

Combating Fake News on Social Media with Source Ratings: The Effects of User and Expert Reputation Ratings

ANTINO KIM

(antino@indiana.edu)

Operations and Decision Technologies Department, Kelley School of Business,
Indiana University, Bloomington IN 47405

PATRICIA L. MORAVEC

(pmoravec@indiana.edu)

Operations and Decision Technologies Department, Kelley School of Business,
Indiana University, Bloomington IN 47405

ALAN R. DENNIS

(ardennis@indiana.edu)

Operations and Decision Technologies Department, Kelley School of Business,
Indiana University, Bloomington IN 47405

ABSTRACT: As a remedy against fake news on social media, we examine the effectiveness of three different mechanisms for *source* ratings that can be applied to articles when they are initially published: *expert rating* (where expert reviewers fact-check articles, which are aggregated to provide a source rating), *user article rating* (where users rate articles, which are aggregated to provide a source rating), and *user source rating* (where users rate the sources themselves). We conducted two experiments and found that source ratings influenced social media users' beliefs in the articles and that the rating mechanisms behind the ratings mattered. Low ratings—which would mark the usual culprits in spreading fake news—had stronger effects than did high ratings. When the ratings were low, users paid more attention to the rating mechanism, and, overall, expert ratings and user article ratings had stronger effects than did user source ratings. We also noticed a second-order effect, where ratings on some sources led users to be more skeptical of sources without ratings, even with instructions to the contrary. A user's belief in an article, in turn, influenced the extent to which users would engage with the article (e.g., read, like, comment and share). Lastly, we found confirmation bias to be prominent; users were more likely to believe—and spread—articles that aligned with their beliefs. Overall, our results show that source rating is a viable measure against fake news and propose how the rating mechanism should be designed.

KEY WORDS AND PHRASES: Fake news, misinformation, social media, article rating, source rating, expert rating, user rating, fact-checking.

Introduction

Fake news rose to global attention in 2016 with the U.S. presidential election, where disinformation campaigns—both domestic and foreign—to influence the election result were widespread across various social media [1, 3, 43, 66]. Moreover, the influence of fake news did not stop at the election. A conspiracy theory that came to be known as “Pizzagate” (alleging that a pizzeria in D.C. was the home of a child abuse ring led by the Democratic Party) [17] went viral on social media [13]. The restaurant owners and employees were harassed [17, 36], and a man visited the restaurant with a rifle, terrifying customers and workers [13]. The article started on Twitter [19], but Facebook played a key role in the spread of this story and many others [74].

More than 60% of adults get news from social media (primarily Facebook), and the proportion is increasing [20], so the problem of fake news is likely to become more serious [59]. The prevalence of fake news has not only shaken the public's trust in journalism but also stirred up criticism towards social media for not taking more proactive countermeasures [3]. More fake news articles are shared on social media than real news [68], and users play a big role in fake news gaining momentum: 23% of social media users report that they have spread fake news [3], and a recent study shows that false articles spread faster than true ones, primarily because of people, not bots [78]. Unfortunately, users tend to believe articles that align with their beliefs [69] due to *confirmation bias* [58], making them more gullible when faced with posts crafted to their point of view. Facebook tried flagging fake news articles, but this proved ineffective and was removed [50].

There are three important differences between news articles on social media and traditional media that make it harder for social media users to recognize fake news. The first is the mindset of users. People visit social media with a hedonic mindset [25] (e.g., to have fun and connect with friends [33]) rather than a utilitarian mindset, as they would when they visit news site or open a newspaper. Individuals in a hedonic mindset are less likely to think critically than those in a utilitarian mindset [26], causing confirmation bias to prevail.

Second, on social media, anyone can create “news”—real or fake—and the news spreads throughout the Internet as social media users read it and share it with their contacts [53]. This “democratization of journalism” through social media “has proven to be a sharp, double-edged sword” [82]. Quality control has moved from journalists with a putative interest in truth to users who have no training and often give little thought to verifying facts before spreading news [37].

Third, on Facebook, users do not choose the source¹ of the articles they see. With traditional media, the user first picks the source (e.g., newspapers, TV, online news sites), and they do so generally cognizant of the nature of the source. This is not true for Facebook. Facebook presents a mix of articles from many different sources, such as friends, sources based on past use, and advertisers who have paid to place their content in the user's newsfeed (some with malicious intent). People are more likely to believe and share articles that align with their beliefs [58] due to confirmation bias, and fake news takes advantage of this.

The spread of fake news on social media is an important problem facing society, and information systems researchers have an obligation to mitigate the problems created by this new information system [cf. 82]. The key issue here is whether we can redesign an information system (social media such as Facebook) to reduce the impact caused by those who intentionally misuse it to spread false information. Our focus is on the provision of additional information in social media, specifically, information about the sources of articles.

In this paper, we examine whether different mechanisms for rating news sources have different effects on users' beliefs in social media articles; i.e., are users more strongly influenced by ratings from experts or

by ratings from other users? We conducted two studies to investigate three mechanisms for developing source ratings that are attached to new articles as they are published. We examined: (i) *expert rating* (where *expert* fact-checkers rate *articles* and these ratings are aggregated to provide an overall source rating), (ii) *user article rating* (where *users* rate *articles* and these ratings are aggregated to provide the source rating), and (iii) *user source rating* (where *users* directly rate the *sources*). The first and second are similar to eBay seller ratings (where purchasers rate each transaction, and ratings are aggregated into an overall seller score) but with experts or users doing the ratings; the third is similar to opinion polls where users rate different media [e.g., 40].

We use reputation theory [9] to argue that the users would perceive expert ratings to have a cognitive basis and user ratings to be driven more by emotion. The results show that all three rating mechanisms influenced users' beliefs, with different mechanisms having different effects. In other words, ratings in general are influential, but the mechanisms by which ratings are created have different impacts. Expert ratings and user article ratings have stronger effects on believability than do user source ratings for low-rated sources, which are the usual suspects responsible in spreading fake news. Believability influenced the extent to which users would engage with the articles (e.g., read, like, comment and share). We found confirmation bias [58] to be prominent—users were more likely to believe and engage with articles that aligned with their beliefs. In the following sections, we first summarize prior research and develop our hypotheses. We then describe the two studies and their results. We conclude with a discussion of the findings and their implications.

Prior Theory and Research

Fake news has been defined as “news articles that are intentionally and verifiably false and could mislead readers” [1].² This is not the first instance of fake news in our society, as news has always been questionable in its reliability [53]. Even before the rise of the Internet, some newspapers were known for their biases and potentially distorted reports [18]. However, the current issue of fake news, spread primarily through social media, became increasingly important during the 2016 presidential election in the U.S. [1, 3, 6]. People had not previously been exposed to such a vast number and variety of fake news stories, or at least had not been aware of it. In response to fake news, a number of fact-checking initiatives have been launched [21, 44]. Fact checking works [84], but it needs to be presented at the time of news consumption [67]. News articles tend to have a short “shelf life,” and by the time the results of fact checking are posted, most people have already read the article; fact checking individual articles can be too slow. As Mark Twain noted “a lie can travel halfway around the world while the truth is putting on its shoes” [35].

Several technical solutions attempt to automate fact-checking, such as Truthy [61] and Hoaxy [67], so that results can be provided more quickly. Hoaxy searches fact-checking sites that verify news articles (e.g., snopes.com, politifact.com, and factcheck.org) and sites that have a history of publishing fake news to build a database of stories. It displays both the spread of stories and their fact-checking. While automated solutions can be faster than manual efforts, fact-checking still happens after much of the consumption—and damage—is done. Aside from speed, there are also other design challenges for fact-checking automation. For instance, research on recommendation agents, which Hoaxy may be considered, indicates that the process used must be sufficiently transparent for users to find their recommendation credible [79].

An alternative solution, or a complement to fact-checking, should be one that is effective at the time and place when the consumption—and spread—of news articles takes place: on social media when the article is first published, not on a different website days after the article has been published. Fact-checking

is most effective when it is presented simultaneously with the article. One solution is *source* ratings (not *article* ratings) applied to articles when they are first published, the same way that sellers on eBay have source ratings applied to all new products they offer. Similar to seller ratings on eBay, the sources can be continually evaluated based on prior articles, and the resulting source ratings can be associated with any subsequent articles posted by the sources, thereby avoiding the delayed aspect of fact-checking solutions. Research shows source ratings influence the extent to which users believe social media stories [37] and online news consumption [1, 6].

One unanswered question is whether the mechanism used to produce source reputation ratings affects the extent to which ratings influence users' beliefs. Does the source of the rating (i.e., the "rater") itself matter? Fact checking has traditionally been done by experts (e.g., Politifact), but e-commerce ratings have been done by users (e.g., eBay). Facebook's fake news flag was produced by experts, but it was deemed ineffective and replaced by user ratings [15, 50]. Is one rating mechanism more effective than others in influencing users' beliefs? This question is the core of this work.

Information Processing in Social Media

People use the Internet for many different purposes, such as accomplishing tasks or seeking hedonic pleasure [87]. Social media use tends to be more hedonic in nature, [25], such as seeking entertainment or connecting with friends [33], rather than utilitarian, such as completing work tasks. Individuals in a hedonic mindset are less likely to think critically than those in a utilitarian mindset [26], and they are less mindful in their actions [72]. Making the matter worse, people generally perceive news to be highly credible, more so than other types of information available online [14], and users rarely verify online information [14]. Combining a hedonic mindset, content formatted as "news," and a low tendency to verify online information, we have a perfect breeding ground for fake news.

Complicating matters, Facebook attempts to present information to users that will keep them on the site. This leads to a user newsfeed that shows information that the user already prefers. Facebook learns users' preferences by tracking what they read and the actions that they take (e.g., read, like, comment and share). As a commercial entity, Facebook aims to maximize user satisfaction, and thus, it deliberately displays more content matching the users' preferences, so the users see more posts that match their existing beliefs [73]. This causes a decrease in the range of information that the users encounter, and, as a result, Facebook users often exist in information bubbles—commonly referred to as echo chambers [6]—that reinforce their beliefs and make them believe that others around the world are more like them [73].

Thus, many articles that users encounter on social media are related to topics they have previously viewed [73], topics on which they have already formed an opinion. When users encounter information that aligns with their pre-existing opinions, they are inclined to believe it [8, 41, 58]. This tendency to favor information that confirms one's pre-existing opinions—and ignore information that challenges them—is called *confirmation bias* [8, 41, 58]; people are more likely to believe information matching their pre-existing opinions [1, 27]. When they encounter information contrary to their opinions and expectations, they experience cognitive dissonance [12]. 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 honest because lying is unrelated to honesty) or by accepting one and rejecting the other (e.g., John is honest, so I do not believe he lied; or John lied, so he is not honest) [12].

Resolving such a cognitive dissonance takes cognitive effort, and humans tend to be cognitive misers who resist expending effort [70]. We prefer to accept the easiest answer and maintain our own internal

beliefs [34]. This tendency is exacerbated when humans are in a hedonic mindset, as their passive pursuit of pleasure causes them to be even less likely to expend cognitive effort than when primed to exert their mental capabilities [26]. Because rejecting the new information is simpler cognitively than reassessing one's pre-existing opinions, most people retain their existing opinion and discard the new information as being false [8, 41, 49]. Therefore:

H1: *Confirmation bias prevails on social media: Pre-existing opinions on a topic directly influence the extent to which a social media user perceives a news article to be believable.*

What H1 implies is that, left to their own devices, social media users will believe information matching their own biases; if a piece of information—true or false—matches their world views, people will accept it without much thought. Our focus is fake news, so this bias is only an issue when the social media article is false but aligns with the user's beliefs; in this case, users are likely to believe the fake article. Given this nature of the problem of fake news on social media, one piece of the solution may be to influence the way users process information by providing them more information about the source.

Source Reputation

In addition to pre-existing opinions, the source of an article can influence the extent to which a user believes it [28, 37, 47]. This assessment of a source's reputation will depend on our prior experience with the source and the new information that we see. When someone tells us a story that challenges our pre-existing opinion, we often consider the source of the story as we assess whether to believe it. Consider your own feelings toward a news story that challenged your opinions about the president (e.g., he did something good or something bad). Would the source of the story influence its believability (e.g., Breitbart, BBC, CNN, Fox News, NPR, or Wall Street Journal)? Do you tend to believe a friend more than a stranger? Sources matter.

Source reputation affects the extent to which we believe a specific article to be credible, in addition to other article-specific factors [54, 85]. Source reputation for news is built gradually through a history of providing information the reader sees as reliable [71]. The reputation that we ascribe each news source is dependent upon our previous experiences with that source and our own feelings of its credibility and reliability. This evaluation of past behavior is an assessment of "truth" but is also partly subjective because, as we argued above, confirmation bias influences our judgement. These two theories are not incompatible, as we are capable of assessing reputation on various attributes, and our own biases influence these determinations. Reputation Theory [5, 9, 75] argues that we evaluate past behavior through three lenses to assess a source's reputation: functional, social and affective. Given the limitations of our cognitive effort during certain tasks, our motivation and ability to assess reputation on these dimensions will vary.

The first lens, the functional dimension, is the "objective" view of reputation [9]. It is the ability to achieve certain aims and goals. This aspect of reputation is competence, instrumentality and purpose-oriented [9, 22]. In order for a functional reputation to be assessed, one must have a set of criteria on which to judge others. These criteria come from publicly understood objectives, such as producing stories that are verifiably true. An increase in meeting these objectives leads to an increase in functional reputation. The main assigners of functional reputation are those agents with the experience to judge others' reputations, such as scientists, journalists, experts and analysts [9].

The second dimension is the social reputation which is "normative" [9]. This dimension is the extent to which one's actions appear to be legitimate when assessed by social norms [9, 81]. Here, reputation is an indicator of the agent's fidelity with social norms and ethical behavior. The key issue, of course, is whose social norms? Different groups of people are likely to assign reputation based on their own norms of religion,

ethics, morals and “good versus evil” societal expectations [9]. To contrast the two dimensions, functional reputation is whether the entity is “doing things right” whereas social reputation is whether it is “doing the right things.” Given our societal norms and individual preferences, people in various societal and cultural groups may assess social reputation differently, but we are all capable of assessing social reputation.

The third dimension is the affective reputation, which is “subjective” and based on the appraiser’s beliefs [9]. In contrast to the functional and social dimensions, which are based on the actions of the person or organization being assessed, with affective reputation, the “inner world” and emotional logic of the appraiser dominates the assessment. The attribution of reputation is done from the perspective of the agent and arises from indications of someone’s attractiveness, sympathy or authenticity [9]. While both functional and social reputation involve expectation management, affective reputation is primarily concerned with identity management. Affective reputation is likely to be highly influenced by our biases and internal expectations for others. Of all dimensions, this is the one that is least likely to be uniform among assigners of reputation. However, it is also the aspect of reputation that we are all most capable of assessing individually.

Each of the three dimensions is distinct, but they are related [9]. The three dimensions can be further organized into two, more fundamental, aspects. The functional and social dimensions form a fundamental *cognitive* aspect of reputation because they rely on cognitive assessments of the actions of the person or organization being assessed [5, 9, 75]. In contrast, affective reputation forms a second fundamental *emotional* aspect because affective reputation is rooted more in the emotions of the assessor than the actions of the assessed [5, 9, 75].

A positive reputation relies on a delicate balance among the different dimensions [9]. One would hope that news source reputation is highly dependent on the functional dimension (i.e., objective truth) and the social dimension (i.e., norms and ethics) so that truth and fairness drive this reputation. Each of the three rating mechanisms we investigate in this study (expert rating, user article rating, and user source rating) places different weights on these three dimensions.

Source Reputation Rating

In cases where users lack past experiences with a source to determine its reputation, they often resort to checking the evaluations of others (i.e., reviews and ratings). Expert and consumer ratings have long been used in e-commerce, and research shows that they have a significant influence on consumers’ behaviors [30, 62, 88]. Reviews and ratings have also been used in other contexts such as online health information within a broader domain of media credibility [e.g., 51, 80, 4]. Rating systems consistently come up as a viable solution in improving the quality of information online and for providing signals of credibility to users. In the context of online health information, a group of researchers proposed a framework to manage, vet and certify information and contributors in order to address the problem of inaccurate health related information on the internet [10]. In their framework, quality seals are assigned by trusted third-party expert raters.

One important difference between online health information and content on social media is that people generally do not read health information for entertainment; they are usually in a utilitarian mindset seeking information, in contrast to the hedonic mindset of social media use. This difference in mindset means that users on social media may be less concerned with source reputation compared to those seeking health information. It is also less likely that people have a strong pre-existing opinion on health information compared to fake news on social media. Given such differences in the contexts and the users’ mindsets, it is not clear whether we can simply extrapolate what we know from the context of e-commerce or even

online health information. Before we start comparing different news source rating mechanisms, we must first consider whether source ratings would be effective at all.

In general, online users use heuristics based on others' evaluations (e.g., reviews) to determine the credibility of information [51]. Ratings of news sources could be displayed to social media users in a manner similar to the ratings commonly used by online retailers (e.g., Amazon's star rating of products, eBay's percent rating of sellers). There is little empirical research on the impact of ratings on the consumption of social media news. Some research has found that fact-checking ratings of individual *articles* influences the perceived believability of those *articles* [84]. However, empirical research on how *source* ratings affect the believability of articles has been scant. Based on the theory and empirical results above, we argue that source ratings will have a direct influence on users' believability when the user is unfamiliar with the source of a news article. High source ratings will increase believability and low ratings will decrease it. Thus:

H2: *Source ratings matter for articles on social media: Higher (lower) source reputation ratings will positively (negatively) affect the believability of news articles.*

Source Reputation Rating Mechanisms

There are many different ways in which reputation ratings could be developed for news sources. An interesting theoretical question is whether they would influence social media users in the same way. In this paper, we examine three different types of rating mechanisms that are widely used online to understand how users perceive and respond to the different rating mechanisms for news sources on social media: expert ratings, user article ratings and user source ratings. Our research focuses on assessing how each mechanism *affects users' beliefs*, not how each mechanism *actually works*. Our goal is to provide guidance to social media developers about the likely effects of implementing one of these mechanisms, not to assess the relative quality of actual ratings produced by the three mechanisms. The consideration of "which solution would be most effective" should come before asking "how should a solution be implemented."

There are two fundamental theoretical differences in the way the three rating mechanisms produce a rating: who performs the assessment (expert or user) and what is assessed (article or source). We consider three combinations of these: expert, user article and user source.³

Expert Rating: One common way to produce reputation ratings is by using the assessments of experts. Consider the many online expert rating and review sites, for instance: Michelin Guide for restaurants, MotorTrend for automobiles, DPRReview for digital cameras and PolitiFact for political claims. A similar mechanism could be developed whereby experts rate articles, and the individual article ratings are aggregated to provide an overall rating for the source, much like how eBay does for sellers. In fact, NewsGuard is a startup that provides source reliability ratings, based on evaluations by a team of experts consisting of educated consultants and journalist, for 4,500—and counting—major news sites that account for 98% of online engagement in the U.S. [57]. Based on nine criteria related to credibility and transparency, NewsGuard provides a holistic evaluation of a source. The source's reputation ratings would gradually converge over time as the experts rate more articles. In the realm of social media, recently, a long list of fact-checking organizations have signed the non-partisan International Fact-Checking Network Code of Principles, and Facebook is now looking into ways to use the reports from those expert organizations [11, 60].

A sharp difference between this expert rating of sources and the flagging approach that Facebook once employed is that expert rating is attached to every article when it is first published, rather than having expert

verification of individual articles attached to articles days after they have been published. Thus, this reputation rating is visible when an article first appears, unlike previous approaches that focus on individual articles and come into effect days after they are published. New articles can be evaluated sometime later, and the source ratings can be updated accordingly.

The impact of expert rating on users' beliefs depends on how the users perceive experts to assess articles. Experts' ratings place the strongest emphasis on functional reputation because experts have the ability, knowledge and criteria for assessing "objective" truth. Functional reputation rests on the assessment of competence, the ability to produce work that is free from error when assessed from an objective external world view [9]. Experts are perceived to be knowledgeable and have greater ability to assess objective truth than non-experts [9]. Among the three dimensions of reputation discussed above, experts are also likely to place greater emphasis on the functional dimension of reputation as they form their assessments [9]. At the same time, experts should assign the least weight to the affective dimension because experts should be objective in their assessment and not swayed by personal emotions [9]. If users accurately perceive the reputation assessment capabilities of experts, then experts should be perceived to be highly credible, with more reliable ratings.

The social dimension may also have some weight because we would expect experts to consider ethical issues [9], but the larger issue is the extent to which the experts are influenced by social norms and, if so, whose norms—liberal, conservative, etc. While it is clear that users would expect a heavier weight on the functional dimension than on the affective dimension from expert ratings, it is not entirely clear what they would expect regarding the social dimension. Some users may believe that experts assign little weight to social norms and only focus on objective truth, while other users may believe that experts give strong weight to liberal or conservative norms and, thus, consider the ratings to be biased. What matters is not how experts actually weight each dimension but how users who view reputation ratings believe the experts weight the dimensions.

User Article Rating: Similar to the method through which experts' ratings of articles can be aggregated to form a source rating, user article ratings may be aggregated to form a source rating that is applied to every news article from that source. Such crowd-sourced rating methods can "replace the authoritative heft of traditional institutions" [46]; in other words, social information pooling would lessen the burden of gatekeeper for social media [51]. At the same time, if users rate the articles instead of experts, we would expect the functional dimension to have less weight because an average user is less capable of providing a valid functional rating [9]. In the case of news articles, how can a user provide a functional rating—one grounded in objective truth? When users rate products or services, we expect them to have actually used the product or service. Unless the user was actually involved in the events in the article (either as a participant or a witness), users would be unable to assess functional reputation because they have little knowledge of the facts. One may argue that users could scour the Internet and combine information from various news sources to develop some quasi-expertise on the event, but this is rare and quite unlikely. As cognitive misers in a hedonic mindset, humans deplore spending cognitive effort when snap judgments are deemed adequate [34].

Therefore, except in exceptionally rare cases where the user was a participant or witness of the events in the article, user ratings will be mostly based on the social and affective dimensions. In other words, the social norms of the users and their affect toward the article will drive the reputation ratings that an article receives. Research shows that social media users are more likely to read and view articles that match their *a priori* opinions [37, 73], so confirmation bias is likely to influence the ratings they provide. Making the

matter worse, users experience greater difficulty in evaluating subjective information than objective information [51]. As a result, ratings are more likely to be based on the article's fit with the user's opinions than on objective truth, and, in turn, users are likely to perceive the ratings to be heavily based on affective dimension.

Past research also shows that user ratings tend to be extreme. Users are more likely to post a highly positive or highly negative rating than a neutral rating because only motivated users take time to post a rating and are usually very happy or very angry [31]. Social (i.e., normative) and affective (i.e., subjective and emotional) perceptions of users lead to extreme ratings, so evaluations based on functional (i.e., objective) aspect will further become muted.

User Source Rating: In this case, users assign a reputation rating directly to the source, not to a specific article that they read. This approach asks users for a gestalt reputation assessment of the source, not the assessment of a specific article with facts. Here, the functional dimension of reputation is even more difficult to assess because there are no specific objective facts to assess. Even if users have direct personal knowledge of the events reported in some stories, they will not have knowledge of many or even most articles. Thus, they are incapable of assessing functional reputation, so it should receive little weight. The social and affective dimensions of reputation dominate, both of which are biased by the viewpoints of the users [9]. As a result, the ratings from this approach will depend more on who does the ratings than on the source being rated. Also, users may be influenced by others' ratings [86], and while it might be desirable to have users rate sources independently (without seeing others' ratings), this would undermine the primary objective (i.e., showing source ratings on all articles to all users).

In summary, we theorize that different rating mechanisms place different weights on the three dimensions of reputation. Reputation Theory argues that experts weight the functional dimension more than the others [9]. Except in rare cases, most social media users are incapable of assessing functional reputation because they lack the expertise and direct personal knowledge of the events described in the articles. Therefore, user ratings will be more heavily driven by the social and affective dimensions, which are strongly influenced by emotion and confirmation bias. If the users are cognizant of these distinctions, the rating mechanism will matter and influence users' beliefs in news articles on social media. This is an assumption, however, one that we explore deeper in the following section, leading to H4. For now, assuming that the users will be cognizant of the differences across the rating mechanisms, we hypothesize that expert reputation ratings will have a stronger effect on believability than user ratings (of either articles or sources directly). Likewise, we hypothesize that user source rating will be perceived to be a more biased assessment than user article rating and thus will be less likely to influence perceived believability. Therefore:

H3: *The rating mechanism matters:*

H3a: *Expert rating will more strongly influence believability than user article rating.*

H3b: *Expert rating will more strongly influence believability than user source rating.*

H3c: *User article rating will more strongly influence believability than user source rating.*

The Differential Importance of Reputation Rating Mechanisms

We began this paper by noting the rise of fake news and deliberate attempts to use social media to spread false information [1, 3, 66]. Most fact checking initiatives have been motivated by the goal of identifying fake news [61, 67], as has Facebook's various responses [15, 89]. Many social media users report they are confused by fake news and thus are concerned about it [3]. From a public policy perspective, reducing the

spread of fake news in social media is arguably more important than encouraging the spread of true news in social media [1, 3, 68]. In terms of source ratings, fake news is directly related to negative ratings.

Research on online reviews has found that negative ratings have a greater impact than positive ratings because low numeric ratings are often used to quickly screen products out of consideration [7, 86]. Users have come to expect a J-shaped pattern of reviews so that most online reviews have a preponderance of high values [48]; the average review on Amazon is 4.4 out of 5 stars [63]. Thus, high ratings confirm expectations while low ratings are a disconfirmation.

As a result, low ratings are unexpected and are more likely to cause users to turn their attention toward the cause of the disconfirmation. Evolutionarily, we are conditioned to be alert to risks to ourselves [77], including risks to our beliefs from our environment. When in a hedonic mindset, we are especially subject to the influence of confirmation bias. Here, the confirmation bias comes in two forms, our own beliefs on the topic in question, and our expectations of high ratings. Having an expectation or experience with a news article topic, such that one has *a priori* beliefs, will cause attention to turn toward the article and rating details when the rating disconfirms what one expects to observe in ratings [86]. Hence, when the ratings are low, disconfirmation will cause users to pay attention to the rating details, including the mechanism that produced the ratings. Disconfirmation bias will make low ratings more influential due to their surprising nature; in contrast, social media users are not incentivized to deeply think about the mechanisms behind a highly rated article, as their hedonic mindset causes them to disregard information that is not surprising [86].

H4: *Low ratings matter more: The relationships in H3 will be stronger when the source is rated low than when the source is rated high.*

Effects on Behavior

Social media users' perceived believability of news articles can affect their actions. They can choose to read the article or not, and they also can choose to provide feedback on the article (e.g., like or comment) as well as contribute to the spread of the article (e.g., share). Each of these actions is separate and distinct; you can like, comment on or share an article without reading it, although there may be some coherent sequencing in behavior—reading before liking, sharing only if one read and liked a story (sometimes without clicking the *Like* button), and so on [38]. Research shows that many people share an article without reading its entirety; they act on the headline alone [16], and those wishing to spread fake news design headlines with this understanding [83]. One important thing to note here is that, while sharing is one obvious form of engagement that contributes to the spread of news—real or fake—it is certainly not the only one; on Facebook, clicking the Like button or commenting on an article also make the article appear on friends' feeds in a similar manner to shared articles.

We argue that pre-existing opinions influence these behaviors; a user is more likely to read an article if it is congruent with his or her prior opinions due to confirmation bias [2]. Confirmation bias often causes selective information search [2, 39] in which people seek information actively that confirms their beliefs and avoid information that does not. Selective information searching will be intensified when people are in a hedonic mindset because they are not seeking to find a correct outcome but rather are seeking enjoyment. Viewing information that supports your opinions is more enjoyable than viewing information that challenges them [12, 52], so people will be more likely to read articles that support their opinions.

Other Facebook actions—liking, commenting and sharing—are unequally distributed [23, 42]. Most users seldom engage in these behaviors, perhaps because they require more effort and more commitment

to the subject than reading [38, 55]. Yet, users who engage in these behaviors do so relatively often [23, 42]. The choice to act can be influenced by emotion or information, with liking being driven more by emotion, commenting more by cognition, and sharing by both [38].

Therefore, we theorize that one important factor influencing the decision to read, like, comment on or share social media articles is how its headline fits with pre-existing opinions. However, we are not testing how emotion and cognition influence the actions, as that result has previously been found [37]. Here, we focus on the influence of confirmation bias on the decision to engage with the article. The stronger the fit, the more likely the article is to trigger an emotional reaction leading to a *Like*, a cognitive reaction leading to a comment, or both emotional and cognitive reactions leading to sharing. Thus:

H5a: *The degree of fit between an article's headline and pre-existing opinions influences the choice to read, like, comment on and share the article.*

The second important factor influencing the decision to act on an article is the extent to which a user believes it to be true. Believability can be an important factor in the use of social media information [33] because, if someone does not perceive information to be true, they are less likely to engage in it or to encourage its spread by sharing it. Thus:

H5b: *The believability of an article's headline influences the choice to read, like, comment on and share it.*

To sum, we hypothesize that the primary factors affecting user behavior are pre-existing beliefs and believability. Source ratings may have additional effects over and above confirmation bias and believability, so we include them in our analyses although we do not hypothesize any effects. For instance, users may be less likely to click like or share an article from a source that has a low rating, even if they believe the article is true, because it may be less socially desirable to show that you believe such an article. Reading is a private activity, so users may feel no social desirability effects in reading an article from a low-rated source. In other words, although you may read tabloids in private, you may be inhibited about showing your friends that you do (i.e., a “guilty pleasure”).

Study 1

To test our hypotheses, we conducted two separate experiments, Study 1 (discussed in this section) and Study 2 (discussed in the next section). Both studies were online surveys assessing the believability of articles given alternative rating mechanisms. Overall, the two studies are similar and arrive at similar conclusions, but the two studies have a number of subtle differences in their experiment designs (e.g., sequence, operationalization of confirmation bias, number of rating levels). These differences enable us to test whether our results remain valid across different settings. We begin with Study 1.

Method

Participants: We recruited 590 participants from a Qualtrics panel of adults in the United States. About 51% were female; about 5% were below 24 years of age, 40% between 25 and 44, 38% between 45 and 64, and 17% above 65. Approximately 42% had not completed college, 13% had an associate's degree, 26% had a bachelor's degree, and 19% had a graduate degree, which is similar to the U.S. population as a whole [64]. For our analysis, we grouped the participants by age, by education level, and by Facebook usage as shown in Table 1.

Insert Table 1

Task: The participants viewed eight news headlines and reported the believability of each article and their likelihood of reading, liking, commenting on, and sharing it. The headlines focused on the issue of abortion, with four designed to appeal to politically left-leaning participants and the other four to right-leaning participants (see Table 2). The specific headlines were drawn from [37] as they fit the purpose of our study, and they were formatted as they might appear as posts on Facebook (see Figure 1; Appendix B presents larger images, as shown to the participants). 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 (“Robin Miller” in Figure 1) for the *poster* who posts the article on his/her social media account—not to be confused with the original source where the article was authored and published—and the comment from the poster was a summary of the headline itself. To minimize any news source specific effect (e.g., trusted sources versus unknown sources), we fabricated eight names that sounded plausible (*HotNews.com*, *NewsHeadlines.com*, *NewsMedia.com*, *NewsStand.com*, *NewsUnion.com*, *SocialNews.com*, *TopNews.com*, and *TodaysNews.com*). All these URLs were verified to be inactive prior to the experiment (i.e., not used by any news provider or anyone else).

Insert Table 2

Insert Figure 1

Treatments: This was a repeated measures study in which participants received all eight headlines; see Figure 2 for a high-level description of the headlines. Participants received two headlines in the control treatment (no reputation ratings) and the remaining six headlines in one of the three randomly assigned between-subjects treatments: expert rating, user article rating, or user source rating. The headlines were randomly assigned to treatments and presented in random order to control for any headline-specific effects.

Insert Figure 2

In a repeated measure design, there is concern about the effects of an earlier treatment bleeding over into a later treatment [29, 65]. 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 spillover between treatments. This is the case in our study. The control treatment 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 rating treatment is likely to have strong influence on the treatments that follow it because, once users recognize the presence of ratings, the users may infer the lack of ratings as unreliable. Thus, placing the control treatment, which lacks ratings, after the rating treatment could confound the control treatment because users would instinctively view stories without ratings in the same way they view eBay or Amazon sellers without ratings—as suspicious, unproven sources.

Therefore, the control treatment was always presented first—right after the initial survey for

demographic information—and was designed to mimic the current Facebook style of presentation as closely as possible (see Figure 1(a)). The control treatment was followed by a description of the rating treatment (see Appendix A), followed by the rating treatment. By displaying this message about the source ratings, we were confident that all subjects presumed the ratings to be legitimate (i.e., offered by Facebook as opposed to some unknown third party). To ensure that the subjects understood the details of the rating mechanisms, we solicited some opinions about the rating mechanisms by asking participants an open-ended question on their understanding of the mechanism, although we do not use the responses in this work. The participants were then shown six headlines in the one rating treatment to which they were randomly assigned: (i) expert rating, (ii) user article rating, or (iii) user source rating (see Figure 1). Two of the headlines had high ratings (4.9 stars out of a maximum of 5 stars), two had medium ratings (2.6 stars), and two had low ratings (1.1 stars). The overall flow of the experiment is shown in Figure 3.

Insert Figure 3

Independent Variable: Confirmation bias was measured using two items from [37] that were assessed for each headline. The first was the participants' 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). The second was their position on the headline (−3= extremely negative to +3= extremely positive). The two items were multiplied together to make a scale from −21 to +21.

Dependent Variables: The believability of each article was measured using three 7-point items [37] (*How believable do you find this article, How truthful do you find this article, How credible do you find this article*). Reliability was adequate ($\alpha = 0.95$). We also measured how likely the participant would be to take actions, each as a separate item: read, like, comment and share. A comment can support or oppose the article, so the four actions become five measurement items: *Read, Like, Post a supporting comment, Post an opposing comment, and Share*.

Results

The mean believability levels for different rating mechanism types are shown in Figure 4. The figure suggests that the three rating mechanisms are perceived relatively similarly when they present high ratings (i.e., mean believability values are close to each other). But as the rating decreases, indicating a disreputable source that may be more likely to produce fake news, the rating mechanisms begin to have different effects on perceived believability (i.e., mean believability values are more spread out).

Insert Figure 4

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

Insert Table 3

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

H2 argued that source ratings would affect believability. The coefficients for all rating treatments are in the expected direction, and all are significant except for high user source ratings. High ratings increase believability, whereas medium and low ratings decrease believability when compared to *no rating* control treatment (i.e., the base case). The low ratings have a stronger negative impact on believability than do the medium ratings. Because low-rated sources indicate disreputable ones (i.e., those most responsible for fake news), we are most interested in the effects of low ratings. We conclude that H2 is partially supported.

H3 posited that the rating mechanism mattered, such that expert ratings would have a stronger impact on believability than would user article ratings, which in turn would be stronger than user source ratings, while H4 argued that this pattern would be stronger for low rather than high ratings. When ratings were high, both expert rating and user article rating had significant effects, but user source ratings had no effect. As can be seen from Table 4, there were no significant differences between expert rating and user article rating (Chi-Squared = 0.00, $p > 0.05$). For low-rated sources, all three mechanisms had significant effects; expert ratings had stronger impacts than did user source ratings (Chi-Squared = 15.39, $p < 0.001$), as did user article ratings (Chi-Squared = 4.87, $p < 0.05$); there was, however, no significant differences between expert rating and user article rating (Chi-Squared = 2.89, $p > 0.05$). We conclude that H3 and H4 are partially supported.

Insert Table 4

H5 argued that confirmation bias and believability would influence users' actions. Across all actions, confirmation bias and believability had a significant impact; users were more likely to read, like and share articles that they believed and matched their point of view. Commenting activities are consistent with other types of activities; users were more likely to leave supporting comments for articles that matched their opinions and leave opposing comments for articles that they disagreed with. Hence, H5a and H5b are supported.

As an aside, we note that the rating mechanisms had no consistent effects across different actions over and above their effects on believability. That is, there are no direct effects of rating mechanisms on actions, but there are indirect effects through mediation by believability.

We included one post hoc test to check our manipulations. We included a question that asked all participants (regardless of treatment) to rate the extent to which the three rating mechanisms would be influenced by either cognitive or emotional aspects using a -5 (emotional) to $+5$ (cognitive) scale, with zero indicating a balance between the two (see Appendix C for details). We found that expert ratings were perceived to be more cognitive than user article ratings ($t(387) = 3.40$, $p < 0.001$) and user source ratings ($t(393) = 3.44$, $p < 0.001$); there was no difference between user article ratings and user source ratings ($t(394) = 0.10$, $p > 0.05$). Thus, our participants perceived the cognitive and emotional differences we expected across the different rating mechanisms.

Discussion

The results provide some general support for the differences we theorized among the three different rating mechanisms, but they also point to some needed refinements to our theorizing. We argued that the different mechanisms would have different strengths (H3) and that these differences would be most pronounced for

headlines receiving low ratings (H4).

In general, the ratings from all three mechanisms had an effect on user beliefs, but they had differential effects depending on whether the rating was high or low. The effects of high ratings (indicating a reputable source) were small or nonexistent, and there were no meaningful differences among the three mechanisms; participants were not strongly swayed by high ratings and tended not to care how the ratings were produced. We theorized that participants expected high ratings, so seeing them had only a small influence on believability and did not trigger much interest in how the ratings were produced. As a result, there were few differences in the effects across the different mechanisms. Our manipulation test suggests that the participants indeed perceived the cognitive and emotional differences across the three rating mechanisms. The participants simply were not compelled to pay attention to the differences.

When the ratings were medium, they did not yet deviate enough from user expectations for ratings, although Figure 4 suggests that the mean believability in the medium rating condition was beginning to diverge somewhat.

In contrast, when the ratings were low, the participants' expectations were disconfirmed; they were surprised, and this triggered more consideration of the ratings and how the ratings were produced. When triggered to consider that an article might be fake, users cared more deeply about the rating mechanism. Thus, our revised hypothesis (replacing H3 and H4) is:

H6: *The rating mechanism matters only when sources have low ratings,*

H6a: *When sources are rated high, there will be no differences in believability among the three different rating mechanisms.*

When sources are rated low,

H6b: *Expert rating will more strongly influence believability than user article rating.*

H6c: *Expert rating will more strongly influence believability than user source rating.*

H6d: *User article rating will more strongly influence believability than user source rating.*

Study 2

We conducted Study 2 to test the new H6 and to reexamine H1, H2 and H5. We also took this opportunity to check if our choice of methods in Study 1 had any undue effects. First, the control treatment in Study 1 always preceded the rating treatment so as to prevent any spillover from the ratings to the control, and it remains an open question as to whether such spillover actually exists. Hence, in Study 2, we first described the source rating mechanism and then randomly ordered the headlines so that control headlines (those without ratings) were randomly interspersed with the treatment headlines (those with ratings). To prevent users from interpreting a lack of rating as a suspicious source (e.g., newly launched site) after seeing sources with ratings, we inserted a message at the beginning of the experiment that source rating was a new feature that Facebook was testing, so some sources might not have the ratings available yet (see Appendix A).

Also, in Study 1, confirmation bias was calculated from two items that were presented together with each headline. Thus, the confirmation bias measure might have been influenced by the source ratings on each headline. In Study 2, we asked participants their position on the overall issue (i.e., abortion) as part of an opinion survey.

Method

We recruited 299 participants from a Qualtrics panel of adults in the United States, and grouped participants

by age, education level, and Facebook use (see Table 5). The method for Study 2 closely followed the method for Study 1; the task, treatments, and the operationalization of dependent variables were the same.

Insert Table 5

We made three changes to the experiment design. First, we randomized the presentation of the control headlines and the treatment headlines with ratings (see Figure 5). We included a control variable to see if there were any differences in believability due to order (i.e., were headlines in the first half more or less likely to be believed than those in the second half). Second, we measured confirmation bias using three 7-point items (*Where do you stand on the issue of abortion?*, *Do you support pro-choice or pro-life?*, *How positive or negative do you feel about access to abortion?*). The first two were presented at the beginning of the experiment and the last at the end. Reliability was acceptable ($\alpha = 0.76$). For left-leaning articles, Confirmation Bias was the average over the three items; for right-leaning articles, it was 8 minus the average (i.e., reversed scale).⁴ Finally, we included only high ratings and low ratings, resulting in two headlines with no ratings (control), two with high ratings, and two with low ratings. We randomly picked six out of the eight headlines used in Study 1 while keeping the equal ratio between left leaning articles and right leaning articles. In practice, we would be most interested in marking the bad sources to curb the spread of fake news, hence a binary rating system might be a more suitable choice [c.f., 57].

Insert Figure 5

Results

H1, which argued that confirmation bias affects the believability of articles, is supported (see Table 6). Articles in the first half were slightly less likely to be believed than those in the second half, suggesting some order effect. High ratings had no significant effect across all rating mechanisms, while low ratings from expert rating and user article rating mechanisms had a negative influence on believability (H2 is partially supported); user source ratings had no effect. Comparing the rating mechanisms that had significant effects, there were no differences in the effects of the expert rating and user article rating mechanisms for low ratings (Chi-Squared = 0.64, $p > .05$). H6, which argued that the rating mechanism does not matter for sources that are rated high but does matter when sources are rated low is partially supported. H5, which argued that confirmation bias and believability influence users' actions, is supported.

Insert Table 6

Discussion

First and foremost, we note that most of our main findings from Study 1 remain consistent in Study 2 despite the changes in the experiment designs, demonstrating the robustness of our results. The pattern of results shows, once again, that confirmation bias plays an important role in users' beliefs of the social media articles and also strongly influences their engagement with the articles (i.e., reading, liking, commenting, and sharing). Ratings affected believability, but only when sources were rated low: articles from low-rated sources were less believable—when the ratings were from experts or users evaluating articles—but articles

from high-rated sources were not more believable. Believability and confirmation bias both affected user actions.

We found an order effect; users were more likely to believe articles that they saw later than the ones that came earlier. In other words, users became *less* critical over the duration of the experiment.

There was also an evidence of a spillover effect from the ratings. The believability of the control articles without ratings in Study 2 (which were intermixed with the treatment articles with ratings) was significantly lower than the believability of control articles in Study 1 (which always preceded the treatment articles) (believability score of 4.42 versus 4.82; $t(1776)=5.22$, $p<.001$). Note that the order effect we found suggests that this should have been the other way around given the sequences of Study 1 and Study 2 (compare Figure 3 and Figure 5)—without any spillover effect, that is. Hence, the finding suggests that ratings made users more skeptical of the control headlines that lacked ratings. This was despite specific instructions that Facebook had not had time to rate all news sources, so a lack of ratings did not have any meaning (see Appendix A).

This spillover shows that there is a significant second-order effect of ratings. The first-order effect of ratings is what we hypothesized in H2 and H6; that low ratings directly influence believability of the rated articles that the users see. Study 1 was designed to prevent spillover effects sometimes found in repeated measures designs where one treatment (the ratings in this case) spills over to the control treatment because the control treatment comes later in the experiment. In Study 2, the fully randomized order of the control headlines and rated headlines allowed spillover effects to occur. Because the control headlines in Study 2 had significantly different effects on believability than did those in Study 1, we conclude that spillover effects occurred. This second-order spillover effect is that, once the users become aware of ratings, they start thinking more critically about the believability of all articles, even those without ratings. Perhaps, this is because ratings draw attention to the source, nudging the readers to evaluate articles more critically [37].

General Discussion

Overall, our results show that presenting source reputation ratings directly influences the extent to which users believe articles on social media. In general, low ratings decrease believability, while high ratings do not. Confirmation bias is also alive and well; our findings show that participants were more likely to believe articles that they agreed with. Likewise, believability and confirmation bias had strong and consistent effects on actions such as reading, liking, commenting, and sharing. Ratings did not have a direct effect on these behaviors beyond their effects through believability.

We also found an unexpected second-order effect. Once users saw ratings on some sources, they became more skeptical of articles from sources without ratings, despite instructions that the lack of ratings had no meaning. This finding also has an important implication. Many fake news sources pop-up to spread a few fake stories and then quickly disappear; thus, they will have no ratings. If a lack of ratings increases skepticism, then widespread application of ratings to reputable sources may help limit the impact of fake news from these pop-up sources which will lack ratings. We believe these results have implications for both research and practice.

Implications for Research

First and foremost, our research shows that reputation ratings matter and that different mechanisms for creating the ratings have different effects on believability for low-rated sources that help spread fake news.

Mechanisms that are perceived to be cognitive in nature and emphasize functional reputation (expert rating) have the most influence. We need more research on the different reputation rating methods, and especially research on ways to leverage social reputation to identify fake news more effectively as some sources are well known for the social norms they use in developing articles (e.g., CNN, Fox News). We studied the effect of reputed credibility, not experienced credibility (which is based on a user's experience with a source over time [76]). Some sources have a reputation for social norms that lean left (e.g., CNN) or right (e.g., Fox News), so users tend to perceive articles from these sources as biased [53]. When the ratings differ from the users' experience, can the ratings effectively persuade the users, or will the users disregard them?

The second-order effect of ratings (i.e., the effect of a lack of ratings when other sources have ratings) also deserves further investigation. Our initial results are promising; not only do the ratings have the desirable effect of alarming users against the negatively rated sources but also have the second-order effect of stimulating users to think more critically about the truthfulness of the articles they see when no ratings are present. Does this second-order effect persist over time? Is there a threshold for the percentage of rated sources for the second-order effect to manifest? Or, is it the mere existence of source ratings that makes the users think more critically?

We also need more research on the role of confirmation bias in fake news. Tables 3 and 6 show that the coefficients on confirmation bias (which are standardized) are comparable to the coefficients of low ratings from expert and user article rating mechanisms. This means that a one standard deviation change in confirmation bias has the same effect as the presence of low ratings. In other words, a user's belief in a social media article is influenced as much by his/her desire for it to be true as by experts' (or other users') ratings saying the article comes from a suspicious source. Similarly, the results in Tables 3 and 6 also show that confirmation bias has a strong effect on the actions that users take, such as reading, liking, commenting, and sharing. Confirmation bias has the strongest effects on clicking like, which is primarily driven by emotion [38]. The relative strengths of confirmation bias and believability present an interesting contrast. For reading, believability is more important than confirmation bias. For other actions, the effect of confirmation bias is comparable to that of believability; users care about whether the article supports (or opposes) their opinions as much as whether it is true or false. We need more research on how confirmation bias affects user actions, particularly those actions that help spread fake news.

Because people typically use social media for hedonic purposes [33] than utilitarian ones, they are less likely to think critically [26], and this causes confirmation bias to prevail, exacerbating the issue of fake news. It is important to have further investigation that ties together different contexts (e.g., social media versus news sites), users' motivation (hedonic versus utilitarian), effect of confirmation bias, and source ratings. In the context of news sites or search engines, what is the extent of change in the users' mindset and the effect of confirmation bias? Would source ratings have different effects across different contexts?

As mentioned previously, on Facebook, clicking the Like button or commenting on an article has similar effects as sharing; such actions make the article appear on friends' feeds in a similar manner to shared articles. Do users know they could be spreading fake news by liking and commenting? If users understand who sees their activities, would that also make a difference? More research on understanding what users believe may be the key in designing strategies to reduce the spread of fake news.

We focused on the consumption of ratings, how users would respond to source ratings produced by three different mechanisms. While this work answers "which solution would be most effective," we must also consider "how a solution should be implemented." Hence, we need more research on the supply side of ratings, the procedures and costs associated with producing expert ratings and user article ratings (e.g., collecting rating information, safeguarding against rating manipulation, correcting for biases and noises,

etc.). Though expert ratings are (and perceived to be) more cognitive than emotional, it is possible to erode people's trust in experts by labeling them as biased with agenda. Hence, experts need to manage their reputation, and the entity proving the ratings must be judicious in choosing the experts. Also, while expert ratings can produce more reliable and better quality information compared to user ratings, they can be more expensive as well. In an attempt to put together the best of both worlds, some online forums and review sites allow peers to rate contributors, and the sites label top contributors (e.g., super reviewers). Can a similar mechanism be used in the context of fake news, or will it be simply introducing yet another problem of confirmation bias in peer rating?

User rating of articles needs more investigation because past research has largely focused on the consumption of reputation ratings but has not investigated the quality differences in the production of expert and user reputation ratings. 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 [30]. While there are many quality assessment instruments available (e.g., popularity, awards, etc.) [4], it is important to find out which of them are practically usable for general social media users facing the problem of fake news. Poorly designed rating instruments may end up being useless or actually do more harm than good [32].

One aspect of reputation is social, how the target is perceived to comply with social norms. Some social norms are widespread (e.g., ethical behavior, fairness), but other norms differ by culture. Thus, we need research into ways to integrate reputation ratings produced by conformance with different social norms and research into how to communicate such ratings. It is also important to investigate how the biases of the reviewers may affect the way they rate sources. Does the wisdom of the crowd suffer from the biases of the contributing crowd? It is also crucial to understand how confirmation bias would affect source ratings.

Finally, we conducted an online experiment with random assignment to control for the myriad of factors other than our treatments that could influence the outcomes. A more dynamic environment where participants could interact with the system may reveal additional insights [24]. If more information about the source is available to users, would they click to view that additional information? For those who do click to view additional information, how do their behaviors differ from those who do not? Field research is needed to observe over time whether users rely on source ratings and how different factors (e.g., article type, time of day, etc.) influence their effectiveness.

Implications for Practice

The public is starting to recognize the role of social media in the spread of fake news and is calling for action. We approached the problem from the opposite direction from other researchers who have started by first building prototypes and then testing if they affect believability [61, 67]; we started by first testing different source reputation mechanisms to see how they affect believability and users' actions, so that we could provide evidence-based design advice.

Our results show that the source reputation rating can play an important role in how people evaluate the believability of the posts they see, which, in turn, affects their engagement with those posts (i.e., read, share, etc.). We know ratings have an impact on the sale of products and services, and our research shows that they are also likely to have an impact on the consumption of news, real or fake. As merchants and service providers accumulate poor ratings, they will eventually get pushed out of the market, at which point it is difficult for them to re-enter without improving their quality. Likewise, new merchants lacking ratings are regarded with caution, and some users simply do not use them until they acquire some basic level of good ratings. We would expect to see the same basic patterns with news providers. Therefore, we

recommend that social media establish source reputation ratings based on either expert or user ratings of past articles.

There are several arguments in favor of user-based ratings (i.e., crowd-sourced). 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 incur little cost, as we see from various review sites on the Internet. Facebook decided that this was their next move [15], hoping to add credibility to news consumption by prioritizing articles from sources that the users have deemed to be reputable. Unfortunately, our findings suggest that this will not be as effective as having experts provide the ratings. One fundamental concern about user-based ratings is that users lack the knowledge needed to verify the facts in articles because they have not personally witnessed the events. They cannot assess the functional dimension of reputation and, therefore, will rely on the social and affective dimensions. The result will be articles rated by their fit with users' pre-existing beliefs and users being shown more articles that fit their beliefs; the filter bubbles will get worse.

Although expert ratings may be more effective than user ratings, there are challenges to that approach as well. A company undertaking the task of providing expert ratings needs to form a team of experts consisting of educated consultants and journalist, and it must also create a set of standards to answer the fundamental question of what constitutes a credible source. NewsGuard is a startup doing exactly this [57], and it will be interesting to observe its development.

We believe that the effect of source ratings may go above and beyond what we find in this study because, in practice, many of the fake news articles are from sources that are unknown or even deceptive. The culprits interested in luring traffic to their sites for advertisement revenue may post fake news as "click bait," and to make their scheme more effective, they may intentionally choose names that are very similar to legitimate ones, such as ABCnews.com.co [45, 56]. We believe source rating can be an effective countermeasure against such deceptions.

In this paper, we investigated the effects of three different reputation rating mechanisms for social media news sources on the extent to which users believed stories and would take further action to spread them (e.g., like, comment, and share). Our research shows that confirmation bias has strong effects on both the belief in fake news and the actions that users take to spread it. However, the results also show that two reputation rating mechanisms (expert or user ratings of past articles attached to new articles) had significant effects on believability and, through it, on the actions that users would take. Who is doing the ratings and how they are doing it matters to users for sources with negative ratings, suggesting that reputation ratings have the potential to slow the acceptance and spread of fake news. It also suggests some second-order effects, where knowing that ratings exist makes users more cautious about unrated sources.

Conclusion

One of our primary objectives was comparing the effects of different reputation rating mechanisms on users' belief of articles. We found no meaningful differences among rating mechanisms for highly rated sources; all three had similar modest or nonexistent effects. However, for sources with low ratings, expert ratings and user article ratings had a stronger impact on believability than user source ratings. As theorized, users perceived expert ratings to be more cognitive in nature than both user article ratings and user source ratings. These findings are critical given Facebook's recently announced plans to have users rate the credibility of sources [15].

This is an interesting pattern as it implies that, when sources are highly rated, users pay little attention

to the rating mechanism. Users are accustomed to high ratings (e.g., from product and service reviews) and, in their hedonic mindset, they do not exert much effort on critical thinking. The differences among the rating mechanisms do not rise to the users' consciousness. However, low-rated sources present a surprise (a disconfirmation of assumptions), and they nudge users to pay closer attention to the rating and rating mechanism. The difference in impact for low-rated sources is important because low-rated sources are the usual culprits in spreading fake news. In conclusion, we demonstrate that source ratings can be an effective measure against fake news and that the mechanism behind the ratings matter.

Notes

¹ In this work, by "source," we mean any sites where "news" articles are published. Similar to how a "seller" on e-commerce site may be an individual or a corporation, we do not make the distinction between a site run by an individual or an institution. For instance, *infowars.com* is a well-known fake news site founded and run by a radio host Alex Jones. On the other hand, *rt.com* is a website run by the Russian government to spread propaganda.

² We note that the term "fake news" has recently been employed by some as part of their strategy to discredit credible news or scientific sources. However, here, we use its proper definition of misinformation and disinformation.

³ We do not make the distinction between expert article and expert source as we do for user-rating methods. By definition, "experts" are professionals who conduct an extensive research to learn about the sources. Hence, an expert rates sources by considering many different aspects such as the sources' backgrounds and the accuracy and reliability of their past articles [58]; a rating without considering the past articles would not fall into the category of "expert" rating. Therefore, it is not appropriate, nor meaningful, to have expert source ratings that ignore past articles.

⁴ All analyses were also checked by operationalizing confirmation bias as we did in Study 1. We observed no changes to our statistical conclusions.

References

1. Allcott, H.; and Gentzkow, M. Social media and fake news in the 2016 election, *Journal of Economic Perspectives*, 31, 2 (2017), 211–236.
2. Ask, K.; and Granhag, P.A. Motivational sources of confirmation bias in criminal investigations: The need for cognitive closure, *Journal of Investigative Psychology and Offender Profiling*, 2, 1 (2005), 43–63.
3. Barthel, M.; Mitchell, A.; and Holcomb, J. Many americans believe fake news is sowing confusion, *Pew Research Center*, 15 (2016).
4. Bernstam, E.V.; Shelton, D.M.; Walji, M.; and Meric-Bernstam, F. Instruments to assess the quality of health information on the world wide web: What can our patients actually use?, *International Journal of Medical Informatics*, 74, 1 (2005), 13–19.
5. Bitektine, A. Toward a theory of social judgments of organizations: The case of legitimacy, reputation, and status, *Academy of Management Review*, 36, 1 (2011), 151–179.
6. Cerf, V.G. Information and misinformation on the internet, *Communications of the ACM*, 60, 1 (2016), 9.
7. Chevalier, J.A.; and Mayzlin, D. The effect of word of mouth on sales: Online book reviews, *Journal of Marketing Research*, 43, 3 (2006), 345–354.
8. Devine, P.G.; Hirt, E.R.; and Gehrke, E.M. Diagnostic and confirmation strategies in trait hypothesis testing, *Journal of Personality and Social Psychology*, 58, 6 (1990), 952.
9. Eisenecker, M.; and Imhof, K. The true, the good and the beautiful: Reputation management in the media society, *Public Relations Research* (2008), 125–146.
10. Eysenbach, G.; Yihune, G.; Lampe, K.; Cross, P.; and Brickley, D. Medcertain: Quality management, certification and rating of health information on the net, *Proceedings of the AMIA Symposium*, 2000, 230–234.
11. Facebook Help Center Facebook, how are third-party fact-checkers selected?, Facebook Help Center, 2018.
12. Festinger, L. A theory of cognitive dissonance, Stanford University Press, California, 1957.
13. Fisher, M.; Cox, J.W.; and Hermann, P. Pizzagate: From rumor, to hashtag, to gunfire in d.C., *The Washington Post*, 2016.
14. Flanagin, A.J.; and Metzger, M.J. Perceptions of internet information credibility, *Journalism & Mass Communication Quarterly*, 77, 3 (2000), 515–540.

15. Frenkel, S.; and Maheshwari, S. Facebook to let users rank credibility of news, 2018.
16. Gabielkov, M.; Ramachandran, A.; Chaintreau, A.; and Legout, A. Social clicks: What and who gets read on twitter?, *ACM SIGMETRICS / IFIP Performance 2016*, Antibes Juan-les-Pins, France, 2016.
17. Gajanan, M. Now pizzagate conspiracy theorists are targeting a pizzeria in new york city, *Time*, 2016.
18. Gaziano, C.; and McGrath, K. Measuring the concept of credibility, *Journalism Quarterly*, 63, 3 (1986), 451-462.
19. Gillin, J. How pizzagate went from fake news to a real problem for a d.C. Business, Politifact, 2016.
20. Gottfried, J.; and Shearer, E. News use across social medial platforms 2016, Pew Research Center, 2016.
21. Graves, L. Boundaries not drawn: Mapping the institutional roots of the global fact-checking movement, *Journalism Studies* (2016), 1-19.
22. Habermas, J. *Theorie des kommunikativen handelns*. Erster band., Suhrkamp, Frankfurt am Main, 1988.
23. Hampton, K.N.; Goulet, L.S.; Marlow, C.; and Rainie, L. Why most facebook users get more than they give, *Pew Internet & American Life Project*, 3 (2012), 1-40.
24. Hardin, A.; Looney, C.A.; and Moody, G.D. Assessing the credibility of decisional guidance delivered by information systems, *Journal of Management Information Systems*, 34, 4 (2017), 1143-1168.
25. Harsanyi, J.C. Morality and the theory of rational behavior, *Social Research*, 44, 4 (1977), 24.
26. Hirschman, E.C.; and Holbrook, M.B. Hedonic consumption: Emerging concepts, methods and propositions, *Journal of Marketing*, 46, 3 (1982), 92-101.
27. Housholder, E.E.; and LaMarre, H.L. Facebook politics: Toward a process model for achieving political source credibility through social media, *Journal of Information Technology & Politics*, 11, 4 (2014), 368-382.
28. Hovland, C.I.; Janis, I.L.; and Kelley, H.H. *Communication and persuasion; psychological studies of opinion change*, Yale University Press, New Haven, CT, US, 1953.
29. Howell, D.C. *Statistical methods for psychology*, Cengage Learning, 2012.
30. Hu, N.; Liu, L.; and Zhang, J.J. Do online reviews affect product sales? The role of reviewer characteristics and temporal effects, *Information Technology and Management*, 9, 3 (2008), 201-214.
31. Hu, N.; Pavlou, P.A.; and Zhang, J. Can online reviews reveal a product's true quality?: Empirical findings and analytical modeling of online word-of-mouth communication, *Proceedings of the 7th ACM Conference on Electronic Commerce*, ACM, 2006, 324-330.
32. Jadad, A.R.; and Gagliardi, A. Rating health information on the internet: Navigating to knowledge or to babel?, *JAMA*, 279, 8 (1998), 611-614.
33. Johnson, T.J.; and Kaye, B.K. Reasons to believe: Influence of credibility on motivations for using social networks, *Computers in Human Behavior*, 50 (2015), 544-555.
34. Kahneman, D. Maps of bounded rationality: Psychology for behavioral economics, *The American Economic Review*, 93, 5 (2003), 1449-1475.
35. Kaid, L.L. Ethics and political advertising, in Denton Jr., R.E., ed., *Political communication ethics: An oxymoron?*, Praeger, 2000, 147-178.
36. Kang, C. Fake news onslaught targets pizzeria as nest of child-trafficking, *The New York Times*, 2016.
37. Kim, A.; and Dennis, A.R. Says who? The effects of presentation format and source rating on fake news in social media, *MIS Quarterly* (in press).
38. Kim, C.; and Yang, S.-U. Like, comment, and share on facebook: How each behavior differs from the other, *Public Relations Review*, 43, 2 (2017), 441-449.
39. Klayman, J. Varieties of confirmation bias, *Psychology of Learning and Motivation*, 32 (1995), 385-418.
40. Knight Foundation American views: Trust, media and democracy, Gallup, 2018.
41. Koriati, A.; Lichtenstein, S.; and Fischhoff, B. Reasons for confidence, *Journal of Experimental Psychology: Human Learning and Memory*, 6, 2 (1980), 107-118.
42. Lee, S.-Y.; Hansen, S.S.; and Lee, J.K. What makes us click "like" on facebook? Examining psychological, technological, and motivational factors on virtual endorsement, *Computer Communications*, 73 (2016), 332-341.
43. Levin, S. Mark Zuckerberg: I regret ridiculing fears over facebook's effect on election, *The Guardian*, 2017.
44. Lowrey, W. The emergence and development of news fact-checking sites: Institutional logics and population ecology, *Journalism Studies*, 18, 3 (2017), 376-394.
45. Lu, J. What is 'click bait' and why facebook wants to display less of it, *The Washington Post*, 2014.
46. Madden, M.; and Fox, S. Riding the waves of "web 2.0", *Pew Internet & American Life Project* (2006), 1-6.
47. McCracken, G. Who is the celebrity endorser? Cultural foundations of the endorsement process, *Journal of Consumer Research*, 16, 3 (1989), 310-321.
48. McGlohon, M.; Glance, N.S.; and Reiter, Z. Star quality: Aggregating reviews to rank products and merchants, *International Conference on Weblogs and Social Media*, 2010, 114-121.

49. McKenzie, C.R. Increased sensitivity to differentially diagnostic answers using familiar materials: Implications for confirmation bias, *Memory & Cognition*, 34, 3 (2006), 577-588.
50. Meixler, E. Facebook is dropping its fake news red flag warning after finding it had the opposite effect, *Time* (2017).
51. Metzger, M.J.; Flanagin, A.J.; and Medders, R.B. Social and heuristic approaches to credibility evaluation online, *Journal of Communication*, 60 (2010), 413-439.
52. Minas, R.K.; Potter, R.F.; Dennis, A.R.; Bartelt, V.; and Bae, S. Putting on the thinking cap: Using neurois to understand information processing biases in virtual teams, *Journal of Management Information Systems*, 30, 4 (2014), 49-82.
53. Mohammed, S.N. The (dis)information age: The persistence of ignorance, Peter Lang Inc., 2012.
54. Morris, J.D.; Choi, Y.; and Ju, I. Are social marketing and advertising communications (smacs) meaningful?: A survey of facebook user emotional responses, source credibility, personal relevance, and perceived intrusiveness, *Journal of Current Issues & Research in Advertising*, 37, 2 (2016), 165-182.
55. Muntinga, D.G.; Moorman, M.; and Smit, E.G. Introducing cobras: Exploring motivations for brand-related social media use, *International Journal of Advertising*, 30, 1 (2011), 13-46.
56. Murtha, J. How fake news sites frequently trick big-time journalists, *Columbia Journalism Review* (2016).
57. NewsGuard NewsGuard: Criteria for and explanation of ratings, 2018.
58. Nickerson, R.S. Confirmation bias: A ubiquitous phenomenon in many guises, *Review of General Psychology*, 2, 2 (1998), 175-220.
59. Pan, Z.; Lu, Y.; Wang, B.; and Chau, P.Y. Who do you think you are? Common and differential effects of social self-identity on social media usage, *Journal of Management Information Systems*, 34, 1 (2017), 71-101.
60. Poynter International fact-checking network fact-checkers' code of principles, Poynter, 2018.
61. Ratkiewicz, J.; Conover, M.; Meiss, M.; Gonçalves, B.; Patil, S.; Flammini, A.; and Menczer, F. Truthy: Mapping the spread of astroturf in microblog streams, *Proceedings of the 20th International Conference Companion on World Wide Web*, ACM, 2011, 249-252.
62. Reinstein, D.A.; and Snyder, C.M. The influence of expert reviews on consumer demand for experience goods: A case study of movie critics, *The Journal of Industrial Economics*, 53, 1 (2005), 27-51.
63. ReviewMeta Analysis of 7 million amazon reviews: Customers who receive free or discounted item much more likely to write positive review, reviewmeta.com, 2016.
64. Ryan, C.; and Bauman, K. Educational attainment in the united states: 2015, United States Census Bureau, 2016.
65. Salkind, N.J. Encyclopedia of research design, Sage, 2010.
66. Shane, S. The fake americans russia created to influence the election, *The New York Times*, 2017.
67. Shao, C.; Ciampaglia, G.L.; Flammini, A.; and Menczer, F. Hoaxy: A platform for tracking online misinformation, *Proceedings of the 25th International Conference Companion on World Wide Web*, 2016, 745-750.
68. Silverman, C. This analysis shows how viral fake election news stories outperformed real news on facebook, *Buzzfeed News*, 16 (2016).
69. Silverman, C.; and Singer-Vine, J. Most americans who see fake news believe it, new survey says, *BuzzFeed News* (2016).
70. Simon, H.A. Rational decision making in business organizations, *The American Economic Review*, 69, 4 (1979), 493-513.
71. Teng, S.; Khong, K.W.; Chong, A.Y.L.; and Lin, B. Persuasive electronic word-of-mouth messages in social media, *Journal of Computer Information Systems*, 57, 1 (2017), 76-88.
72. Thatcher, J.; Wright, R.; Sun, H.; Zagenczyk, T.; and Klein, R. Mindfulness in information technology use: Definitions, distinctions, and a new measure, *MIS Quarterly*, 42, 3 (2018), 831-847.
73. The Wall Street Journal. Blue feed, red feed, *The Wall Street Journal*, 2016.
74. The Washington Post. 'Pizzagate' shows how fake news hurts real people, *The Washington Post*, 2016.
75. Thiessen, A.; and Ingenhoff, D. Safeguarding reputation through strategic, integrated and situational crisis communication management: Development of the integrative model of crisis communication, *Corporate Communications: An International Journal*, 16, 1 (2011), 8-26.
76. Tseng, S.; and Fogg, B. Credibility and computing technology, *Communications of the ACM*, 42, 5 (1999), 39-44.
77. Vaish, A.; Grossmann, T.; and Woodward, A. Not all emotions are created equal: The negativity bias in social-emotional development, *Psychological Bulletin*, 134, 3 (2008), 383.
78. Vosoughi, S.; Roy, D.; and Aral, S. The spread of true and false news online, *Science*, 359, 6380 (2018), 1146-1151.

79. Wang, W.; and Benbasat, I. Empirical assessment of alternative designs for enhancing different types of trusting beliefs in online recommendation agents, *Journal of Management Information Systems*, 33, 3 (2016), 744-775.
80. Wathen, C.N.; and Burkell, J. Believe it or not: Factors influencing credibility on the web, *Journal of the American Society for Information Science and Technology*, 53, 2 (2002), 134-144.
81. Weber, M. *Wirtschaft und gesellschaft. Grundriss der verstehenden soziologie*, University of California Press, Berkeley, CA, 1978.
82. Weiss, R. Nip misinformation in the bud, *Science*, 358, 6362 (2017), 427-427.
83. West, D.M. How to combat fake news and disinformation, *Brookings*, 2017.
84. Wintersieck, A.L. Debating the truth: The impact of fact-checking during electoral debates, *American Politics Research*, 45, 2 (2017), 304-331.
85. Yang, T. The decision behavior of facebook users, *Journal of Computer Information Systems*, 52, 3 (2012), 50-59.
86. Yin, D.; Mitra, S.; and Zhang, H. Research note—when do consumers value positive vs. Negative reviews? An empirical investigation of confirmation bias in online word of mouth, *Information Systems Research*, 27, 1 (2016), 131-144.
87. Zhou, Z.; Jin, X.-L.; Vogel, D.R.; Fang, Y.; and Chen, X. Individual motivations and demographic differences in social virtual world uses: An exploratory investigation in second life, *International Journal of Information Management*, 31, 3 (2011), 261-271.
88. Zhu, F.; and Zhang, X. Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics, *Journal of Marketing*, 74, 2 (2010), 133-148.
89. Zuckerberg, M. News feed fyi: More local news on facebook, Facebook Newsroom, 2018.

Figures and Tables



(a)



(b)



(c)



(d)

Figure 1: A sample of article with (a) the control treatment, (b) expert rating treatment, (c) user article rating treatment and (d) user source rating treatment.

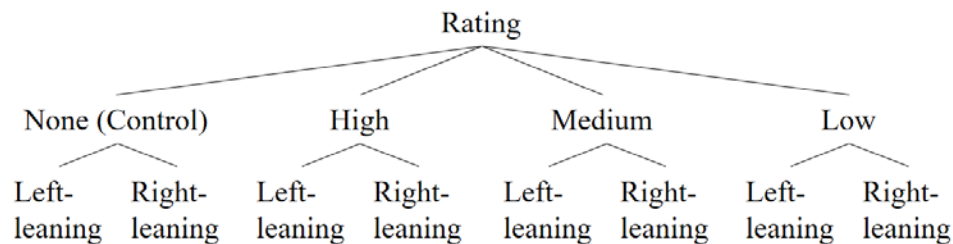


Figure 2: Description of 8 Headlines Used in Study 1

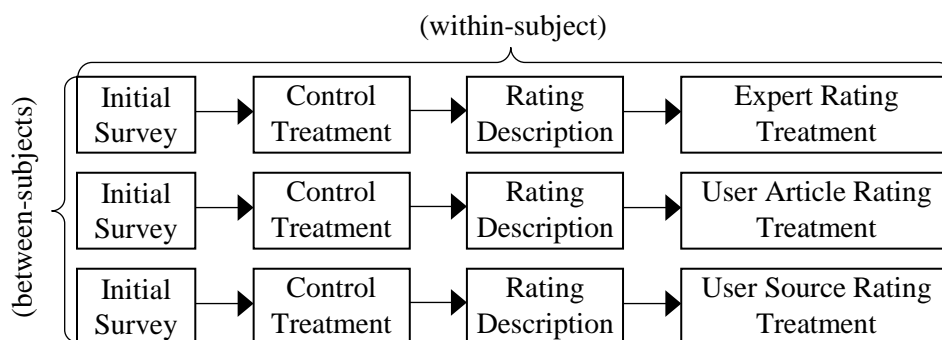


Figure 3: Study 1 Sequence of the Experiment

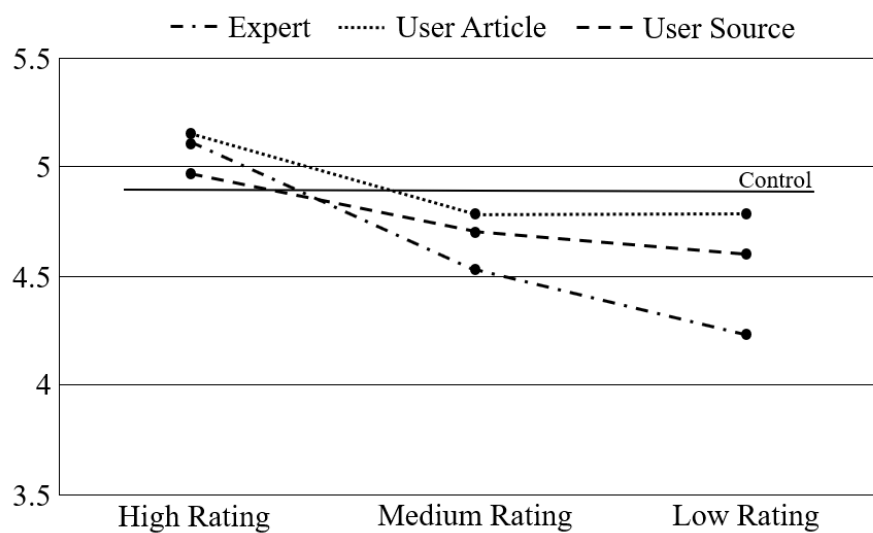


Figure 4: Study 1 Mean Believability by Rating Mechanisms and Levels

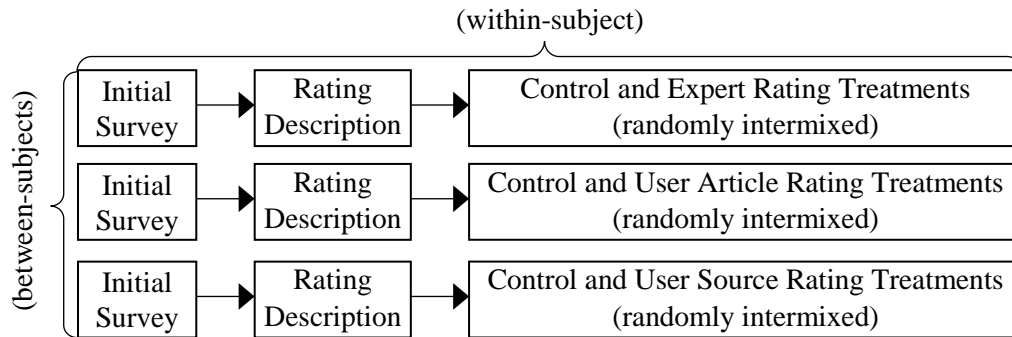


Figure 5: Study 2 Sequence of the Experiment

Table 1: Study 1 Participant Group Description

Group Variable	Description	Approximate Percentage (%)
Age	College age (18–24)	5
	Working age (25–64)	78
	Retirement age (≥ 65)	17
Education	Below bachelor's degree	55
	Bachelor's degree	26
	Graduate degree	19
Facebook Use	Once a week or less	12
	More than once a week	22
	More than once a day	66

Table 2: The 8 News Headlines Used in the Experiment

-
- Universities Connect their Healthcare Systems with Planned Parenthood to Provide Better Care to Students
 - Pro-Life Supporters Rally in Front of Planned Parenthood Nationwide
 - Girl Scouts are Planning an Organization-Wide Fundraiser for Planned Parenthood
 - Republicans Pledge to Only Fund National Pregnancy Care Center That Does Not Perform Abortions
 - Planned Parenthood Receives a Sum of \$1,000,000 Donation from Crowd Sourcing
 - On-Campus Pro-Life Supporters Significantly Reduce the Number of Abortions among Students
 - Planned Parenthood Visits Campuses to Educate Young Women about the Importance of Having a Choice
 - Planned Parenthood Now Required to Provide Classes on Abortion Before Getting Consent for the Procedure
-

Note: The formats of these headlines were randomized to minimize any headline-specific effects.

Table 3: Study 1 Results for Believability and User Actions

Independent Variables		(1) Believability	(2) Read	(3) Like	(4) Support	(5) Oppose	(6) Share
Believability			0.631*** (0.023)	0.444*** (0.021)	0.367*** (0.020)	0.358*** (0.024)	0.499*** (0.021)
Expert Rating	High Rating	0.177* (0.081)	-0.232** (0.076)	-0.123 (0.069)	-0.092 (0.068)	-0.087 (0.080)	-0.081 (0.070)
	Medium Rating	-0.454*** (0.081)	-0.234** (0.076)	-0.106 (0.069)	-0.067 (0.068)	-0.060 (0.080)	-0.107 (0.071)
	Low Rating	-0.784*** (0.081)	-0.420*** (0.076)	-0.122 (0.070)	-0.073 (0.069)	-0.109 (0.081)	-0.079 (0.071)
User Article Rating	High Rating	0.180* (0.081)	0.125* (0.075)	-0.013 (0.069)	-0.125 (0.068)	-0.039 (0.080)	-0.086 (0.070)
	Medium Rating	-0.176* (0.081)	0.048 (0.075)	-0.003 (0.069)	-0.029 (0.068)	-0.131 (0.080)	-0.130 (0.070)
	Low Rating	-0.601*** (0.081)	-0.039 (0.076)	-0.069** (0.069)	-0.090 (0.068)	-0.047 (0.081)	-0.125 (0.071)
User Source Rating	High Rating	0.020 (0.080)	-0.205** (0.074)	-0.214** (0.068)	-0.104 (0.067)	-0.024 (0.079)	-0.140* (0.070)
	Medium Rating	-0.265*** (0.080)	-0.189** (0.074)	-0.198** (0.068)	-0.088 (0.067)	-0.117 (0.079)	-0.113 (0.070)
	Low Rating	-0.366*** (0.080)	-0.042 (0.074)	-0.156* (0.068)	-0.007 (0.067)	-0.092 (0.079)	-0.084 (0.070)
Confirmation Bias		0.585*** (0.021)	0.253*** (0.020)	1.022*** (0.018)	0.817*** (0.018)	-0.616*** (0.021)	0.437*** (0.019)
Female		-0.150* (0.076)	-0.097 (0.128)	-0.277** (0.102)	-0.463*** (0.110)	-0.702*** (0.144)	-0.474*** (0.124)
Working Age		-0.026 (0.171)	-0.537* (0.286)	-0.410 (0.227)	-0.220 (0.247)	-0.426 (0.322)	-0.449 (0.279)
Retirement Age		-0.116 (0.190)	-0.577* (0.317)	-1.035** (0.252)	-1.008*** (0.274)	-1.458*** (0.357)	-1.249*** (0.309)
Bachelor's Degree		0.020 (0.088)	-0.171 (0.147)	-0.228* (0.117)	-0.259* (0.127)	-0.388* (0.165)	-0.300* (0.143)
Graduate Degree		0.333*** (0.100)	0.375* (0.167)	0.253 (0.133)	0.279 (0.144)	0.734*** (0.187)	0.310 (0.162)
Facebook Use More Than Once a Week		-0.004 (0.133)	0.196 (0.222)	0.307 (0.177)	-0.274 (0.192)	-0.244 (0.250)	0.372 (0.217)
Facebook Use More Than Once a Day		-0.014 (0.117)	0.595** (0.195)	-0.490** (0.155)	-0.332* (0.169)	-0.364 (0.220)	0.613*** (0.190)

Note: Estimated coefficients (and standard errors). Among the independent variables, believability and confirmation bias are standardized; indicator variables are not. (N=590)

*** $p \leq 0.001$, ** $p \leq 0.01$, * $p \leq 0.05$.

Table 4: Study 1 Wald Tests for H3

Rating Tier	Comparison	Chi-Squared
High Rating	Expert Article Rating (0.177) vs. User Article Rating (0.180)	0.00
	Expert Article Rating (0.180) vs. User Source Rating (0.020)	2.17
	User Article Rating (0.180) vs. User Source Rating (0.020)	2.25
Medium Rating	Expert Article Rating (−0.454) vs. User Article Rating (−0.176)	6.69**
	Expert Article Rating (−0.454) vs. User Source Rating (−0.265)	3.12
	User Article Rating (−0.176) vs. User Article Rating (−0.265)	0.71
Low Rating	Expert Article Rating (−0.784) vs. User Article Rating (−0.601)	2.89
	Expert Article Rating (−0.784) vs. User Source Rating (−0.366)	15.39***
	User Article Rating (−0.601) vs. User Article Rating (−0.366)	4.87*

Note: **p ≤ 0.001, **p ≤ 0.01, *p ≤ 0.05.

Table 5: Study 2 Participant Group Description

Group Variable	Description	Approximate Percentage (%)
Age	College age (18–24)	11
	Working age (25–64)	78
	Retirement age (≥65)	11
Education	Below bachelor's degree	66
	Bachelor's degree	21
	Graduate degree	13
Facebook Use	Once a week or less	13
	More than once a week	32
	More than once a day	55

Table 6: Study 2 Results for Believability and User Actions

Independent Variables		(1) Believability	(2) Read	(3) Like	(4) Support	(5) Oppose	(6) Share
Believability			0.597*** (0.037)	0.532*** (0.038)	0.409*** (0.036)	0.192*** (0.039)	0.414*** (0.032)
Expert Rating	High Rating	0.026 (0.102)	0.122 (0.105)	0.105 (0.109)	0.014 (0.103)	0.077 (0.111)	0.029 (0.092)
	Low Rating	-0.365*** (0.102)	-0.133 (0.105)	-0.025 (0.109)	-0.030 (0.103)	0.089 (0.111)	-0.066 (0.092)
User Article Rating	High Rating	0.182 (0.108)	-0.006 (0.111)	0.063 (0.114)	0.067 (0.108)	-0.018 (0.117)	0.015 (0.097)
	Low Rating	-0.250* (0.108)	-0.023 (0.111)	0.079 (0.114)	0.143 (0.108)	-0.152 (0.117)	0.034 (0.097)
User Source Rating	High Rating	-0.089 (0.106)	0.303** (0.109)	0.217 (0.113)	0.120 (0.107)	-0.084 (0.115)	0.141 (0.095)
	Low Rating	-0.077 (0.106)	0.165 (0.109)	0.187 (0.113)	0.103 (0.107)	0.005 (0.115)	-0.004 (0.023)
Confirmation Bias		0.153*** (0.027)	0.181*** (0.027)	0.382*** (0.028)	0.275*** (0.026)	-0.244*** (0.028)	0.119*** (0.184)
Female		-0.320* (0.127)	-0.091 (0.165)	0.009 (0.169)	-0.079 (0.179)	-0.098 (0.185)	-0.152 (0.290)
Working Age		0.092 (0.201)	0.006 (0.260)	-0.348 (0.266)	-0.319 (0.282)	-0.385 (0.291)	-0.122 (0.390)
Retirement Age		-0.071 (0.271)	-0.126 (0.350)	-0.972** (0.359)	-1.206*** (0.380)	-1.204** (0.392)	-1.247*** (0.229)
Bachelor's Degree		-0.022 (0.159)	-0.115 (0.205)	-0.287 (0.210)	-0.337 (0.223)	-0.440 (0.230)	-0.438 (0.276)
Graduate Degree		0.189 (0.192)	0.157 (0.248)	-0.056 (0.254)	-0.055 (0.270)	-0.019 (0.278)	-0.225 (0.298)
Facebook Use More Than Once a Week		0.043 (0.206)	0.196 (0.267)	0.008 (0.274)	0.041 (0.290)	0.075 (0.300)	-0.068 (0.285)
Facebook Use More Than Once a Day		0.316 (0.198)	0.625* (0.256)	0.208 (0.262)	0.106 (0.277)	0.175 (0.286)	-0.116 (0.049)
First Half Articles		-0.129* (0.057)	0.083 (0.057)	0.039 (0.059)	0.023 (0.055)	0.049 (0.060)	0.108* (0.351)

Note: Estimated coefficients (and standard errors). Among the independent variables, believability and confirmation bias are standardized; indicator variables are not. (N=299)

*** $p \leq 0.001$, ** $p \leq 0.01$, * $p \leq 0.05$.

Appendix A. Rating Mechanism Descriptions

A.1.1. Study 1: For Expert Rating

For the remaining set of 6 articles, please assume the following scenario:

Facebook has formed a group of experts to assess the reliability of different news sources. These expert reviewers rate the sources in an objective, unbiased manner based on facts and data that may or may not support the claims made by the sources. Once a news source has posted enough articles to make a reasonable assessment, Facebook displays the Source Reliability Scale.

This new expert rating offered by Facebook is like the ones offered by MotorTrend (for cars), Michelin Guide (for restaurants), and PolitiFact (for the accuracy of claims by elected officials and other).

A.1.2. Study 2: For Expert Rating

Please assume the following scenario:

Facebook has formed a group of experts to assess the reliability of different news sources. These expert reviewers rate the sources in an objective, unbiased manner based on facts and data that may or may not support the claims made by the sources. Once a news source has posted enough articles to make a reasonable assessment, Facebook displays the Source Reliability Scale.

This new expert rating offered by Facebook is like the ones offered by MotorTrend (for cars), Michelin Guide (for restaurants), and PolitiFact (for the accuracy of claims by elected officials and other).

We will now show you 6 news articles (their headlines) and ask you some questions about them. Please take a look at each headline at your usual pace and work through the questions.

Since the news source rating is a new feature that Facebook is rolling out, some sources may not have the ratings available yet.

A.2.1. Study 1: For User Article Rating

For the remaining set of 6 articles, please assume the following scenario:

Facebook has introduced a new feature where users can rate the reliability of the individual articles they see. That is, this crowd-sourced rating feature collects the ratings from users like you. Facebook displays the Source Reliability Scale based on an average of the users' ratings of the sources' articles. Note that, to guard against a large volume of frivolous rating submissions, only verified active users can participate in rating, and they can rate a given article only once.

This new article rating feature offered by Facebook is like the ones offered by many online retailers, such as Amazon, for rating individual products, but provides a rating by averaging all the individual article reviews to provide an overall source rating.

A.2.2. Study 2: For User Article Rating

Please assume the following scenario:

Facebook has introduced a new feature where users can rate the reliability of the individual articles they see. That is, this crowd-sourced rating feature collects the ratings from users like you. Facebook displays the Source Reliability Scale based on an average of the users' ratings of the sources' articles. Note that, to guard against a large volume of frivolous rating submissions, only verified active users can participate in rating, and they can rate a given article only once.

This new article rating feature offered by Facebook is like the ones offered by many online retailers, such as Amazon, for rating individual products, but provides a rating by averaging all the individual article reviews to provide an overall source rating.

We will now show you 6 news articles (their headlines) and ask you some questions about them. Please take a look at each headline at your usual pace and work through the questions.

Since the news source rating is a new feature that Facebook is rolling out, some sources may not have the ratings available yet.

A.3.1. Study 1: For User Source Rating

For the remaining set of 6 articles, please assume the following scenario:

Facebook has introduced a new feature where users can rate the reliability of the source they see. That is, this crowd-sourced rating feature collects the ratings from users like you. Facebook displays the Source Reliability Scale based on the users' source ratings. Note that, to guard against a large volume of frivolous rating submissions, only verified active users can participate in rating, and they can rate a given source only once.

This new user rating feature offered by Facebook is like the ones offered by Yelp (for rating restaurants) and eBay (for rating sellers).

A.3.2. Study 2: For User Source Rating

Please assume the following scenario:

Facebook has introduced a new feature where users can rate the reliability of the source they see. That is, this crowd-sourced rating feature collects the ratings from users like you. Facebook displays the Source Reliability Scale based on the users' source ratings. Note that, to guard against a large volume of frivolous rating submissions, only verified active users can participate in rating, and they can rate a given source only once.

This new user rating feature offered by Facebook is like the ones offered by Yelp (for rating restaurants) and eBay (for rating sellers).

We will now show you 6 news articles (their headlines) and ask you some questions about them. Please take a look at each headline at your usual pace and work through the questions.

Since the news source rating is a new feature that Facebook is rolling out, some sources may not have the ratings available yet.

Appendix B. Experiment Snapshots



(a) With the control treatment

Robin Miller
4 hrs · 🌐

Universities Connect their Healthcare with Planned Parenthood to Provide Better Care to Students



Universities Connect their Healthcare Systems with Planned Parenthood to Provide Better Care to Students

One university spokesperson says that the new alliance will provide easier access to much needed health care services to female students.

NEWSUNION.COM

Source Reliability
★★★★★
4.9 out of 5

From Facebook's Fact-Checking Board of Experts.
[See how we rate](#)

Like Comment Share

(b) With expert rating treatment

Robin Miller
4 hrs · 🌐

Universities Connect their Healthcare with Planned Parenthood to Provide Better Care to Students



Universities Connect their Healthcare Systems with Planned Parenthood to Provide Better Care to Students

One university spokesperson says that the new alliance will provide easier access to much needed health care services to female students.

NEWSUNION.COM

Based on 15,138 article ratings from users like you.
[See rating details](#)

Source Reliability
★★★★★
4.9 out of 5
[Rate this article](#)

Like Comment Share

(c) With user article rating treatment

 **Robin Miller**
4 hrs · 🌐

Universities Connect their Healthcare with Planned Parenthood to Provide Better Care to Students



Universities Connect their Healthcare Systems with Planned Parenthood to Provide Better Care to Students

One university spokesperson says that the new alliance will provide easier access to much needed health care services to female students.

NEWSUNION.COM

Based on 15,138 ratings from users like you.
[See rating details](#)

Source Reliability
★★★★★
4.9 out of 5
[Rate this source](#)

👍 Like 💬 Comment ➦ Share

(d) With user source rating treatment

Appendix C. Messages for Rating Mechanisms' Emotional-Cognitive Scale Test

C.1. For Expert Rating

Some decisions use a detailed step-by-step analysis, others are based on gut-level instinct, and some are combination of the both. Do you think the expert panel rating reliability would use detailed analysis, instinct, or a combination of both? Please move the slider below to show how you think the panel's rating would balance these approaches.

C.2. For User Article and User Source Ratings

Some decisions use a detailed step-by-step analysis, others are based on gut-level instinct, and some are combination of the both. Do you think the users rating reliability would use detailed analysis, instinct, or a combination of both? Please move the slider below to show how you think users' ratings would balance these approaches.

Appendix D. Summary Statistics

D.1. For Study 1

Variable	Mean (Std)	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Believability	4.609 (1.692)													
2. Conf. Bias	2.977 (10.673)	0.480***												
3. Rated High	0.250 (0.433)	0.122***	0.044**											
4. Rated Med	0.250 (0.433)	-0.043**	-0.017	-0.333***										
5. Rated Low	0.250 (0.433)	-0.152***	-0.050***	-0.333***	-0.333***									
6. Age	3.832 (1.533)	-0.058***	-0.098***	0.000	0.000	0.000								
7. Degree	4.015 (1.534)	0.073***	0.057***	0.000	0.000	0.000	-0.019							
8. FB Use	5.498 (1.876)	0.015	0.039**	0.000	0.000	0.000	-0.092***	-0.034*						
9. Female	1.505 (0.500)	-0.086***	-0.085***	0.000	0.000	0.000	-0.142***	-0.224***	0.111***					
10. Read	4.373 (2.113)	0.511***	0.383***	0.040**	-0.022	-0.069***	-0.088***	0.074***	0.107***	-0.047**				
11. Like	3.378 (2.183)	0.537***	0.675***	0.037*	-0.022	-0.063***	-0.202***	0.046**	0.097***	-0.099***	0.655***			
12. Support	3.132 (2.112)	0.497***	0.620***	0.024	-0.016	-0.046**	-0.210***	0.050***	0.073***	-0.139***	0.592***	0.881***		
13. Oppose	2.939 (2.066)	0.297***	0.079***	0.014	-0.016	-0.019	-0.212***	0.060***	0.062***	-0.145***	0.420***	0.444***	0.529***	
14. Share	3.206 (2.134)	0.509***	0.505***	0.031*	-0.024	-0.050***	-0.216***	0.037*	0.122***	-0.122***	0.633***	0.820***	0.842***	0.618***

Note: ***p ≤ 0.001, **p ≤ 0.01, *p ≤ 0.05.

D.2. For Study 2

Variable	Mean (Std)	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Believability	4.358 (1.501)													
2. Conf. Bias	1.932 (9.460)	0.497***												
3. Rated High	0.333 (0.472)	0.050*	0.037											
4. Rated Low	0.333 (0.472)	−0.079***	−0.020	−0.500***										
5. Age	3.291 (1.599)	0.004	−0.070**	0.000	0.000									
6. Degree	3.605 (1.506)	−0.007	−0.047*	0.000	0.000	0.205***								
7. FB Use	5.217 (1.925)	0.104***	0.024	0.000	0.000	0.062**	−0.085***							
8. Female	1.502 (0.500)	−0.090***	−0.022	0.000	0.000	−0.170***	−0.172***	0.120***						
9. Read	3.971 (1.950)	0.486***	0.435***	0.049*	−0.042	−0.053*	−0.008	0.160***	−0.027					
10. Like	3.562 (1.987)	0.435***	0.539***	0.034	−0.018	−0.169***	−0.084***	0.078***	0.012	0.637***				
11. Support	3.327 (1.960)	0.402***	0.485***	0.018	−0.008	−0.183***	−0.103***	0.048*	−0.002	0.571***	0.828***			
12. Oppose	3.303 (1.958)	0.265***	0.121***	0.005	−0.011	−0.205***	−0.120***	0.049*	0.004	0.407***	0.508***	0.600***		
13. Share	3.391 (1.939)	0.410***	0.401***	0.026	−0.028	−0.199***	−0.130***	0.008	−0.020	0.577***	0.756***	0.792***	0.688***	
14. First Half Articles	0.667 (0.472)	−0.038	−0.010	−0.039	−0.019	0.000	0.000	0.000	0.000	0.008	−0.002	−0.003	0.008	0.018

Note: *** $p \leq 0.001$, ** $p \leq 0.01$, * $p \leq 0.05$.