# Believability of News

## Understanding Users Perceptions of Fake News and the Effectiveness of Fact-Checking Badges

André Calero Valdez[1[0000-0002-6214-1461]] and Martina Ziefle[1[0000-0002-6105-4729]]

[1] Human-Computer Interaction Center, RWTH Aachen University, Campus Boulevard 57, 52074 Aachen, Germany
calero-valdez@comm.rwth-aachen.de

**Abstract.** When social media becomes a dominant channel for the distribution of news, manipulation of the news agenda and news content can be achieved by anyone who is hosting a website with access to social media APIs. Falsehoods masked as legitimate news with the intent to manipulate the public are called Fake News. This type of propaganda is disseminated by sharing of individual social media users. Fake news pose a threat to democracies as they influence the public agenda and contribute to polarization of opinions. To limit the dissemination of fake news, social media websites utilize fact-checking badges to flag possibly fabricated content. It has however not been investigated whether these badges are effective and who responds to them. In a survey study with 120 participants we found little evidence for the effectiveness of such badges. However, believability of news in a social media sites were generally rated rather low.

**Keywords:** Fake news, user study, fact checking, personality, social media, opinion formation.

## 1 Introduction

In June 2016, the population of Great Britain decided on whether to remain a part of the European Union or not. About 51.9% no longer felt to be a part of EU and voted with leave. Shortly before, all opinion polls suggested the opposite outcome. A similar mistake was repeated on November 8th, 2016 during the presidential election of the United States of America. Donald Trump was elected president unexpected to polls and media alike. Opinion polls were reported on heavily in the last few hours before the vote ended, possibly causing more voters to go to the booth. One explanation is looked for in the different uses of social media by Trump and Clinton during the campaign. Trump's campaign relied heavily on social media and more postings associated with Trump's hashtags were found online than for Clinton [1].

With the increasing importance of social media, two phenomena have become increasingly important as well. The use of social bots to promote posts beyond humanly possible activity levels is one way of shifting the perception of majority and minority in social media. Another way of affecting the public opinion is the utilization of Fake

news [2]. Fake news are news posts that seem like authentic news posts, but have been fabricated to affect the public opinion. Such posts often address controversial topics and intentionally report falsehoods to manipulate opinions. The bigger problem with these posts is that not only social bots share them, but also users. Users seem to share fake news to a large extent—up to 20% more than regular news posts [3]. But, who are those users that repost fake news? Do users verify content before sharing? Do services such as "fact-checker" from snopes or the associated press help users in determining what is true and what is fabricated?

In this paper, we investigate how fake and authentic news regarding refugees in Germany affect opinion forming by measuring the impact of different fabricated news articles on different topics in comparison to authentic articles. In an experimental study with 142 participants we measure the believability of fake news and how the knowledge of the news being fake influences the believability afterwards.

## 2 Related Work

In the time of digitization, media, and mass communication have such short-circuit feedback loops, where people respond to content of news quickly, thus affecting the content being reported on. When Social Media is used, these feedback loops become ever shorter and feedback becomes almost immediate. The public's reaction to news itself becomes important on its own. In the case of fake news, this can cause problems. We define fake news in accordance with Allcot et al. [1], namely as deliberate falsehoods being reported seemingly objectively to manipulate the publics opinion with the purpose of disinformation by misinterpreting, manipulating or fabricating content. But who reads these "obvious" or not so obvious fabrication? And who interacts with it? Does it affect opinion?

How opinions are formed has been studied since the 1960s extensively in subjects such as political opinion formation and the formation of opinions on novel products. The theory of opinion leadership by Rogers & Cartano [4] postulates that a certain subgroup of the population characterized by high levels of domain knowledge and social status determine opinion formation in the population. By addressing opinion leaders first, a shift in the public opinion can be achieved most easily. Resulting from this research Childers [5] investigated the psychometric properties of opinion leaders to allow the identification of opinion leaders more easily.

The theories of opinion formation have become increasingly important with the rise of mass media and the public news broadcast and their perception in the population [6]. The media has the potential to shift opinion not only by setting the agenda, but also by creating a spiral of silence by not reporting minority opinions. In the digital age, digital media has similar influences. Yet, content creators are no longer people with broadcasting licenses, but everyone.

Dimitrova [7] found that an increase in usage of digital media is not necessarily linked to an increase in political knowledge, while the contrary was true for classical media. Social media comes with the challenges and opportunities that a decentralized, global system brings along [8]. Small minorities find likeminded others even when they

are scattered across the globe. While from the perspective of, e.g., a self-help group this is beneficial, it is problematic when enemies of democracy organize themselves on social media. One of the dangers stems from the so-called filter bubble proposed by Pariser [9]. The filter bubble is the phenomenon that algorithms affect the information that is being presented to users in social media in such a way that new content matches the preferences of the users. Right-wing radicals are exposed to right-wing fake news by liking right-wing content. Bozdag & Hoven found [10] that the filter bubble poses several threats to democracy and that users should aim at breaking their filter bubble.

How can these challenges be addressed? Kosinski et al. [11] found that algorithms are able to determine private traits such as personality from user behavior online and might consequently adapt their own behavior to provide content that matches the receptive patterns of users. If social media providers understood how these patterns can trigger, e.g., fact checking behavior, social media providers could steer users to refrain from reposting fake news. The open questions are: What determines the believability of news posts and who are the users that repost such content. In this paper we investigate the effect of fact-checking badges on believability and how their effectiveness is affected by user factors.

## 3 Method

In order to understand who is affected by fact-checking badges such as the ones provided by snopes or the associated press, we conducted an online survey that was distributed using surveymonkey.com. We recruited 142 participants using convenience sampling and asked them to rate the believability of 13 news items embedded in a social media context (see Fig. 1). Participants were asked to rate the believability using a slider on a 1-100 scale. If the news items was considered fake news by snopes we would then show a banner (see Fig. 2) indicating this. We would then ask the participant to reconsider their judgment. In order to not make the participants assume, all news items are fake, half of the items were real news items.

The set of news messages were all selected from current political topics, addressing the refugee "crisis" in Germany during the time of conducting the survey in early 2017. The different news messages that were used contained the following headlines (F marks faked articles):

1. Lesbos: Refugees set refugee camp on fire (F)
2. The amount of refugees applying for asylum has been stagnating for 2 years despite the refugee crises. (F)
3. Jamaica-Coalition agrees on an upper limit of 1 Mio. refugees. (F)
4. Merkel: 100,000 refugees will leave Germany voluntarily. (F)
5. Integration commissioners demand integration tax (F)
6. 700 Euro Christmas bonus for refugees (F)
7. France wants to take in 10,000 refugees in the next 2 years under the UN programme.
8. Refugees increase economic growth (F)

9. The number of asylum proceedings in the administrative courts has increased fivefold in recent years.
10. The Aachen police donated lost bicycles to refugees.
11. Security risk: 20 disguised jihadists came to Germany as refugees
12. Refugees visit their home countries despite their asylum status.
13. Refugees thank Aacheners, give roses, say thank you.

**Fig. 1.** Example news post: The headline reads "Lesbos: Refugees set refugee camp on fire" (translated from German).

Special effort was taken to not pick fake and real news from a certain political angle. The political "tone" of the news does not help in estimating, whether the news are correct or fake. The order was randomized for all participants.

**Fig. 2.** Banner indicating the previous entry was considered to be fake.

To investigate who believes what type of news and who changes their opinion about the believability of a given news post, we assessed demographic variables such as age, gender, income and education and we asked participants how often they used different types of media (social media, online news, online yellow press, message boards, radio, political blogs, local newspaper, newspaper, TV, discussion groups). We further asked them to complete a short personality test called BFI-10, measured their political interest using 4 items, and measured two of the BIDR social desirability scales (self-deceptive enhancement and impression management) as well as political judgements regarding refugees (nationalistic judgments, consequential judgments). As a last step we asked users, whether they had fact-checked some of the news while answering the survey.

# 4 Results

Results were generated using R, the tidyverse, and the psych package. The data is analyzed using descriptive statistics with 95% confidence intervals and correlation analysis. Factor structures are tested using minimum residuals using ordinal least squares. Factor counts are decided using the Bayesian information criterion (BIC). Reliability of resulting scales was measured using Cronbach's alpha. We report t-tests and correlation tests as null-hypothesis significance tests and set the level of significance to $\alpha = .05$.

## 4.1 Description of the sample

From the 142 participants, 120 participants completed the survey with sensible results. Twenty-two were removed because their believability ratings were always 1. Of the remaining 120, 60 were male, and 60 were female. The average age was $M = 29.6$ years ($SD = 11.4$). Thirty-seven reported to have 700€ or less, 23 between 700€ and 1500€, 20 between 1500€ and 2500€, 19 between 2500€ and 3500€, 12 between 3500€ and 5000€ and 9 over 5000€ of monthly gross salary.

Men showed higher conscientiousness ($t(116.02) = -3.8, p < .01, M_{male} = 3.3; M_{female} = 2.65$), higher neuroticism (t$(116.86) = -4.05, p < .01, M_{male} = 4; M_{female} = 3.24$), and a lower political interest ($t(112.98) = 4.79, p < .001, M_{male} = 2.33; M_{female} = 3.3$) than women in our sample. Men further used social media to a larger extent than women ($t(117.76) = 2.3481, p < .05, M_{male} = 3.63; M_{female} = 2.86$).

We found that age negatively correlated with conscientiousness ($r = -.22, p <.05$), but also negatively with nationalistic judgments ($r = -.26, p < .01$), and positively with consequentialist judgements ($r = .27, p < .01$). Older users used classical media more frequently ($r = -.27, p < .01$) and had larger incomes ($r = .31, p<.001$).

Classical media consumption is associated with lower nationalistic judgments ($r = -.2, p < .01$), with lower levels of education ($r = -.29, p < .01$), with higher levels of income ($r = .2, p < .05$), and social media consumption ($r = .19, p < .05$). Social media consumption on the other hand is highly negatively correlated with political interest ($r = -.46, p < .001$). Political interest is negatively correlated with nationalistic judgments ($r = -.33, p <.001$).

From the social desirability scales impression management correlates negatively with conscientiousness ($r = -.44, p < .001$) and agreeableness ($r = -.2, p < .05$). Self-deceptive enhancement correlates with nationalistic judgments ($r = .23, p < .05$), consequentialist judgments ($r = .24, p < .05$), extraversion ($r = .35, p < .001$), conscientiousness ($r = .32, p <.001$), and negatively with neuroticism ($r = -.29, p < .01$).

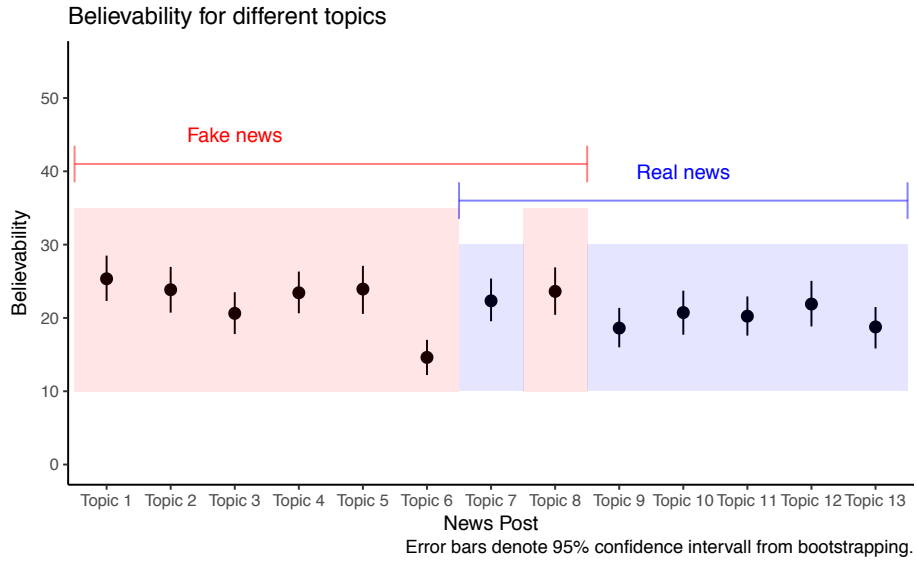


**Fig. 3.** Believability of our thirteen news posts and their confidence intervals. Topic 1 has the highest believability and topic 6 the lowest. The y-axis scale extends up to 100.

### 4.2 Believability of News Posts

We first look at how the believability of our 13 news posts was seen by our participants. For this purpose we plot the 95% confidence intervals on the half scale range of the slider in the survey (see Fig. 3). Interestingly, most believability ratings are rather low—lower than a third of the scale. The believability of fake news was, generally speaking, no higher or lower than authentic news posts. The most dramatic difference in believability can be seen between post 1 ("Lesbos: Refugees set refugee camp on fire") and post 6 ("700 Euro Christmas bonus for refugees").

When looking at how the fact-checking badge affects the believability we look at fake news only. Here we see that only a few news posts are affected by the badge at all (see Fig. 4). In particular post 1 sees a shift of believability that is downwards, as does post 8 ("Refugees increase economic growth").

Lastly, only eight participants reported to have checked the facts during filling out the survey. Thus, we can assume that a large proportion of the outcome is due to the spontaneous reaction of the participants filling out the questionnaire.

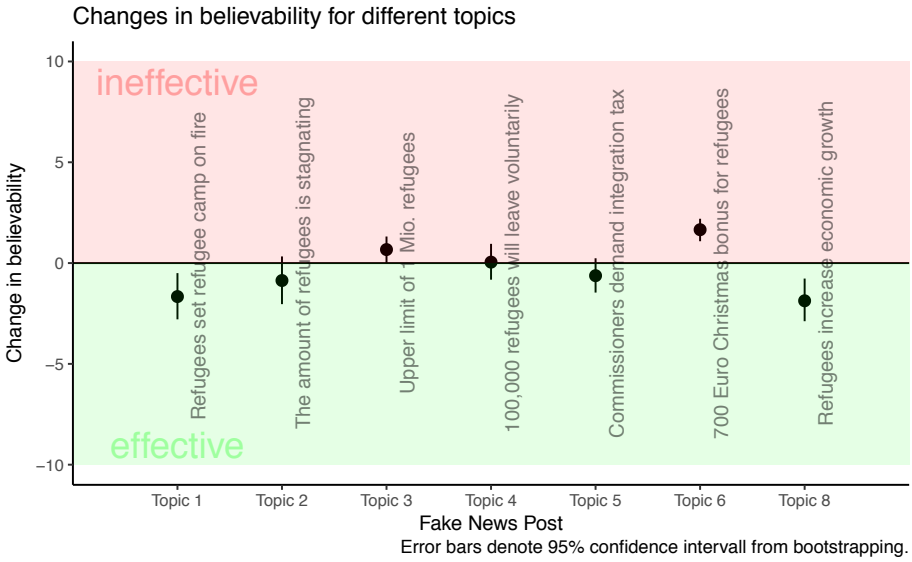


**Fig. 4.** Changes in believability depending on topic after a fact-checking badge was shown. Only two topics are positively affected by the badge (topic 1 and 8).

### 4.3 Effects of User Diversity on Believability

We found only little evidence of impact of user diversity factors on the believability ratings of our fake news topics. Of all our independent variables only nationalistic judgments, political interest, education, and self-deceptive enhancement affect the rating of the believability of our topics. Nationalistic judgments increases the believability of topics 1 and 5 ("refugees set camp on fire": $r = .23, p < .05,$ and "integration tax": $r = .20, p < .05$). Political interest increases the believability of topic 2 ("stagnating refugee counts": $r = .25, p < .01$), and education of topic 1 ($r = .2, p < .05$). Self-deceptive enhancement increases the believability of topic 5 ($r = .25, p < .01$). Overall, there is only little evidence for a systematic influence of believability of fake news.

When looking at the *changes of believability* individually, we found that the strongest correlations was between self-deceptive enhancement and the changes regarding topic 5 ($r = -.27, p < .01$), indicating that people who tend deceive themselves about their own behavior change their believability of a news post in accordance with the badge (i.e., they decrease the believability)—yet, only if the topic is the introduction of an integration tax.

Similarly, we find that education affects believability of topic 1 ($r = -.18, p < .05$, higher education, stronger decrease in believability), classic media usage affects topic 4 ($r = -.21, p < .05$), conscientiousness affects topic 2 and 3 ($r = -.18, p < .05$ and $r = -.19, p < .05$), agreeableness affects topic 3 ($r = -.21, p < .05$), and nationalistic judgments affect topic 6 ($r = -.19, p < .05$). These findings, however, provide only little

evidence for the influence of user-diversity on shifts in opinions with a fact-checking badge, when looking at the set of all possible interactions.

## 5 Discussion

We have seen that the participants in our study rated most of our news posts—both fake and authentic—relatively low in terms of believability. The changes in believability after adding a fact-checking badge were also very small and in some cases even increased believability on our scales. What does that mean?

First of all, we should look at the way our measurements were conducted. We used a 100-point slider, which defaulted to 1. This could make it very easy for users, who distrusted the news to leave the slider where it was. Believability in our experiment was a conscious decision to change the slider. In other words, we nudged participants to rate our news posts as not believable. In reality the nudging is different. Users can share a post using two clicks (share, ok) to spread fake news. Moving a slider all the way from 1 to 100 across the screen requires a stronger reflection of the content. The interface and representation of the study could bias the outcome of the study [13].

Nevertheless, users did move the sliders, some even up to 65 depending on the topic. So even when our mean is biased towards zero the correlations between scales, which are agnostic to the numerical values, do indicate some but only little evidence for an influence of user-diversity regarding both initial believability and post-badge-exposure believability. There could be several explanations. Cognitive dissonance could nudge users to not adjust their believability rating after seeing the badge. When committing to certain value, showing a small badge will not affect the "naturally certain judgment" of a user. Users might not find it easy to admit that their initial assessment was poor. In consequence, later judgments would all be rather low.

Another reason for such weak effects can be derived from the sample description. The sample we achieved was of rather high education, young age, and low income. The proportion of students in the sample was probably rather high. This might dramatically shift the preconceptions about the topic refugees and thus could have caused the influence of a confirmation bias. Also most participants did have very low nationalistic judgments, so even if the variance in this variable did influence some of the believability ratings, the variance was in the low end of the scale. To improve the data here, extending the sample to a larger and more diverse group of people would be necessary.

Methodologically speaking, we must consider that the procedure (i.e., within-subject, pre- and post-badge judgment) could not be ideal for such an experiment. The error rates are lower in within-subject designs, however, effects of order and across trials are also expected. An improved study would rather rely on fewer judgments, with a larger sample size. Still, a shift in believe change can then not be traced back to individual differences or user diversity.

The results do have an interesting implication. If setting a slider across the screen, causes users to reflect on the content they are about to "believe" or "share", such input metaphors could be used to cause reflection in a social media website. Similar to the "slide to unlock" metaphor a "slide to share" metaphor could very implicitly reduce

sharing of fake news. To investigate, whether this effect is present, one could compare a slider-based version of this research with a single-click-based version and compare the tendency to rate believability between these versions.

The larger question, how does a decision-support system ("is this fake news?") help users in making better judgments and how responds to such systems still requires further research efforts that link the findings to user-factors and factors of interface design [14].

### 5.1 Conclusion

In this article we have shown that fact-checking badges have little effect on the believability of news posts in social media sites. Generally, news that are posted on social media are distrusted by a young, educated, and generally left-oriented group of users. The little effects that we found were related to agreeableness and conscientiousness. Putting a fact-checking badge on social media could be seen as preaching to the choir for those who already rate fake news as non-believable, and be seen as paternalism by the left-wing media by those who don't. Testing such badges in real-life conditions should shed light onto this question.

## Acknowledgements

We would like to thank the anonymous participants for sharing this intimate data with us and Isabell Busch, Claudia Gerke, Vanessa Götzl, Nicole Kuska and Vivian Lotz for designing the news items and helping with sampling. This research was supported by the Digital Society research program funded by the Ministry of Culture and Science of the German State of North Rhine-Westphalia. The authors thank the German Research Council DFG for the friendly support of the research in the excellence cluster "Integrative Production Technology in High Wage Countries".

### References

1. Allcott, Hunt and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives*, 31 (2): 211-36.
2. Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... & Schudson, M. (2018). The science of fake news. *Science*, 359(6380), 1094-1096.
3. Berghel, H. (2017). Lies, damn lies, and fake news. Computer, 50(2), 80-85.
4. E. M. Rogers and D. G. Cartano. "Methods of measuring opinion leadership". In: *Public Opinion Q.* (1962), pp. 435–441.
5. T. L. Childers. "Assessment of the psychometric properties of an opinion leadership scale". In: *Journal of marketing research* (1986), pp. 184–188.
6. J. Zaller. "Political awareness, elite opinion leadership, and the mass survey response". In: *Soc. Cogn.* 8.1 (1990), pp. 125–153.
7. D. V. Dimitrova et al. "The effects of digital media on political knowledge and participation in election campaigns: Evidence from panel data". In: *Communication Research* 41.1 (2014), pp. 95–118.

8. A. M. Kaplan und M. Haenlein. “Users of the world, unite! The challenges and opportunities of Social Media”. In: *Business horizons* 53.1 (2010), pp. 59–68.
9. E. Pariser. The filter bubble: What the Internet is hiding from you. Penguin UK, 2011.
   E. Bozdag and J. van den Hoven. “Breaking the filter bubble: democracy and design”. In: *Ethics and Information Technology* 17.4 (2015), pp. 249–265.
10. M. Kosinski, D. Stillwell, and T. Graepel. “Private traits and attributes are predictable from digital records of human behavior”. In: *PNAS* 110.15 (2013), pp. 5802–5805.
11. Hart, C. M., Ritchie, T. D., Hepper, E. G., & Gebauer, J. E. (2015). The balanced inventory of desirable responding short form (BIDR-16). *Sage Open*, 5(4), 2158244015621113.
12. Calero Valdez, A., Ziefle, M., & Sedlmair, M. (2018). Priming and anchoring effects in visualization. *IEEE transactions on visualization and computer graphics*, 24(1), pp. 584-594.
13. Calero Valdez, A., Brauner, P., Ziefle, M., Kuhlen, T. W., & Sedlmair, M. (2016). Human factors in information visualization and decision support systems. Mensch und Computer 2016–Workshopband.