

# Understanding How Readers Determine the Legitimacy of Online News Articles in the Era of Fake News

Srihaasa Pidikiti, Jason Shuo Zhang, Richard Han, Tamara Lehman, Qin Lv, and Shivakant Mishra  
*University of Colorado Boulder, Boulder, CO, USA*  
 {srihaasa.pidikiti, jasonzhang}@colorado.edu

**Abstract**—Internet users are routinely exposed to fake news in their social media feeds. The main goal of this paper is to identify the factors readers consider important in discriminating against fake news from true news when reading an online news article. We design and conduct three surveys using Amazon Mechanical Turk to identify the top factors and rate them under diverse scenarios. Our results suggest that people perceive news Source and Content to be the most important factors, in general, to distinguish fake news from true news, however, their importance reduces in practice when people actually read a news article. Furthermore, the importance of different factors in the credibility determination of a news article varies with people’s political leanings. Our work is the first of its kind and offers new insights into how people determine the legitimacy of online news articles.

**Index Terms**—fake news, critical thinking, source, social media

## I. INTRODUCTION

The emergence of the 24-hour news cycle, citizen journalism, and the abundant information available at our fingertips on social media has had a profound impact on not only how we consume news, but also how we trust the news. Besides, in the digital era, there are ecosystems created by domestic groups or foreign actors to intentionally promote fake news and conspiracy theories, making this problem much more complicated. In November 2016, an analysis conducted by BuzzFeed News found that the top fake election news stories generated more engagements (likes and shares) on Facebook than the top election stories from nineteen major news outlets combined [1]. The spread of misinformation is a real threat to our democracy, as it can disrupt the public trust of legitimate news sources and undermine our political spectrum.

To prevent the spread of misinformation, fact-checking has become an article of faith in the era of dueling facts. For example, social media platforms like Facebook and Twitter have teamed up with fact-checking organizations to label fake stories and remove them from the platform [2]. However, the effectiveness of fact-checking is being questioned, as many citizens may resist fact-checking messages due to their prior beliefs. Moreover, in a world in which an entire industry exists to deceive consumers of information, it is hard for these service providers to scale up and automatically fact-check the plethora of information coming from different news

sources. With all these challenges, developing people’s ability to discern fact from fiction and responding productively to those who counter data with a belief is critical for fighting misinformation in this new era.

In this work, we seek to systematically understand how people determine the legitimacy of online news articles. We explore factors that play important roles in this determination and examine how they vary in different scenarios. While some researchers explain the role of analytical thinking in detecting fake news [3]–[6], and other studies analyze how the presentation of an article helps in detecting if it is fake [7], [8], there is no prior work to analyze which factors of the article representation make a larger impact on the reader in discerning fact from fiction. Our contribution is the shift of the analysis of the article representation to focus on the readers. In particular, we aim to answer the following three research questions:

- **RQ1:** What are the most important factors that people consider in deciding if an online news article is fake?
- **RQ2:** Do people rate these factors differently in general compared to when they actually read a news article?
- **RQ3:** How does political leaning affect people’s judgment in differentiating representation of fake and true news articles?

To answer these questions, we have designed several surveys and used Amazon Mechanical Turk to administer them and collect results. We first conduct a pre-survey to identify the factors that people consider important in determining whether or not a news article is fake. Next, for each of our research questions, we conduct a separate online survey. In the first survey, we ask our participants to rate the importance of each factor based on their prior experience, without reading any news article. This survey reveals that Content and Source are rated as the most important factors, while Picture and Date are rated as the least important. In the second survey, we present each participant with 20 news articles and ask them to assess for each news article the importance of each factor in determining whether that article is fake or not. Interestingly, when reading these news articles, Content and Source turn out to be much less important in determining the legitimacy of the news article whereas Title turns out to be an important factor. These results suggest that the factors that affect people’s judgment in determining whether a news article is fake or not

vary in actuality. In the third survey, we select a balanced set of news articles in terms of the political leaning of the news content and explore how people’s political ideologies affect their ability to discriminate fake news articles from true ones. We find that both conservatives and liberals are more likely to correctly identify political news credibility that is not fake and consistent with their beliefs.

This work provides one of the first studies addressing the consumer side of online news, namely what factors news readers consider to decide whether an online news article is fake or not. Our study makes the following findings:

- We show that Content and Source are the most important factors in deciding the credibility of online news articles when asked in the abstract.
- We observe that people’s perception of what matters in discerning truth from fiction differs when presented with articles. Content and Source are considered less important and Title is also rated as one of the important factors when people actually read a news article compared with when they are being asked abstractly based on prior experience.
- We analyze the differences in identifying fake news among political leanings. Both conservatives and liberals are more likely to accurately determine the credibility of a political news article that is not fake and consistent with their beliefs.
- Finally, we conduct qualitative analysis to understand user’s confidence or uncertainty about an article’s legitimacy, which are further used to corroborate the importance of certain factors in the article’s representation.

## II. BACKGROUND AND RELATED WORK

Studying the recognition of fake articles can be broken down into several categories. In this section, we break down existing work in each category and explain how our study fits in the big picture. Most of these pieces work complement our study and in some cases corroborate findings.

### A. Beliefs, Familiarity and Confirmation Bias

Several pieces of work have contradicting conclusions on whether a person’s beliefs affect their ability to identify fake articles. According to a study conducted on Facebook, users are so susceptible to confirmation bias, that they would be better off tossing a coin to determine what is true and what is false [3]. In this study, authors conclude that if the article aligns with the users’ prior beliefs there is greater cognitive activity. On the other hand, when the article opposes the user’s beliefs there is less cognitive activity and sometimes the article is simply ignored. Harper *et.al.* suggests that party identification is a major factor in building people’s perceptions of and reactions to political news [9]. Another study established that delusion-prone individuals, dogmatic individuals, and religious fundamentalists are more likely to believe fake news, which is explained by suggestions that such individuals have reduced involvement in analytical thinking applications [6]. In contrast, another research study found

that the analytical thinking abilities of individuals help in determining fake political articles irrespective of them aligning with their beliefs [5].

Besides a person’s own prior beliefs, repeated exposure to a fake article may ultimately lead to developing a belief that it is true. This effect is often referred to as the continued influence effect of the misinformation phenomenon. In a prior study, this phenomenon was studied in relation to the correction of misinformation [4]. Swire *et.al.* found that in fact, this phenomenon impacts a person’s ability to detect fake information, even after it has been corrected.

Studies of factors affecting a person’s ability to discern fake news, such as familiarity and a person’s own beliefs, are complementary to our work. Understanding how a person comes to a decision regarding an article’s truthfulness is a complex system that combines multiples factors. To aid in a better understanding of this complex system, our research is the first work that analyses the relative importance of an article’s representation factors like Title, Content, Source, Picture, Authors and Date in how readers determine the legitimacy of online news articles.

### B. Fake News Dissemination in Social Media

A majority of U.S. adults, about 62 percent, get news on social media, and 18 percent do so often, according to a survey [13]. Social confirmation in such platforms plays a major role in evaluating the credibility of the article by the users [14]. If several people consume some information, recommend it, and agree with it, then users assume it is credible information. There is also a high chance that a person might not be exposed to opposing beliefs and, as a result, experience echo chamber effects [15]. Prior studies made successful attempts in using user profile and sentiment score of user comments on a news article to detect fake news [16], [17]. While our study is limited to presenting news articles in a vacuum, the article’s representation in social media can still have an impact on the user’s ability to discern fake news and thus we find our work to be complementary to these social media studies.

### C. Fake News Detection

Fake news articles are represented in a way that favors their dissemination and credibility in comparison to true articles. In one study, it is shown that fake and true news articles are notably distinguishable, specifically in the title of the articles [7]. Money making clickbait articles, unlike true articles, have lengthier titles with both content and functional words and frequently use words misleading the users [18].

Another notable distinction between true and fake news is in the way the content is presented. In fake news articles, the content is presented similar to that of a satire. There is evidence that fake news articles contain content that is less complex, less self-relevant, and more characterized by negativity [8]. Content’s sentiment score has effective usage in fake news detection [19].

The source of the article can also be a good indicator of fake news. It was found that articles that highlight the source make readers critical about their credibility, which implies a correlation between the source and the article’s credibility [20]. Indeed, Baly *et.al.* found that the source of an article is one of the most important factors for credibility [21].

This prior work focuses on the unique differences between true and fake news articles and thereby aids in automatic fake news detection, essentially focusing on the producer side of online news dissemination. This paper, on the other hand, explores the consumer side of online news dissemination. In particular, we focus on what factors consumers of online news consider in determining whether a news article is fake or not.

### III. METHODS

We have designed and conducted several surveys to answer our research questions. First, to determine the most important factors that users consider for evaluating the credibility of news articles, we conduct a preliminary survey asking participants to rate the importance of six commonly considered factors (i.e., Title, Picture, Content, Source, Author and Date). In addition, we allow them to suggest any other factors that they deem critical in making their judgment. The survey results show that other than the six factors, survey participants mentioned two other factors, namely news article Popularity, and news article Recommendation. Based on this feedback, we conduct a second survey that includes these two factors: Popularity and Recommendation. However, the results show that Popularity and Recommendation are the least important factors compared to the other six factors. Thus, we decide to focus on the original six factors (Title, Picture, Content, Source, Author, and Date) in determining the credibility of a news article. We conduct Survey 1, Survey 2, and Survey 3 to answer research questions RQ1, RQ2 and RQ3, respectively. We use the Amazon Mechanical Turk Platform to run these surveys. Our experimental protocol is reviewed and conducted under IRB Protocol 19-0610.

*A. Survey 1: What are the most important factors that people consider in deciding if an online news article is fake?*

The design of Survey 1 aims to find out the most important factors that users consider in determining the credibility of online news articles. In this survey, we merely ask participants to rate the importance of the six factors without showing any specific news article. The goal is to understand the relative importance of the six factors in general, irrespective of the specifics of a news article.

This question includes a rating scale High, Moderately High, Moderately Low, and Low for each factor. Demographic information like Gender, Political Leaning of each respondent is also collected as part of the survey. To ensure the quality of the survey, we include some test questions as well. We pay 50 cents to each participant to complete the whole survey. A total of 100 responses are collected that passed the quality check. These 100 respondents are from the US and have an equal age distribution, with 20% in each of the five age groups. The

**When reading the below news article, to determine whether or not this news article is likely to be false, please rate each of the given factors below based on their importance for you?**

**Breaking News: China will admit coronavirus coming from its P4 lab**

JANUARY 25, 2020 BY GM35



Source: GNEWS

As the “novel” coronavirus originated in Wuhan is spreading to ten countries, more and more people including international bio-weapon experts are questioning its link to the Wuhan P4 lab located about 20 miles from a seafood market where the first few cases of human infections were found

A reliable source told *Miles Guo* today that the Chinese Communist Party (CCP) will admit to the public of an “accidental” leak of lab-created virus from a P4 lab in Wuhan to put blames on “human errors”. But the official announcement is still being finalized initially. The Chinese communist propaganda machines were blaming the virus on wild animals like bats by showing many videos of people eating bats.

In January 2018, a biosafety level four (BSL-4) laboratory was built in the city of Wuhan, which focuses on the control of emerging diseases and stores purified SARS and other types of viruses. It is supposed to act as a WHO reference laboratory linked to similar labs around the world.

Fig. 1: In Survey 2 we show our participants 20 different news articles and ask them to rate the importance of each factor for each article.

distribution of political leaning of respondents is 41 % liberals, 25% conservatives and 33% moderates, and the remaining 1% choose “other.”

*B. Survey 2: Do people rate factors differently in general compared to when they actually read a news article?*

The motivation of Survey 2 is to understand whether the relative importance of factors to evaluate the legitimacy of an article changes with a concrete article compared to Survey 1. In this survey, we present 20 different news articles (See Fig. 1) to the participants, and for each of the 20 articles, ask them to rate the importance of the six factors that influence their decision for each article. These 20 articles are randomly taken from the fact-checking organizations Snopes [22] and Politifact [23], and the number of true and fake news articles are balanced. Among these articles, 11 of these articles are published in 2020, 5 in 2019, 3 in 2018, and 1 in 2014.

Similar to Survey 1, the rating choices are High, Moderately High, Moderately low, and low for each factor and demographic information like Gender, Political Leaning, Age of each respondent is also collected. Each participant is paid 50 cents to complete the whole survey. In total, 100 responses are collected that passed our quality checks. These 100 respondents are from the US, and there is an equal distribution of respondents in each of the five age groups. The distribution of the political leaning of respondents is 51% liberals, 25 % conservatives, and 21% moderates, and the remaining 3% choose “other.”

*C. Survey 3: How do political leanings affect people’s judgment in distinguishing a fake news article from a true one?*

The goal of Survey 3 is to understand how political leanings affect the users’ judgment in deciding whether a news article is fake or not. Similar to Survey 2, we present 20 different news articles (See Fig. 2) and for each of them, we ask whether the article is fake or not in addition to the question to rate the importance of the factors that influenced their decision making. These articles are taken from fact-checking organizations Snopes [22] and Politifact [23]. Of these 20 articles, we choose five true articles that are liberal-leaning, five true articles that are conservative-leaning, five fake articles that

Do you think this news article is fake?

**When Iran Took Americans Hostage, Bernie Backed Iran's Defenders**

JANUARY 17, 2020 BY RONALD RADOSH



Source: The Daily Beast

Bernie Sanders, a top competitor in the Democratic primaries, has attacked Joe Biden for bringing "just a lot of baggage" into the race. But if past views are a major consideration, consider the baggage that Sanders drags into the campaign. Go back over 40 years, to the start of Iran's long conflict with the United States. On April 1, 1979, the theocratic Islamic Republic of Iran was proclaimed. Ayatollah Ruhollah Khomeini, who had returned to Iran from exile to assume command of the revolt, became Supreme Leader in December of that year. His rise was accelerated by the seizure on Nov. 4 of 52 American diplomats and citizens, and citizens of other countries, at the U.S. Embassy in Tehran. The hostage crisis became the means by which the Ayatollah crushed political opponents in Iran. Dealing with the hostage taking became the overwhelming political crisis for President Jimmy Carter. It lasted 444 days. Virtually all Americans—Democrats, Republicans and independents—united in support of the hostages and the international call for their freedom. One prominent political figure on the 2020 stage, then almost completely unknown, stood apart by joining a Marxist-Leninist party that not only pledged support for the Iranian theocracy, but also justified the hostage taking by insisting the hostages were all likely CIA agents. Who was that person? It was Bernie Sanders. Sanders would like the public to believe, as an *ABC News* put it, that "Democratic socialism [is] the economic philosophy that has guided his political career." But that has not always been the case. In 1977, he left the tiny left-wing Liberty Union Party of Vermont that he'd co-founded, and in 1980 instead aligned himself with the Socialist Workers Party (SWP), the self-proclaimed Trotskyist revolutionary party, became its presidential elector in Vermont, and campaigned for its candidates and platform that defended the Iranian hostage seizure.

Fig. 2: The template of Survey 3 is the same as Survey 2, except we also ask if the news article is fake.

are liberal-leaning, and five fake articles that are conservative-leaning. Among these articles, 7 of these articles are published in 2020, 8 in 2019, 2 in 2018, 2 in 2016, and 1 in 2014. At the end of the survey, we inquire about the legitimacy of the article that participants are completely confident about and the article they are highly uncertain about and ask them to provide a specific explanation of why those articles have been chosen out of the shown twenty articles. We pay \$1.5 for each participant to complete the whole survey. In total, 100 responses are collected that passed our quality check. These 100 respondents are from the US. The distribution of political leaning of respondents is 40% liberals, 36% conservatives, and 23% moderates, and the remaining 1% choose "other."

**D. Clustering Analysis**

For surveys 1 and 2, we employ the k-means [24] algorithm for clustering respondents into three distinct groups using  $k=3$ . We specifically use  $k=3$  to see if clusters align with participant's political leaning (liberals, moderates, and conservatives). First, we transform each respondent's answers to the survey into a vector to apply the k-means algorithm. For Survey 1, we convert ratings given to each of the six factors (Title, Picture, Content, Source, Author and Date) into a six-dimensional vector for clustering. Similarly, for Survey 2, we cluster using participant's ratings of each of the six factors for each of the 20 articles, i.e., a 120-dimensional vector representation of each participant. Finally for Survey 3, to identify any specific patterns concerning the accuracy of credibility determination, we cluster participants using k-means based on their response to whether or not an article in the survey is fake and the k-value is chosen using elbow method. To do this, we represent each participant as a 20-dimensional vector formed by their responses to the article's credibility.

**IV. RESULTS**

We now present a detailed analysis of the survey results. Each survey builds upon the last to better understand the results. The ratings of each factor have been converted to a numerical format (High- 3, Moderately high- 2, Moderately low- 1, Low-0) for data analysis.

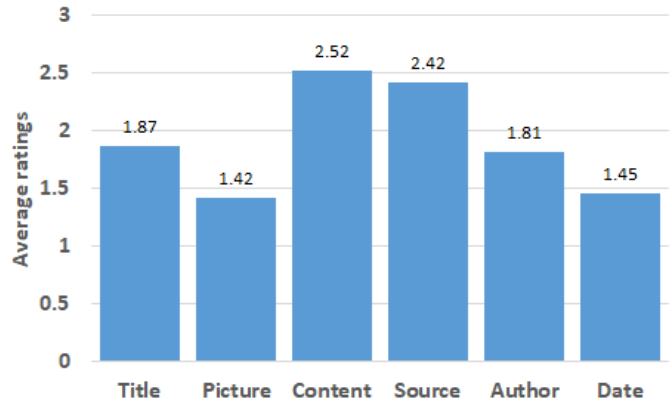


Fig. 3: Average importance ratings of factors from Survey 1. These results indicate that Title, Content, and Source are the most important factors.

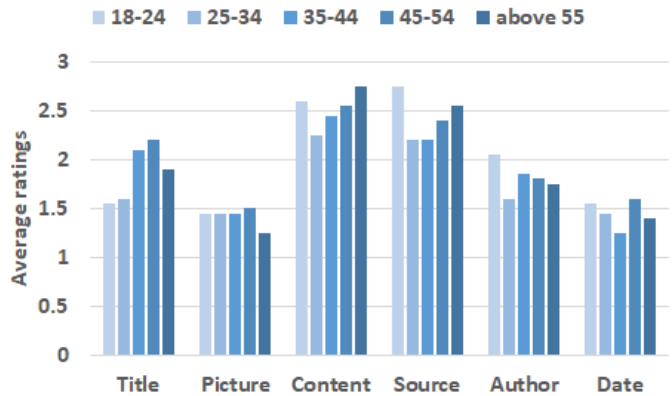


Fig. 4: Average importance ratings of factors grouped by age in Survey 1.

**A. Survey 1**

We calculate the average factors' ratings based on a participant's prior experience of reading online news articles. As shown in Fig. 3, Content and Source are rated as the most important factors in determining the credibility of political news. Whereas Title, Picture, Author and Date are less important factors in comparison. The observed mean differences of the six factors are statistically significant ( $p < 0.05$  according to one-way ANOVA test results). Based on further posthoc analysis there is a significant mean difference ( $p < 0.05$ ) except for between the pairs- (Title, Author), (Content, Source) and (Picture, Date).

These results answer research question RQ1 that Content and Source are the most important factors when deciding the credibility of a news article. We further examine how the relative importance of factors differs across age groups. As shown in Fig. 4 Content and Source are the most important factors irrespective of the age of respondents. We observe the differences in means of the six factors are statistically significant ( $p < 0.05$  according to one-way ANOVA test results) in each of the five age groups. In addition to Content and Source, Title and Author are the important factors in age groups 25-34 and 45-54 whereas Author and Title are the important factors in age groups 18-24 and 45-54 respectively based on the results of posthoc analysis.

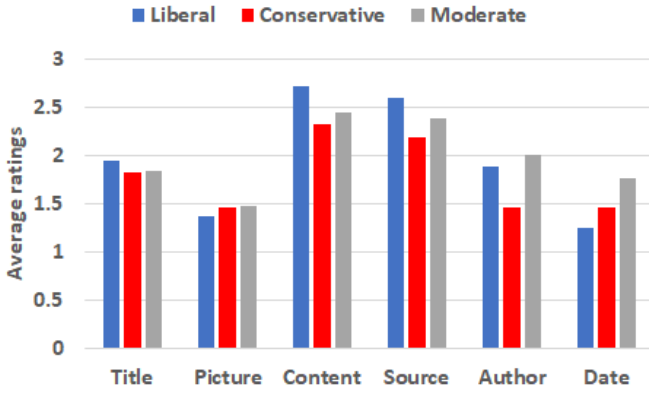


Fig. 5: Average importance ratings of factors grouped by political leaning for Survey 1. Content and Source are the most important factors, irrespective of political leaning of respondents.

TABLE I: Political leanings of participants in Survey 1: Cluster 1 has the highest percentage of liberals, Cluster 2 has the highest percentage of moderates, and Cluster 3 has the highest percentage of conservatives.

	Cluster 1	Cluster 2	Cluster 3
# participants	28	38	34
Leaning	Liberal- 46.43%	Liberal- 40.54%	Liberal- 41.18%
	Conservative- 25%	Conservative- 24.32%	Conservative- 35.29%
	Moderate- 28.57%	Moderate- 35.14%	Moderate- 23.53%

Finally, we analyze the impact of the political leaning of respondents in the determination of ratings. As shown in Fig. 5, political leaning does not have any significant effect on the participant’s response. Content and Source are the most important factors for the credibility determination irrespective of political leaning. We observe the differences in means of the six factors are statistically significant ( $p < 0.05$  according to one-way ANOVA test results) regardless of the political leaning. After posthoc analysis, Title is considered one another important factor for moderates whereas, for conservatives, Authors is considered as more important.

We perform k-means ( $k=3$ ) clustering with ratings for each of the six factors into 6-dimensional data. Fig. 6 shows each cluster’s average ratings. The political leaning break-out of three clusters are shown in Table I. We note that Cluster 1 has the highest percentage of liberals (46%), Cluster 2 has the highest percentage of moderates (35%), and Cluster 3 has the highest percentage of conservatives (35%). Respondents in Cluster 1 tend to rate every factor higher in comparison to respondents in other clusters. Content, Title and Source are the most important factors in this cluster. Respondents in Cluster 2 tend to rate Content, Source and Authors as the most important factors. In comparison, respondents in Cluster 3 tend to rate every factor lower, but similarly to Cluster 1, Content, Title and Source are still the most important ones.

### B. Survey 2

We evaluate the accuracy of people’s perceptions of which factors affect their ability to discern fake news. We compare

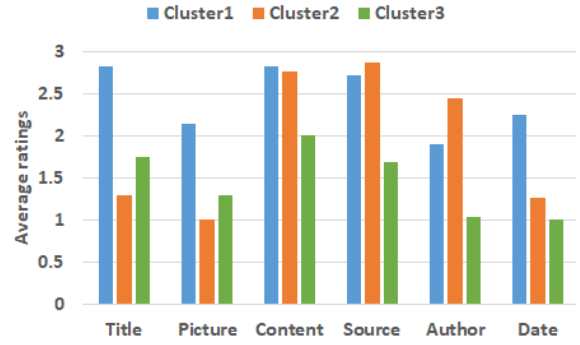


Fig. 6: Clustering analysis of Survey 1: A comparison of the average rating of each factor between three clusters.

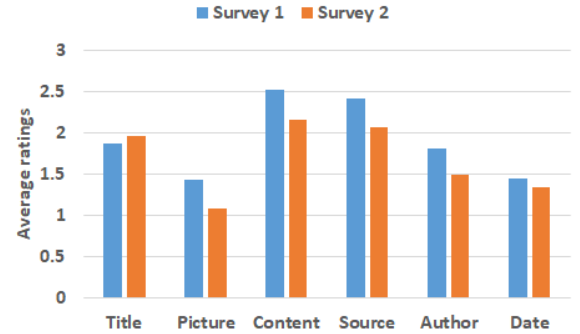


Fig. 7: A comparison between the average importance rating of each factor between Survey 1 and 2.

the relative importance of factors with and without news articles (Survey 1 vs. Survey 2). While there are several differences between Survey 1 and Survey 2, the statistically significant ones occur with respect to Picture, Content and Source ( $p < 0.05$  according to Mann-Whitney test results), as shown in Fig. 7. All three factors—Picture, Content and Source—are considered less important in determining legitimacy when reading actual articles. Nonetheless, Survey 2 shows consistent results with Survey 1 in that Content and Source remain the two most important factors for respondents in determining credibility. Alongside these, we identify Title as another important factor based on posthoc analysis results at a significance level of 5%. These results address the research question RQ2.

Fig. 8 shows the change in the importance of factors with respect to the age of respondents. We observe that except for the age 18-24 group, there is a decreasing trend in Picture’s importance and an increasing one in Content’s importance with increasing age and these correlations are statistically significant ( $p < 0.05$  according to spearman’s rank-order test results). Content, Source and Title are the three most important factors in credibility detection in Survey 2 irrespective of their ages.

We perform k-means ( $k=3$ ) clustering with ratings for each of the six factors of each of the 20 articles, i.e., 120-dimensional vector representation of each participant. Fig. 9 shows each cluster’s average ratings of factors. Respondents in Cluster 1 tend to rate Content, Title and Source as the



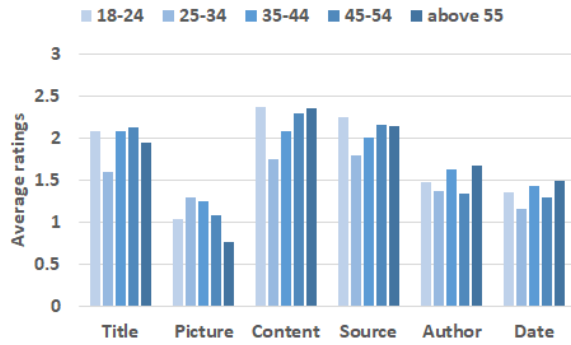


Fig. 8: Average importance rating of each factor with respondents grouped by age for Survey 2.

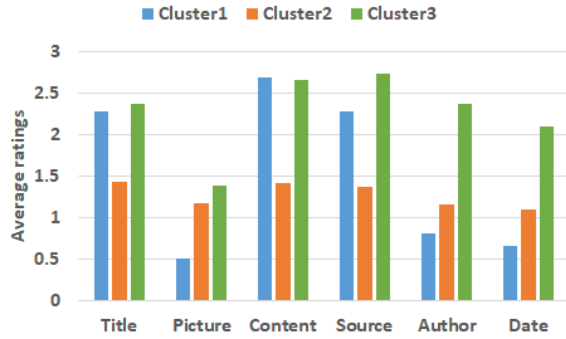


Fig. 9: A comparison of the average factors' ratings between three k-means clusters for Survey 2.

most important factors. In comparison, respondents in Cluster 2 rated every factor lower, but Content, Title and Source are still the most important ones. Respondents in Cluster 3 tend to rate every factor higher except Picture in comparison to respondents in other clusters; however, still Content and Source are the most important factors.

Table II gives the political leanings of each cluster. Among participants in Cluster 1, 78% of the respondents are liberals, while Cluster 2 and Cluster 3 have a more balanced quantity of each group. Note that Cluster 1 of Survey 2 has Picture, Author and Date with the least ratings of all (the average rating is lower than 1). Among the other clusters, no factor is rated less than 1.

### C. Survey 3

Given a balanced set of true and fake news articles for different political leanings, we seek to understand to what extent readers with different political leanings can accurately distinguish the true/fake articles. Fig. 10 shows that those with liberal leanings perform better in comparison to conservatives in detecting true news articles. This difference between liberals and conservatives is statistically significant ( $p < 0.05$  according to Mann-Whitney test results). On a whole, liberals perform better than conservatives in determining the credibility of a news article, and this difference is statistically significant ( $p < 0.05$ ).

We also observe that respondents are generally more accurate in determining the legitimacy of articles that align with

TABLE II: Clustering Analysis of Survey 2. Cluster 1 is overwhelmingly liberals. In comparison, Conservatives and moderates are more likely to be in Cluster 2 and Cluster 3.

	Cluster 1	Cluster 2	Cluster 3
# participants	24	41	35
Leaning	Liberal- 78.26%	Liberal- 42.50%	Liberal- 47.06%
	Conservative- 8.69%	Conservative- 30.00%	Conservative- 20.59%
	Moderate- 13.04%	Moderate- 27.50%	Moderate- 32.35%

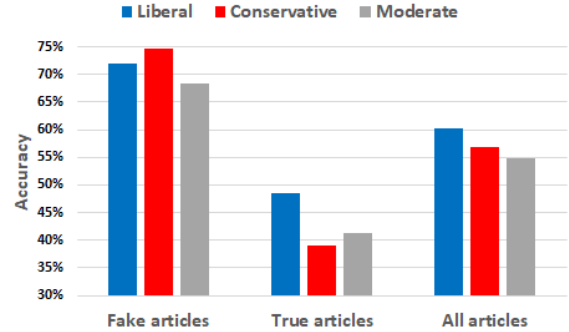


Fig. 10: There is a statistically significant difference in the accuracy of determining whether a news article is fake or not between conservatives, liberals, and moderates.

their political leaning. Conservatives achieve higher accuracy (74% vs. 63%) in identifying fake articles that are conservative in nature, while liberals achieve higher accuracy (81% vs. 75%) in identifying fake articles that are liberal.

The accuracy of detecting true articles is surprisingly lower than the fake article detection across all the political leaning groups. However, there are differences when comparing accuracies of the opposite groups. Specifically, liberals perform better than conservatives (44% vs. 42% accuracy) in identifying true conservative articles and true liberal articles (53% vs. 36% accuracy).

We perform k-means clustering ( $k=3$  optimal choice based on elbow method) with participants' responses of credibility determination for the 20 articles (20-dimensional data). We observe that respondents in Cluster 1 have higher fake news detection accuracy than true. In comparison to respondents in Cluster 1 and Cluster 3, respondents in Cluster 2 have lower true news detection accuracy. Respondents in Cluster 3 have higher true news accuracy than fake. Note that 56% of the respondents in Cluster 1 are liberals and 61% of the respondents in Cluster 2 are conservatives. In summary, with regard to RQ3, all these results indicate that political leaning is a potential factor that may influence people's accuracy in true/fake news detection.

### Qualitative Analysis

In Survey 3, we further conduct qualitative analysis to understand why our users are confident or uncertain about an article's legitimacy and how these factors play out in their thought process. Each participant is asked to choose one most-confident article and one least-confident article, and explain his/her decision for choosing each article. We examine

these responses and divide them into several categories. The top categories and sample responses for each category are summarized in Table III and Table IV.

First, we ask our participants why they are highly confident about the legitimacy of a news article they read. As shown in Table III, 21 respondents find the content of a certain article is untrustworthy, which lead them to determine that article as a fake one. Being already familiar with the article helped 20 respondents in determining its legitimacy confidently. The reputation of and familiarity with the source is a driving factor for 11 respondents in determining the article as fake. After that, eight respondents mention that the title of the article is highly misleading, which makes them highly confident that the article is fake. Responses in these four categories, especially about Title, Content and Source, are consistent with results in Survey 1 and Survey 2 that these three factors are rated highly important in determining the legitimacy of an article.

We also ask our participants why they are least certain about a news article they read. As indicated in Table IV, unfamiliarity with the content of the article is mentioned by 18 respondents as the primary reason for being uncertain about the article’s credibility. Also, ten respondents write that they trust the source but are not sure about the credibility of content. Interestingly, four respondents mention that content sounds credible, but they have no trust in the source. Six respondents are uncertain because some events mentioned in the article happened a long time ago. All these mentions specifically about Source and Content can be explained with the observations from previous surveys. Since Source and Content are rated important in determining an article’s legitimacy, one of them appearing credible and the other seeming untrustworthy may have prompted respondents to mention being uncertain of that article. Moreover, we observe that being familiar or unfamiliar with the story in the article has given confidence or caused uncertainty in the determination of the article’s legitimacy, respectively. In summary, all these responses about the decision-making process of respondents for the news articles they read provide qualitative evidence for the factors that rated high in all three surveys.

## V. DISCUSSION

Prior work shows that people, in general, make judgments about a news article’s legitimacy very quickly [25]. Far from carefully reading the source and content and possibly cross-checking the content, which are supposed to be the most important factors, many people make their judgment immediately after reading the title. This problem is prevalent in social media. Earlier research points out that nearly 80% of people will share news articles online right after reading the headline [26], and only the remaining 20% will read the rest. Even though our participants rate Content and Source as the most important factors, it is not reflected in their real-life behavior. This observation highlights the challenge for people in identifying fake news.

Furthermore, “filter bubbles” that social media spawns and “fake news” continue to have drastic consequences for political

TABLE III: Reasons for respondents being highly confident about a news article they read.

Category	#Tuckers	Sample responses
Unreliable Content	21	<i>Incredible stupidity of the content. Content was implausible with visual style similar to other fake websites. Poorly written and highly biased content.</i>
Familiarity	20	<i>I heard about this from multiple sources. I remember when the article was debunked in 2016. I recall reading in a well known newspaper the Washington examiner.</i>
Unreliable Source	11	<i>Name of the source website and ludicrous claims in the article. Source seemed particularly iffy. Source sounds unknown not reputable fake.</i>
Misleading Title	8	<i>Title seemed completely click-bait. Content had an incendiary title. Headline seemed fishy to me.</i>

TABLE IV: Reasons for respondents being highly uncertain about a news article they read.

Category	#Tuckers	Sample responses
Unfamiliar with the story	18	<i>I haven't heard the story before today. This was unfamiliar to me therefore I wasn't sure about its accuracy. It looked real but I was unsure because I wasn't aware of the alleged facts.</i>
Trust source but not content	10	<i>It didn't sound credible but came from credible source. It came from reliable source but content seemed suspicious. I highly trust National Public Radio but the content would be weird for a possible presidential candidate.</i>
Event date	6	<i>I do not know what happened in 1995 and trump helping out. It is hard to know because it was so long ago. Because the date of news is outdated it should be in 2020.</i>
Trust content but not source	4	<i>Content seemed legitimate but the source did not. Content seemed credible but I could not get past the source. Fox News is generally untrustworthy but they do have some news people that are actual journalists despite their conservative news.</i>

partisanship. Recent studies show that both Democrats and Republicans are 15% more likely to believe “ideologically aligned headlines” [27]. The results in Survey 3 echo these findings and show that people are more likely to believe political news that is consistent with their political leaning, regardless of whether the news is fake or true.

The findings in our work raise fruitful directions for future research. First, social media platforms may consider the factors identified here in the design of their news-forwarding mechanisms to minimize the spread of misinformation. Second,

future work may investigate not only the perception of fake news but also the consequences of reading news that changes beliefs and behavior in real life. It is an important research question in understanding to what extent and how fake news impacts previous and ongoing elections worldwide.

**Limitations.** There are several limitations in the current study that are important to acknowledge. First, our survey results rely on Amazon Mechanical Turk workers and the sample size is limited to 100 responses per survey. The representation of this sample should not be over-claimed. However, previous work suggests that Turk workers are a reliable source for studying behaviors associated with fake news determination, which gives us more confidence [5]. Moreover, all the discussed results are proven to be statistically significant. Second, the user interface of news articles in our study is a confounding factor and difficult to rule out. However, the survey is designed in such a way as to reflect how people read diverse news articles on social media. Thirdly, our study environment is not a perfect simulation of a real-world setting. The participants are asked to rate a stream of news articles with no posts from friends, family members, or advertisements. Finally, participants are asked to rate each of the article's representation factors' importance before determining the credibility of the article. This could have prompted participants to think more analytically. In the future, we plan to design an application to monitor participants' real-time online news consumption and query them regarding the most important factors aiding their judgment about each article's credibility to further explore and address these problems.

## VI. CONCLUSION

Guided by the open question of what factors contribute to the determination of the legitimacy of online news articles, we have conducted empirical surveys and identified that Content and Source are the most important factors. We further identify that when shown actual news content, Content and Source remain the most important factors, but their importance is less than when asked to rate these factors in the abstract. Our study, with an equal number of true and fake articles about liberals and conservatives, reveals differences in accuracy between two ideological groups. We also identify that these differences are more significant in true news articles than fake ones. Furthermore, we perform a qualitative analysis to understand why our users are confident or uncertain about an article's legitimacy and how these factors play out in their thought process. All these results shed light on readers' thinking process when deciding the credibility of an article and serve to extend the work on improving an individual's critical thinking skills to combat fake news online.

## REFERENCES

- [1] C. Silverman, "This analysis shows how viral fake election news stories outperformed real news on facebook," *BuzzFeed news*, vol. 16, 2016.
- [2] M. Anany, "The partnership press: Lessons for platform-publisher collaborations as facebook and news outlets team to fight misinformation," *Tow Center for Digital Journalism*, 2018.

- [3] P. Moravec, R. Minas, and A. R. Dennis, "Fake news on social media: People believe what they want to believe when it makes no sense at all," *Kelley School of Business Research Paper*, no. 18-87, 2018.
- [4] B. Swire, U. K. Ecker, and S. Lewandowsky, "The role of familiarity in correcting inaccurate information," *J. of experimental psychology: learning, memory, and cognition*, vol. 43, no. 12, p. 1948, 2017.
- [5] G. Pennycook and D. G. Rand, "Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning," *Cognition*, vol. 188, pp. 39–50, 2019.
- [6] M. V. Bronstein, G. Pennycook, A. Bear, D. G. Rand, and T. D. Cannon, "Belief in fake news is associated with delusionality, dogmatism, religious fundamentalism, and reduced analytic thinking," *J. of Applied Research in Memory and Cognition*, vol. 8, no. 1, pp. 108–117, 2019.
- [7] B. D. Horne and S. Adali, "This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news," in *ICWSM*, 2017, pp. 759–766.
- [8] M. L. Newman, J. W. Pennebaker, D. S. Berry, and J. M. Richards, "Lying words: Predicting deception from linguistic styles," *Personality and social psychology bulletin*, vol. 29, no. 5, pp. 665–675, 2003.
- [9] C. A. Harper and T. Baguley, "'you are fake news': Ideological (a) symmetries in perceptions of media legitimacy," 2019.
- [10] G. L. Cohen, "Party over policy: The dominating impact of group influence on political beliefs," *J. of personality and social psychology*, vol. 85, no. 5, p. 808, 2003.
- [11] A. N. Washburn and L. J. Skitka, "Science denial across the political divide: Liberals and conservatives are similarly motivated to deny attitude-inconsistent science," *Social Psychological and Personality Science*, vol. 9, no. 8, pp. 972–980, 2018.
- [12] A. Guess, J. Nagler, and J. Tucker, "Less than you think: Prevalence and predictors of fake news dissemination on facebook," *Science advances*, vol. 5, no. 1, p. eaau4586, 2019.
- [13] J. Gottfried and E. Shearer, *News Use Across Social Media Platforms 2016*. Pew Research Center, 2016.
- [14] M. J. Metzger, A. J. Flanagin, and R. B. Medders, "Social and heuristic approaches to credibility evaluation online," *J. of communication*, vol. 60, no. 3, pp. 413–439, 2010.
- [15] V. G. Cerf, "Information and misinformation on the internet," *Communications of the ACM*, vol. 60, no. 1, pp. 9–9, 2016.
- [16] L. Cui, S. Wang, and D. Lee, "Same: sentiment-aware multi-modal embedding for detecting fake news," in *ASONAM*, 2019, pp. 41–48.
- [17] K. Shu, X. Zhou, S. Wang, R. Zafarani, and H. Liu, "The role of user profiles for fake news detection," in *ASONAM*, 2019, pp. 436–439.
- [18] A. Chakraborty, B. Paranjape, S. Kakarla, and N. Ganguly, "Stop clickbait: Detecting and preventing clickbaits in online news media," in *ASONAM*. IEEE, 2016, pp. 9–16.
- [19] J. P. Dickerson, V. Kagan, and V. Subrahmanian, "Using sentiment to detect bots on twitter: Are humans more opinionated than bots?" in *ASONAM*. IEEE, 2014, pp. 620–627.
- [20] S. Helmstetter and H. Paulheim, "Weakly supervised learning for fake news detection on twitter," in *ASONAM*. IEEE, 2018, pp. 274–277.
- [21] R. Baly, G. Karadzhov, D. Alexandrov, J. Glass, and P. Nakov, "Predicting factuality of reporting and bias of news media sources," *arXiv preprint arXiv:1810.01765*, 2018.
- [22] Snopes, "Snopes," <http://www.snopes.com>, 2019, [Online; accessed 10-January-2019].
- [23] Poltifact, "Poltifact," <http://www.poltifact.com>, 2019, [Online; accessed 10-January-2019].
- [24] K. Krishna and M. N. Murty, "Genetic k-means algorithm," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 29, no. 3, pp. 433–439, 1999.
- [25] N. Klein and E. O'Brien, "People use less information than they think to make up their minds," *PNAS*, vol. 115, no. 52, pp. 13 222–13 227, 2018.
- [26] M. Gabelkov, A. Ramachandran, A. Chaintreau, and A. Legout, "Social clicks: What and who gets read on twitter?" in *SIGMETRICS*, 2016, pp. 179–192.
- [27] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *J. of economic perspectives*, vol. 31, no. 2, pp. 211–36, 2017.
- [28] A. Coppock, "Generalizing from survey experiments conducted on mechanical turk: A replication approach," *Political Science Research and Methods*, vol. 7, no. 3, pp. 613–628, 2019.