriate prompt, we propose a rating intervention that will explicitly ask users if they have personal knowledge of the events. This explicit question will trigger the recognition from users that they lack the knowledge needed to rate the truthfulness of the article and may result in deeper processing of the underlying issue.

As argued in H1, when users first see an article, their System 1 will produce an instant assessment of its believability based on confirmation bias. When users are then asked to provide a truthfulness rating that asks about personal knowledge of the event reported in the article, their System 1 will produce an instant answer that the user has no such knowledge of the event.

There are several events that can trigger System 2 cognition, and one of the most powerful is a discrepancy from normal expectations [14, 28]. When System 1 detects that something is not normal, it produces a low “Feeling of Rightness” (FOR) that indicates something is amiss [7, 46]. A low FOR is an alert that System 1 believes that System 2 cognition should be invoked for a deeper thinking. We theorize that when an individual’s System 1 is asked whether the user has personal knowledge of the events in the article and returns a negative answer, it will produce a low FOR, which would invoke System 2 cognition.

Different individuals may choose to respond to a low

FOR in different ways. Some individuals may have a high need for cognition and therefore engage System 2 in situations when the FOR is only slightly low, while others may be reluctant to engage in System 2 cognition even when the FOR is very low [12, 22]. Empirical evidence suggests that most people’s System 1 is relatively good at producing an accurate FOR [22]. Whether System 2 is invoked or not depends on the individual’s personality, though a low FOR will be sufficient to invoke at least some amount of System 2 cognition under normal conditions [22]. These users’ System 2 cognition will realize that they lack direct knowledge of the events in the story and thus cannot produce a truthfulness rating that is based on their own experiences as they would a product or service they have used.

Instead, they will realize that any initial assessment of truthfulness will be based on their perception of the credibility of the article’s source and/or on the fit of the article’s content with their prior opinions and indirect knowledge about what is true or false. Both of these are weaker criteria than the individual’s own direct personal knowledge. Thus, users who invoke their System 2 cognition will be less certain of the accuracy of their truthfulness assessment. They will be less likely to believe the news article. As mentioned before, this effect will be most pronounced for fake articles, which is where we most need it to address the issue of misinformation.

An alternate mechanism for the decreased perception of believability of an article comes from increased mindfulness caused by our prompt to rate some headlines. Research on mindfulness suggests that IT mindfulness can be manipulated [43]. The most influential dimension of mindfulness in this research is the “orientation in the present” dimension. Once users become aware that information on social media may be false, they become more mindful of the potential fiction of the news that they see.

In summary, we argue that changing the design of the current Facebook newsfeed will encourage critical thought. Specifically, asking users to evaluate a news article in a way that causes them to consider if they have personal knowledge of the events in the article will trigger a recognition that they do not have direct personal knowledge of the article and thus cannot provide an independent rating. This recognition will increase the likelihood that users will engage in System 2 cognition to assess the article, and the resulting uncertainty will reduce the extent to which they believe the article. Thus:

**H2:** *Asking users to rate a news headline will reduce the extent to which they perceive it to be believable.*

The arguments above consider the direct effects of being asked to rate an article on the believability of that specific article, but are there carry-over effects? Does a simple design change on one article influence believability of future articles? Once a user realizes that he or she lacks the personal knowledge needed to rate an article, will this realization carry over to the articles that follow, regardless of whether the user is asked to rate them or not? Once an individual realizes that he or she lacks the direct personal knowledge to rate a specific article, the individual may generalize this realization to all articles that do not concern the individual's own life. This realization cannot be "unlearned," so once internalized, it will become part of the normal assessment of future articles. Thus, once the user has some experience with being asked to rate articles (and realizing that he or she lacks personal knowledge), then he or she will be less likely to believe any future articles (whether asked to rate or not) because the user will recognize the basis on which one's belief is formed is less certain than a normal assessment. Thus:

**H3:** *There will be a carry-over effect such that, once users have been asked to rate some headlines, they will be less likely to perceive any headlines as believable.*

### 3. Methodology

#### 3.1. Participants

We recruited a total of 68 undergraduate participants from a large business core course. All were between the age of 18 and 24, and about 35% of them were female. We also collected Facebook usage and political affiliation information; see Table 1.

**Table 1: Participant group description**

|                       | Categories            | Percentage |
|-----------------------|-----------------------|------------|
| Facebook Use          | Once a week or less   | 22%        |
|                       | More than once a week | 47%        |
|                       | More than once a day  | 31%        |
| Political Affiliation | Democrat              | 18%        |
|                       | Moderate/Independent  | 57%        |
|                       | Republican            | 25%        |

#### 3.2. Task

The participants viewed 42 news headlines and reported the believability of each article. Of the headlines, 16 were designed to appeal to politically left-leaning participants and 16 to right-leaning participants (see Table 2 for examples). There were also 10 that may be of special interest to our participants because of the locality of the issues the headlines described. All the headlines were formatted as they might appear as posts

on Facebook, some with the user-rating treatment and some without (see Figure 1).

**Table 2: Examples of news headlines used in the experiment**

- Robert Mueller Encouraged to Resign over Affair with Nancy Pelosi
- NFL Players Union Votes to Encourage Kneeling During National Anthem
- Nancy Pelosi Said Building a Wall Will Violate Rights of 'Millions of Illegals'
- Melania Trump Bans White House Staff from Taking Flu Shot
- Senator Tom Cotton Called for Drug Testing for Social Security Recipients
- Trump's Doctor Said the President Was 'Too Sick to Talk to Mueller'



**Figure 1: An example of a news headline (with the user-rating treatment in the dotted box)**

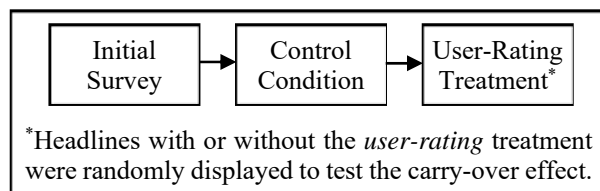
The headlines and images were designed to avoid major differences in the type and magnitude of feelings they would generate. We used a gender-neutral name for the person posting—not to be confused with the original source who authored the article—and the comment from the person posting was a summary of the headline itself. To minimize any news source specific effect (e.g., some

sources well-known and trusted by some users while other sources are not), we fabricated a source name that sounded plausible (*NewsUnion.com*). The URL was verified to be inactive prior to the experiment (i.e., not used by any news provider or anyone else).

### 3.3. Treatments

This is a repeated measures study in which participants received all 42 headlines. Participants received both the control condition (no user-rating) and the experimental treatment (with user-rating and no-rating intermixed). The headlines were randomly assigned to treatments and presented in random order within treatment to control for any headline- and image-specific effects. In a repeated measure design, there is a concern about the effects of an early treatment bleeding over into a later treatment [21, 36]. This is usually controlled by random treatment order or a fully crossed design in which all treatments orders are used equally, except in cases where there are likely to be meaningful theoretical differences in the bleed-over between treatments. This is the case in our study. The control condition is the current Facebook format, so it is unlikely to influence later treatments because it is the normal interface that users regularly use. In contrast, the user-rating treatment is likely to have strong influence on the treatments that follow it, because once users are asked to rate an article, we theorize that they will become more discerning in their consumption of articles; see H3. Thus, randomizing the order of appearance of the rating manipulation and controls would confound the later treatment.

Therefore, the control condition was always presented first, followed by a description of the user-rating treatment, followed by the user-rating treatment. The overall flow of the experiment is shown in Figure 2. As a robustness check, we also test for any ordering effect that may stem from our experiment design; see our results section.



**Figure 2: Sequence of the experiment**

Participants completed an initial survey for demographic information. The first treatment was the control condition (no user ratings) designed to mimic the current Facebook style of presentation as closely as possible. This treatment had 14 randomly assigned headlines.

To test for the carry-over effect (H3), as well as for the user-rating effect (H2), the remaining 28 headlines were randomly presented with or without the user-rating treatment.

### 3.4. Independent variables

Confirmation bias was assessed using two items that were self-reported by participants [25]. The first was the participant's position on the headline ( $-3 =$  extremely negative to  $+3 =$  extremely positive). The second was the participant's perceived importance of the headline (using a 7-point scale: Do you find the issue described in the article important?  $1 =$  not at all,  $7 =$  extremely). In other words, the former represents the *direction* of confirmation bias whereas the latter represent the *magnitude*. Putting the two together, we are able to measure the degree of fit between a headline and a participant's prior position. The two items were multiplied together to form our *Confirmation Bias* variable, which ranges from  $-21$  to  $+21$ . This was used to test H1.

To test our hypotheses, we indicated which articles were displayed with the user-rating prompt (*User Rating*), and which were displayed without it but after our participants had been exposed to the rating prompt (*After Exposure Without Rating*). These headlines appeared exactly as those in the control condition. The former variable was to test H2, whereas the latter was to test H3.

Finally, we also controlled for our participants' demographic factors such as gender, Facebook usage level, and political affiliation.

### 3.5. Dependent variable

The believability of each article was measured using three 7-point items (How believable do you find this article, How truthful do you find this article, How credible do you find this article) [25]. Reliability (Cronbach's alpha) was 0.96, which is adequate.

## 4. Results

To test our hypotheses, we performed multilevel mixed-effects linear regression with random intercepts in Stata. The base case was the control condition. The results are reported in Table 3.

H1 argued that confirmation bias has a positive effect on the believability of articles. Table 3 shows that confirmation bias has a significant positive effect. H1 is supported.

H2 posited that user rating would affect believability. Table 3 shows that User Rating has a negative and significant effect on Believability, indicating that the user-rating treatment influenced the

participants to become more critical about the truthfulness of the headlines they saw. We conclude that H2 is supported.

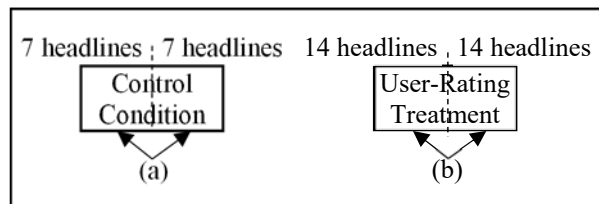
We also hypothesized in H3 that user rating would also have a carry-over effect. On Table 3, *After Exposure Without Rating* variable measures this carry-over effect, and it turns out to be negative and significant. Hence, H3 is also supported. The difference between the coefficient for *After Exposure Without Rating* and that for *User Rating* is statistically insignificant ( $\chi^2(1) = 0.40, p < 0.527$ ), indicating that the carry-over effect is as strong as the effect from the rating prompt. We note that gender, Facebook usage level, and political affiliation had no significant influence on believability.

**Table 3: Estimation Results for Believability<sup>1</sup>**

| Independent Variables                      |           |
|--|-----------|
| User Rating                                | -0.358*** |
| After Exposure Without Rating <sup>†</sup> | -0.312*** |
| Confirmation Bias                          | 0.431***  |
| Female                                     | 0.264     |
| FB Use: More than once a week              | -0.054    |
| FB Use: More than once a day               | 0.076     |
| Democrat                                   | -0.171    |
| Republican                                 | 0.211     |

<sup>†</sup>After Exposure Without Rating refers to those headlines displayed without asking users to rate them (i.e., the control condition) after a headline asking users to rate it was presented.

We also included post hoc tests to provide robustness checks for our results. As described earlier, we could not adopt a fully crossed design in our experiment because of the asymmetric bleed-over effect; placing any control condition after the user-rating treatment is likely to confound the later treatment. The sequence of our experiment (Figure 2) may cast doubts as to whether our findings stem from other ordering effects. For instance, some may wonder whether users simply become more skeptical the more articles they read. To test this, we divided each treatment group into two parts—first half of the headlines and second half—and compared believability; see Figure 3.



**Figure 3: Testing for ordering effect**

<sup>1</sup> Confirmation Bias is standardized for all results.  
\*\*\* $p \leq 0.001$ , \*\* $p \leq 0.01$ , \* $p \leq 0.05$ .

As we can see from Table 4, though the effect of confirmation bias was still very much present, there was no evidence of any ordering effect. This conclusion was consistent be it in the control condition group (see (a) in Table 4) or the user-rating treatment group (see (b) in Table 4). Hence, these tests further strengthen our finding that it was the user-rating treatment that influenced users to think more critically about the truthfulness of the articles, not some other ordering effect.

**Table 4: Estimation Results for Believability**

| Independent Variables         | (a)      | (b)      |
|-------------------------------|----------|----------|
| First Half                    | 0.104    | 0.102    |
| Confirmation Bias             | 0.339*** | 0.483*** |
| Female                        | 0.378*   | 0.213    |
| FB Use: More than once a week | 0.026    | -0.104   |
| FB Use: More than once a day  | 0.220    | -0.001   |
| Democrat                      | -0.219   | -0.149   |
| Republican                    | 0.321    | 0.158    |

## 5. Discussion

From our analyses, we found that participants were more likely to believe articles that they agreed with, proving that confirmation bias is a significant factor that amplifies the effect of fake news on social media. As a possible remedy, our results show that prompting users to evaluate the truthfulness of articles directly influences the extent to which users believe articles on social media. As theorized, the rating prompt was enough to nudge users to think more critically about the articles by gently asking them whether they *really* knew—from personal experience—if the events described in the articles were true or false. And once this critical filter was activated, it stayed active even for articles without the user-rating treatment. These findings are especially relevant given Facebook’s recent decision to have users to rate the credibility of articles [10]. Though knowledge of external facts could influence believability, the randomized experimental design in our study reduces the potential negative influence of missing variable bias.

There are several arguments in favor of user-based ratings. Developing ratings directly from users may be easier than finding appropriate experts and compensating them for their ratings. There are more users available to rate articles than experts, and voluntary ratings from consumers will cost nothing. After all, that is how we do ratings on the Internet for

products and services. Facebook decided that this is their next move [10], hopefully adding credibility to news consumption by prioritizing articles from sources that the users have deemed to be reputable. Unfortunately, such crowdsourced rating approaches also have several problems that need to be ironed out before the rating methods can produce a useful reference that can influence users' behavior in a meaningful way [8]. After all, while people rate products and services based on their personal experience, the same cannot be expected for news.

However, this very limitation of crowdsourced rating of news—i.e., the lack of personal experience of the events described in the news—may have a positive upside. Our work shows that the prompt of asking users to evaluate the truthfulness of the articles itself can remind the users that they do not have the necessary personal experience to judge whether the articles are true or not, helping them to think more critically about the articles. This effect is important because, as shown in prior research [25], believability has strong effects on the actions that users take, such as reading, liking, commenting, and sharing. Just as sharing of articles contributes to news—fake or not—going viral, clicking the *Like* button or commenting on an article may also spread the news farther and faster due to the appearance of the post in other users' news feeds. Hence, invoking the critical filter in users' minds—and encouraging them to be more careful in their sharing of information—is an important first step in tackling the issue of fake news.

### 5.1. Implications for future research

The prevalence of news consumption on social media [15], combined with the hedonistic, entertainment-seeking goals of social media use [23] that trigger confirmation bias, suggests that news consumption on social media may not be mindful. Mindfulness can have significant impacts on the way we use technology, for better or worse. In this research, we begin to address the issue of mindfulness in social media consumption. We believe that asking users to evaluate articles is one step towards inducing more mindful consumption, but more research needs to be done on other ways to trigger a more mindful behavior.

Largely, we can categorize the issues around the rating approaches for the problem of fake news into the production of ratings and the consumption of them. In this paper, we focused on the production side of the ratings and how the action of collecting the rating information from users may actually influence their behaviors.

We need more research on the production side of ratings, the different procedures, and pros and cons associated with those procedures. One fundamental

concern about crowdsourced ratings is that users may lack the knowledge needed to verify the facts in articles because they may not have personally witnessed the events in the article. User ratings are common on the web, and we generally accept them as somewhat reliable indicators of underlying quality, but it is not entirely clear whether this assumption is true, especially in the context of news. Since most users simply lack the expertise and experience to judge the truthfulness of articles, the result may be articles rated by their fit with users' pre-existing beliefs and users being shown more articles that fit their beliefs, causing the filter bubbles to worsen.

Also, for crowdsourced ratings to be successful, the level of participation is crucial. For product and service ratings, consumers have various incentives to leave their feedback. They feel invested as they have paid for the products and services. However, unlike product and service ratings where the process of rating is easier to understand, news article ratings may not be as easily understood. When fake news articles continually show up, will the users' participation in rating news stories remain high or will the users feel flustered and develop a sense of helplessness and stop contributing? If user attrition is a concern, perhaps expert rating may be a better solution in the long run.

Though we are mainly concerned with user article rating in this paper, there may be many other ways of collecting rating information. For instance, compared to Facebook's proposed user rating approach, expert rating approaches have some obvious benefits [8]. Expert ratings of articles (i.e., fact checking) are fairly well understood with many services available (e.g., PolitiFact). Aggregating those services to produce source level ratings is straightforward, though the timeliness of the rating is still a concern.

On the consumption side of rating information, given that there are different ways of producing source ratings, we should investigate whether different rating approaches have differing impacts on the users. If the ratings came from regular users assessing older stories (as Amazon and eBay for past purchases), would they have a stronger or weaker effect on believability than ratings from a panel of experts? If the ratings came from the users, would the mechanisms of rating matter? Would user ratings of the source be perceived differently than user ratings of individual articles? Would there be meaningful differences in how the users interpret these ratings from different sources and mechanisms? We need more research on ways to compile and present ratings of social media stories.

Presentation format of rating information is also important in influencing users' behaviors. For instance, Facebook had to roll back their flagging approach because it proved to be ineffective. One possible reason



is that the users simply did not understand what the flags meant. If Facebook educated users on the flags (through banners and TV campaigns), would they have been more effective? What if the flags had a more eye-catching icon and stronger wording to trigger System 1 and 2 cognitions? Research has shown that users' behavior may change even with a subtle change in the interface, such as displaying the news source before the headline [25]. Therefore, it is important to investigate how the rating information should be presented to maximize its influence on the users.

Another important implication from our research is the role of confirmation bias. From Table 3, we see that a one standard deviation change in confirmation bias (which is standardized) has a larger effect than the user-rating treatment. In other words, a user's belief in a social media article is influenced more by his/her desire for it to be true than by the push from the rating prompt. Fortunately, the rating prompt nearly compensates for the influence of confirmation bias. In addition to investigating other ways to help users think more critically when consuming news on social media, we need more research on confirmation bias and how it affects our beliefs in news stories, especially fake news stories.

Finally, in this work, we held constant the person sharing the article with the participant. Would it matter who shared the article and their relationship to the user, who liked it, and how many people commented on it? More research is needed to better understand how who shares an article influences its believability and the actions users take.

## 5.2. Implications for practice

The public is starting to recognize the role of social media providers and search engine providers in the spread of fake news and is calling for more proactive measures. We approached the problem of fake news from the opposite direction from many other researchers, who have started by first building prototypes and then testing if they affect believability [35, 38]; we started by first testing if the prompt of asking users to rate the articles they see on social media would influence them to think more critically about the truthfulness of those articles. In so doing, we aim to provide evidence-based design advice.

As discussed earlier, figuring out how to make crowdsourced rating is going to be challenging, and that may be the basis to push for expert rating instead [8]. However, different rating methods need not be mutually exclusive, and we can adopt different approaches together (e.g., user rating with expert rating). Figuring out the best way to collect and provide rating information must be a long-term endeavor. However,

we find that user rating has a rather immediate benefit, and it only requires a cosmetic change to the interface which is likely to be low cost.

The key idea is to momentarily stop users from mindlessly consuming news on social media by asking them to evaluate the articles. Perhaps, other ways to stop the stream of mindless consumption may also prove to be effective. Currently, most of (if not all) Facebook surveys and rating prompts (e.g., Would you recommend this place?) do not come with any submit button or confirmation prompt. All that users have to do is to click on an option, and the response is submitted. Following Facebook's current practice, we also designed this experiment in that fashion. However, if there was an extra step that further prompted users to think about their response (e.g., Are you sure?), it may have a greater impact on the users' behavior.

In this work, we investigated a possible positive side effect in asking users to rate news articles on social media. The jury is still out on whether such a crowdsourced rating mechanism would be effective in generating useful information for readers to evaluate the truthfulness of the articles they see. That being said, our work shows that there may be merits in asking users to rate the news articles they see on social media. It is a low hanging fruit that Facebook (and other social media for that matter) may want to consider regardless of their long-term plans and alternative options.

## References

- [1] A. Achtziger and C. Alós-Ferrer, "Fast or rational? A response-times study of Bayesian updating", *Management Science*, 60 (2013), pp. 923-938.
- [2] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election", *Journal of Economic Perspectives*, 31 (2017), pp. 211-236.
- [3] M. Barthel, A. Mitchell and J. Holcomb, "Many Americans believe fake news is sowing confusion", *Pew Research Center*, 15 (2016).
- [4] V. G. Cerf, "Information and misinformation on the internet", *Communications of the ACM*, 60 (2016), pp. 9.
- [5] J. A. Chevalier and D. Mayzlin, "The effect of word of mouth on sales: Online book reviews", *Journal of marketing research*, 43 (2006), pp. 345-354.
- [6] S. de Castro Bellini-Leite, "The embodied embedded character of system 1 processing", *Mens sana monographs*, 11 (2013), pp. 239.
- [7] W. De Neys, "Conflict detection, dual processes, and logical intuitions: Some clarifications", *Thinking & Reasoning*, 20 (2014), pp. 169-187.
- [8] A. R. Dennis, A. Kim and P. Moravec, *Facebook's Bad Idea: Crowdsourced Ratings Work For Toasters, But Not News*, BuzzFeed BuzzFeed 2018.
- [9] P. G. Devine, E. R. Hirt and E. M. Gehrke, "Diagnostic and confirmation strategies in trait hypothesis testing", *Journal of Personality and Social Psychology*, 58 (1990), pp. 952.

- [10] E. Dwoskin and H. Shaban, *Facebook will now ask users to rank news organizations they trust*, *The Washington Post*, 2018.
- [11] J. S. B. T. Evans, "Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition", *Annual Review of Psychology*, 59 (2008), pp. 255-278.
- [12] J. S. B. T. Evans and K. E. Stanovich, "Dual-Process Theories of Higher Cognition: Advancing the Debate", *Perspectives on Psychological Science*, 8 (2013), pp. 223-241.
- [13] L. Festinger, *A theory of cognitive dissonance* Evanston, Stanford University Press, California, 1957.
- [14] C. J. G. Gersick and J. R. Hackman, "Habitual routines in task-performing groups", *Organizational Behavior and Human Decision Processes*, 47 (1990), pp. 65-97.
- [15] J. Gottfried and E. Shearer, *News Use Across Social Media Platforms 2016*, Pew Research Center, 2016.
- [16] L. Graves, "Boundaries Not Drawn: Mapping the institutional roots of the global fact-checking movement", *Journalism Studies* (2016), pp. 1-19.
- [17] L. Guo, J. S. Trueblood and A. Diederich, "Thinking Fast Increases Framing Effects in Risky Decision Making", *Psychological Science*, 28 (2017), pp. 530-543.
- [18] J. C. Harsanyi, "Morality and the Theory of Rational Behavior", *Social Research*, 44 (1977), pp. 24.
- [19] E. C. Hirschman and M. B. Holbrook, "Hedonic Consumption: Emerging Concepts, Methods and Propositions", *Journal of Marketing*, 46 (1982), pp. 92-101.
- [20] E. E. Housholder and H. L. LaMarre, "Facebook politics: Toward a process model for achieving political source credibility through social media", *Journal of Information Technology & Politics*, 11 (2014), pp. 368-382.
- [21] D. C. Howell, *Statistical methods for psychology*, Cengage Learning, 2012.
- [22] E. D. Johnson, E. Tubau and W. De Neys, "The Doubting System 1: Evidence for automatic substitution sensitivity", *Acta Psychologica*, 164 (2016), pp. 56-64.
- [23] T. J. Johnson and B. K. Kaye, "Reasons to believe: Influence of credibility on motivations for using social networks", *Computers in Human Behavior*, 50 (2015), pp. 544-555.
- [24] D. Kahneman, *Thinking, fast and slow*, Macmillan, 2011.
- [25] A. Kim and A. R. Dennis, "Says Who?: How News Presentation Format Influences Perceived Believability and the Engagement Level of Social Media Users", *Proceedings of the Hawaii International Conference on System Sciences*, Waikoloa, HI. (2018).
- [26] A. Koriati, S. Lichtenstein and B. Fischhoff, "Reasons for confidence", *Journal of Experimental Psychology: Human Learning and Memory*, 6 (1980), pp. 107-118.
- [27] G. Loewenstein, T. O'Donoghue and S. Bhatia, "Modeling the interplay between affect and deliberation", *Decision*, 2 (2015), pp. 55.
- [28] M. R. Louis and R. I. Sutton, "Switching Cognitive Gears: From Habits of Mind to Active Thinking", *Human Relations*, 44 (1991), pp. 55-76.
- [29] W. Lowrey, "The Emergence and Development of News Fact-checking Sites: Institutional logics and population ecology", *Journalism Studies*, 18 (2017), pp. 376-394.
- [30] C. R. McKenzie, "Increased sensitivity to differentially diagnostic answers using familiar materials: Implications for confirmation bias", *Memory & Cognition*, 34 (2006), pp. 577-588.
- [31] E. Meixler, "Facebook Is Dropping Its Fake News Red Flag Warning After Finding It Had the Opposite Effect", *Time* (2017).
- [32] G. J. C. Mograbi, "Neural basis of decision-making and assessment: issues on testability and philosophical relevance", *Mens sana monographs*, 9 (2011), pp. 251.
- [33] R. S. Nickerson, "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises", *Review of General Psychology*, 2 (1998), pp. 175-220.
- [34] Z. Pan, Y. Lu, B. Wang and P. Y. Chau, "Who Do You Think You Are? Common and Differential Effects of Social Self-Identity on Social Media Usage", *Journal of Management Information Systems*, 34 (2017), pp. 71-101.
- [35] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, S. Patil, A. Flammini and F. Menczer, *Truthy: mapping the spread of astroturf in microblog streams*, *Proceedings of the 20th international conference companion on world wide web*, 2011.
- [36] N. J. Salkind, *Encyclopedia of research design*, Sage, 2010.
- [37] S. Shane, *The Fake Americans Russia Created to Influence the Election*, *The New York Times*, 2017.
- [38] C. Shao, G. L. Ciampaglia, A. Flammini and F. Menczer, *Hoaxy: A platform for tracking online misinformation*, *Proceedings of the 25th International Conference Companion on World Wide Web*, 2016.
- [39] C. Silverman, "This analysis shows how viral fake election news stories outperformed real news on Facebook", *Buzzfeed News*, 16 (2016).
- [40] H. A. Simon, "Rational decision making in business organizations", *The American economic review*, 69 (1979), pp. 493-513.
- [41] K. E. Stanovich, *Who is rational?: Studies of individual differences in reasoning*, Psychology Press, 1999.
- [42] K. E. Stanovich and R. F. West, "Individual differences in reasoning: Implications for the rationality debate?", *Behavioral and brain sciences*, 23 (2000), pp. 645-665.
- [43] J. B. Thatcher, R. T. Wright, H. Sun, T. J. Zagenczyk and R. Klein, "Mindfulness in Information Technology Use: Definitions, Distinctions, and a New Measure", *MIS Quarterly*, 42 (2018), pp. 831-847.
- [44] *The Wall Street Journal*, *Blue Feed, Red Feed*, 2016.
- [45] *The Washington Post*, *'Pizzagate' shows how fake news hurts real people*, 2016.
- [46] V. A. Thompson, J. A. Prowse Turner and G. Pennycook, "Intuition, reason, and metacognition", *Cognitive Psychology*, 63 (2011), pp. 107-140.
- [47] S. Vosoughi, D. Roy and S. Aral, "The spread of true and false news online", *Science*, 359 (2018), pp. 1146-1151.
- [48] A. L. Wintersieck, "Debating the Truth: The Impact of Fact-Checking During Electoral Debates", *American Politics Research*, 45 (2017), pp. 304-331.
- [49] Z. Zhou, X.-L. Jin, D. R. Vogel, Y. Fang and X. Chen, "Individual motivations and demographic differences in social virtual world uses: An exploratory investigation in Second Life", *International Journal of Information Management*, 31 (2011), pp. 261-271.