If You Have a Reliable Source, Say Something: Effects of Correction Comments on COVID-19 Misinformation

Haeseung Seo, Aiping Xiong, Sian Lee, Dongwon Lee
The Pennsylvania State University, USA
{hxs378, axx29, szl43, dongwon}@psu.edu

Abstract

In the post-truth era, particularly during the COVID-19 pandemic, an effective correction on misinformation is necessary to promote personal and public health. To better understand the effect of "correcting" misinformation, therefore, we investigated correction from different users on social media (e.g., individual users, fact-checking websites, and health organizations) and the frequency of correction (e.g., once vs. twice) in three online experiments. In each experiment, we evaluated participants' perceived accuracy and willingness to share in terms of real and fake news of COVID-19, respectively. Across all experiments, a single correction from the health organizations effectively reduced participants' perceived accuracy rating on the COVID-19 fake news. Experiments 2 and 3 revealed the effects of a single correction from individual users and fact-checking websites. Moreover, results of post-session questionnaires indicated that participants counted on the reliability of the sources in the correction. We did not obtain the consistent effects of frequent correction but verified the vulnerability of participants with high health anxiety to the COVID-19 fake news across all experiments. Overall, our study highlights the effects of user-initiated correction regardless of whether the user is an individual or an organization, as long as the correction contains a reliable source.

Introduction

The explosion of fake news is one of the problems that social media has triggered (Ha, Andreu Perez, and Ray 2021). On social media platforms, people can easily get and share news even before checking its veracity. Thus, platforms such as Twitter and Facebook have attempted to contrive ways to curtail misinformation, such as computationally detecting fake news (Facebook 2020) or correcting misinformation (Roth and Pickles 2021). The focus of our work is on the latter.

Despite considerable research on correcting misinformation (Vraga and Bode 2017; Bode and Vraga 2018; Pennycook, Cannon, and Rand 2018; Seo, Xiong, and Lee 2019; Seo et al. 2021), it is still premature to conclude the most effective way to correct misinformation. In addition to the methods to correct misinformation by platforms (e.g., the popping-up warning message for questionable contents), research on *user-initiated* correction has started recently (Vraga and Bode 2017, 2018). It is critical to investigate effective user-initiated methods to correct misinformation as users are the main characters sharing information on social media. Moreover, prior studies demonstrated that users have actively participated in the corrections on social media (Bode and Vraga 2021).

During the COVID-19 pandemic, a great deal of misinformation about the virus and treatments has been pouring out on social media, threatening personal and public health worldwide. In reality, there were many cases where people died from wrong treatments for COVID-19 due to fake news.[1] Since people with high health anxiety are more associated with seeking online health information (Starcevic and Berle 2013; McMullan et al. 2019), they can be more vulnerable to COVID-19 misinformation and more resistant to the correction compared to people with low health anxiety. Therefore, it is imperative to examine the effective correction on COVID-19 misinformation, and understand how people's health anxiety level impacts the correction effects.

Focusing on the correction on COVID-19 misinformation, we investigated the following research questions (RQs) in current work.

- **RQ 1.** Will the correction from an individual user or an organization user (e.g., a health organization or a fact-checking website) reduce participants' susceptibility to fake news relative to a control condition in which there is no correction?
- **RQ 2.** Will more frequent correction reduce participants' susceptibility to fake news more?
- **RQ 3.** Will people's health anxiety level have an impact on their susceptibility to misinformation and the effect of misinformation correction?

We conducted three online experiments ($N = 2,841$) on Amazon Mechanical Turk, examining correction from three types of users (e.g., health organizations, fact-checking websites, and individual users) (**RQ1**). We verified that correction from all three types of users could reduce participants' perceived accuracy rating on the COVID-19 fake news. Critically, we unearthed that participants counted on the reliability of the sources (e.g., social media users or URLs in the correction). We did not obtain the frequency effect (**RQ2**). However, we discovered that participants having high health anxiety were more likely to believe fake news than low anxiety participants. We also obtained evidence showing the correction effect only for participants with low health anxiety. Those results imply that people with high health anxiety are more susceptible to the COVID-19 misinformation and more resistant to the misinformation correction (**RQ3**).

[1] https://www.bbc.com/news/world-53755067

To the best of our knowledge, the current paper is the first to experimentally investigate the effective user-initiated correction on COVID-19 misinformation. Our experiments improve the understanding of user-initiated correction on misinformation through the following contributions.

- Via systematically-designed experiments, we obtained correction effect on fake news from three different types of users: individual users, health organizations, and fact-checking websites.
- We unearthed that people weigh the reliable sources in correction the most when deciding perceived accuracy rating.
- We found that individuals with high health anxiety are more susceptible to health-related misinformation than those with low health anxiety and correction can work better for individuals with low health anxiety.

Related Work

Misinformation

Despite many definitions, the term "misinformation" generally refers to falsely fabricated information regardless of whether it includes intent to mislead (Cook, Ecker, and Lewandowsky 2015; Ha, Andreu Perez, and Ray 2021). According to one of the commonly used definitions, it indicates information without clear evidence and expert opinion (Nyhan and Reifler 2010; Vraga and Bode 2017). Meanwhile, "fake news," as a type of misinformation (Lazer et al. 2018), has begun to attract the public's attention with the growth of social media and mobile media since 2008, and have become popularized around the world since the 2016 U.S. presidential election (Quandt et al. 2019). In this work, we call interchangeably false information on social media either misinformation or fake news.

Misinformation and Correction

With the dissemination of misinformation online, researchers are increasingly interested in correcting misinformation (Seifert 2002; Nyhan and Reifler 2010; Gordon and Shapiro 2012; Cook, Ecker, and Lewandowsky 2015; Thorson 2016; Ha, Andreu Perez, and Ray 2021). Nevertheless, empirical studies revealed mixed results on the effect of correction (Walter and Murphy 2018). While some studies showed that correction can significantly reduce participants' misinformation belief (Vraga and Bode 2017; Bode and Vraga 2018), the other studies showed negative effects such as backfire (Nyhan and Reifler 2010; Mosleh et al. 2021), or combined effects simultaneously (Nyhan and Reifler 2015; Jiang and Wilson 2018).

Lewandowsky and his colleagues analyzed the reasons for failure of misinformation correction comprehensively from a psychological perspective and suggested solutions, such as alternative account, emphasis on facts, and simple rebuttal (Lewandowsky et al. 2012). Walter and Murphy (2018) conducted a meta-analysis of empirical studies and categorized promising ways of correction, including source credibility, fact-checking, and providing general warnings.

Due to the proliferation of misinformation on social media (Allcott and Gentzkow 2017), recent studies on correction effects have mainly centered on platforms such as Twitter and Facebook. Depending on who initiated the correction, we can divide those studies into two approaches: *platform-driven* correction and *user-initiated* correction.

Platform-driven Correction. Studies of platform-driven correction evaluated corrections, including fact-checking warnings (Pennycook, Cannon, and Rand 2018; Seo, Xiong, and Lee 2019) or related articles (Smith and Seitz 2019). Due to the inclusion of news sources, the effect of platform-driven corrections was somewhat inconclusive (Smith and Seitz 2019). After controlling the news source, Seo, Xiong, and Lee (2019) found that the effect of fact-checking warning became non-significantly different from a control condition. Moreover, the effect of platform-driven corrections may also be impacted by the public's increasing distrust of social media platforms due to events such as personal information leakage or excessive control over information.[2]

User-initiated Correction. Since users are the main actors of information sharing on social media (Boyd and Ellison 2007; Bechmann and Lomborg 2013), the other approach concentrates upon user-initiated correction using user comments or posts (Vraga and Bode 2017; Gesser-Edelsburg et al. 2018). Vraga and Bode investigated the effect of user-initiated correction on Zika virus misinformation using users' comments on social media (Vraga and Bode 2017, 2018; Bode and Vraga 2018). They manipulated one post within a simulated Twitter or Facebook feed. In the control condition, there was no misinformation. In the treatment conditions, a piece of fake news about the Zika outbreak in the U.S. was presented, with the image and the headline claiming that the outbreak was caused by the release of GMO mosquitoes. Following misinformation, a debunking sentence with a reference link from the Centers for Disease Control and Prevention (CDC) was presented. Across conditions, the source of correcting comments was varied from an organization (e.g., CDC), an individual, or both.

They recruited undergraduate students and measured participants' misperception about the Zika virus before and after the correction of misinformation. Across studies, they demonstrated that social correction could work if it includes sufficient source information, including organization logos (such as *snopes.com* and CDC) and reference links from the organizations. Their results also revealed that platforms (Facebook or Twitter) did not matter (Vraga and Bode 2018) and social and algorithmic corrections were equally effective in mitigating misperceptions (Bode and Vraga 2018). One of their studies found the correction effect from a reputable organization (e.g., CDC) but not from an individual user (Vraga and Bode 2017).

Moreover, recent work showed that people experience both misinformation and its correction on social media. Bode and Vraga (2021) conducted an online survey about people's correction experience regarding COVID-19 misin-

[2] http://tiny.cc/at1juz;http://tiny.cc/zugxtz;http://tiny.cc/1vgxtz

formation on social media. Out of 1,094 participants, 34% witnessed that others' wrong beliefs were corrected and 22% corrected others' misinformation, demonstrating users' active participation in correction. Given the impacts of user-initiative comments, it is critical to examine the effective correction in the setting of COVID-19. In our current work, we evaluated a single correction comment initiated by different social media users (e.g., an organization user or an individual user) with a more structured experimental design using COVID-19 news.

Health Anxiety

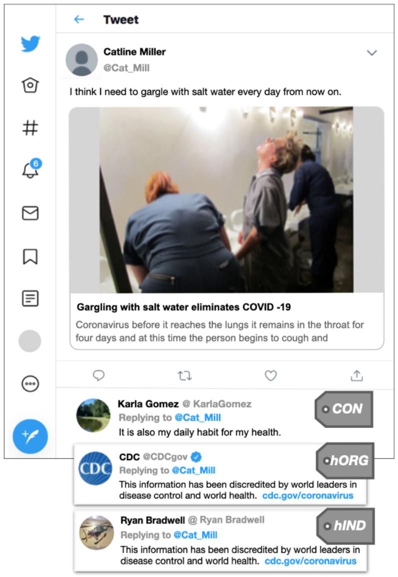
Besides an innumerable amount of COVID-19 fake news being spread online (Tasnim, Hossain, and Mazumder 2020), many cases related to health anxiety (i.e., hypochondria) and mental health issues have been reported during COVID-19 pandemic (Jungmann and Witthöft 2020). Prior studies (Asmundson and Taylor 2020; Banerjee, Rao et al. 2020) explain abnormally increased fear and anxiety for health as one reason for information seeking about the COVID-19, which is also called "Cyberchondria" (McMullan et al. 2019; Laato et al. 2020). Considering the health news topic, we also measured participants' health anxiety (Lucock and Morley 1996) and analyzed its impact on the effect of correction for COVID-19 misinformation.

The Present Study

We investigated a single user-initiated correction comment using tweets format in three online experiments. In contrast to prior works (Bode and Vraga 2015; Vraga and Bode 2017), we recruited Amazon Mechanical Turk (MTurk) workers who are more demographically diverse (Berinsky, Huber, and Lenz 2012; Briones and Benham 2017; Weigold and Weigold 2021) than college students. In each experiment, participants evaluated eight to twelve pieces of fake and real news about COVID-19 instead of evaluating a single piece of misinformation. We examined the effect of correction comments on people's acceptance of fake news about COVID-19 compared to a control in which comments without correction were presented, rather than examining the correction effect by comparing participants' pre- and post-perception of misinformation (Vraga and Bode 2017).

We examined COVID-19 misinformation, which is most timely. Moreover, we ran our experiments at three different time points over six months during the COVID-19 pandemic. We updated the evaluated news set from Experiment 2 to reflect the rapidly pouring news. We proposed and evaluated different types of users on social media (e.g., individual users, health organizations, and fact-checking websites).

Experiment 1 examined the effect of a single correction comment about COVID-19 fake news by individual users or organization users. Experiment 2 was conducted to replicate the findings of Experiment 1 using a new set of COVID-19 real and fake news. Experiment 3 increased the types of correction to be investigated for a more systematic understanding of the correction effect.



Tweet

Catline Miller
@Cat_Mill

I think I need to gargle with salt water every day from now on.

Gargling with salt water eliminates COVID -19

Coronavirus before it reaches the lungs it remains in the throat for four days and at this time the person begins to cough and

Karla Gomez @ KarlaGomez
Replying to @Cat_Mill
It is also my daily habit for my health.

CDC @CDCgov
Replying to @Cat_Mill
This information has been discredited by world leaders in disease control and world health. cdc.gov/coronavirus

Ryan Bradwell @ Ryan Bradwell
Replying to @Cat_Mill
This information has been discredited by world leaders in disease control and world health. cdc.gov/coronavirus

Figure 1: An example of COVID fake news stimuli across the three between-subject conditions composed of an image, a headline, and a snippet of the news article shown under a message tweeting the fake news. Following the tweet, a comment was presented. The comment sentences of *hORG* and *hIND* are the same, correcting the tweet by indicating the falsity of the fake news with an identical reference link from a health organization (CDC or WHO). *CON* has a comment without a correction message or a reference link.

Experiment 1

Using a between-subject design, we investigated the effect of correction comments on COVID-19 misinformation across three conditions: no correction (*CON*), correction by an individual user (*hIND*), and correction by a health organization (*hORG*). Specifically, fake news stimuli of *hIND* and *hORG* had the same correcting comment, including a reference link from a health organization. "*h*" indicates the identical link from health organizations (CDC or WHO). Moreover, to investigate the frequency effect of correction, we composed Phases 1 and 2. Half of the fake and real news of Phase 1 were presented again at Phase 2, each of which was with a similar correcting comment but from a different user and a different organizational source.

Participants

In this and the following experiments, we recruited participants by posting the Human Intelligent Task (HIT) on MTurk. We restricted the workers to those who (1) were at least 18 years old; (2) were located in the U.S.; and (3) completed more than 100 HITs with a HIT approval rate of at least 95%. Qualtrics was used to program our online studies. Our study was approved by the institutional review board (IRB) office at our institution.

Materials

We selected eight news articles about COVID-19 released between March and May 2020 from *snopes.com* or *poli-*

tifact.com, both of which are well-regarded fact-checking websites. Four pieces of the news were fake and the other four pieces were real. Also, we used another piece of real news for an attention check (Hauser and Schwarz 2016).

As shown in Figure 1, we created a simulated Twitter interface, in which each piece of news was embedded within a tweet message. For each stimulus, a tweet message from a fictional user was presented above the COVID-19 news. The tweet message was a short sentence related to the news without any correcting message. The embedded news was composed of an image, a headline, and a snippet of the news article. Following the news article, a comment from another user was presented.

For the fake news, each comment in the *hIND* and *hORG* conditions included a sentence pointing out the falsity of the fake news articles with a reference link from an authoritative organization (CDC for Phase 1, WHO for Phase2). The correction messages and the reference links were the same between *hORG* and *hIND*.

In contrast, each comment of the fake news in *CON* did not contain a correcting sentence or reference link. Instead, the comment included a commenter's plausible but non-correcting messages varied according to the contents of each piece of news. Likewise, the real news comments had non-correcting messages, which were constructed in the same way as the fake news in *CON*. We used the same set of real news and its comments across the three conditions.[3]

When we presented half of the stimuli again at Phase 2, we varied the user of the comment for both fake and real news regardless of conditions. Furthermore, we presented a different reference link for fake news of *hORG* and *hIND* at Phase 2 (e.g., the comment from CDC and a reference link, *cdc.gov/coronavirus*, in Phase 1, the comment from World Health Organization (WHO) and a reference link, *who.int/coronavirus*, in Phase 2). The expressions of comments at both phases conveyed the same content but with some wording changes. All authors reached a consensus on the contents of all messages we used for experiments.

Procedure

Figure 2 illustrates the flow chart of Experiment 1. Participants were randomly assigned to one of three conditions. After participants provided informed consent, Phase 1 started, in which the eight pieces of stimuli were presented in a randomized order. Half of them included fake news and the other half included real news. We asked two questions to examine participant's susceptibility to the "claim" embedded in the news article of each stimulus. First, participants answered, "How accurate is the claim in the above news?" on a 7-point scale with "1" meaning "Very inaccurate" and "7" meaning "Very accurate." Then they rated their willingness to share the news by answering, "Would you consider sharing this news online (for example, through Facebook or Twitter)?" using another 7-point scale with "1" meaning "Never" and "7" meaning "Always."

Following Phase 1, two pieces of the real news stimuli and two pieces of the fake news stimuli were presented once more in a randomized order at Phase 2 to investigate the effect of correction frequency. We used a semi Latin-square design for a better-balanced assignment of news shown at Phase 2. Participants answered the same two questions for each piece of stimuli as Phase 1.

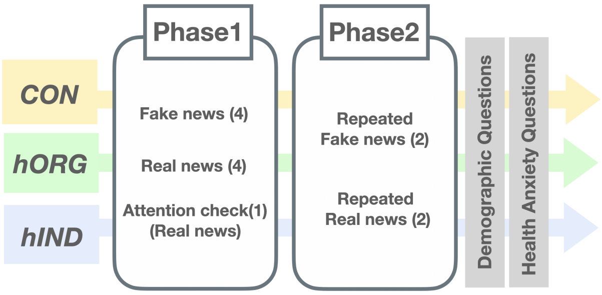


Figure 2: A flow chart of Experiment 1. *CON*, *hORG*, and *hIND* refer to the three between-subject conditions. At Phase 1, four pieces of fake news stimuli and four pieces of real news stimuli were shown in a randomized order for participants. One piece of real news stimulus was presented for an attention check. At Phase 2, half of the fake news stimuli and half of the real news stimuli from Phase 1 were shown again. We used a semi Latin-square design for a better-balanced assignment of news shown at Phase 2. All stimuli at Phase 2 were randomly presented as well. After Phase 2, questions of demographic information and health anxiety were asked as post-session questions in Experiment 1.

After Phase 2, there was a post-session questionnaire. Participants first filled in their demographic information, including age, gender, ethnicity, and education. Then we measured participants' health anxiety level using four representative questions (Lucock and Morley 1996). A 5-point scale with "1" meaning "None at all" and "5" meaning "A great deal" was used for the first two questions: 1) "How much do you usually worry about your health?" and 2) "How much are you ever worried that you may get a serious illness in the future?" Another 5-point scale with "1" meaning "Rarely" and "5" meaning "Usually" was used for the latter two questions: 3) "How often do you tend to read up about illness and disease to see if you may be suffering from one?" and 4) "How often do your bodily symptoms stop you from concentrating on what you are doing?" Participants were allowed to choose "Prefer not to answer" for all post-session questionnaires except for age.

An extra piece of real news was included at Phase 1 to exclude inattentive participants. We presented specific instructions about how to answer the attention-check question to the participants. For any participants who failed to follow the instruction, their survey was terminated immediately.

Results

We recruited $1,275$ MTurk workers in July 2020. We accepted 907 participants' answers after removing five responses submitted out of the U.S., two responses submitted within two minutes (median completion time was about six minutes), 103 responses who failed an attention check, and

[3] https://osf.io/dxs9c/

Item	Options	Exp.1	Exp.2	Exp.3
Gender	Female	49.4%	51.8%	55.7%
	Male	49.6%	47.8%	44.2%
	Prefer not to answer	1.0%	0.4%	0.2%
Age	18-27	20.6%	19.0%	16.4%
	28-37	33.8%	33.6%	39.5%
	38-47	21.1%	20.8%	23.3%
	48-57	13.3%	13.9%	11.9%
	58-67	8.2%	9.6%	6.9%
	Over 67	3.0%	3.0%	2.1%
Ethnicity	Asian	8.9%	7.3%	5.9%
	African American	12.4%	9.8%	11.8%
	Hispanic/Latino	4.5%	5.5%	6.6%
	Caucasian	70.3%	74.4%	73.5%
	Other	2.8%	2.3%	1.6%
	Prefer not to answer	1.1%	0.8%	0.5%
Education	High school	6.3%	7.9%	7.0%
	Bachelor's degree	48.5%	44.9%	47.4%
	Master's degree	20.3%	19.0%	19.2%
	Doctorate degree	2.9%	2.1%	3.3%
	Other	21.8%	25.5%	22.9%
	Prefer not to answer	0.2%	0.5%	0.3%

Table 1: Demographic information of the participants in the three experiments.

258 duplicated submissions. The numbers of participants of the three conditions included in the data analysis are as follows: 295 (*CON*), 305 (*hORG*), and 307 (*hIND*). We paid \$0.75 for participants who completed the task based on an hourly payment of \$7.5. Participants' demographic information is shown in Table 1.

For data analysis, we used three levels of news frequency : *ONCE* (Phase 1 results of news shown at Phase 1 only), TWO_{1st} (Phase 1 results of news shown at both phases), TWO_{2nd} (Phase 2 results of news shown at both phases).

Perceived accuracy rating and willingness-to-share measures were entered into 3 (condition: *CON*, *hORG*, *hIND*) × 2 (veracity: *fake*, *real*) × 2 (frequency: *ONCE*, TWO_{2nd}) mixed analysis of variances (ANOVAs) (Herzog, Francis, and Clarke 2019) with a significance level of .05, respectively. Post-hoc tests with Bonferroni correction were performed. We report the effect size using η_p^2 reported by SPSS (Ecker, Lewandowsky, and Apai 2011; Vraga and Bode 2017).[4]

To understand the frequency effect, we also analyzed the results between *ONCE* and TWO_{1st}, and between TWO_{1st} and TWO_{2nd}. Across the three experiments, the analysis results did not show any significant difference between *ONCE* and TWO_{1st}. Also, the results of the analysis between TWO_{1st} and TWO_{2nd} showed similar patterns and had only marginal differences compared to the results between *ONCE* and TWO_{2nd}. Thus, we evaluated the main analysis between *ONCE* and TWO_{2nd} in this and the following experiments.

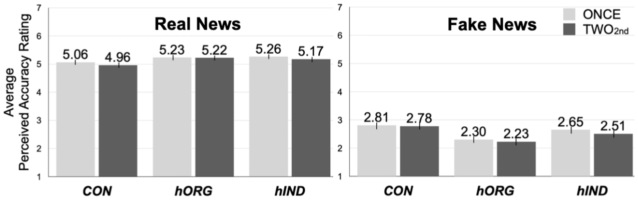


Figure 3: The average values of perceived accuracy ratings as a function of frequency × condition for real news (left panel) and fake news (right panel) with one standard error.

Perceived Accuracy Rating. Results of average the perceived accuracy rating are shown in Figure 3. Participants clearly distinguished real news (5.15) from fake news (2.55), $F_{(1,904)}{}^{5} = 1914.98$, $p < .001$, $\eta_p^2 = .679$. The two-way interaction of news veracity × condition was also significant, $F_{(2,904)} = 12.85$, $p < .001$, $\eta_p^2 = .028$. Post-hoc tests revealed that the effect of condition was significant for both fake news, $F_{(2,904)} = 6.65$, $p = .001$, $\eta_p^2 = .014$, and real news, $F_{(2,904)} = 4.29$, $p = .014$, $\eta_p^2 = .009$. Nevertheless, the effect of condition revealed different patterns. For fake news, only participants in the *hORG* condition (2.27) gave lower accuracy ratings relative to the *CON* condition (2.79), $p_{adj.} = .001$. The other two pairwise comparisons (i.e., *CON* vs. *hIND* (2.58), $p_{adj.} = .427$, and *hORG* vs. *hIND*, $p_{adj.} = .090$) were not significant. For real news, relative to *CON* (5.01), participants gave higher accuracy ratings for both *hORG* (5.22), $p_{adj.} = .029$, and *hIND* (5.22), $p_{adj.} = .039$. However, the perceived accuracy ratings between *hORG* and *hIND* conditions were not significantly different, $p_{adj.} > .999$.

Thus, we obtained the correction effect of comment from an organization user for COVID-19 fake news, which is in agreement with the prior work about Zika virus misinformation (Vraga and Bode 2017). Also, we found the positive impact of the news correction on real news.

Sharing Decisions. Results of willingness-to-share measure are presented in Figure 4. Participants' willingness to share real news (3.68) was higher than that of fake news (2.35), $F_{(1,904)} = 708.13$, $p < .001$, $\eta_p^2 = .439$. The interaction between veracity × condition only showed a trend to be significant, $F_{(2,904)} = 2.87$, $p = .057$, $\eta_p^2 = .006$. Moreover, participants' overall rating for the willingness-to-share measure (*real*: 3.68, *fake*: 2.35) was lower than that of the perceived accuracy rate (*real*: 5.15, *fake*: 2.55), indicating that they tended to be conservative in sharing decisions than perceived accuracy evaluation.

Health Anxiety. We calculated a mean score of the four questions about health anxiety after removing the results of two participants who refused to answer all of the questions (*CON*:294, *hORG*:305, *hIND*:306). We then classified the

[4] We are aware of η_G^2, which was recommended to report (Bakeman 2005; Lakens 2013) considering factors manipulated and measured between subjects (Olejnik and Algina 2003). To make the results comparable to the literature (Vraga and Bode 2017), we report the η_p^2 to show the effect size based on SPSS analysis.

[5] *F* value equals to variance estimate based on variability among group means divided by variance estimate based on variability within groups. Hence, the larger *F* value indicates that the variability in the measurements is mostly determined by the group differences (Herzog, Francis, and Clarke 2019).

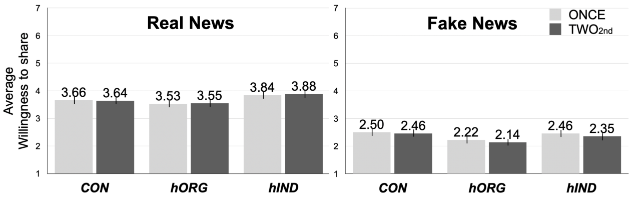


Figure 4: The average values of willingness-to-share as a function of frequency × condition for real news (left panel) and fake news (right panel) with one standard error.

results into two groups: *low health anxiety* (scores from 1 to 2) and *high health anxiety* (scores from 3 to 5). For the statistical tests, we added *health anxiety* as an additional between-subject factor into the main analysis. For perceived accuracy rating, participants with *high health anxiety* (4.30) gave a higher accuracy rating than those with *low health anxiety* (3.58), $F_{(1,899)} = 93.24, p < .001, \eta_p^2 = .094$, and its interaction with veracity, $F_{(1,899)} = 71.67, p < .001, \eta_p^2 = .074$, were significant. The effect of health anxiety was more evident for the fake news (*high health anxiety*: 3.28; *low health anxiety*: 2.09), $F_{(1,903)} = 104.84$, $p < .001$, $\eta_p^2 = .104$ than for the real news (*high health anxiety*: 5.29; *low health anxiety*: 5.06), $F_{(1,903)} = 10.99, p < .001, \eta_p^2 = .012$.

Likewise, participants with *high health anxiety* (3.78) showed more willingness to share news than those with *low health anxiety* (2.57), $F_{(1,899)} = 125.79$, $p < .001$, $\eta_p^2 = .123$. The two-way interaction of veracity × health anxiety was also significant, $F_{(1,899)} = 9.76$, $p = .002$, $\eta_p^2 = .011$. The effect of health anxiety was also more evident for the fake news (*high health anxiety*: 3.19; *low health anxiety*: 1.84), $F_{(1,903)} = 124.00$, $p < .001$, $\eta_p^2 = .121$, than for the real news (*high health anxiety*: 4.33; *low health anxiety*: 3.29), $F_{(1,903)} = 78.45$, $p < .001$, $\eta_p^2 = .080$. We also obtained a four-way interaction of veracity × frequency × conditions × health, $F_{(2,899)} = 3.13$, $p = .044$, $\eta_p^2 = .007$. The post-hoc test presented that it was mainly due to a non-significant trend of correction effect on fake news for the *low health anxiety* group, $F_{(2,557)} = 2.81$, $p = .061$, $\eta_p^2 = .010$, suggesting the misinformation susceptibility of people with high health anxiety was somewhat difficult to mitigate.

Summary

In Experiment 1, we examined the effects of correcting comments from health organizations (*hORG*) and individual users (*hIND*) (**RQ1**), as well as the effect of correction frequency (**RQ2**) in helping users mitigate fake news. Consequently, we verified the effect of correction from a health organization (*hORG*) for reducing perceived accuracy rating on fake news as in the prior study (Vraga and Bode 2017) but did not find a frequency effect. Furthermore, we discovered perceived accuracy rating of real news was higher given correction compared to *CON*, which indicates correction increased participants' confidence in real news through learning effects from the correction on fake news.

In addition, we found that participants with high health anxiety were more susceptible to COVID-19 misinformation than those with low health anxiety (**RQ3**): People with high health anxiety gave higher perceived accuracy ratings and were more willing to share health-related news than those with low heath anxiety; and such pattern was more evident for the fake news than for the real news.

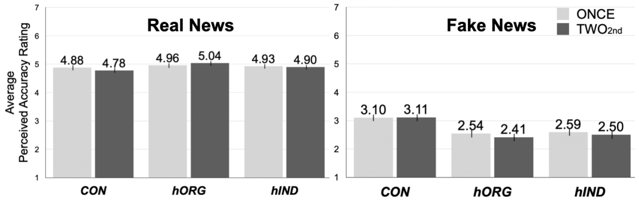


Figure 5: The average values of perceived accuracy ratings in Experiment 2 as a function of frequency × condition for real news (left panel) and fake news (right panel) with one standard error.

Experiment 2

To replicate the findings of Experiment 1, we conducted Experiment 2 with up-to-date COVID-19 news articles released from May to July 2020. The experimental setting was the same as Experiment 1 except as noted. We created twelve stimuli, half about fake news and the other half about real news, on the simulated Twitter interface of Experiment 1. The attention check was between Phases 1 and 2. We added two political-stance related questions in the post-session questions due to the impact of political ideology on people's susceptibility to COVID-19 misinformation (Calvillo et al. 2020). Moreover, we included follow-up questions to identify which factors among the given stimuli influenced participants' perceived accuracy rating.

Results

We recruited 1,255 MTurk workers in December 2020. We accepted 768 participants' answers after removing two responses submitted out of the U.S., 291 who failed an attention check, 186 duplicated submissions, and eight responses submitted less than three minutes (median completion time is about ten minutes). The numbers of participants included for data analysis are as follows: 253 (*CON*), 261 (*hORG*), and 254 (*hIND*). The base payment was $0.50. There was a bonus of $0.75 for participants who passed the attention check and completed the task. The payment rate ($7.5/hr) is the same as Experiment 1. Participants' demographic information is shown in Table 1. We analyzed the data in the same way as Experiment 1.

Perceived Accuracy Rating. Average results of the real and fake news for each condition are shown in Figure 5. The main effects of news veracity, $F_{(1,765)} = 1411.78, p < .001$, $\eta_p = .649$, condition, $F_{(2,765)} = 4.15, p = .016, \eta_p^2 = .011$, as well as the two-way interaction of news veracity × condition, $F_{(2,765)} = 17.53$, $p < .001$, $\eta_p^2 = .044$, were significant. Same as in Experiment 1, participants can distinguish real news (4.91) from fake news (2.71). Post-hoc analysis

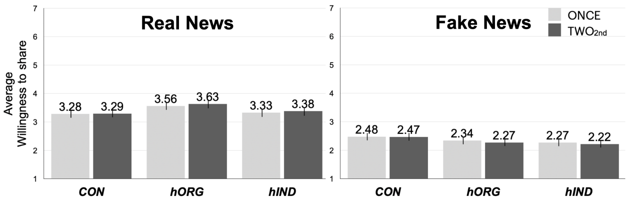


Figure 6: The average values of willingness-to-share as a function of frequency × condition for real news (left panel) and fake news (right panel) with one standard error.

revealed that the perceived accuracy ratings across the conditions were similar for real news but different for fake news. Compared to *CON* (3.11), lower accuracy rating was evident for fake news at *hORG* (2.48), $p_{adj.} < .001$, and *hIND* (2.55), $p_{adj.} < .001$, respectively. Thus, the correction effect was evident for both *hORG* and *hIND* conditions. In addition, there was a three-way interaction of veracity × frequency × conditions, $F_{(2,765)} = 3.45, p = .032$, $\eta_p^2 = .009$, showing decreased perceived accuracy rating for fake news from *ONCE* to TWO_{2nd} in *hORG* (2.54 → 2.41) and *hIND* (2.59 → 2.50) but not in *CON* (3.10 → 3.11).

Sharing Decisions. Results of willingness-to-share measure are presented in Figure 6. Participants showed more willingness to share real news (3.41) than fake news (2.34), $F_{(1,765)} = 480.27$, $p < .001$, $\eta_p^2 = .386$. The interaction of veracity × condition was significant, $F_{(2,765)} = 8.25$, $p < .001$, $\eta_p^2 = .021$. The main effect of condition was not significant at each veracity level. However, participants' willingness-to-share for the *hORG* and *hIND* conditions showed a trend to be larger than that of *CON* for the real news, while an opposite pattern was revealed for fake news.

Moreover, the two-way interaction of veracity × frequency approached significance, $F_{(1,765)}$ = 3.76, $p = .053$, $\eta_p^2 = .005$, suggesting an increase of willingness to share for real news but a decreasing trend for fake news from *ONCE* to TWO_{2nd}. As in Experiment 1, the average of willingness-to-share measure was lower than that of perceived accuracy rating, indicating that participants tended to be conservative in sharing news regardless of news accuracy.

Health Anxiety. After removing the results of three participants who did not complete the questions, we analyzed 765 (*CON*:251, *hORG*:261, *hIND*:253) participants' results by adding *health anxiety* in the main analyses. For perceived accuracy rating, participants with *high health anxiety* (4.09) gave a higher accuracy rating than those with *low health anxiety* (3.63), $F_{(1,759)} = 37.62$, $p < .001$, $\eta_p^2 = .047$. And its interaction with veracity, $F_{(1,759)} = 7.41$, p = .007, $\eta_p^2 = .010$ was also significant. As in Experiment 1, the effect of health anxiety was more evident for the fake news (*high health anxiety*: 3.09; *low health anxiety*: 2.46), $F_{(1,763)} = 28.07$, $p < .001$, $\eta_p^2 = .035$, than for the real news (*high health anxiety*: 5.10; *low health anxiety*: 4.80), $F_{(1,763)} = 18.73, p < .001, \eta_p^2 = .024$. Moreover, the three-way interaction of veracity× condition × health anxiety was significant, $F_{(1,759)} = 3.38, p = .034$, $\eta_p^2 = .009$. Post-hoc comparison showed that the correction on fake news turned out to be effective for participants with *low health anxiety* in the *hORG* (2.09), $p_{adj.} < .001$, and the *hIND* (2.27), $p_{adj.} < .001$, than those in the *CON* (3.03), respectively.

Participants with *high health anxiety* (3.44) gave higher willingness-to-share score than those with *low health anxiety* (2.51), $F_{(1,759)} = 67.99$, $p < .001$, $\eta_p^2 = .082$. Thus, people who are highly anxious about their health tended to share more health-related news regardless of news veracity.

Political Stance. At the post-session questions, we also measured participants' political stance with a 5-point scale ("1" meaning "very liberal," "5" meaning "very conservative"). Participants who gave a rating of "1" or "2" were categorized as liberals (336), and those who gave a rating of "4" or "5" were categorized as conservatives (228). We excluded moderates (204), i.e., who gave a rating of "3" from the data analysis. We added political stance (*liberals, conservatives*) as another factor into ANOVAs of perceived accuracy rating and willingness-to-share measure, respectively.

For both perceived accuracy rating and willingness-to-share measure, the main effect of political stance, $Fs_{(1,558)} = 52.89$ and 30.50, $ps < .001$, $\eta_{ps}^2 = .087$ and .052, and its interaction with veracity, $Fs_{(1,558)} = 135.63$ and 65.09, $ps < .001$, $\eta_{ps}^2 = .196$ and .104, were significant. Specifically, for perceived accuracy rating, the effect of political stance was only significant for the fake news (*liberals*: 2.18, *conservatives*: 3.58), $F_{(1,562)} = 108.85, p < .001$, $\eta_p^2 = .162$, but not for the real news (*liberals*: 5.0, *conservatives*: 4.91), $F_{(1,562)} = 1.306$, $p = .254$, $\eta_p^2 = .002$. For the willingness-to-share measure, the effect of political stance was more evident for the fake news (*liberals*:1.90, *conservatives*: 3.12), $F_{(1,562)} = 66.94$, $p < .001$,$\eta_p^2 = .106$, than for the real news (*liberals*:3.32, *conservatives*: 3.65), $F_{(1,562)} = 4.84, p = .028$, $\eta_p^2 = .009$.

Yet, we did not obtain the three-way interaction of political stance × veracity × condition for perceived accuracy rating or willingness-to-share measure, $Fs_{(1,558)} = 2.18$ and 1.82, $ps = .114$ and .162, $\eta_{ps}^2 = .008$ and .006, indicating minimal impacts of correction on addressing conservatives' higher susceptibility to COVID-19 misinformation.

Influential Factors. In the post-session question, we also asked participants to specify factors that impacted their perceived accuracy rating, including "user tweet (text)," "users tweet (image)," "comments," and "other." For participants who selected "comments," we further asked them to select the parts of comments affected their decision the most among the options "who wrote the comment," "how persuasively the comment was written," "whether the comment included a reference URL," and "other."

As shown in Figure 7 top panel, the majority of the participants chose "other's comments" as the most influential factors, and the results were similar between *hORG* (36%) and *hIND* (33.9%). For participants who chose "comment," those in the *hORG* condition chose "who wrote" the most (67.6%) while those in the *hIND* condition chose "URL" the most (49.6%), $\chi^2_{(3)} = 103.54$, $p < .001$, revealing the influence of reliable sources.

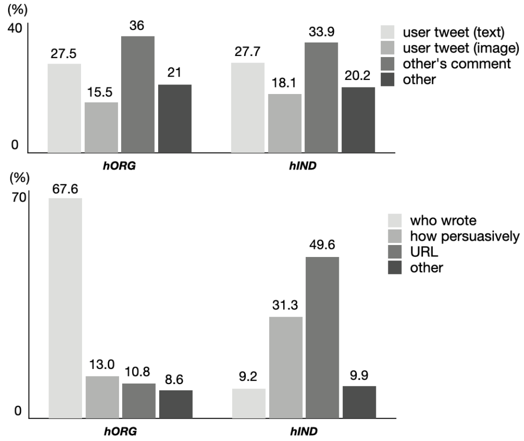


Figure 7: The top panel shows the response rate of the follow-up question asking the most influential factors in participants' perceived accuracy rating, and the bottom panel shows that of the most influential factors in the comment.

Summary

In Experiment 2, we not only replicated the effects of a correcting comment from health organizations (*hORG*) but also verified the correction effect from individual users (*hIND*). Both *hORG* and *hIND* reduced perceived accuracy rating on fake news (**RQ1**). Moreover, we obtained that participants relied on reliable sources of correcting comments. Specifically, *hORG* valued "who wrote the comment" the most while *hIND* valued "URL" the most. We also obtained evidence of the frequency effect showing decreased perceived accuracy rating for the fake news at Phase 2 (**RQ2**). Meanwhile, we found both *hORG* and *hIND* were effective for the *low anxiety* group to reduce their perceived accuracy rating on fake news (**RQ3**).

Experiment 3

Besides health organizations, fact-checking websites investigated false claims about COVID-19 from the beginning of the pandemic (Brennen et al. 2020). Thus, we ran Experiment 3 not only to verify again the correction effects of existing conditions (*hORG*, *hIND*), but also to examine the correction effects with two new conditions that are relevant to fact-checking websites (*fcORG*, *fcIND*, where "*fc*" indicates fact-checking websites). All experiment designs and news contents were the same as Experiment 2 except as noted. *fcORG* condition included correction from a fact-checking website with a reference link from the site. *fcIND* condition included the identical reference link as the *fcORG* condition but the correction was from an individual user. Regarding frequency effect, for both the *fcORG* and *fcIND* conditions, the link of *snopes.com* was used for the correction at Phase 1, and the link of *politifact.com* was used for the correction at Phase 2.

Results

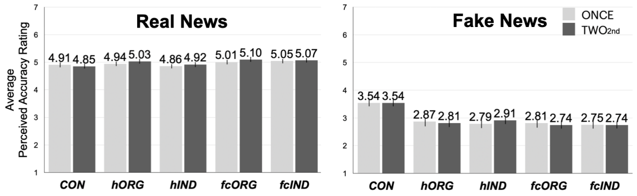


Figure 8: The average values of perceived accuracy ratings as a function of frequency × condition for real news (left panel) and fake news (right panel) with one standard error.

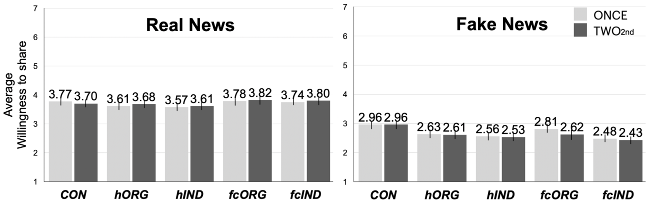


Figure 9: The average values of willingness-to-share as a function of frequency × condition for real news (left panel) and fake news (right panel) with one standard error.

We recruited $2,060$ MTurk workers from November to December, 2020. We accepted $1,166$ participants' answers after removing one incomplete submission, four responses submitted out of the U.S., 406 who failed an attention check, 473 duplicated submissions, and ten responses submitted less than three minutes (median completion time is about 10 minutes). The number of participants for each condition is as follows: 250 (*CON*), 236 (*hORG*), 227 (*hIND*), 214 (*fcORG*), and 239 (*fcIND*). The payment was the same as Experiment 2. Participants' demographic information is as Table 1.

Perceived accuracy rating and willingness-to-share measure were entered into 5 (condition: *CON*, *hORG*, *hIND*, *fcORG*, *fcIND*) × 2 (veracity: *fake*, *real*) × 2 (frequency: *ONCE*, TWO_{2nd}) mixed ANOVAs with a significance level of .05, respectively. Post-hoc tests with Bonferroni correction were performed.

Perceived Accuracy Rating. Results of perceived accuracy rating are shown in Figure 8. Same as prior two experiments, participants can distinguish real news (4.97) from fake news (2.95), $F_{(1,1161)} = 1577.27$, $p < .001$, $\eta^2 = .576$. The main effect of condition was also significant, $F_{(4,1161)} = 3.76$, $p = .005$, $\eta_p^2 = .013$, and the difference among conditions was qualified by the effect of news veracity, $F_{(4,1161)}$ = 12.59, $p < .001$, $\eta_p^2 = .042$. Same as Experiment 2, post-hoc tests revealed that the effect of each treatment condition was significant for the fake news compared to *CON*, $p_{adjs.} < .001$.

Sharing Decisions. Results of willingness-to-share measure are presented in Figure 9. Participants showed more willingness to share real news (3.71) than fake news (2.66), $F_{(1,1161)}$ = 605.38, $p < .001$, $\eta_p^2 = .343$. Also, the interaction of veracity × condition was significant, $F_{(4,1161)} = 4.32$, $p = .002$, $\eta_p^2 = .015$. In the post-hoc comparisons, only the gap between *fcIND* and *CON* for fake news was significant, $p_{adj.} = .033$.

Health Anxiety. We analyzed 1166 participants' results as in previous experiments. For perceived accuracy rating, participants with *high health anxiety* (4.32) gave a higher accuracy rating than those with *low health anxiety* (3.72) $F_{(1,1156)} = 84.42$, $p < .001$, $\eta_p^2 = .068$, and such pattern was more evident for fake news (*high health anxiety*: 3.45; *low health anxiety*: 2.63), $F_{(1,1164)} = 67.58$, $p < .001$, $\eta_p^2 = .051$ than for real news (*high health anxiety*: 5.20; *low health anxiety*: 4.82), $F_{(1,1164)} = 41.63$, $p < .001$, $\eta_p^2 = .035$.

For willingness-to-share, participants with *high health anxiety* (3.83) gave higher willingness-to-share score than those with *low health anxiety* (2.75), $F_{(1,1156)} = 124.33$, $p < .001$, $\eta_p^2 = .097$, similar to Experiments 1 and 2. Thus, highly anxious people about their health tended to share more health-related news regardless of news veracity.

Political Stance. We analyzed the effect of political stance as in Experiment 2 with 464 of liberals and 377 of conservatives after removing 325 moderates. The main effect of political stance, $Fs_{(1,831)} = 24.53$ and 17.95, $ps < .001$, $\eta_{ps}^2 = .029$ and .021, and its interaction with veracity, $Fs_{(1,831)} = 161.99$ and 80.38, $ps < .001$, $\eta_{ps}^2 = .163$ and .088, were significant for both perceived accuracy rating and willingness-to-share measure. Specifically, for perceived accuracy rating, the effect of political stance was more evident for the fake news (*liberals*: 2.58, *conservatives*: 3.73), $F_{(1,839)} = 86.07$, $p < .001$, $\eta_p^2 = .093$, than for the real news (*liberals*: 5.16, *conservatives*: 4.83), $F_{(1,839)} = 22.61$, $p < .001$, $\eta_p^2 = .026$. For the willingness-to-share measure, the effect of political stance was only significant for the fake news (*liberals*:2.37, *conservatives*: 3.33), $F_{(1,831)} = 48.54$, $p < .001$, $\eta_p^2 = .055$, but not the real news (*liberals*:3.75, *conservatives*: 3.82).

As in Experiment 2, conservatives were more susceptible to COVID-19 misinformation than liberals (Uscinski et al. 2020). However, we again did not obtain the three-way interaction of political stance × veracity × condition for neither measures, $F_s < 1.0$, showing limited impacts of correction on mitigating conservatives' higher susceptibility to misinformation as Experiment 2.

Influential Factors. As in Experiment 2, we asked participants which factors were influential for their decision-making and which parts of comments influenced the most. Across all treatment conditions, participants chose "other's comments" the most except for those in the *hIND* condition (see Figure 10 top panel). For the following question asking the most influential part in the comments, participants in both organization conditions chose "who wrote" the most while those in both individual conditions chose "URL" the most, $\chi^2_{(9)} = 145.80$, $p < .001$ (see Figure 10 bottom panel). Overall, the obtained results were consistent with those found in Experiment 2.

Summary

Findings of Experiment 3 were consistent with the previous two experiments. All types of corrections were effective in reducing participants' perceived accuracy rating of fake news. We verified the effects of a correcting comment (**RQ1**) from fact-checking websites (*fcORG*) as well as the one from health organizations (*hORG*). Also, we found the effects of individual users' correction which has a reference link from either fact-checking websites (*fcIND*) or health organizations (*hIND*). Moreover, we discovered that participants counted on the reliable source of a correcting comment. Specifically, participants in the *hORG* and *fcORG* weighed "who wrote the comment" the most, while those in the *hIND* and *fcIND* counted on "URL" the most. As in Experiment 1, we did not obtain the frequency effect (**RQ2**). Meanwhile, we found the minimal impacts of health anxiety or political stance on the correction effect (**RQ3**).

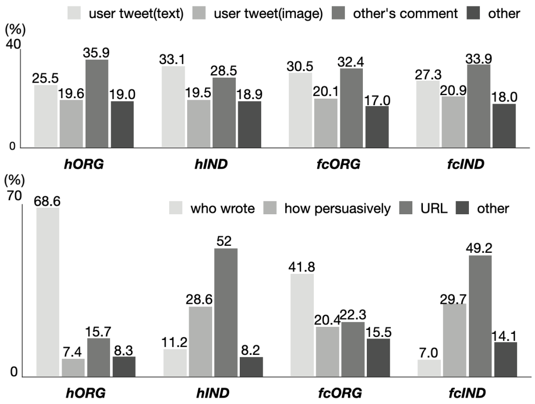


Figure 10: The top panel shows the response rate of the follow-up question asking the most influential factors in participants' perceived accuracy rating, and the bottom panel shows the most influential factors in the comment.

General Discussion

In the current study, we investigated if the correction from organization users or individual users can reduce participants' susceptibility to COVID-19 fake news (**RQ1**). We also examined whether more frequent correction can further reduce the susceptibility (**RQ2**), and whether individuals' health anxiety level has an impact on the effect of correction (**RQ3**). Across the three online experiments with $2,841$ participants, we examined the correction effects of three types of users on social media. We verified the effect of user-initiated correction in general, with the fact that participants counted on the reliability of correction. We also found that participants with high health anxiety were more susceptible to COVID-19 fake news than those with low health anxiety in all experiments.

Effect of Correction from a Single User

Previous work obtained the effect of user-initiated correction on social media by conducting a survey (Bode and Vraga 2021) or analyzing Twitter data (Jiang et al. 2020). Also, other studies (Vraga and Bode 2017, 2018) showed the effect of correcting comments in social media contexts by conducting experiments. Our study extended those works by demonstrating the consistent effects of user-initiated correction with reliable sources in reducing perceived accuracy rating on COVID-19 fake news in a social media context.

We corroborated that people's perceived accuracy rating on fake news could be reduced by a single correction comment by health organizations, fact-checking websites, or individual users. In particular, the correction effect was similar across different types of users. Critically, we unearthed that participants depended on the reliability of sources in the correction to decide their perceived accuracy rating. Participants in *hORG* and *fcORG* chose "who wrote the comment" the most, while those in *hIND* and *fcIND* chose "whether the comment included a reference URL" the most. The only difference between *ORG* and *IND* was whether the correction is directly delivered by reliable users or indirectly delivered through reliable URLs. Thus, our findings on the indirect effect of a reliable source contribute to the literature about the source effect (Vraga and Bode 2017; Seo, Xiong, and Lee 2019).

Although the *hIND* was effective on correction in Experiments 2 and 3, it did not show a significant difference compared to the no correction condition in Experiment 1. One possible explanation for the difference might be related to users' increased knowledge about COVID-19 news (Bode and Vraga 2021), and consequently increased reliance on other individual users beyond health organizations to gain more information. The above reason may also explain the non-significant results of Vraga and Bode's work (2017) since they implemented a relatively unfamiliar topic to the participants in their experiment. Moreover, across the three experiments, we found that participants' perceived accuracy rating on real news in the treatment conditions was increased or similar to that in the control conditions, indicating limited side effects of user-initiated correction compared to platform-driven correction (e.g., fact-checking warnings) (Clayton et al. 2020).

In all experiments, participants were conservative in sharing news than the perceived accuracy rating, which may contribute to the minimal correction effect on misinformation sharing. Various motivations such as information-seeking, socializing, status-seeking, or prior social media sharing experience (Lee and Ma 2012) could drive people's intention of news sharing on social media regardless of correction. Future studies could investigate effective correction to prevent misinformation sharing considering such motivations of sharing behaviors.

Frequency Effect on Correction

Throughout the experiments, the frequency effect on correction was only evident in Experiment 2. Results of Experiments 1 and 3 were consistent with the correction effect but did not show statistical significance: perceived accuracy rating and willingness-to-share measures were numerically smaller for the second correction than the first correction. Those results may be due to the use of the same comment messages across phases, since people typically expect varied comments from different social media users. Future work could consider varying correction messages to understand further the frequency effect of correction.

Correction Effect Depending on Health Anxiety

We discovered that participants with high health anxiety tended to believe and share more news regardless of news veracity than those with low health anxiety, indicating that the highly anxious people might seek reassurance through health information (Starcevic and Berle 2013). In particular, we verified that *hORG* and *hIND* were only effective for the low health anxiety group to reduce their perceived accuracy rating on fake news in Experiment 2. To our best knowledge, our study was the first dealing with the impact of health anxiety on correction for COVID-19 related fake news. Future studies should contrive correction methods, especially for people with high health anxiety, to mitigate their susceptibility to COVID-19 misinformation in particular and fake health news in general.

Correction Effect Depending on Political Stance

Experiments 2 and 3 revealed that conservatives were more susceptible to COVID-19 fake news than liberals. Such results are in agreement with a recent study showing stronger beliefs in COVID-19 fake news by conservatives (Uscinski et al. 2020). In both of our experiments, corrections showed minimal impacts on helping conservatives. Considering the impacts of political stance on various misinformation literature (Frenda et al. 2013; Benegal and Scruggs 2018; Pennycook and Rand 2019), we believe that further investigation on effective correction methods for more vulnerable populations (e.g., conservatives) is essential.

Limitations and Future Work

We discuss a few limitations that could be addressed in future studies. First, we chose MTurk for recruitment to gain a reasonably large sample size as previous misinformation studies did (Pennycook, Cannon, and Rand 2018; Clayton et al. 2020). MTurk workers are more demographically diverse than the college students (Briones and Benham 2017; Weigold and Weigold 2021). However, the MTurk population in general cannot fully represent the whole population. For instance, most MTurk workers tend to be in their 30's (Burnham, Le, and Piedmont 2018). Therefore, a more comprehensive recruiting method could be used to generalize our findings to other samples in future studies. In addition, in terms of materials, we used a more recent news set in Experiments 2 and 3 since COVID-19 news has been quickly updated and diversified. This change seemed to lead to different average gaps of perceived accuracy ratings between real and fake news among experiments: Exp.1 (2.59), Exp.2 (2.20), Exp3 (2.02). Furthermore, it should be recognized that the effects found in our experiments may not appear in practice due to exclusion of other factors on social media (e.g., multiple replies and social relationships among users, etc.). Moreover, we are aware that participants could not pay attention to the correction in reality because of many distracting factors on social media, such as other postings and interactions with other users in real time. Also, we did not measure participants' prior beliefs in the fake news. Therefore, we can not rule out the possibility of having negative effects from correction (e.g., backfire effects) (Nyhan and Reifler 2010; Mosleh et al. 2021). Future works could develop experimental designs with more ecological validity and evaluate the generalizability of our findings.

Conclusion

In this work, we carried out three online experiments with a more systematic design to comprehend the impact of a single correction comment on mitigating users' fake news susceptibility on social media. In total, three types of users were investigated across the experiments. We verified the correction effects on reducing the user's perceived accuracy ratings from individual users, health organizations, and fact-checking websites. Moreover, our study revealed that participants counted on the reliability of correction sources for their decision-making. We also found that high health anxiety people could be more susceptible to COVID-19 misinformation. Additionally, our results showed that conservatives are more susceptible to COVID-19 fake news than liberals. In conclusion, our findings highlight 1) the importance of encouraging social media users to leave correcting comments on fake news, as long as they have reliable sources, and 2) the necessity to develop effective correction methods considering individual differences (e.g., health anxiety level and political stance).

Ethical Statement

Our research protocol was approved by the Institutional Review Board (IRB) at The Pennsylvania State University. We asked for informed consent from participants. We made sure to take suitable steps in our data collection and analysis to ensure an ethical study and preserve user privacy. Additionally, in order to avoid any issues of account or user identification and to protect user privacy, we did not name any accounts in this paper. In particular, we note that we did not debrief the participants. Several recent studies did the debriefing to minimize the impacts of misinformation over time (Murphy et al. 2020). Thus, we acknowledge that the lack of debriefing in our experiments could have potentially harmful effects on some participants (e.g., those in the control condition without misinformation correction). However, as a recent study revealed, misinformation study in general does not significantly increase participant's long-term susceptibility to misinformation used in the experiments (Murphy et al. 2020).

Acknowledgements

This research was supported in part by the PSU SSRI seed grant and NSF awards #1742702, #1820609, #1915801, and #2121097.

References

Allcott, H.; and Gentzkow, M. 2017. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2): 211–236.

Asmundson, G. J.; and Taylor, S. 2020. How health anxiety influences responses to viral outbreaks like COVID-19: What all decision-makers, health authorities, and health care professionals need to know. *Journal of Anxiety Disorders*, 71: 102211.

Bakeman, R. 2005. Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, 37(3): 379–384.

Banerjee, D.; Rao, T. S.; et al. 2020. Psychology of misinformation and the media: Insights from the COVID-19 pandemic. *Indian Journal of Social Psychiatry*, 36(5): 131–137.

Bechmann, A.; and Lomborg, S. 2013. Mapping actor roles in social media: Different perspectives on value creation in theories of user participation. *New Media & Society*, 15(5): 765–781.

Benegal, S. D.; and Scruggs, L. A. 2018. Correcting misinformation about climate change: The impact of partisanship in an experimental setting. *Climatic Change*, 148(1): 61–80.

Berinsky, A. J.; Huber, G. A.; and Lenz, G. S. 2012. Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk. *Political Analysis*, 20(3): 351–368.

Bode, L.; and Vraga, E. K. 2015. In related news, that was wrong: The correction of misinformation through related stories functionality in social media. *Journal of Communication*, 65(4): 619–638.

Bode, L.; and Vraga, E. K. 2018. See something, say something: Correction of global health misinformation on social media. *Health Communication*, 33(9): 1131–1140.

Bode, L.; and Vraga, E. K. 2021. Correction Experiences on Social Media During COVID-19. *Social Media + Society*, 7(2).

Boyd, D. M.; and Ellison, N. B. 2007. Social network sites: Definition, history, and scholarship. *Journal of Computer-mediated Communication*, 13(1): 210–230.

Brennen, J. S.; Simon, F.; Howard, P. N.; and Nielsen, R. K. 2020. Types, sources, and claims of COVID-19 misinformation. *Reuters Institute*, 7(3): 1–13.

Briones, E. M.; and Benham, G. 2017. An examination of the equivalency of self-report measures obtained from crowdsourced versus undergraduate student samples. *Behavior Research Methods*, 49(1): 320–334.

Burnham, M. J.; Le, Y. K.; and Piedmont, R. L. 2018. Who is Mturk? Personal characteristics and sample consistency of these online workers. *Mental Health, Religion & Culture*, 21(9-10): 934–944.

Calvillo, D. P.; Ross, B. J.; Garcia, R. J.; Smelter, T. J.; and Rutchick, A. M. 2020. Political ideology predicts perceptions of the threat of covid-19 (and susceptibility to fake news about it). *Social Psychological and Personality Science*, 11(8): 1119–1128.

Clayton, K.; Blair, S.; Busam, J. A.; Forstner, S.; Glance, J.; Green, G.; Kawata, A.; Kovvuri, A.; Martin, J.; Morgan, E.; et al. 2020. Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. *Political Behavior*, 42(4): 1073–1095.

Cook, J.; Ecker, U.; and Lewandowsky, S. 2015. Misinformation and how to correct it. In *Emerging trends in the social and behavioral sciences: An interdisciplinary, searchable, and linkable resource*, 1–17. John Wiley & Sons.

Ecker, U. K.; Lewandowsky, S.; and Apai, J. 2011. Terrorists brought down the plane!—No, actually it was a technical fault: Processing corrections of emotive information. *Quarterly Journal of Experimental Psychology*, 64(2): 283–310.

Facebook. 2020. Here's how we're using AI to help detect misinformation. https://ai.facebook.com/blog/heres-how-were-using-ai-to-help-detect-misinformation. Accessed: 2021-09-15.

Frenda, S. J.; Knowles, E. D.; Saletan, W.; and Loftus, E. F. 2013. False memories of fabricated political events. *Journal of Experimental Social Psychology*, 49(2): 280–286.

Gesser-Edelsburg, A.; Diamant, A.; Hijazi, R.; and Mesch, G. S. 2018. Correcting misinformation by health organizations during measles outbreaks: A controlled experiment. *PLoS One*, 13(12): e0209505.

Gordon, L. T.; and Shapiro, A. M. 2012. Priming correct information reduces the misinformation effect. *Memory & Cognition*, 40(5): 717–726.

Ha, L.; Andreu Perez, L.; and Ray, R. 2021. Mapping recent development in scholarship on fake news and misinformation, 2008 to 2017: Disciplinary contribution, topics, and impact. *American behavioral scientist*, 65(2): 290–315.

Hauser, D. J.; and Schwarz, N. 2016. Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, 48(1): 400–407.

Herzog, M. H.; Francis, G.; and Clarke, A. 2019. ANOVA. In *Understanding Statistics and Experimental Design*, 67–82. Springer.

Jiang, S.; Metzger, M.; Flanagin, A.; and Wilson, C. 2020. Modeling and measuring expressed (dis) belief in (mis) information. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, 315–326.

Jiang, S.; and Wilson, C. 2018. Linguistic signals under misinformation and fact-checking: Evidence from user comments on social media. In *Proceedings of the ACM on Human-Computer Interaction 2 (CSCW)*, 82.

Jungmann, S. M.; and Witthöft, M. 2020. Health anxiety, cyberchondria, and coping in the current COVID-19 pandemic: Which factors are related to coronavirus anxiety? *Journal of Anxiety Disorders*, 35: 102239.

Laato, S.; Islam, A.; Islam, M. N.; and Whelan, E. 2020. Why do people share misinformation during the Covid-19 pandemic? *arXiv preprint arXiv:2004.09600*.

Lakens, D. 2013. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4: 863.

Lazer, D. M.; Baum, M. A.; Benkler, Y.; Berinsky, A. J.; Greenhill, K. M.; Menczer, F.; Metzger, M. J.; Nyhan, B.; Pennycook, G.; Rothschild, D.; et al. 2018. The science of fake news. *Science*, 359(6380): 1094–1096.

Lee, C. S.; and Ma, L. 2012. News sharing in social media: The effect of gratifications and prior experience. *Computers in Human Behavior*, 28(2): 331–339.

Lewandowsky, S.; Ecker, U. K.; Seifert, C. M.; Schwarz, N.; and Cook, J. 2012. Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3): 106–131.

Lucock, M. P.; and Morley, S. 1996. The health anxiety questionnaire. *British Journal of Health Psychology*, 1(2): 137–150.

McMullan, R. D.; Berle, D.; Arnáez, S.; and Starcevic, V. 2019. The relationships between health anxiety, online health information seeking, and cyberchondria: Systematic review and meta-analysis. *Journal of Affective Disorders*, 245: 270–278.

Mosleh, M.; Martel, C.; Eckles, D.; and Rand, D. 2021. Perverse Downstream Consequences of Debunking: Being Corrected by Another User for Posting False Political News Increases Subsequent Sharing of Low Quality, Partisan, and Toxic Content in a Twitter Field Experiment. In *proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–13.

Murphy, G.; Loftus, E.; Grady, R. H.; Levine, L. J.; and Greene, C. M. 2020. Fool me twice: How effective is debriefing in false memory studies? *Memory*, 28(7): 938–949.

Nyhan, B.; and Reifler, J. 2010. When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2): 303–330.

Nyhan, B.; and Reifler, J. 2015. Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information. *Vaccine*, 33(3): 459–464.

Olejnik, S.; and Algina, J. 2003. Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychological Methods*, 8(4): 434–447.

Pennycook, G.; Cannon, T. D.; and Rand, D. G. 2018. Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, 147(12): 1865–1880.

Pennycook, G.; and Rand, D. G. 2019. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188: 39–50.

Quandt, T.; Frischlich, L.; Boberg, S.; and Schatto-Eckrodt, T. 2019. Fake news. *The International Encyclopedia of Journalism Studies*, 1–6.

Roth, Y.; and Pickles, N. 2021. Updating our approach to misleading information. http://tiny.cc/77vquz. Accessed: 2021-09-15.

Seifert, C. M. 2002. The continued influence of misinformation in memory: What makes a correction effective? In *Psychology of Learning and Motivation*, volume 41, 265–292. Elsevier.

Seo, H.; Xiong, A.; and Lee, D. 2019. Trust It or Not: Effects of Machine-Learning Warnings in Helping Individuals Mitigate Misinformation. In *Proceedings of the 10th ACM Conference on Web Science*, 265–274.

Seo, H.; Xiong, A.; Lee, S.; and Lee, D. 2021. (In) effectiveness of Accumulated Correction on COVID-19 Misinformation. In *Proceedings of Technology, Mind & Society*.

Smith, C. N.; and Seitz, H. H. 2019. Correcting misinformation about neuroscience via social media. *Science Communication*, 41(6): 790–819.

Starcevic, V.; and Berle, D. 2013. Cyberchondria: towards a better understanding of excessive health-related Internet use. *Expert Review of Neurotherapeutics*, 13(2): 205–213.

Tasnim, S.; Hossain, M. M.; and Mazumder, H. 2020. Impact of rumors and misinformation on COVID-19 in social media. *Journal of Preventive Medicine and Public Health*, 53(3): 171–174.

Thorson, E. 2016. Belief echoes: The persistent effects of corrected misinformation. *Political Communication*, 33(3): 460–480.

Uscinski, J. E.; Enders, A. M.; Klofstad, C.; Seelig, M.; Funchion, J.; Everett, C.; Wuchty, S.; Premaratne, K.; and Murthi, M. 2020. Why do people believe COVID-19 conspiracy theories? *Harvard Kennedy School Misinformation Review*, 1: 3.

Vraga, E. K.; and Bode, L. 2017. Using expert sources to correct health misinformation in social media. *Science Communication*, 39(5): 621–645.

Vraga, E. K.; and Bode, L. 2018. I do not believe you: how providing a source corrects health misperceptions across social media platforms. *Information, Communication & Society*, 21(10): 1337–1353.

Walter, N.; and Murphy, S. T. 2018. How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs*, 85(3): 423–441.

Weigold, A.; and Weigold, I. K. 2021. Traditional and Modern Convenience Samples: An Investigation of College Student, Mechanical Turk, and Mechanical Turk College Student Samples. *Social Science Computer Review.*