Popups, Credibility, and the Spread of Misinformation

Hollis Greenberg
Wentworth Institute of Technology
greenbergh1@wit.edu

Laurie Dringus
Nova Southeastern University
laurie@nova.edu

Junping Sun
Nova Southeastern University
jps@nova.edu

Ling Wang
Nova Southeastern University
lingwang@nova.edu

ABSTRACT

The desire to "debunk" false information is a shared goal among government, social media platforms, and society. How can organizations determine whether a debunking technique is effective? Noting that simply debunking misinformation may not affect its spread, the user actions (clicking or not clicking on like, share, or comment) on postings need to be studied independently from whether a user finds a posting credible. New debunking techniques are routinely adopted social media platforms. Although popup warnings have been available for quite some time, these warnings have been newly repurposed by some platforms as a debunking tool. As such, this study focused on the effects of message popup warnings on credibility and effectiveness (user actions or inactions).

CCS CONCEPTS

• **Human-centered computing**; • **Collaborative and social computing**; • **Collaborative and social computing systems and tools**; • **Social networking sites**;

KEYWORDS

Misinformation, Disinformation, Fake news, Popup warnings, Effectiveness, Credibility, Accuracy, Debunking techniques

ACM Reference Format:
Hollis Greenberg, Laurie Dringus, Junping Sun, and Ling Wang. 2024. Popups, Credibility, and the Spread of Misinformation. In *The 25th Annual Conference on Information Technology Education (SIGITE '24), October 10–12, 2024, El Paso, TX, USA*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3686852.3687065

1 INTRODUCTION

Misinformation is not a new phenomenon. Throughout time, society has labeled false news as a hoax, propaganda, and fake news. More recently, social media platforms have given misinformation the ability to cast its net far and wide with a click of a button. Society, governments, and social media platforms have struggled with how to debunk and curb the spread of misinformation. Specifically, social media platforms provide users with software functionality to repost or share information, and it is this functionality that may inadvertently spread misinformation to other users [1].

There are two types of false news: misinformation and disinformation. The term misinformation is used to describe non-accurate postings, while the term disinformation implies that the false information was intentional [2, 3]. This study did not focus on intent, rather all false postings were classified as misinformation.

Researchers have faced various limitations when studying the spread of misinformation in a controlled environment. Two of the limitations at play are 1.) there is no common terminology in this field and 2.) there is a lack of interactive interfaces in which to study users' actions. The first limitation, lack of common terminology, affects and confuses readers of extant literature. Terminology, such as effectiveness and credibility, are interchangeable in multiple studies [3]. However, credibility infers the belief or lack of belief in the truthfulness of a posting, while effectiveness studies the user's action or inaction with the posting. The second limitation, lack of interactive interfaces, limits researchers from studying user actions. Additionally, studies have not provided functionally to allow access to the full content of the postings. Existing studies provide only headlines and pictures to determine the credibility of the posting [3, 4]. In this study, effectiveness was defined as the influence of user actions; effectiveness evaluates the user choice response (clicking or not clicking on like, share, or comment) for each posting.

Debunking techniques were developed to deter the spread of misinformation. Social media platforms have employed the following debunking techniques: warning screens covering the posting, labels or warnings, hyperlinks providing fact-checking information, algorithms that move false postings lower in users' newsfeeds, turning off functionality to reply or share, reducing posting's visibility, removing a disputed posting, or providing a message popup warning when a user tries to take action (clicking on like, share, or comment) on a posting [5, 6]. This study focused on the understudied technique of message popup warnings. See Figure 1 for an example of a posting presenting with a message popup warning.

Popup warnings have been around for quite some time, yet not used to combat misinformation until more recently. Historically, popup warnings have served several functions: to interrupt and force a user to interact [7] and to slow down and distract users [8]. Research has shown that message popup warnings may be a useful debunking technique [9]. However, the wording within the warning needs to be crafted carefully to elicit the desired response [9].

This study was developed to answer the overarching research question of: Are message popup warnings effective for deterring the

SIGITE '24, October 10–12, 2024, El Paso, TX, USA

ACM ISBN 979-8-4007-1106-0/24/10
https://doi.org/10.1145/3686852.3687065

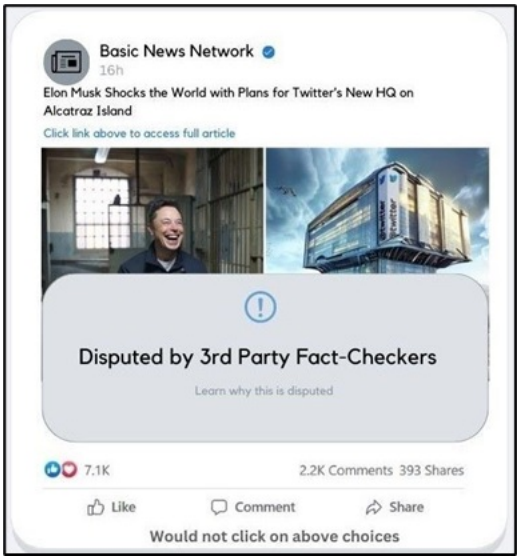

Figure 1: False News Posting with Message Popup Warning, Photograph by ©Architizer. (https://architizer.com/blog/inspiration/stories/elon-musk-twitter-hq-alcatraz-island)

spread of misinformation? This question focuses on the user actions separately from the user's belief of the accuracy of each posting. In some instances, users have chosen to spread misinformation even when they believed the posting to be false [10].

2 RELATED RESEARCH

Previous studies have co-mingled the terms of credibility and effectiveness [11]. There has been a generalized belief that if a posting can be debunked (and users convinced that it is untrue) that the disbelief in the posting would deter the spread of misinformation. Debunking research has tended to focus on the change of users' beliefs, attitudes, and preferences, not on deterring the spread of false news [3, 12]. Various debunking techniques have been found to be effective when users doubt the reliability of the posting [12]. Other researchers believe in *pre-bunking*, as a type of misinformation "vaccination", by presenting users with labeled false news before they inadvertently come across it on their own [13].

Participants in misinformation studies are provided with limited information. Users are often given a picture, headline, and fake or missing news source identifiers to evaluate whether to believe or share the posting with others. Pennycook et al. (2020) suggested that debunking research should provide users with more information than a story headline [4]. Links to full posting content would allow users to assess posting credibility on their own. Real social media platforms provide much more information for user evaluation, such as headlines, full posting content, the poster, and a variety of information provided by debunking techniques [3, 14].

To date, message popup warnings as a debunking technique have only been evaluated in static surveys or through interviews [3, 4, 12, 14-16]. Effectiveness, defined as the user's choice response (clicking or not clicking on like, share, or comment) to a social media posting, is difficult to capture without an interactive survey or simulation. Consequently, there is little research on the effectiveness of message popup warnings. Alternatively, interactive simulations allow for the tracking of mouse clicks and allow researchers to compare user behavior to user perceptions [2].

Only a few studies recognize that users may choose to share false postings even when they know the postings to be false. [9]. Studies need to focus on debunking tools that correct false information, aid users to recognize misinformation, and deter the spread of these postings [14].

3 METHODOLOGY

In 2024, a study was launched to address several of the shortcomings from existing misinformation and debunking surveys. This study included an interactive survey, the ISS (Interactive Scenario and Survey), to allow participants to fully engage by reading the full content of the postings, rate the credibility of the posting, and give participants the ability to make a choice response (click or not click on like, share, or comment) for each posting [10]. This study extended studies by Pennycook et al. (2020) and Kirchner and Reuter (2020). Like these studies, the ISS displayed postings in a Facebook-like format, with half of the postings true and other the half false. Unlike Pennycook et al.'s (2020) and Kirchner and Reuter's (2020) studies, the ISS was interactive; participants were given access to the full posting's content through a hyperlink, and participants could choose to click (or not click) on like, share, or comment on each posting. The question of credibility or accuracy was separate from the decision to like, share, or comment on the posting.

To evaluate credibility and effectiveness, a between-subjects experimental design was employed. A research question and hypothesis was crafted to assess each construct. The ISS was created to evaluate both the credibility and effectiveness of message popup warnings on false news postings. This instrument was tested for reliability and validity by instrumentation experts; functionality was also tested by the experts and a pilot test group [10]. Participants (N = 109) were divided into two groups: control and treatment. The control group (n = 54) randomly received 12 true postings and 12 false postings all without message popup warnings [10]. The treatment group (n = 55) randomly received 12 true postings, six false postings with a message popup warning, and six false postings without a message popup warning [10]. The same postings (articles) were used in both groups.

4 FINDINGS

Focusing on credibility, the following research question and hypotheses were crafted:

RQ1: Regarding credibility, how does the debunking technique of message popup warnings influence the users' credibility rating of the posting, compared to ratings given for postings without message popup warnings? [10]

H1: Regarding credibility, there will be no significant differences in how users rate the credibility of the posting when receiving the message popup warning, compared to ratings given for postings without message popup warnings [10].

To evaluate credibility, this study adapted Pennycook et al. (2020) and Kirchner and Reuter's (2020) accuracy prompt and 4-point

Table 1: Means, Standard Deviations, and One-Way Analyses of Variance for Credibility Among Posting Types

		Post-Treatment				Φ
		No Action		Action		
		N	%	*N*	%	
Pre-Treatment	No Action	169	78.6%	12	10.4%	0.653*
	Action	46	21.4%	103	89.6%	

* $p < .001$

Likert scale responses [3, 4]. After each posting, users were asked, "To the best of your knowledge, how accurate is this article?" and responded using a 4-point Likert scale (1="not at all accurate", 2="not very accurate", 3="somewhat accurate", 4="very accurate") [10].

For credibility, this study compared the same six false postings from the control and treatment groups; postings presented without message popup warnings for the control group and postings presented with a message popup warning for the treatment group. Within these groups, data was compared by averaging the scores and then performing an independent samples *t*-test. A significant difference, $t(107) = 2.46$, $p = .015$, was found between the control ($M = 2.33$, SD = .84) and treatment groups ($M = 1.99$, SD = .6) [10]. Results indicated that the effect size was medium with a Cohen's d of $d = 0.47$ [10]. Thus, H1 was rejected.

As the first statistical test (*t*-test) demonstrated a significant difference, further analysis was performed. The postings were broken down into five posting types: Control False (control group postings #1-12 with false postings with no warnings), Control True (control group postings #13-24 with true postings), Treatment False-Warning (treatment group postings #1-6 with false postings and message popup warnings), Treatment False-No Warning (treatment group postings #7-12 with false postings and no warnings), and Treatment True (treatment group postings #13-24 with true postings) [10]. Each posting type contained six or 12 postings and these user responses were then averaged. An ANOVA was performed to further analyze the differences. The one-way ANOVA demonstrated a statistically significant difference in credibility ratings between the five posting types ($F(4, 268) = 4.516$, $p = .002$), with a medium effect size ($\eta^2 < .06$) [10]. See Table 1.

Delving further, Tukey's HSD Test for multiple comparisons was used to determine which posting types were significantly different. The results determined that the mean value of credibility was significantly different between Treatment False-Warning and Treatment True ($p = .004$, 95% C.I. = -.761, -1.000) and Treatment False-Warning and Control True ($p = .003$, 95% C.I. = -.774, -0.111) [10].

4.1 Effectiveness

Turning the focus to effectiveness, the following research question and hypotheses were crafted:

RQ2: Regarding effectiveness, what is the user choice response (actions or inactions) to postings when comparing postings that are presented with or without message popup warnings? Note: actions (1=liking, sharing, or commenting on a false posting) or inactions (0=not liking, sharing, or commenting on a false posting) [10].

H2: Regarding effectiveness, there will be no significant differences in user choice response (actions or inactions) to postings, when comparing postings that are presented with or without message popup warnings [10].

The interactive interface of the ISS was used to capture user choice responses, a/k/a effectiveness. The effectiveness of message popup warnings was determined by the user response to the prompt, "If you were to see this article on Facebook or another social media platform, would you consider liking, sharing, or commenting on it?" [4, 10]. Effectiveness data was collected as the participant chose (clicked on) one of the following user options: like, share, comment, or would not click on above choices.

To analyze effectiveness, the same six false postings were compared in the control (false postings with no warnings) and treatment (false postings with popup warnings) groups, similar to credibility. Data was analyzed as a binary variable. If the participant chose to like, share, or comment on the posting, the response equaled one. If the participant chose to not click on like, share, or comment (clicked on "would not click on above choices"), the response equaled zero. A Chi-square test of independence was employed and found that the relationship between message popup warnings and effectiveness was not significant and had a negligible association, $\chi 2$ (1, 654) = .03, $p = .863$, $\Phi = -.007$ [10]. Thus, H2 was failed to reject. These results indicate that the participants in neither the control or treatment groups were more likely to click on like, share, or comment than participants in the other group.

Further analysis was necessary to determine whether message popup warnings were effective. The ISS posed the effectiveness prompt a single time for all postings in posting types, except for the posting type of Treatment False-Warning. Within the Treatment False-Warning posting type, participants first viewed the false posting without a warning and were prompted to answer the effectiveness question. Once the prompt was answered, a message popup warning appeared and then the participants were re-prompted ("Regarding the previous posting, would you reconsider liking, sharing, or commenting on it?" [4, 10]) to again answer the effectiveness question. A McNemar's test was used to compare the pre- and post-test effectiveness questions. These results displayed a significant difference, $p < .001$, in the user responses between the pre- and post-test questions. See Table 2. Additionally, a strong association ($\Phi = .653$) between message popup warnings and effectiveness was found [17].

Table 2: Crosstabulation and McNemar's Test of Effectiveness Comparing Pre- and Post-Test Results

Posting Type	*N*	Mean	SD	Mean Square	$F(4, 268)$	η^2
Between Groups				1.796	4.516*	0.063
Treatment False-Warning	55	1.991	0.601			
Treatment False-No Warning	55	2.264	0.607			
Treatment True	55	2.421	0.651			
Control False	54	2.208	0.697			
Control True	54	2.434	0.590			
Within Groups				0.398		

* $p < .01$

Table 3: Crosstabulation and Chi-Square Results for Effectiveness for All Posting Types

		No Action (*N*)	Action (*N*)	$\chi 2$	Φ
Treatment False-No Warning	No Action	145	43	27.600*	0.289
	Action	70	72		
Treatment True	No Action	126	37	20.939*	0.252
	Action	89	78		
Control False	No Action	143	70	1.042	0.056
	Action	72	45		
Control True	No Action	115	59	.143	0.021
	Action	100	56		

* $p < .01$

Table 4: Frequencies for Effectiveness by Posting Type

	No Action		Action	
Posting Type	*N*	%	*N*	%
Treatment False-Warning	215	65.2%	115	34.8%*
Treatment False-No Warning	188	57.0%	142	43.0%*
Treatment True	338	51.2%	322	48.8%*
Control False	414	63.9%	234	36.1%
Control True	344	53.1%	304	46.9%

* $p < .01$

A final series of Chi-Square tests were run to compare effectiveness by posting type. The analysis revealed that there was a significant difference ($p < .01$) of user responses between all the treatment posting types (Treatment False-Warning, Treatment False-No Warning, Treatment True) [10]. See Table 3.

Additionally, Table 4 further demonstrates the significant differences in user choice responses between the treatment posting types ($p < .01$) [10]. Comparatively, no significant difference was shown between the Treatment False-Warning posting type and the Control False or Control True postings ($p = 0.307$ and $p = 0.705$, respectively) [10]. These results suggest that when message popup warnings display on some postings, the user choice responses are impacted on all postings.

* $p < .01$

5 DISCUSSION AND FUTURE WORK

Data from the ISS provided additional insights into the credibility and effectiveness of message popup warnings on social media postings.

5.1 Credibility

For the construct of credibility, the initial *t*-test determined that there was a statistically significant difference ($p = .015$) between credibility ratings for the same six false postings in the control group and the treatment group. This data corroborated the findings in Ardèvol-Abreu et al.'s (2020) qualitative study, where users doubted the accuracy of a posting when a popup warning appears [9]. Other studies have experienced similar outcomes with warning labels [3, 4, 18].

Further analysis with ANOVA and Tukey's HSD tests found significant differences ($p < .01$) between the Treatment False-Warning posting type and both Treatment True and Control True posting

types [10]. This result is important as it demonstrates that the popup warning on a false posting is more credible than a true posting without confirmation or warning. In this study, participants chose to trust the warning label more than their own judgment [10].

It is important to note that this study did not use political postings, thereby avoiding the real-world issue of pre-existing beliefs. Individuals tend to trust postings that confirm their beliefs [19]. Without pre-existing beliefs present, the message popup warnings appear to influence the posting's credibility rating [10]. It is unknown whether message popup warnings would be effective in negating previously held beliefs.

5.2 Effectiveness

Regarding effectiveness, the initial Chi-square tests between the same six false postings in the control group and the treatment group did not find statistically significant differences ($p > .05$) [10]. The only difference among the postings was that the treatment group was presented with a message popup warning on their postings and the control group did not receive any warnings. Since each of the postings were false, one could infer that the participants were able to deduce that the false postings were indeed false. Alternatively, the participants may not have been interested in these non-political postings and chose to not take action (click on like, share, or comment).

Delving further, data demonstrated a significant difference ($p < .01$) in user actions between the treatment posting types. The data suggests that when message popup warnings are present, users are less likely to take action on known false postings than on postings with unknown accuracy. This analysis is in line with the same findings from Pennycook et al.'s (2020) study.

More insights were found when analyzing the pre- and post-test results. By comparing the pre- and post-treatment results for the treatment group, it resulted in a significant difference ($p < .001$) between these effectiveness prompts [10]. This analysis clearly demonstrates that message popup warnings affected the user choice response to take action or not take action. Data gathered for pre-warning and post-warning must be collected through an interactive scenario. The ISS was able to gather the user responses to the posting, both before and after the appearance of the popup warning.

5.3 Limitations

Although the interactive nature of the survey and simulation was a success, this study faced several limitations:

- The sample used in the study consisted primarily of college students of typical college age. As the institution is STEM focused, many of the respondents were male-identifying. Furthermore, the number of students who qualified to take the survey by passing the pre-screening questions (Students needed to answer in the affirmative to both questions: 1.) "Do you have a social media account? (Binary answer: Yes or No)" and 2.) "Have you liked, shared, or commented on a news article on social media? (Binary answer: Yes or No)") was a small sample of 109 [10].
- A main testing effect may have influenced the participants answers to the effectiveness prompt [10]. Participants in the treatment group were asked the effectiveness question twice, once before the message popup warning and a second time after the warning was presented.
- Facebook was used as the social media platform for the posting template. This was an unexpected limitation and weakness as many of the participants preferred to use other social media platforms and were less prone to like, share, or comment on Facebook.
- All postings were non-political. While avoiding the effects of pre-existing beliefs, participants may not have been interested in all the survey's postings.

5.4 Future Research

There is ample opportunity to extend research on misinformation and debunking techniques. First, the ISS should be extended to a more diverse (age and otherwise) group of participants. Older participants may react differently to false postings or the Facebook-like platform. Different groups of people may respond in various ways to false postings as well. Second, other social media platforms should be used as separate treatment groups to compare the platform's effect on user choice behaviors and credibility ratings. Third, this study coded user choice responses with a binary scale ("like", "share", and "comment" were equal to one and "wouldn't click on above choices" was equal to zero.) By separating out the different responses, those results would add to the body of knowledge, as like, share, and comment trigger non-equal rates of propagation to others' newsfeeds [20]. Fourth, researchers could study different messaging on the popup warning. The messaging itself may affect the user's perception of credibility and user actions (effectiveness). Finally, research could explore a possible correlation between message popup warning's credibility and effectiveness.

REFERENCES

[1] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," Journal of Economic Perspectives vol. 31, no. 2, pp. 211-36, 2017.
[2] C. Geeng, S. Yee, and F. Roesner, "Fake news on Facebook and Twitter: Investigating how people (don't) investigate," Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1-14, 2020.
[3] J. Kirchner and C. Reuter, "Countering fake news: A comparison of possible solutions regarding user acceptance and effectiveness," Proceedings of ACM Human-Computer Interaction, vol. 4, no. CSCW2, p. Article 140, 2020, doi: 10.1145/3415211.
[4] G. Pennycook, A. Bear, E. T. Collins, and D. G. Rand, "The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings," Management Science, vol. 66, no. 11, pp. 4944-4957, 2020.
[5] Meta. "Transparency Center policy details." Facebook. https://transparency.fb.com/policies/community-standards/false-news/ (accessed June 19, 2022, 2022).
[6] Twitter. "Civic integrity policy." https://help.twitter.com/en/rules-and-policies/election-integrity-policy (accessed June 19, 2022, 2022).
[7] R. Wash and C. Lampe, "The power of the ask in social media," Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, pp. 1187–1190, 2012, doi: 10.1145/2145204.2145381.
[8] E. Abdulin and D. Billman, "Confirmation responses: In-context, visible, predictable design versus popup windows," Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, pp. 2969–2975, 2016, doi: 10.1145/2851581.2892403.
[9] Ardèvol-Abreu, P. Delponti, and C. Rodríguez-Wangüemert, "Intentional or inadvertent fake news sharing? Fact-checking warnings and users' interaction with social media content," Profesional de la Información, vol. 29, no. 5, 2020.
[10] H. Greenberg, "Heed the Warning Signs: The Effectiveness of Message Popup Warnings for Deterring the Spread of Misinformation," NSUWorks, vol. 1192,

2024.

[11] J. J. Argo and K. J. Main, "Meta-analyses of the effectiveness of warning labels," Journal of Public Policy & Marketing, research-article vol. 23, no. 2, pp. 193-208, 10/01/ 2004. [Online]. Available: https://ezproxywit.flo.org/login?url = https://search.ebscohost.com/login.aspx?direct$=$true&db$=$edsjsr&AN$= $edsjsr.30000760&site$=$eds-live&scope$=$site.

[12] J. Colliander, ""This is fake news": Investigating the role of conformity to other users' views when commenting on and spreading disinformation in social media," Computers in Human Behavior, vol. 97, pp. 202-215, 2019/08/01/ 2019, doi: 10.1016/j.chb.2019.03.032.

[13] S. Van der Linden, Foolproof: Why misinformation infects our minds and how to build immunity. WW Norton & Company, 2023.

[14] P. L. Moravec, A. Kim, and A. R. Dennis, "Appealing to sense and sensibility: System 1 and system 2 interventions for fake news on social media," Information Systems Research, vol. 31, no. 3, pp. 987-1006, 2020.

[15] M. Gao, Z. Xiao, K. Karahalios, and W.-T. Fu, "To label or not to label: The effect of stance and credibility labels on readers' selection and perception of news articles," Proceedings of the ACM Human-Computer Interaction, vol. 2, no. CSCW, p. Article 55, 2018, doi: 10.1145/3274324.

[16] J. Fan, M. Shao, Y. Li, and J. Wang, "The impact of different design characteristics of warning message on users' perceived usefulness and perceived effectiveness of health information searching," Proceedings of the 15th International Conference on Service Systems and Service Management (ICSSSM), pp. 1-6, 2018.

[17] L. M. Rea and R. A. Parker, Designing and conducting survey research: A comprehensive guide. John Wiley & Sons, 2014.

[18] T. Koch, L. Frischlich, and E. Lermer, "The effects of warning labels and social endorsement cues on credibility perceptions of and engagement intentions with fake news [Preprint]," PsyArXiv, 2021.

[19] F. Spezzano, A. Shrestha, J. A. Fails, and B. W. Stone, "That's fake news! Reliability of news when provided title, image, source bias," Proceedings of ACM Human-Computer Interaction, vol. 5, no. CSCW1, p. Article 109, 2021, doi: 10.1145/3449183.

[20] Kim and S.-U. Yang, "Like, comment, and share on Facebook: How each behavior differs from the other," Public relations review, vol. 43, no. 2, pp. 441-449, 2017.