



Research Article

Misinformation more likely to use non-specific authority references: Twitter analysis of two COVID-19 myths

This research examines the content, timing, and spread of COVID-19 misinformation and subsequent debunking efforts for two COVID-19 myths. COVID-19 misinformation tweets included more non-specific authority references (e.g., “Taiwanese experts”, “a doctor friend”), while debunking tweets included more specific and verifiable authority references (e.g., the CDC, the World Health Organization, Snopes). Findings illustrate a delayed debunking response to COVID-19 misinformation, as it took seven days for debunking tweets to match the quantity of misinformation tweets. The use of non-specific authority references in tweets was associated with decreased tweet engagement, suggesting the importance of citing specific sources when refuting health misinformation.

Authors: Joseph McGlynn (1), Maxim Baryshevsev (2), Zane A. Dayton (3)

Affiliations: (1) Department of Communication Studies, University of North Texas, USA, (2) Department of Communication Studies, The University of Texas at Austin, USA, (3) Department of Communication, University of Kentucky, USA

How to cite: McGlynn, J., Baryshevsev, M., & Dayton, Z. A. (2020). Misinformation more likely to use non-specific authority references: Twitter analysis of two COVID-19 myths. *Harvard Kennedy School (HKS) Misinformation Review*, 1(3).

Received: June 4th, 2020. Accepted: August 1st, 2020. Published: September 4th, 2020.

Research questions

- How did COVID-19 misinformation tweets and debunking tweets differ in their use of authority references?
- How long did it take COVID-19 debunking tweets to match the quantity of COVID-19 misinformation tweets?
- How did the use of authority references relate to the spread of COVID-19 misinformation?

Essay summary

- Twitter posts related to COVID-19 misinformation were collected, coded, and analyzed.
- Misinformation tweets received a head start that provided early momentum. It took seven days for debunking tweets to equal the quantity of misinformation tweets, a delay that provided opportunity for misinformation to spread on social networks. After the inflection point at seven days, the number of debunking tweets accelerated and outpaced misinformation.
- The use of authority references in COVID-19 tweets related to user engagement. Tweets that

¹ A publication of the Shorenstein Center for Media, Politics and Public Policy, at Harvard University, John F. Kennedy School of Government.

included references to non-specific authority sources were less likely to receive likes or retweets from other users.

- Misinformation tweets were more likely to reference non-specific authority sources (e.g., “Taiwanese researchers”, “top medical officials”). Debunking tweets were more likely to reference specific or verifiable authority sources (e.g., the CDC, the World Health Organization, a Snopes article debunking the claim).
- This research suggests references to non-specific authority sources should present a warning sign to suspect misinformation. If a recommendation is attributed to a general non-specific authority source, users should heighten their suspicion for misinformation.

Implications

As coronavirus infections surged globally, people turned to social media to learn how to protect themselves from the emergent threat (Volkin, 2020). Social media is an important tool for communicating information during times of crisis (Austin et al., 2012; Houston et al., 2014). However, misinformation for COVID-19 emulated the trail of the virus and spread person-to-person through online social networks (Frenkel et al., 2020). Misinformation related to COVID-19 included misleading, false, and manipulated content (Brennen et al., 2020). Misleading information exacerbates health problems and increases the workload of health organizations who must exert resources to squashing the false claims (van der Meer & Jin, 2020).

The COVID-19 “infodemic” created a multi-layered problem for risk communication (Krause et al., 2020). In addition to the health risk of contracting COVID-19, searching online for information became a risk in itself through the possibility of encountering false claims (Krause et al., 2020). Twitter users often fail to detect rumors (Wang & Zhuang, 2018). Misinformation for infectious diseases has focused primarily on issues related to transmission, prevention, treatment, and vaccination (Nsoesie & Oladeji, 2020). The false claims in misinformation can produce negative health outcomes and lead people to take harmful actions in search of a remedy, such as gargling bleach in attempt to eliminate a virus.

Misinformation tweets and debunking tweets differed in their use of authority references. Misinformation tweets were more likely to attribute information to non-verifiable authority sources, such as “Taiwanese researchers” or “health experts.” Debunking efforts included more specific and verifiable authority appeals, such as providing a link to a fact-checking article from Snopes, or attributing the information to a specific organization, such as the CDC. In addition, debunking tweets were more likely to come from verified users. In our sample, 93 of 645 (14.4%) debunking tweets came from verified users, compared to seven of 577 (1.2%) misinformation tweets, indicating verified users were mostly not responsible for posting the misinformation. However, verified users often have larger follower counts and may create greater total impressions, or tweet views, due to the size of their following (Wang & Zhuang, 2017). During crisis events, users interact most frequently with verified users (Hunt et al., 2020). Future research should examine specifically the relationship between verified users, tweet timing, and total impressions during surging misinformation events.

Our results indicate that references to non-specific authority sources should act as a signal to increase one’s guard against potential misinformation attempts. Users should view general attributions to authority references as a guideline for raising their suspicion of the accuracy of the tweet, as misinformation tweets were more likely to include ambiguous, non-specific cues to authority references (e.g., “medical professionals”, “a nurse friend”). Intentional deliberation can reduce misinformation spread (Bago et al., 2020). To contrast these findings, research should examine the prevalence of non-specific and specific authority appeals in different health contexts. Current results suggest the appearance

of an ambiguous authority appeal in health messages should lead people to pause and deliberate on the accuracy of the claim.

The use of authority appeals in COVID-19 tweets related to user engagement. Misinformation tweets with non-specific authority references were less likely to receive engagement from other users in the form of likes or retweets. The inclusion of specific authority references did not increase or decrease the likelihood of tweet engagement by other users. Future work can advance these findings by examining categories of abstract and specific authority appeals and their effect on tweet engagement. Likes on social media influence credibility perceptions of health information (Borah & Xiao, 2018), and credibility can affect intentions to adopt healthy behaviors (Hu & Sundar, 2010). The inclusion of authority appeals may also influence credibility judgments of misinformation and debunking tweets. Source credibility is paramount when attempting to correct inaccurate health information (Bode & Vraga, 2018).

Debunking represents an important public strategy to reduce the persistence of misinformation (Chan et al., 2017). For debunking to be necessary, the misinformation must first occur. However, findings indicate that debunking tweets lagged behind misinformation by several days before accelerating in frequency. The delay between misinformation emergence and debunking response provided an opportunity for the misinformation to spread. The observed delay in debunking response is problematic because even minimally repetitive exposure to fake news can increase the perceived accuracy of the information (Pennycook et al., 2018). Once people believe misinformation, they tend to do so with strong conviction and correcting the information becomes more difficult (Lewandowsky et al., 2012).

Improving the efficacy of debunking efforts remains a priority (van der Meer & Jin, 2020). The current results indicate avoiding non-specific authority references is a good first step for debunking efforts, as non-specific authority references had a negative relationship with tweet engagement. In our sample, only 1.7% of debunking tweets included a non-specific authority appeal. Twitter users frequently provide website links to debunk false rumors (Hunt et al., 2020). Debunking efforts should focus on including links and appeals to specific authority references. Of note, the quantity of debunking efforts differed between the two myths analyzed. Specifically, users focused greater attention to debunking the gargle myth than the hold breath myth, perhaps because gargling liquids such as bleach presents a greater immediate threat to people's health. This result suggests the perceived danger of health misinformation may affect likelihood of debunking response. When perceived danger of the misinformation is high, people may be more motivated to debunk the myth.

Proactive communication is vital for the successful reduction of COVID-19 misinformation (Jamieson & Albarracín, 2020). Prompting users to ask why they believe the information to be true can be an effective gatekeeping device to reduce sharing of false claims (Fazio, 2020). However, misinformation is often blended with accurate information, making it difficult for users to discern myth from fact. For example, the claim that people should gargle a liquid to avoid COVID-19 was frequently embedded with other health recommendations, such as getting adequate sleep or exercise. Understanding the effects of messages with complete misinformation versus partial misinformation remains a priority for future research.

Users and organizations must keep in mind that misinformation campaigns related to COVID-19 are far from over. Rather, organizations should anticipate and expect misinformation attempts in regard to future COVID-19 issues such as vaccine development, testing, and treatment. A faster response time for debunking efforts will be critical for halting future misinformation efforts before they build momentum. Organizations should treat misinformation campaigns as a known threat, fully expecting for misinformation to spread during a crisis rather than being surprised by the surplus of myths and false claims. Anticipating the existence of misinformation attempts may hasten the efficiency and strengthen the efficacy of future debunking efforts.

Findings

Finding 1: Misinformation and debunking tweets included two types of authority appeals: non-specific authority references and specific authority references. Misinformation tweets included more appeals to non-specific authority references. Debunking tweets included more appeals to specific authority references.

Misinformation tweets and debunking tweets differed in their use of authority appeals. Two types of authority references appeared in the tweets, including appeals to non-specific authority references and appeals to specific authority references. For example, tweets could include a non-specific authority reference (e.g., “Taiwanese experts”, “top medical officials”) or a specific authority reference (e.g., the CDC, Factcheck.org). Misinformation tweets were significantly more likely than debunking tweets to include a non-specific authority reference (12.3% misinformation tweets vs. 1.7% debunking tweets). Debunking tweets were significantly more likely than misinformation tweets to include appeals to specific authority references (45.4% of debunking tweets vs. 9.5% of misinformation tweets), $\chi^2(2, n = 1222) = 218.84, p < .001$.

Misinformation tweets and debunking tweets also differed in how frequently they included an authority reference. Overall, debunking tweets were more likely to include an authority reference than misinformation tweets. Misinformation tweets used an authority appeal in 126 of 577 tweets (21.8%). Debunking tweets included an authority appeal in 304 of 645 tweets (47.1%), which is significantly more often ($\chi^2[1, n = 1222] = 85.44, p < .001$).

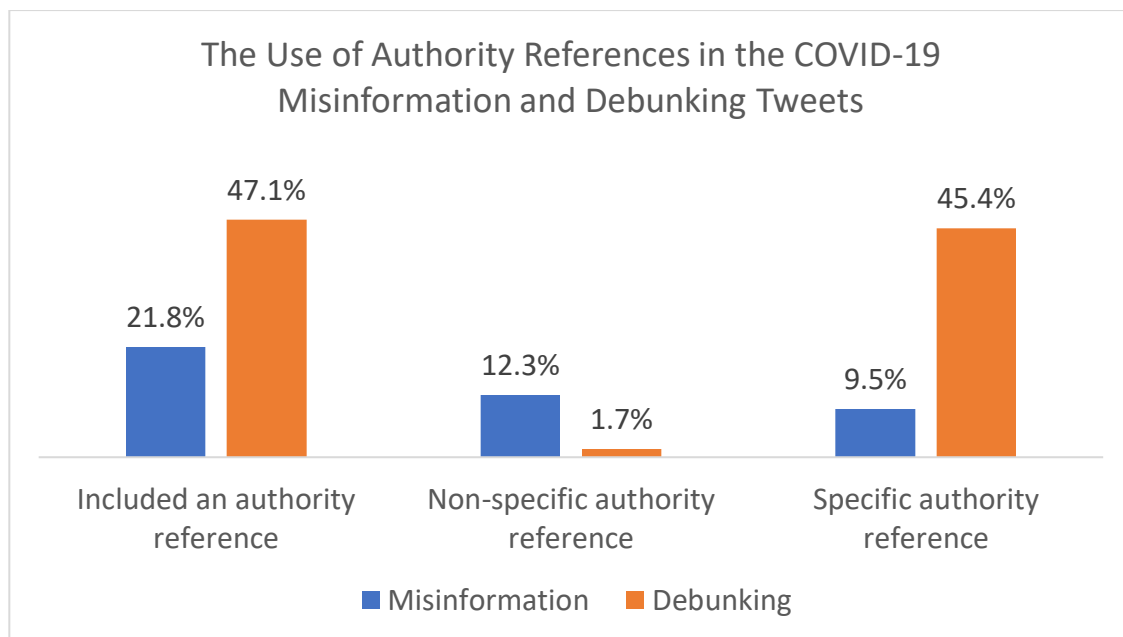


Figure 1. The use of authority references in the COVID-19 misinformation and debunking tweets. Debunking tweets were more likely than misinformation tweets to include an authority reference. Misinformation tweets were more likely than debunking tweets to include appeals to non-specific authority references. Debunking tweets were more likely than misinformation tweets to include appeals to specific authority references. Differences significant at $p < .001$.

Table 1. Descriptive statistics of the analyzed tweets

Variable	Misinformation tweets <i>n</i> = 577				Debunking tweets <i>n</i> = 645			
	Statistics							
Hold Breath	300	52.0%			199	30.9%		
Gargle	277	48.0%			446	69.1%		
Verified User	7	1.2%			93	14.4%		
Debunking (overt)	-	-			506	78.4%		
Debunking (suspicion)	-	-			139	21.6%		
Authority Appeal	126	21.8%			304	47.1%		
Non-specific appeal	71	12.3%			11	1.7%		
Specific appeal	55	9.5%			293	45.4%		
	<i>Mean</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>Mean</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
Followers Count	3,285	12,037	0	162,632	160,965	1,878,676	0	45,907,596
Following Count	1,782	4,513	0	52,634	1,505	2,939	0	28,216
Status Count	21,912	45004	2	423,842	47,300	158,953	2	3,401,865
Retweets	11	141	0	2,836	29	291	0	6375
Likes	30.5	374	0	8,270	90	1025	0	24320

Finding 2: Debunking tweets lagged behind misinformation tweets before accelerating in frequency. Seven days passed before debunking tweets matched the quantity of misinformation tweets.

Examining the two myths combined, the results indicate it took seven days for debunking tweets to match the quantity of COVID-19 misinformation tweets. Misinformation for COVID-19 began picking up speed on 3/7/20, while debunking tweets did not match the total frequency of misinformation tweets until seven days later on 3/14/20. After the inflection point, debunking tweets accelerated and reached similar levels to the peak of COVID-19 misinformation tweets. In our sample, peak levels of debunking trailed the misinformation peak by five days. After that point, misinformation and debunking tweets receded to similar levels. A cumulative analysis of misinformation and debunking for the two myths shows an acceleration of total misinformation on 3/11/20 and a gradual rise in debunking thereafter, with an increase in aggregate debunking between 3/13/20 and 3/19/20.

Examining the two myths separately, both cases illustrate a sharp initial ramp of misinformation, followed by debunking responses that trailed the misinformation attempts. Notably, despite our data including similar amounts of misinformation tweets for both the “hold breath” and “gargle” myths, debunking tweets focused greater effort on refuting the gargle myth. The cumulative analysis for the hold breath myth indicates total debunking efforts did not overtake misinformation for the hold breath myth at any point. In comparison, debunking tweets for the gargle myth rose sharply on 3/14/20 and exceeded the cumulative misinformation total for the gargle myth on 3/18/20. Figures two through seven illustrate the relative frequency of misinformation tweets and debunking tweets over time, from both a per day and a cumulative perspective.

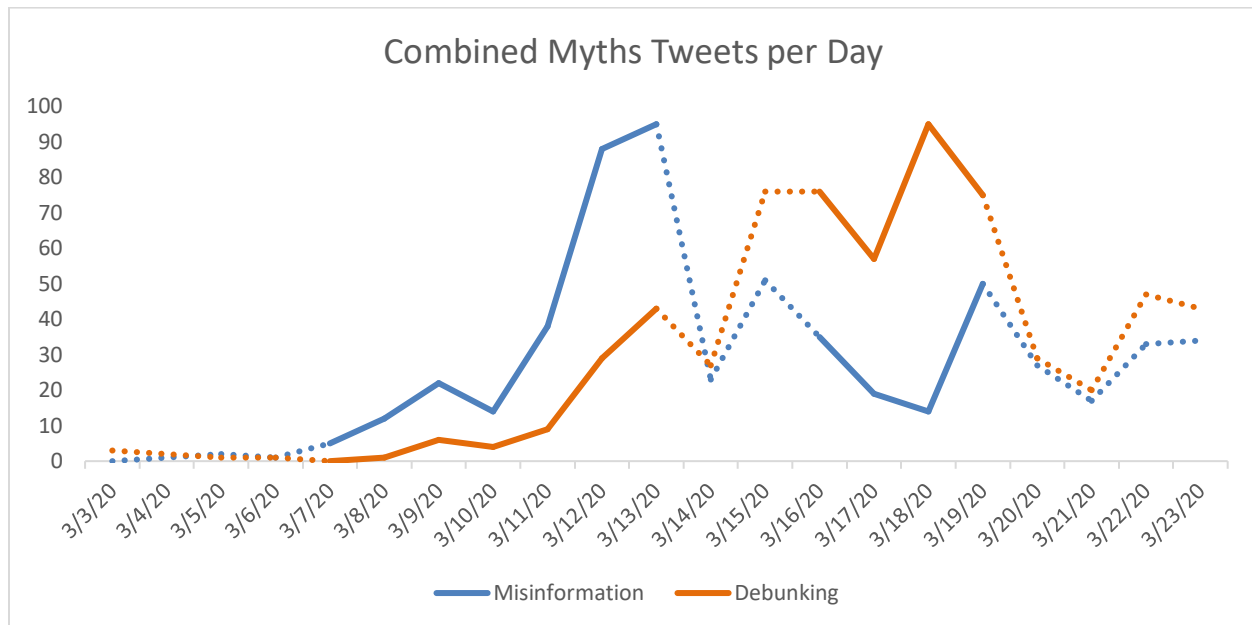


Figure 2. COVID-19 gargle myth and hold breath myth aggregate totals of misinformation tweets vs. debunking tweets over time. The solid line indicates the days when the difference in proportion between misinformation and debunking tweets was significant ($p < .05$) according to a Chi-square Test for Independence. The dotted line indicates the days when the proportion of the two types of tweets did not differ significantly.

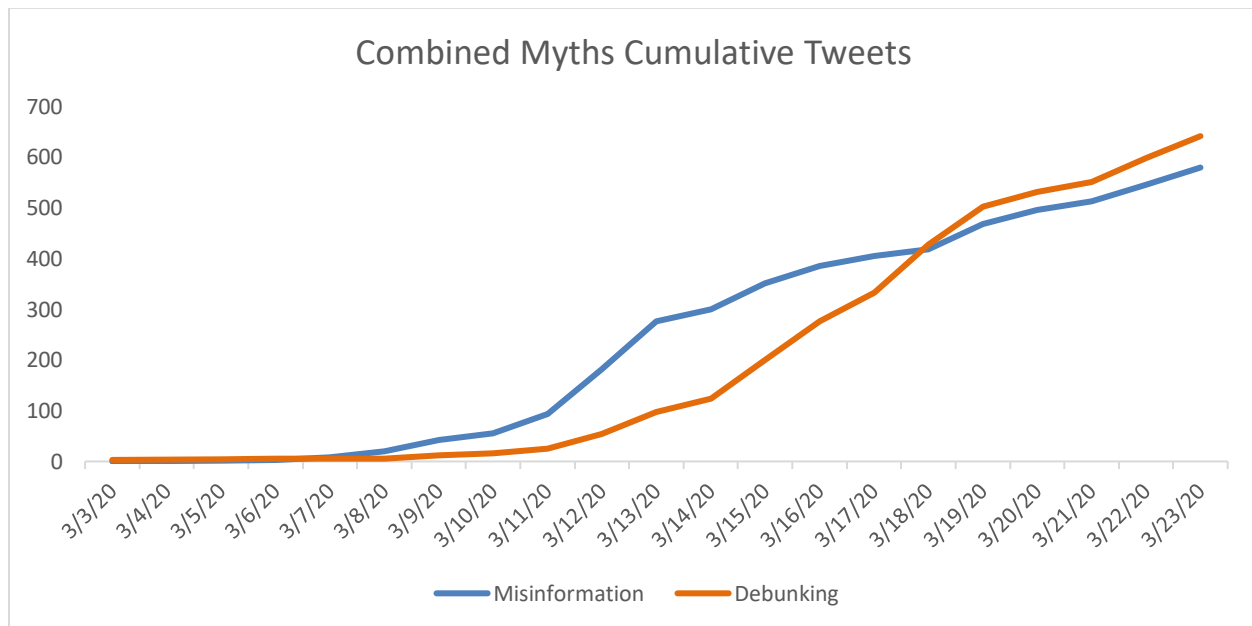


Figure 3. Combined myths cumulative misinformation tweets vs. debunking tweets over time.

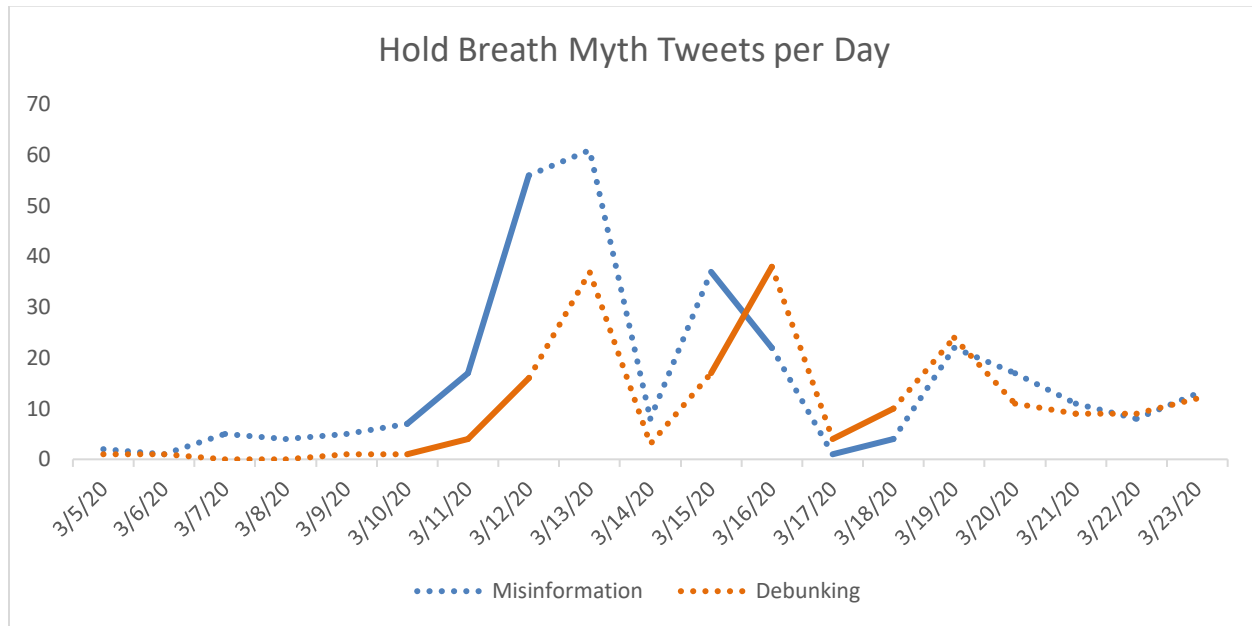


Figure 4. Hold breath misinformation tweets vs. hold breath debunking tweets over time. The solid line indicates the days when the difference in proportion between misinformation and debunking tweets was significant ($p < .05$) according to a Chi-square Test for Independence. The dotted line indicates the days when the proportion of the two types of tweets did not differ significantly.

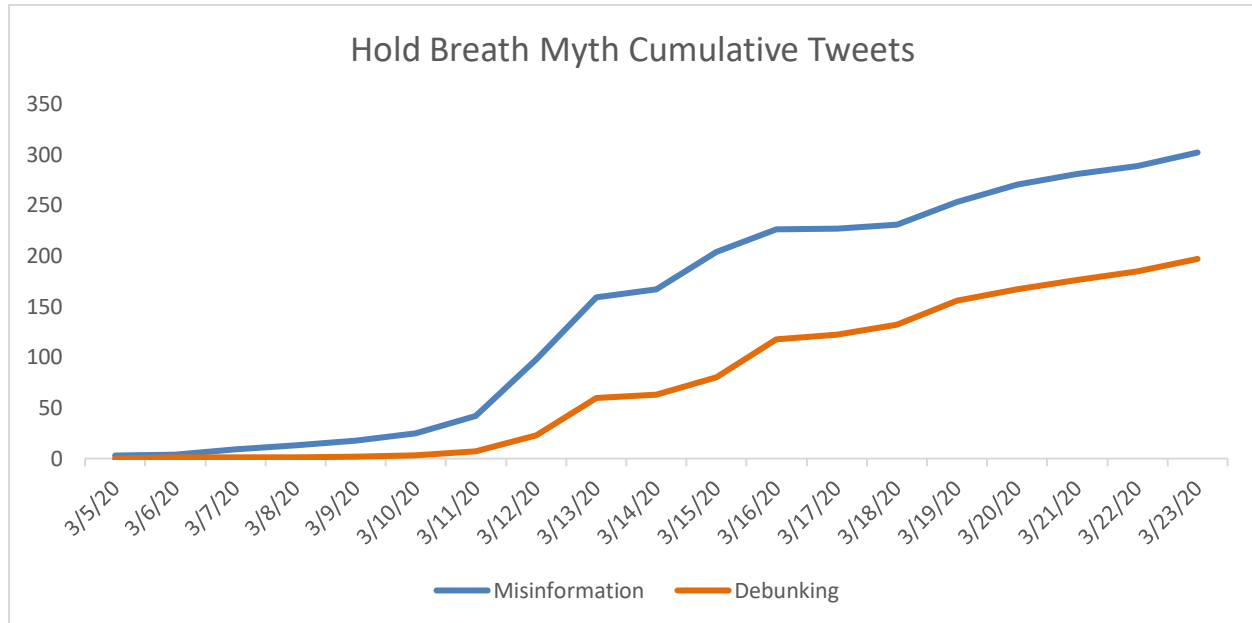


Figure 5. Hold breath myth cumulative misinformation tweets vs. debunking tweets over time.

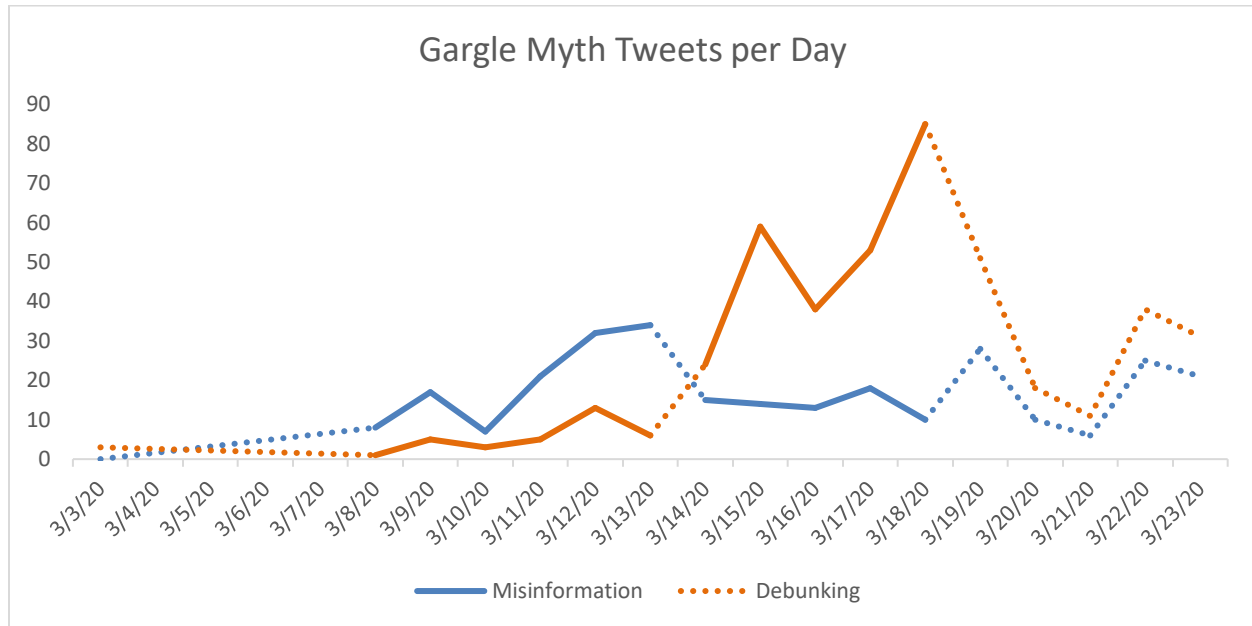


Figure 6. Gargle misinformation tweets vs. gargle debunking tweets over time. The solid line indicates the days when the difference in proportion between misinformation and debunking tweets was significant ($p < .05$) according to a Chi-square Test for Independence. The dotted line indicates the days when the proportion of the two types of tweets did not differ significantly.

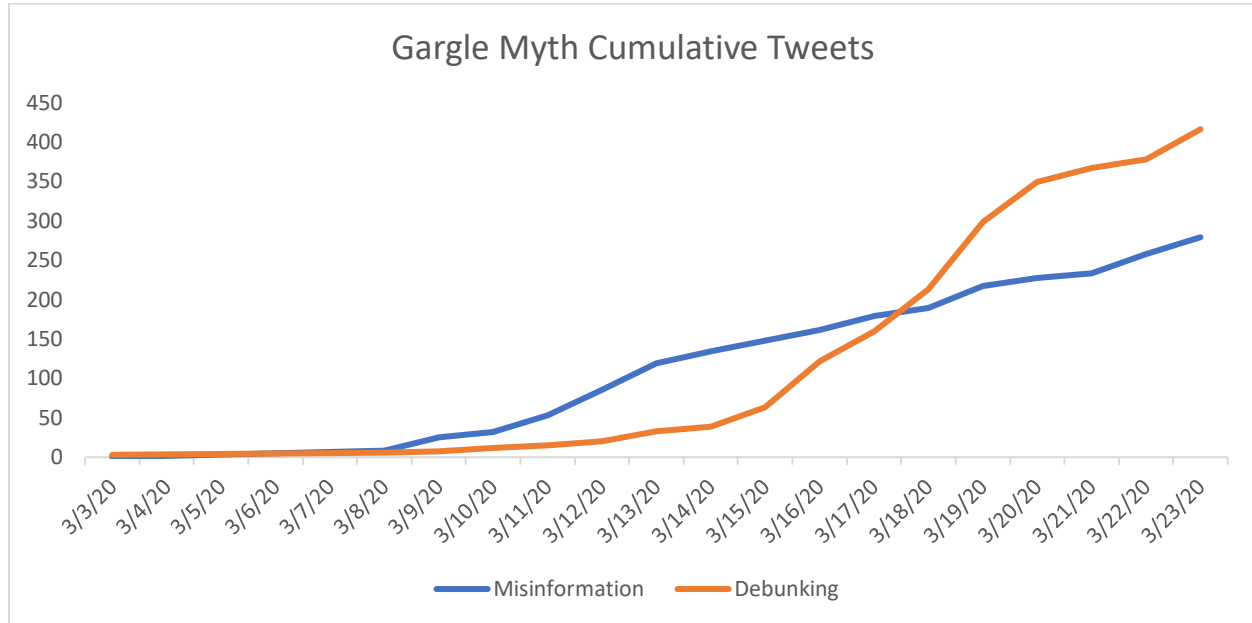


Figure 7. Gargle myth cumulative misinformation tweets vs. debunking tweets over time.

Finding 3: The inclusion of non-specific authority references in tweets was associated with decreased user engagement. Tweets that included non-specific authority references were less likely to receive likes or retweets from other users.

The use of non-specific authority references in COVID-19 tweets was related to decreased tweet engagement. Users were less likely to engage with tweets that included non-specific authority references, indicating that references to ambiguous authority cues may have negative effects on user interaction. This effect was driven by misinformation tweets, as the effect was not present when testing debunking tweets separately. Although non-specific authority references had a negative relationship with tweet engagement, including a specific authority appeal did not increase or decrease the likelihood of engagement. Table 2 provides the model statistics and odds ratios for the likelihood of tweet engagement for the misinformation tweets in the form of retweets and likes.

Table 2. Misinformation GEE logistic regression coefficients for likelihood of retweets and likes

QIC	Model Fit	
	Retweets	Likes
	3235.88	3593.05
Variable	Odds Ratio [95% CI]	Odds Ratio [95% CI]
Verified Tweet	0.10* [0.02, 0.46]	0.63 [0.12, 3.43]
Follower Count	1.10** [1.08, 1.12]	1.08** [1.07, 1.10]
Non-specific Authority	0.10* [0.02, 0.51]	0.07* [0.02, 0.37]

Note: If the odds ratio is below 1, then as the predictor variable increases or is present, the likelihood of tweet engagement decreases. If the odds ratio is above 1, then as the predictor variable increases or is present, the likelihood of tweet engagement also increases. Follower Count odds ratio should be interpreted as an increase in the odds for an increase of 1000 followers. QIC = Quasi-Likelihood under Independence Model

Criterion. * $p < .01$. ** $p < .001$.

Methods

Data Collection

Tweets related to two false COVID-19 pieces of advice (henceforth referred to as *myths*) were collected via Twitter's search API. A total of 1,493 tweets were collected for this study. The two COVID-19 myths were selected among several false claims reported by factchecking organizations (e.g., Snopes.com, Factcheck.org). The myths were selected because they were the least ambiguous (i.e., all or most of its elements were agreed upon by experts as clearly false) and the myths were findable on Twitter with a keyword-based query. The first myth claimed that gargling a liquid (e.g., warm water, saltwater, vinegar, mouthwash, bleach) would kill the coronavirus when the liquid entered the throat. The second myth claimed that a person could self-diagnose COVID-19 infection by holding their breath for a certain period of time. Typically, the claim was to hold your breath for 10 seconds, but the length of time varied slightly across tweets.

Once the two misinformation claims for the analysis were selected, we created search queries for each myth. To collect the tweets, we used the premium 30-day sandbox (free) search API. This specific API allowed us to collect tweets within the previous 30 days, which permitted reaching back to March 3,

2020. At the time of data collection, we were limited to collecting 7,000 historical tweets. Therefore, we restricted the search parameters to a maximum of 200 tweets per day. The search API returned both retweets and quote tweets, which provided the ability to collect the original tweets shared by users. Retweets were removed from the final sample, but the original tweet was retained. For quote tweets, we preserved both the user-entered text and the text from the original tweet. These criteria resulted in the total of 1,493 collected tweets. Next we used the Python package *Tweet Parser*, provided by Twitter, to extract the necessary elements of each tweet.

Content Analysis

After collecting the tweets, we engaged in content analysis. In the first round of coding, each individual tweet was labeled *misinformation*, *debunking*, or *irrelevant*. A *misinformation* tweet was defined as a message that stated the myth, urged others to adopt the myth's recommendation, or otherwise made no attempt at disconfirming or denying the myth. If a tweet indicated skepticism of the claim or disbelief of the myth, the tweet was labeled *debunking*. *Debunking* tweets were defined as messages designed to correct or refute misinformation. Debunking tweets frequently referenced media articles to refute the claim (e.g., Snopes, Factcheck.org), or otherwise addressed the inaccuracy of the myth (e.g., "Fact check: Holding your breath for 10 seconds is not a coronavirus test", "Gargling vinegar will not protect you from COVID-19"). Within the label of debunking, for descriptive purposes, we then specified if the debunking tweet represented *overt debunking* or merely *expressed suspicion* of the myth. The two types of debunking tweets were combined for the current analysis.

Throughout this analysis, we use the terms *authority appeal* and *authority reference* interchangeably. To identify the inclusion of authority appeals in the COVID-19 tweets, we conducted closed coding of each individual tweet, identifying the presence or absence of authority appeals. An *authority appeal* was defined as referencing an individual or organization that has knowledge about COVID-19. Although some individuals or organizations referenced may not be universally considered an authority on a viral outbreak (e.g., political leaders), we decided to code such references as an authority appeal because the authority referenced may have influence on their supporters. Examples of authorities referenced in the tweets included emergency room doctors, nurses, research centers, empirical studies, news outlets, political leaders, and universities. If a tweet attributed the information to such an authority, we coded the tweet as including an authority appeal.

The manner in which authority appeals were referenced varied across tweets. At times the authority references reflected direct statements, such as, "According to medical experts at Stanford, you can self-diagnose COVID-19 by holding your breath for 10 seconds." Other tweets included an article link to a specific authority reference. In that case, the news outlet associated with the article acted as the authority appeal (e.g., Snopes, BBC, World Health Organization). We did not consider links to blogs to be an authority appeal. For certain tweets, images were included in the original post, which we also examined. If the tweet included an image of a medical professional or the logo of an organization, we coded the tweet as including an authority appeal.

In addition to identifying the presence or absence of authority appeals in the tweets, we coded each appeal as either a specific authority appeal or a non-specific authority appeal. A specific authority appeal was defined as a message that included a name or a verifiable authority, such as Factcheck.org or the CDC. Non-specific authority appeals included messages that did not identify a specific person or organization, relying instead on ambiguous authority references. Examples of non-specific authority references included "Japanese doctors" and "advice from a nurse friend."

Finally, we removed any irrelevant tweets from the sample. Irrelevant tweets included terms from the search query but were unrelated to the myth. For example, some tweets described general tips about breathing exercises to boost immunity against COVID-19, but were not claiming that the breathing

exercises were a self-diagnosis tool. Tweets asking an open-ended question of whether or not the myth was accurate were marked irrelevant, as they were not clearly misinformation or debunking.

Table 3. COVID-19 example tweets by category.

Category of tweet	COVID-19 myth	
	Gargle myth	Hold breath myth
Misinformation	Gargles with warm water & salt or vinegar and drinking lots of warm water can wash away #Coronavirus when its clinging to the throat in first 4 days after entering the body, So gargle everyday pl spread this info through retweets #CoronaOutbreak	A simple coronavirus test you can do yourself ... Take a deep breath and hold your breath for 10 seconds ... After you have held your breath for 10 seconds and your not coughing, experiencing chest pains, tightness or any discomfort your probably not infected with coronavirus.
Debunking (overt)	WARNING!! FAKE! While gargling some combination of warm water, salt, and vinegar has long been used as a means of relieving symptoms related to colds and flus, such as sore throats, there's No evidence that it can help drive out infections from the COVID-19.	If anyone is saying to hold your breath for 10 seconds as a test for coronavirus... this has been debunked.
Debunking (expressed suspicion)	Meanwhile I saw someone we went to high school with share something that gargling with salt water would kill the coronavirus and it wasn't a joke	Anyone hear about the 10 second test? If you can hold your breath for 10 seconds you don't or possibly don't have the coronavirus. Different theories about this on social media. I'm not buying it personally. I stopped after 15 seconds but it proves nothing imo.
Non-specific authority appeal (misinformation)	From a medical friend in the US. Pls share! Coronavirus before it reaches lungs it remains in the throat for 4 days n at this time he/she begins to cough n has throat pains. Drinks a lot of water n gargling with warm water and salt or vinegar eliminates the virus. Share pls.	Taiwan experts provide a simple self-check. Take a deep breath, hold your breath for more than 10 seconds. If you complete it without coughing, without discomfort, stiffness or tightness, etc, it proves there is no Fibrosis in the lungs, basically indicates no infection. #COVID19
Specific authority appeal (debunking)	Sorry, gargling with salt water won't kill off the virus, it's a myth: https://t.co/dlieS6ysWq	You can hold your breath for ten seconds? That's nice — but you might be infected with COVID-19 anyway. Good ol' @snopes says so. https://t.co/MbcuwuBqUG

Bibliography

- Austin, L., Fisher, B., & Jin, Y. (2012). How audiences seek out crisis information: Exploring the social-mediated crisis communication model. *Journal of Applied Communication Research*, 40(2), 188-207. <https://doi.org/10.1080/00909882.2012.654498>
- Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology: General*, 149(8), 1608-1613. <https://doi.org/10.1037/xge0000729>
- Bode, L., & Vraga, E. K. (2018). See something, say something: Correction of global health misinformation on social media. *Health Communication*, 33(9), 1131-1140. <https://doi.org/10.1080/10410236.2017.1331312>
- Borah, P., & Xiao, X. (2018). The importance of 'likes': The interplay of message framing, source, and social endorsement on credibility perceptions of health information on Facebook. *Journal of Health Communication*, 23(4), 399-411. <https://doi.org/10.1080/10810730.2018.1455770>
- Chan, M. S., Jones, C. R., Jamieson, K. H., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, 28(11), 1531-1546. <https://doi.org/10.1177/0956797617714579>
- Fazio, L. K. (2020). Pausing to consider why a headline is true or false can help reduce the sharing of false news. *Harvard Kennedy School (HKS) Misinformation Review*, 1(2), 1-8. <https://doi.org/10.37016/mr-2020-009>
- Freelon, D. (2010). ReCal: Intercoder reliability calculation as a web service. *International Journal of Internet Science*, 5, 20-33.
- Frenkel, S., Alba, D. & Zhong, R. (2020, March 8). Surge of virus misinformation stumps Facebook and Twitter. *The New York Times*. <https://www.nytimes.com/2020/03/08/technology/coronavirus-misinformation-social-media.html>
- Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1(1), 77-89. <https://doi.org/10.1080/19312450709336664>
- Houston, J. B., Hawthorne, J., Perreault, M. F., Park, E. H., Goldstein Hode, M., Halliwell, M. R., Turner McGowen, S. E., Davis, R., Vaid, S., McElderry, J. A., & Griffith, S. A. (2014). Social media and disasters: A functional framework for social media use in disaster planning, response, and research. *Disasters*, 39(1), 1-22. <https://doi.org/10.1111/disa.12092>
- Hu, Y., & Sundar, S. S. (2010). Effects of online health sources on credibility and behavioral intentions. *Communication Research*, 37(1), 105-132. <https://doi.org/10.1177/0093650209351512>
- Hunt, K., Wang, B., & Zhuang, J. (2020). Misinformation debunking and cross-platform information sharing through Twitter during Hurricanes Harvey and Irma: A case study on shelters and ID checks. *Natural Hazards*, 103, 861-883. <https://doi.org/10.1007/s11069-020-04016-6>
- Jamieson, K. H., & Albarracín, A. (2020). The relation between media consumption and misinformation at the outset of the SARS-CoV-2 pandemic in the US. *Harvard Kennedy School (HKS) Misinformation Review*, 1(2). <https://doi.org/10.37016/mr-2020-012>
- Krause, N. M., Freiling, I., Beets, B., & Brossard, D. (2020). Fact-checking as risk communication: The multi-layered risk of misinformation in times of COVID-19. *Journal of Risk Research*. <https://doi.org/10.1080/13669877.2020.1756385>
- Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research*, 30(3), 411-433. <https://doi.org/10.1111/j.1468-2958.2004.tb00738.x>

- Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3), 106-131. <https://doi.org/10.1177/1529100612451018>
- Makhortykh, M., Urman, A., & Ulloa, R. (2020). How search engines disseminate information about COVID-19 and why they should do better. *Harvard Kennedy School (HKS) Misinformation Review*, 1(3), 1-11. <https://doi.org/10.37016/mr-2020-017>
- Nsoesie, E. O., & Oladeji, O. (2020). Identifying patterns to prevent the spread of misinformation during epidemics. *Harvard Kennedy School (HKS) Misinformation Review*, 1(3) 1-6. <https://doi.org/10.37016/mr-2020-014>
- Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, 147(12), 1865-1880. <https://doi.org/10.1037/xge0000465>
- van der Meer, T. G. L. A., & Jin, Y. (2020). Seeking formula for misinformation treatment in public health crises: The effects of corrective information type and source. *Health Communication*, 35(5), 560-575. <https://doi.org/10.1080/10410236.2019.1573295>
- Volkin, S. (2020, March 27). Social media fuels spread of COVID-19 information-and misinformation. *Johns Hopkins University*. <https://hub.jhu.edu/2020/03/27/mark-dredze-social-media-misinformation/>
- Wang, B., & Zhuang, J. (2017). Crisis information distribution on Twitter: A content analysis of tweets during Hurricane Sandy. *Natural Hazards*, 89, 161-181. <https://doi.org/10.1007/s11069-017-2960-x>
- Wang, B., & Zhuang, J. (2018). Rumor response, debunking response, and decision makings of misinformed Twitter users during disasters. *Natural Hazards*, 93, 1145-1162. <https://doi.org/10.1007/s11069-018-3344-6>

Funding

The authors report no receipt of funding associated with this research.

Competing interests

The authors declare no potential conflicts of interests with respect to the research, authorship, or publication of this article.

Ethics

The data for the project were obtained from publicly available sources and thus were exempt from IRB review.

Copyright

This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.

Data availability

All materials needed to replicate this study are available via the Harvard Dataverse: <https://doi.org/10.7910/DVN/GSFFFP>

Appendix: Supplementary methods

Twitter Search Query

We limited our search to English tweets only and variations of the term “COVID-19”. Specifically, the API queries we used were:

Gargle Myth

lang:en (gargle OR gargling)

(covid19 OR coronavirus OR covid-19 OR c19 OR cov OR covid_19 OR covid)

Hold Breath Myth

lang:en hold (breath OR breathe) seconds

(covid19 OR coronavirus OR covid-19 OR c19 OR cov OR covid_19 OR covid)

Data Analysis

To test whether authority appeals predicted tweet engagement, we fit a negative binomial model. The negative binomial model accounted for the large variance (i.e., overdispersion) of retweet and like counts. Part of the overdispersion was caused by one extreme outlier (i.e., one tweet with greater than 10,000 likes), which caused the model fit to be unreliable. As a result, we restricted the number of likes to fewer than 10,000. Control variables for the analysis included whether or not the account was verified and the number of followers for the account.

Coding Process and Creation of the Codebook

With regard to the coding of the tweets for content analysis, the first two authors coded the first 300 tweets collaboratively (approximately 20% of the dataset). During the initial coding of the tweets, a codebook was developed to create categories, definitions, and examples for each code. Upon creation of the codebook, the first 300 tweets were then recoded to account for any potential changes in coding decisions that may have developed throughout the coding process. This step of recoding helps to promote stability (Krippendorff, 2004) and ensure the coding decisions did not change over time. The first two researchers then divided the remainder of the dataset evenly and coded the tweets individually. If one of the researchers was uncertain on how to code a particular tweet, the two researchers discussed the coding of the tweet collaboratively. All disagreements were discussed until agreement was met. This process created the set of coded tweets for the analysis.

Intercoder Reliability

To assess intercoder reliability, a researcher blind to the purpose of the study was trained using the codebook created for the study. Reliability between coders was then assessed through an open-source calculation tool for reliability analysis (Freelon, 2010). As our data included two sets of observations and binary coding categories, scores for Scott’s pi, Cohen’s kappa, and Krippendorff’s alpha were asymptotically equal (Hayes & Krippendorff, 2007). We used percent agreement and Krippendorff’s alpha as a guideline for acceptable reliability (Hayes & Krippendorff, 2007). Percent agreement was above 90% for each category, and Krippendorff’s alpha indicated acceptable levels of reliability, ranging from .70 to .89 (Krippendorff, 2004). Krippendorff’s alpha accounts for the possibility of chance agreements, which is a reason why a coding category with a higher percent agreement may have a lower Krippendorff’s alpha. Table 4 illustrates the reliability statistics for each tweet category. The codes assigned to each tweet are

available in the replication materials. The codebook created for the current study is also available in the replication materials. The codebook includes descriptions of the coding categories, definitions for each category, and example tweets for each coding category.

Table 4. *Intercoder reliability statistics for the coded tweets (N = 1,493)*

Category of Tweet	Percent Agreement	N Agreements	N Disagreements	Scott's π	Cohen's κ	Krippendorff's α
Irrelevant	93.64%	1398	95	0.77	0.77	0.77
Misinformation	92.63%	1383	110	0.85	0.85	0.85
Debunking – Overt	94.98%	1418	75	0.89	0.89	0.89
Debunking – Suspicion	94.31%	1408	85	0.88	0.88	0.88
Authority Appeal	90.82%	1356	137	0.79	0.79	0.79
Non-specific Authority Appeal	96.45%	1440	53	0.70	0.70	0.70
Specific Authority Appeal	93.77%	1400	93	0.83	0.83	0.83