University of Texas Rio Grande Valley

ScholarWorks @ UTRGV

You've been fact-checked! Examining the effectiveness of social media fact-checking against the spread of misinformation

Ben Wasike
University of Texas Rio Grande Valley, ben.wasike@utrgv.edu

Follow this and additional works at: https://scholarworks.utrgv.edu/com_fac
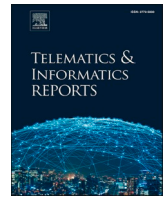
Part of the Communication Commons

Recommended Citation

You've been fact-checked! Examining the effectiveness of social media fact-checking against the spread of misinformation

Ben Wasike

Department of Communication, University of Texas Rio Grande Valley, Ph.D. granting institution: Louisiana State University (2005) One West University Boulevard, Brownsville, TX 78520, United States

ARTICLE INFO

ABSTRACT

Using a random sample of active social media users ($N = 1,156$), this study examined the effectiveness of social media fact-checking against online misinformation sharing. Data indicates that these fact-checks are minimally effective in stopping the spread of misinformation on social media. Being aware of the fact-checks, being fact-checked, or even having content deleted from one's account were not deterrents to sharing misinformation. The fear of isolation was the strongest deterrent, suggesting that account freezes, suspensions, or bans were the most effective ways to curtail the spread of misinformation. The study contributes to research on fact-checking, to research on online surveillance, and to research on online expression and the spiral of silence theory.

Introduction

"The bottom line is: we take misinformation seriously," wrote Facebook's CEO Mark Zuckerberg in a post that launched a series of initiatives aimed at combating falsehoods posted on the social media site ([1], para. 2). These initiatives led to social media fact-checking as we know it today. This is where various social networking sites (SNSs) take an active role in vetting the information posted on their platforms using human fact-checkers as well as using mechanized methods such as algorithms and machine learning systems ([2], para. 2). Until then, fact-checking was traditionally a journalistic practice used to vet questionable published or spoken claims [3]. The erstwhile version of fact-checking was done by news agencies and other independent vetting organizations such as Snopes, PolitiFact, Factcheck.org, and various blogs, among others, who posted fact-check notices on misleading claims [4]. These notices did not appear in situ alongside the original misinformation, but appeared on the vetting agency's own website alongside samples of the misinformation. Additionally, most such vetting was done manually and focused on institutions and political figures [3,4].

Unlike traditional fact-checking, which mostly targeted public figures, social media fact-checking affects anyone using a particular social media site ([5], para. 1). Also, the consequences for rules violations are as dire as having an account frozen or suspended, or the outright expulsion from the site ([2], para. 2; [6], para. 1; [7], para. 1; [5], para. 1). Research shows that any type of online surveillance, including social

media fact-checking, negatively affects expression [8]. This effect, combined with those from the mentioned consequences of rule violations and the visibility of fact-checks via obvious warning labels that accompany questionable posts [9,10], make social media fact-checking the more unique.

Therefore, it is the uniqueness of social media fact-checking and its effectiveness in curbing the spread of misinformation that is the focus of this study. Generally, the study examines the effectiveness of these fact-checks on stopping the spread of misinformation. Specifically, the study examines whether the *awareness* of fact-checking, *experience* with fact-checking, having *content deleted* from one's account, opinion *congruency* with others, and the *fear of isolation* affect the likelihood to: A) *Post* misinformation, B) *Delete* fact-checked information, and C) Engage with fact-checked information via *commenting* on it. To differentiate it from the traditional fact-checking, this study refers to the newer version of fact-checking as social media fact-checking, or simply fact-checking when appropriate.

Literature review

How social media fact-checking works

Fact-checking as a traditional journalistic practice dates to 1912 with the establishment of the Bureau of Accuracy and Fair Play at Joseph Pulitzer's publication, the *World* ([11], para. 4). Modern journalistic fact-checking precedes the internet era and can be traced to the Reagan

era when the practice emerged as a form of corrective action against factual inaccuracies uttered by politicians. The internet subsequently expanded this practice not only among the news media, but also among bloggers [4]. Social media fact-checking by SNSs like Facebook, Twitter, and YouTube follows this long tradition of information vetting [12]. In the second quarter of 2020 for example, Facebook deleted over 7 million COVID-19-related misinformation posts ([5], para. 1). By February 2021, YouTube had deleted over half a million videos that contained COVID-19 misinformation ([6], para. 1). In June 2020, Twitter announced that it had deleted over 170,000 accounts that spread falsities about COVID-19 and the Hong Kong protests ([7], para. 1).

SNSs use a variety of methods to fact-check information on their sites, such as when users flag a post or when an algorithm detects many comments flagging a post ([2], para. 2). Facebook uses artificial intelligence, keyword searches, and independent third-party fact-checkers who must be certified by the International Fact-checking Network ([13], para. 1; [14], para. 1). IFCN is a unit of the Poynter Institute and connects fact-checkers worldwide ([15], para. 1). At the time of writing, IFCN has 88 signatories from various media organizations, academia, research institutes, and independent fact-checking organizations. Examples include the AFP fact-checking unit (France), TV Today Network (India), PesaCheck (Kenya), and Colombiacheck (Colombia), among others ([16], para. 1). Unlike Facebook and Google (which also partner with the IFCN), Twitter, before going under private ownership [17], used its own staff or crowdsourced among its users [18]. YouTube also uses a combination of human evaluators and machine-learning systems to identify misinformation ([19], para. 2). Regardless of the methods used, the results are usually apparent to users and they manifest in the form of warning flags and labels alongside the questionable posts [10]. See Figs. 1–4 for sample notices.

Opinion congruence, fear of isolation, and social media

Among other variables, this study examines how congruence of opinion and the fear of isolation (FOI) affect how people post misinformation online. Both variables derive from the spiral of silence theory. This theory describes scenarios where people gauge prevailing opinions, and if their opinion does not align with the majority opinion, they refrain from expressing it due to FOI. This process climaxes to a point where only the dominant opinion prevails as opposing opinions spiral into silence [20,21]. Research shows that a spiral of silence exists within offline and online communication [22,23]. Research also suggests a confluence among fact-checking, opinion congruence, and the spiral of silence.

Under surveillance, people are less likely to discuss issues on social media as they would do face-to-face. People are also more likely to share their views based on how compatible they deem their views to align with those of others around them ([24], para. 8). This mirrors the perception of opinion climate aspect of the spiral of silence theory. For instance, Stoycheff [8] found that surveillance reduced the likelihood of people to
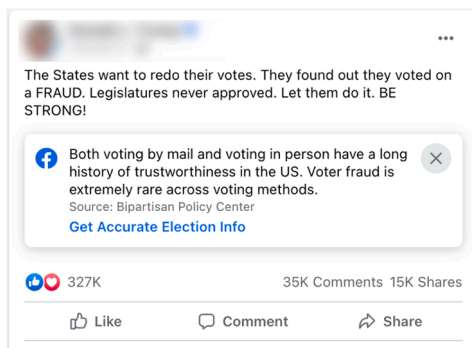

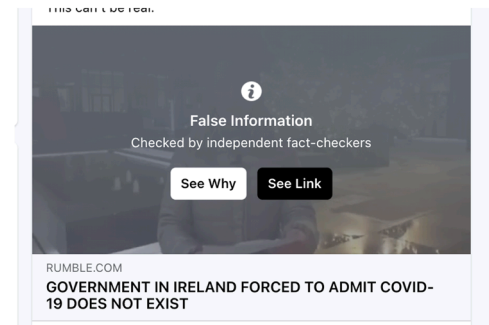**Fig. 2.** A Twitter notice on a suspended account.

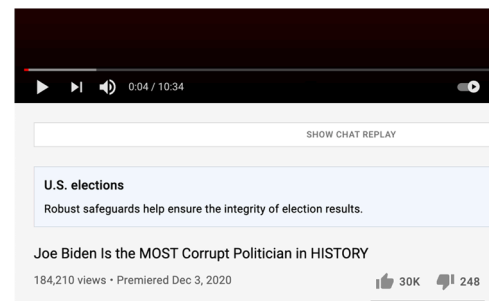**Fig. 3.** A Facebook notice alongside a COVID-19 post.

Fig. 4. A YouTube notice alongside an election fraud post.

speak up on social media regardless of the hostility or friendliness of the opinion climate. Also, a person's perception of the opinion climate affected the likelihood to speak up. This suggests that online surveillance and the perception of the opinion climate affect the likelihood to express oneself on social media.

Research on the spiral of silence and social media shows a strong association between the perception of the opinion climate and the likelihood to self-censor, with FOI as the motivator. Self-censorship is when a person withholds their opinion if they think that others may disagree with it [25]. FOI is "a psychological variable representing a negative emotional state associated with the prospect of voicing one's opinion about a given topic" ([26], p. 452). FOI leads to self-censorship on social media, especially when one opines on controversial topics such as police brutality [27], gay bullying and queer identity [28], and the expression of support for certain political candidates [29].

The effects of fact-checking

Research shows that both traditional and social media fact-checking affects human communication. Traditional fact-checking has mixed results. For instance, the awareness of a fact-check may cause a politician to make less false statements in public [30]. Traditional fact-checking also affects the favorability rating of a politician among voters even

Fig. 1. A Facebook notice alongside an election fraud post.

though it does not change how voters view the fact-checking process itself [31]. However, fact-checking improves perceptions of information accuracy among people who encounter vetted misinformation [32,33], even though other research indicates that perceptions of accuracy may depend on a user's ideological orientation [34].

Burgeoning research on social media fact-checking shows such effects. For one, the overt display of the warning labels [10] has the same shaming effect that traditional journalistic fact-checking had on public figures ([35], para. 9). These warning labels also affect perceptions off accuracy. [9] found that two commonly used Facebook warning labels, "rated false" and "disputed," reduced the perceptions of the accuracy of the false headlines that the labels accompanied [36] found that both a false information label and a detailed refutation accompanying a misleading tweet reduced belief in the tweet, even though the detailed refutation had a bigger effect. Also, the mere presence of a warning label on a social media post may affect perceptions of the credibility of an entire SNS site. Specifically, Oeldorf-Hirsch et al. [37] found that the presence of a "disputed" label on a story in a social media post improved user perception of the SNS's credibility. Additionally, research shows that the mere presence of a warning label alongside a false report reduces the chances of users sharing it with others on social media [38]. However, research also shows that social media fact-checking may be hampered by distrust towards the fact-checkers [39], and the ideological leaning of the users affects their trust in the vetting process [40].

Fact-Checking and online surveillance

Social media fact-checking and the spiral of silence cannot be discussed in the absence of surveillance, and in particular, online surveillance. Fact-checking, broadly defined, is a byproduct of the technology-driven surveillance systems that emerged in the late Twentieth Century [41,42]. These systems used information technology to systematically monitor people's actions and communications [43]. In calling this type of surveillance the "new surveillance," Marx defined it as the "use of technical means to extract or create personal data" (2002, p. 20). Unlike traditional surveillance methods that use human labor and visible means to collect data, the "new surveillance" easily collects and manipulates data using technology. It also differs from traditional surveillance because it is less visible, is masked and belied by routine activities, it is continuous, and it assumes more involuntary consent than traditional surveillance (Marx). Two examples of such surveillance include the Cambridge Analytica's secret data harvesting of Facebook user profiles in 2016 ([44], paras. 2 & 3) and the United States' NSA's secret data collection of phone records in 2013 ([45], para. 1).

Such surveillance is not without consequence, and research indicates that users who are aware of online surveillance are less likely to express their opinions, are less likely to disclose their ideological views, and will only express themselves when they perceive that their opinions match the majority opinion, thus mimicking the spiral of silence [8]. These and the other effects of fact-checking on expression discussed earlier suggest that the awareness of fact-checking may negatively affect online expression. Given the discussion above, this study queries about the following:

RQ1: Does the awareness of fact-checks, experience with fact-checks, having content deleted, congruence of opinion, and FOI affect the likelihood of posting misinformation online?

RQ2: Does the awareness of fact-checks, experience with fact-checks, having content deleted, congruence of opinion, and FOI affect the likelihood of deleting fact-checked content from one's social media account?

RQ3: Does the awareness of fact-checks, experience with fact-checks, having content deleted, congruence of opinion, and FOI affect the likelihood of commenting on a fact-checked post?

Method

This study used a survey for data collection based on a random sample ($N = 1156$) of active social media users derived from a Qualtrics panel. Qualtrics panels are a proven method of deriving representative random samples [46,47]. In order to get active social media users only, respondents answered a screening question about how often they used social media, i.e., multiple times a day, once a day, often but not daily, and rarely/never. Those who indicated rarely or never were dropped from the survey. The sample was 51% female with an average age of 28.26 years. The demographic breakdown was: Non-Hispanic white = 67.5%, non-Hispanic Black = 11.4%, Hispanic = 9.7%, Asian = 2.9% Native Hawaiian of Pacific Islander = 0.4%, American Indian or Alaska Native = 3.6%, Other = 1.95%, and mixed race = 2.55%. Ideologically, the sample was: Lean Liberal = 26%, lean Independent = 28.2%, lean Conservative = 35.8%, and other = 10%. The study was approved by the author's IRB prior to data collection. Data were collected between November 5–10, 2021.

Measurement of variables

The dependent variables were: 1) The likelihood to *post* fact-checked information. 2) The likelihood to *delete* fact-checked information from one's account. 3) The likelihood to *comment* on a post containing fact-checked information. The independent variables were: 1) The *awareness* of fact-checking. 2) *Experience* with fact-checking. 3) Experience with having *content deleted* from one's account. 4) Opinion *congruency*. 5) *Fear of isolation.* Click here for survey questions.

Likelihood to Post Misinformation. This variable measured the likelihood of users to post information that had previously been fact-checked elsewhere. The question was measured on a 0–10 scale. Where 0 = not at all likely and 10 = very likely.

Likelihood to Delete Fact-check Content. This variable was measured by a question asking respondents about the likelihood of deleting information that had been fact-checked on their social media account, and was also based on a 0–10 scale.

Commenting on Fact-checked Posts. This variable measured the likelihood of a user to comment on someone else's fact-checked post. Respondents were asked two questions to gage whether they would comment in support of (positively) or in opposition (negatively) to a fact-checked post. An overall commenting score was then computed based on the average of responses to both questions. The questions, based on a 0–10 scale, were adapted from Gearhart and Zhang [28].

Awareness of Fact-checks. Here, respondents were asked a dichotomous (yes/no) question about whether they were familiar with social media fact-checking. They answered the question after reading a brief description of social media fact-checking and viewing samples of fact-checked posts.

Experience with Fact-checks. Like the awareness question, respondents were asked a dichotomous (yes/no) question whether any information they had posted on social media had been fact-checked.

Experience with Deletion. Like the above questions, respondents were asked a dichotomous (yes/no) question whether any information they had posted on social media had been deleted for any reason.

Fear of Isolation. This variable was measured by a scale adapted from previous studies [27,25] and was modified to suit this study's context. Therefore, questions were edited to reflect a social media context, for example: *It would bother me if no one wanted to be around me on social media.* Respondents were asked to indicate the extent to which the questions described them based on a 0–10 scale where 0 = does not describe me at all and 10 = completely describes me. A composite FOI score was then computed based on the average of the responses (Cronbach's alpha = 0.94, items in scale = 9).

Opinion Congruency. This variable measured the extent to which respondents' opinion matched or deviated from the prevailing majority opinion. Using methods from previous studies [28,8], respondents were

first asked a set of questions to gauge their estimation of public opinion on mask-wearing and the COVID-19 vaccines respectively. They were then asked about their personal opinion on the two issues. Two composite scores were then computed for other and personal opinions respectively. The congruency score was simply the difference between the two scores, with a high score indicating more discrepancy between a subject and other people. For instance, a subject who estimated that the majority opinion on an issue was 7 and reported their own opinion to be 5 scored a 2 on opinion congruency. Likewise, a score of zero indicates perfect congruence with the majority opinion. Because the direction of the difference is meaningless, the absolute values of the differences were used for data analysis (Cronbach's alpha = 0.90, items in scale = 6).

Results

On average, respondents reported having a high level of opinion congruence with others, 1.75 (on a 0–10 scale, where low numbers denote higher congruence). Respondents also reported a relatively low FOI on average, 4.05 (on a 0–10 scale, where low numbers denote a low FOI). Other summary statistics appear in Table 1. Research question 1 queried the effect of the awareness of fact-checks, experience with fact-checks, having content deleted, congruence of opinion, and FOI on the likelihood to post misinformation. For comprehensive analysis, a t-test and OLS regression were used. The t-test compared means of the three dichotomous variables (awareness of fact-checks, experience with fact-checks, and experience with content deletion), as well as median-split versions of the two interval/ratio variables (congruence and the FOI). The regression analysis used the two interval-ratio variables in their original format. As Table 2 shows, the t-test analysis returned significant differences for all independent variables. However, the totality of the results indicate that fact-checks did not reduce the likelihood to post misinformation. For instance, the likelihood of posting misinformation was higher among those who were aware of fact-checks, as well as among those who had experienced being fact-checked. Even having had content deleted from one's account did not reduce the likelihood of posting misinformation. The same pattern emerged with congruence and FOI. Here, those reporting high levels of opinion congruence were more likely to post misinformation than those reporting low levels of congruence. The FOI did not improve matters either, as those reporting high FOI were also more likely to post misinformation which suggests a counterintuitive effect regarding the repercussions of the fact-checking process. The regression analysis results shown in Table 3 largely reflect the t-test results. First, the awareness of fact-checks did not have any effect on the likelihood to post misinformation. Second, all other variables except for opinion congruency increased such likelihood. Additionally, FOI was the most impactful among all the variables in increasing the likelihood to post misinformation with a moderately high effect ($\beta = 0.31$, $p < .000$).

Research question two queried about the effect of the five

independent variables on the likelihood to delete from one's account, content that had been fact-checked. The t-test analysis (using median-splits for the two interval-ratio variables) returned no statistically significant differences between the independent variables and the likelihood to delete fact-checked content except for FOI. This means that being aware of fact checks, being fact-checked, and having content deleted did not improve the likelihood of deleting misinformation. This also means that perceiving whether one's opinion matched the popular opinion or not did not play a role in the process. However, FOI had an effect. Respondents who reported high FOI were likelier to delete fact-checked content from their accounts ($FOI_{High} = 6.06$; $FOI_{Low} = 4.22$; $p < .001$, $t = 9.58$, Cohen's $d = 0.57$). The same emerged with the OLS regression where FOI was the only independent variable to predict the likelihood to delete fact-checked content ($\beta = 0.36$; $p < .001$; $t = 12.76$).

Research question 3 queried about the effect of the five independent variables on the likelihood to comment on fact-checked information. As shown in Table 4, those aware of fact-checks, had experience with fact-checks, and had content deleted were more likely to comment on posts containing misinformation. Opinion congruency also increased this likelihood. Likewise, those reporting higher FOI were more likely to comment. These FOI results largely mirror those reported under RQ2 above. The OLS regression results shown in Table 5 reflect these

Table 1
Summary statistics.

Awareness of fact-checks	
Yes	82%
No	18%
Experience with fact-checks	
Yes	65%
No	35%
Experience with deletion	
Yes	67%
No	33%
Opinion congruency[1]	1.75 (1.59)
Fear of isolation	4.05 (2.78)
Likelihood to post misinformation	4.63 (3.20)
Likelihood to delete a fact-checked post	5.14 (3.39)
Likelihood to comment on a fact-checked post	4.96 (2.70)

[1] Means are shown with standard deviations in parentheses.

Table 2
T-test results for the likelihood to post fact-checked information.

		Mean	S.D.	t	p	d[1]
Awareness of fact-checking	Yes	4.75	3.25	2.71	.001	.21
	No	4.08	2.88			
Experience with fact-checking	Yes	5.83	3.16	9.7	.001	.6
	No	3.98	3.03			
Experience with content deletion	Yes	5.85	3.14	9.5	.001	.59
	No	4.02	3.05			
Congruence of opinion	High	5.18	3.1	7.31	.001	.44
	Low	3.81	3.17			
Fear of Isolation	High	5.61	2.88	10.83	.001	.64
	Low	3.66	3.2			

[1] Represents Cohen's *d* for the effect size.

Table 3
OLS regression predicting the likelihood to post fact-checked information.

	β	S.E.	t	p	VIF
Constant		.49	8.91	.001	
Awareness of fact-checks	−0.02	.22	−0.71	.478	1.05
Experience with fact-checks	.15	.21	4.78	.001	1.52
Experience with deletion of content	.10	.22	3.25	.001	1.54
Congruence of opinion	−0.11	.06	−4.25	.001	1.07
Fear of isolation	.31	.03	10.99	.001	1.15
Gender	−0.08	.18	−2.77	.01	1.12
Age	−0.08	.01	−2.72	.01	1.28

$R^2 = 0.23$, Adjusted $R^2 = 0.23$.

Table 4
T-test results for the likelihood to comment on fact-checked information.

		Mean	S.D.	t	p	d[1]
Awareness of fact-checking	Yes	5.06	2.72	2.59	.01	.20
	No	4.52	2.62			
Experience with fact-checking	Yes	5.94	2.61	9.30	.001	−0.57
	No	4.44	2.61			
Experience with content deletion	Yes	5.90	2.53	−8.61	.001	−0.54
	No	4.50	2.67			
Congruence of opinion	High	5.51	2.64	7.41	.001	−0.44
	Low	4.36	2.66			
Fear of Isolation	High	6.0	2.43	.13.90	.001	−0.82
	Low	3.94	2.59			

[1] Represents Cohen's *d* for the effect size.

Table 5

OLS regression predicting the likelihood to comment on a post with misinformation.

	β	S.E.	t	p	VIF
(Constant)		.30	11.89	< 0.001	
Awareness of fact-checks	−0.02	.18	−0.72	.48	1.05
Experience with fact-checks	.16	.18	5.10	< 0.001	1.52
Experience with deletion of content	.07	.18	2.30	.022	1.52
Congruence of opinion	−0.10	.04	−3.65	< 0.001	1.06
Fear of isolation	.41	.03	15.51	< 0.001	1.15
Gender	.09	.14	−3.35	< 0.001	1.12
Age	.05	.01	−1.60	.11	1.28

$R^2 = 0.29$, Adjusted $R^2 = 0.29$.

patterns. Here, the awareness of fact-checks had no impact on commenting and experience with content deletion had a marginal effect. Experience with fact-checks increased the likelihood of commenting while opinion congruency reduced it, even though both had low to moderate effects at the best. However, FOI had a higher impact on increasing the likelihood of commenting ($\beta = 0.41$, $p < .001$).

Discussion

Overall, data show that the social media fact-checks had a minimal impact on the likelihood to share misinformation. This is a departure from extant research, which shows that fact-checks, both traditional and contemporary, improve perceptions of accuracy among online users [9, 32,33,36]. Data showed a disjunction between user perception of misinformation and the likelihood to share that content. This suggests that users may be sharing misinformation knowing about fact-checks and related consequences. For one, those who were aware of the social media fact-checks were more likely to share misinformation than those who were not aware of them. Those who had been fact-checked before were also more likely to post misinformation than those who had not be fact-checked. The same occurred among users who had experienced content deletion. Additionally, the same pattern manifested regarding commenting on flagged posts.

Another important finding is the likelihood of deleting fact-checked information from one's account. Here too, only one of the five independent variables returned substantive differences. Only those who reported high FOI were more likely to delete fact-checked information than those who reported low FOI. As discussed, FOI generally curtails online expression [27,28,25]. Given that the social media fact-checks carry severe penalties such as account freezes, suspensions, or bans, it makes sense that this variable uniquely increased the likelihood to delete fact-checked information, given that the loss of a social media account would lead to the ultimate isolation online.

To contextualize the findings, it is worth discussing the role of the frequency of social media use and the possibility that the dependency on social media may explain the results reported here. The author therefore ran further analysis based on the frequency of social media use. Using responses from the social media use screening question asked at the beginning of the survey (multiple times a day, once a day, and often, but not daily), respondents were classified into three ordinal categories reflecting usage, thus, high, medium, and low. High social media users, or those who used it multiple times a day, made up most of the sample (78%). These users were more likely to post misinformation than those who used social media once a day or those who used it often but not daily. These heavy users also reported a higher FOI even though they were not likelier than others to delete fact-checked information from their accounts – see Table 6 below. The higher likelihood to post misinformation may simply be the result of the voluminous use of social media, which may be a symptom of a dependency on social media denoted by heavy daily use. This may explain the high FOI among heavy users who stand to lose the most with an account freeze, suspension, or

Table 6

One-way ANOVA analysis based on the frequency of social media use.

Frequency of social media use[1]				
	High	Medium	Low	F
Likelihood to post**	4.82 (3.21)	4.11 (3.11)	3.66 (2.95)	8.40
Likelihood to delete	5.26 (3.53)	4.78 (3.41)	4.61 (3.58)	2.65
Fear of isolation*	4.17 (2.80)	3.81 (2.73)	3.27 (2.54)	5.86

[1] High = Multiple times a day; Medium = Once a day; Low = Often but not daily

**$p < .001$; *$p < .01$.

ban. After all, the loss of access to one's social media account is the ultimate isolation for one so heavily invested in that platform.

Practical and theoretical implications

The findings have both practical and theoretical implications. Practically, the social media fact-checks alone may not be enough of a deterrent to stop the spread of misinformation unless they are accompanied by further sanctions. Awareness of fact-checks, experiencing fact-checks, or even having content deleted from one's account did not lower the likelihood to post misinformation. FOI was the most likely deterrent to sharing misinformation and this fear was more pronounced among heavy social media users who use social media multiple times a day. This variable was measured by questions asking about such fears as losing friends, being left of out of things, wanting to be around others, etc. Losing access to one's social media account due to a freeze, suspension, or ban would bring these fears to fruition.

Theoretically, the study advances research in several ways. First, it contributes to research that has examined the effect of fact-checks on attitudes and perceptions of misinformation, and especially to burgeoning research examining the social media fact-checking system. The study also revealed nuances regarding the sharing of misinformation on social media. Not all users are likely to do this, but heavy users are, even with the full knowledge of fact-checks and with experience with lighter sanctions such as fact-checks labels on their accounts or content deletion by various SNSs. While this study suggests that dependency on social media (evidenced by the frequency of use) may drive the violation of fact-checks among heavy users, future research may further explore patterns of use among heavy users. The study also adds to the spiral of silence literature. This is important because social media giants have been accused of using the fact-checking system to selectively silence certain voices online, even though some of these accusations are unfounded ([48,49], para 3). Instead, the data paints a nuanced picture of how the spiral of silence manifests on social media in the face of social media fact-checking. First, there isn't much silencing regarding the effects of the social media fact-checks. For instance, opinion congruence, a key component of the spiral of silence, did not play much of a role in silencing the respondents. FOI, another key component of the spiral of silence theory, was more effective, but even then both components of the theory were subject to how frequently someone used social media.

Limitations

One limitation is the use of self-reported responses, which are susceptible to desirability responses and other biases [50,51]. For instance, no respondent reported that they had experienced an account suspension or ban, which is curious among a large random sample as the one used here. Additionally, even though this study used a survey with an embedded stimulus, a full experimental design may have detected more effects than those reported here. Also, data were collected during the COVID-19 pandemic, a period that has seen a deluge of COVID-19-related misinformation. Respondents were asked generally about fact-checked information but not about a particular issue. It is possible that some issues are more likely to be shared than others.

Funding

College of Liberal Arts, University of Texas Rio Grande Valley

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Ben Wasike reports financial support was provided by The University of Texas Rio Grande Valley.

Data availability

Data will be made available on request.

References

[1] Zuckerberg, M. (2016, November 8). *No title.* Meta. https://www.facebook.com/zuck/posts/a-lot-of-you-have-asked-what-were-doing-about-misinformation-so-i-wanted-to-give/10103269806149061/.

[2] Conklin, A. (2020, July 13). How Facebook fact checks work. Fox Business. https://www.foxbusiness.com/technology/how-facebook-fact-checks-work.

[3] L. Graves, M. Amazeen, Fact-checking as idea and practice in journalism, Oxford Res. Encycl. Commun. (2019). Retrieved 15 Nov from, https://oxfordre.com/communication/view/10.1093/acrefore/9780190228613.001.0001/acrefore-9780190228613-e-808.

[4] M. Dobbs, The rise of political fact-checking how Reagan inspired a journalistic movement: a reporter's eye view, New Am. Found. (2012). https://www.issuelab.org/resources/15318/15318.pdf.

[5] Reuters, Facebook removes seven million posts for sharing false information on coronavirus. Reuters, NBC News (2020). https://www.nbcnews.com/tech/tech-news/facebook-removes-seven-million-posts-sharing-false-information-coronav-rcna77.

[6] C.M. Marcos, YouTube has removed more than 500,000 COVID-19 misinformation videos since February, USA Today (2021). https://www.usatoday.com/story/tech/2021/01/27/youtube-removed-half-million-covid-19-misinformation-videos/4278489001/.

[7] D. O'Sullivan, Twitter Deletes China-linked accounts That Spread False Information About Hong Kong and Covid-19, June 12, CNN Business, 2020, https://www.cnn.com/2020/06/11/tech/twitter-manipulation-account-removal/index.html.

[8] E. Stoycheff, Under surveillance: examining Facebook's spiral of silence effects in the wake of NSA internet monitoring, J. Mass Commun. Q. 93 (2) (2016) 296–311, https://doi.org/10.1177/1077699016630255.

[9] K. Clayton, S. Blair, J.A. Busam, S. Forstner, J. Glance, G. Green, A. Kawata, A. Kovvuri, J. Martin, E. Morgan, M. Sandhu, R. Sang, R. Scholz-Bright, A.T. Welch, A.G. Wolff, A. Zhou, B. Nyhan, Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media, Polit. Behav. 42 (2020) 1073–1095, https://doi.org/10.1007/s11109-019-09533-0.

[10] C. Thorbecke, What to know about Twitter's fact-checking labels, ABC News (2020). https://abcnews.go.com/Business/twitters-fact-checking-labels/story?id=70903715.

[11] C. Dickey, The rise and fall of facts, Columbia J. Rev. (2019). https://www.cjr.org/special_report/rise-and-fall-of-fact-checking.php.

[12] A. Bovet, H.A. Maske, Influence of fake news in Twitter during the 2016 US presidential election, Nat. Commun. 10 (7) (2019) 1–13, https://doi.org/10.1038/s41467-018-07761-2.

[13] Meta, How Facebook's third-party fact-checking program works, Meta (2021). https://www.facebook.com/journalismproject/programs/third-party-fact-checking/how-it-works.

[14] Meta. (n.d.). About fact-checking on Facebook. *Meta.* https://www.facebook.com/business/help/2593586717571940?id=673052479947730.

[15] Poynter Institute. (n.d.). What is the international fact-checking network? https://www.poynter.org/ifcn/.

[16] IFCN (2021). About. Poynter Institute. https://ifcncodeofprinciples.poynter.org/signatories.

[17] Conger, K. (2022). How Twitter will change as a private company. The New York Times. https://www.nytimes.com/2022/10/28/technology/twitter-changes.html.

[18] Culliford, E., & Paul, K. (2020, May 30). *Reuters.* https://www.reuters.com/article/us-twitter-factcheck/with-fact-checks-twitter-takes-on-a-new-kind-of-task-idUSKBN2360U0.

[19] YouTube, How Does YouTube address misinformation? (n.d.) YouTube, 2024, https://www.youtube.com/howyoutubeworks/our-commitments/fighting-misinformation/.

[20] E. Noelle-Neumann, Turbulences in the climate of opinion: methodological applications of the spiral of silence theory, The Public Opin. Q. 41 (2) (1977) 143–158, https://doi.org/10.1086/268371.

[21] E. Noelle-Neumann, The Spiral of silence: Public Opinion - our social Skin, 2nd. Ed., The University of Chicago Press, Chicago, 1984.

[22] C.J. Glynn, M.E. Huge, Speaking in spirals: an updated meta-analysis of the spiral of silence (Eds.), in: W. Donsbach, C.T. Salmon, Y. Tsfati (Eds.), The Spiral of silence: New perspectives On Communication and Public Opinion, Routledge, New York, NY, 2014, pp. 65–72.

[23] J. Matthes, J. Knoll, C. von Sikorski, The "spiral of silence" revisited: a meta-analysis on the relationship between perceptions of opinion support and political opinion expression, Communic. Res. 45 (1) (2018) 3–33, https://doi.org/10.1177/0093650217745429.

[24] K. Hampton, L.R. Raine, W. Lu, M. Dwyer, I. Shin, K. Purcell, Social media and the 'spiral of silence', Pew Res. Center (2014). https://www.pewresearch.org/internet/2014/08/26/social-media-and-the-spiral-of-silence/.

[25] A.F. Hayes, Willingness to Self-Censor: a construct and measurement tool for public opinion research, Int. J. Public. Opin. Res. 17 (3) (2005) 298–323, https://doi.org/10.1093/ijpor/edh073.

[26] K. Neuwirth, E. Frederick, C. Mayo, The spiral of silence and fear of isolation, J. Commun. 57 (3) (2007) 450–468, https://doi.org/10.1111/j.1460-2466.2007.00352.x.

[27] J. Fox, L.F. Holt, Fear of isolation and perceived affordances: the spiral of silence on social networking sites regarding police discrimination, Mass Commun. Soc. 21 (5) (2018) 533–554, https://doi.org/10.1080/15205436.2018.1442480.

[28] S. Gearhart, W. Zhang, Gay bullying and online opinion expression: testing spiral of silence in the social media environment, Soc. Sci. Comput. Rev. 32 (1) (2014) 18–36, https://doi.org/10.1177/0894439313504261.

[29] M.J. Kushin, M. Yamamoto, F. Dalisay, Societal majority, Facebook, and the spiral of silence in the 2016 US presidential election, Social Media + Society 5 (2) (2019), https://doi.org/10.1177/2056305119855139, 205630511985513–.

[30] B. Nyhan, J. Reifler, The effect of fact-checking on elites: a field experiment on U.S. state legislators, am j pol sci 59 (3) (2015) 628–640.

[31] A. Agadjanian, N. Bakhru, V. Chi, D. Greenberg, B. Hollander, A. Hurt, J. Kind, R. Lu, A. Ma, B. Nyhan, D. Pham, M. Qian, M. Tan, C. Wang, A. Wasdahl, A. Woodruff, Counting the Pinocchios: the effect of summary fact-checking data on perceived accuracy and favorability of politicians, Res. Polit. 6 (3) (2019), https://doi.org/10.1177/2053168019870351.

[32] B. Nyhan, E. Porter, J. Reifler, T.J. Wood, Taking fact-checks literally but not seriously? The effects of journalistic fact-checking on factual beliefs and candidate favorability, Polit. Behav. 42 (3) (2020) 939–960, https://doi.org/10.1007/s11109-019-09528-x.

[33] C. York, J.D. Ponder, Z. Humphries, C. Goodall, M. Beam, C. Winters, Effects of fact-checking political misinformation on perceptual accuracy and epistemic political efficacy, J. Mass Commun. Q. 97 (4) (2020) 958–980, https://doi.org/10.1177/1077699019890119.

[34] N. Walter, J. Cohen, R.L. Holbert, Y. Morag, Fact-checking: a meta-analysis of what works and for whom, Polit. Commun. 37 (3) (2020) 350–375, https://doi.org/10.1080/10584609.2019.1668894.

[35] B. Nyhan, Does fact-checking work? False statements are wrong metric, Columbia J. Rev. (2012). https://archives.cjr.org/united_states_project/does_fact-checking_work_false.php.

[36] U.K.H. Ecker, Z. O'Reilly, J.S. Reid, E.P Chang, The effectiveness of short-format refutational fact-checks, The Br. J. Psychol. 111 (1) (2020) 36–54, https://doi.org/10.1111/bjop.12383.

[37] A. Oeldorf-Hirsch, M. Schmierbach, A. Appelman, M.P. Boyle, The ineffectiveness of fact-checking labels on news memes and articles, Mass Commun. Soc. 23 (5) (2020) 682–704, https://doi.org/10.1080/15205436.2020.1733613.

[38] M. Chung, N. Kim, When I learn the news is false: how fact-checking information stems the spread of fake news via third-person perception, Hum. Commun. Res. 47 (1) (2021) 1–24, https://doi.org/10.1093/hcr/hqaa010.

[39] A. Ardèvol-Abreu, P. Delponti, C. Rodríguez-Wangüemert, Intentional or inadvertent fake news sharing? Fact-checking warnings and users' interaction with social media content, El Profesional de La Informacion (5) (2020) 29, https://doi.org/10.3145/epi.2020.sep.07.

[40] M. Walker, J. Gottfried, Republicans far more likely than Democrats to say fact-checkers tend to favor one side, Pew Res. Center (2019). https://www.pewresearch.org/fact-tank/2019/06/27/republicans-far-more-likely-than-democrats-to-say-fact-checkers-tend-to-favor-one-side/.

[41] K. Macnish, Just surveillance? Towards a normative theory of surveillance, Surveill. Soc. 12 (1) (2014) 142–153, https://doi.org/10.24908/ss.v12i1.4515.

[42] G.T. Marx, What's new about the "new surveillance"? Classifying for change and continuity, Knowledge, Technol. Pol. 17 (1) (2004) 18–37, https://doi.org/10.1007/BF02687074.

[43] R. Clarke, Information technology and dataveillance, Commun. ACM 31 (5) (1988) 498–512, https://doi.org/10.1145/42411.42413.

[44] N. Confessore, Cambridge Analytica and Facebook: the scandal and the fallout so far, New York Times (2018). https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html.

[45] Black, I. (2013, June 10). NSA spying scandal: what we have learned. *The Guardian.* https://www.theguardian.com/world/2013/jun/10/nsa-spying-scandal-what-we-have-learned.

[46] H. Gil de Zúñiga, M. Barnidge, A. Scherman, Social media social capital, offline social capital, and citizenship: exploring asymmetrical social capital effects, Polit. Commun. 34 (1) (2017) 44–68, https://doi.org/10.1080/10584609.2016.1227000.

[47] N. Kalmoe, J. Gubler, D. Wood, Toward conflict or compromise? how violent metaphors polarize partisan issue attitudes, Polit. Commun. 35 (3) (2018) 333–352, https://doi.org/10.1080/10584609.2017.1341965.

[48] Barrett, P.M. & Sims, J.G. (2021). False accusation: the unfounded claim that social media companies censor conservatives. https://static1.squarespace.com/static/5

b6df958f8370af3217d4178/t/6011e68dec2c7013d3caf3cb/1611785871154/NYU±False±Accusation±report_FINAL.pdf.

[49] E.A. Vogels, A. Perrin, M. Anderson, Most Americans think social media sites censor political viewpoints, Pew Res. Center (2020). https://www.pewresearch.org/internet/2020/08/19/most-americans-think-social-media-sites-censor-political-viewpoints/.

[50] M. Devaux, F. Sassi, Social disparities in hazardous alcohol use: self-report bias may lead to incorrect estimates, Eur. J. Public Health 26 (1) (2016) 129–134, https://doi.org/10.1093/eurpub/ckv190.

[51] S. Müller, M. Moshagen, Controlling for response bias in self-ratings of personality: a comparison of impression management scales and the overclaiming technique, J. Pers. Assess. 101 (3) (2019) 229–236, https://doi.org/10.1080/00223891.2018.1451870.

Ben Wasike is a professor in the Communication Department at the University of Texas Rio Grande Valley. He teaches classes in visual communication, communication theory, and research methods, among others. His-research focusses on social media communication, political communication, visual communication, and law and policy issues.