Harvard Kennedy School Misinformation Review[1]
May 2021, Volume 2, Issue 3

DOI: https://doi.org/10.37016/mr-2020-69
Website: misinforeview.hks.harvard.edu

Research Article

How COVID drove the evolution of fact-checking

With the outbreak of the coronavirus pandemic came a flood of novel misinformation. Ranging from harmless false cures to dangerous rhetoric targeting minorities, coronavirus-related misinformation spread quickly wherever the virus itself did. Fact-checking organizations around the world took up the charge against misinformation, essentially crowdsourcing the task of debunking false narratives. In many places, engagement with coronavirus-related content drove a large percentage of overall user engagement with fact-checking content, and the capacity organizations developed to address coronavirus-related misinformation was later deployed to debunk misinformation on other topics.

Authors: Samikshya Siwakoti (1), Kamya Yadav (1), Nicola Bariletto (2), Luca Zanotti (2), Ulaş Erdoğdu (3), Jacob N. Shapiro (4)
Affiliations: (1) ESOC lab, Princeton University, USA, (2) Department of Economics, Bocconi University, Italy, (3) Department of Political Science and International Relations, Boğaziçi University, Turkey, (4) Princeton School of Public and International Affairs, Princeton University, USA
How to cite: Siwakoti, S., Yadav, K., Bariletto, N., Zanotti, L., Erdoğdu, U., & Shapiro, J. N. (2021). How COVID drove the evolution of fact-checking. *Harvard Kennedy School (HKS) Misinformation Review*, *2*(3).
Received: January 21st, 2021. Accepted: March 22nd, 2021. Published: May 6th, 2021.

Research questions

- Did fact-checking organizations scale up efforts to debunk misinformation in 2020?
- Did Twitter users engage more with fact-checking in 2020 than in previous years?
- Did engagement with coronavirus-related content drive overall engagement with fact-checking content?

Essay summary

- Using Twitter's new research API, we collected the engagement metrics and Twitter activity of fifteen locally-focused fact-checking organizations around the world for 2019 and 2020. The 15 organizations were selected for variation by geography and popularity in our data tracking COVID misinformation narratives since March 2020 (Shapiro et al., 2020). Thirteen of them are independent fact-checking organizations while two, Liputan6 and Agencia Lupa, are fact-checking bodies within larger news networks.[2]

[1] *A publication of the Shorenstein Center for Media, Politics and Public Policy, at Harvard University, John F. Kennedy School of Government.*
[2] Although Liputan6 is a larger news program, we only study the Twitter activity relevant to its fact-checking body by specifying the URL to <https://www.liputan6.com/cek-fakta> in the Twitter API query parameter.

- We examined the scale and composition of fact-checking activity as well as user engagement across 2019 and 2020. We found a significant increase in the debunking efforts of fact-checking organizations in 2020 driven largely by fact-checking of COVID misinformation. Much of this activity was not new; it did not appear to substitute for fact-checking on other topics.
- User engagement did not follow as clear a pattern as the increase in fact checkers' activity. Coronavirus may have driven user engagement in the early months of the pandemic, but region-specific salient events and one-off viral tweets influenced user engagement in the later months of 2020. There is substantial heterogeneity across fact-checking organizations and regions in how engagement with coronavirus-related content correlates with overall engagement on fact-checks.
- Our findings contribute to the literature on the potential for fact-checking and pre-bunking to improve individuals' perceptual accuracy of political and medical (mis)information (Clayton et al., 2019; Roozenbeek et al., 2020; Walter et al., 2019; Walter et al., 2020). We show that the grassroots fact-checking community has the ability to respond to sudden changes in the misinformation environment. However, our findings also indicate that fact-checking alone may not be the solution to rampant misinformation.

Implications

Since March 2020, the role of fact-checking organizations has become increasingly important due to the widespread prevalence of misinformation about coronavirus (Siwakoti et al., 2021). The impact of debunking activity on various aspects of people's life has been discussed in an ever-expanding literature.

Fact-checking activity has been shown to have relevant effects on political beliefs. Walter et al. (2020) found that fact-checking has an overall positive influence on political beliefs, but the ability to correct misinformation through such activity is significantly weakened by individuals' preexisting beliefs, ideology, and knowledge. Along the same lines, York et al. (2019) used survey data to show that fact-checking boosts accuracy in a specific political issue perception, but it decreases the overall confidence in the ability to correctly recognize what is true and what is not. Research by Wintersieck (2017) revealed that fact checks that show a political candidate is telling the truth can improve the candidate's popularity, thus underlining the willingness of respondents to vote for honest candidates. Furthermore, Nyhan and Reifler (2014) suggested that debunking activity can have a deterrent effect on candidates, since it raises the reputational costs associated with lies, thus improving information accuracy.

Although this literature shows the beneficial effects of fact-checks, Robertson et al. (2020) argued that fact-checking organizations in the United States are absorbed into a wider ideological debate: users categorized as liberals are more aware of the positive effects of fact-checking activity, and thus regard it as useful for the political debate; in contrast, users with a more conservative political ideology see these websites as useless. In the same vein, Jakesch et al. (2018) showed that alignment with participants' political ideologies plays a more significant role than the brand of the publisher (e.g., the New York Times versus Fox News) when evaluating partisan claims. This highlights how political ideology and polarization can play a huge role in people's perceptions.

Emerging evidence suggests that pre-emptive fact-checking, often termed "pre-bunking," can have a great impact on individuals' ability to recognize the truth when exposed to fake news (Pennycook et al., 2020; Roozenbeek et al., 2020). Silverman et al. (2016) highlighted that funding and guidance to counter-narrative campaigners can improve the efficacy of and engagement with counter-narratives to violent extremism, which in turn, can help in deradicalizing individuals. Clayton et al. (2019) showed that tags such as *disputed* or *rated fals*e make users perceive false headlines as less accurate. Walter et al. (2020) found that correction of health-related misinformation has positive effects, and these can be higher the more individuals are involved with topics covered by the fake content.

Yet, others highlight risks and drawbacks associated with fact-checking activity. Repetition of either fake or real news tends to increase the perceived accuracy of that news (Pennycook et al., 2018; Skurnik et al., 2005). Sally Chan et al. (2017) argued that a detailed debunking message surprisingly positively correlates with the persistence of the misinformation effect. Moreover, the impact of fact-checking is seriously threatened by three factors: the different communities of fact-check sharers versus misinformation sharers; the short period of time in which fact-checks are likely to spread; and, importantly, the amount of shared misinformation which is disproportionately higher than fact-checking content (Burel et al., 2020). Taken as a whole, this literature suggests that if misinformation can be fact-checked at scale, then delivering those checks in an appropriate manner could ameliorate the misinformation challenge.

In this paper, we show that the grassroots fact-checking community is delivering on the first part of that process. Globally and locally salient events can drive spikes in both fact-checking activity and user engagement with that activity. However, these events do not sustain user engagement over time, though fact-checking activity is often sustained after these events. We study how sudden shocks to the misinformation environment shape the supply of and interest in fact checks by assessing the scale and the composition of fact-checking activity by fact-checkers and users on Twitter before and after the outbreak of the pandemic.

We find that fact checkers increased their debunking activity in the early months of the pandemic. Most covered COVID without dropping other topics, and many sustained their momentum throughout 2020. In contrast, we found that while users increased engagement with fact-checking content in the early months of the pandemic as well, this engagement was often not sustained throughout 2020. Sporadic and locally salient events or occasional viral tweets sometimes drove spikes in user engagement, but these spikes could not be attributed to the pandemic.

Our findings are particularly important when understood in the context of empirical work surrounding fact-checking and its impacts. These findings have important implications which can be extended beyond the coronavirus context. The pandemic serves as an interesting case study to address some of the pressing questions for fact-checking organizations. First, the responsibilities of fact-checking organizations increase during crises and misinformation-prone events. This implies that fact-checking organizations could do more with increased funding and capacity that could be available for mobilization during crises when people are most likely to engage with fact-checking, as suggested by higher engagement during the early months of the pandemic.

Second, there is great variability in the activity and the engagement with fact-checking content. This suggests that advertising revenues for these organizations will be highly variable. In turn, this could imply that fact-checking organizations may need some form of subsidization to be able to scale if another crisis of this sort breaks out. Moreover, this variability suggests that future research could move to analyze what behaviors and strategies are able to sustain a high level of engagement over time. Third, given that user engagement may not be sustained after such events, governments and civil society organizations must find alternative methods to debunk misinformation, such as public awareness campaigns and sharing of fact-checking content by influencers along with citing scientific sources (Chen et al., 2020). Fourth, since there is empirical evidence for partisan and ideological differences in engagement with misinformation and fact-checking, future research must focus on how to bridge this divide so that accurate information reaches beyond partisan and ideological differences. Though our study does not delve into it, future research can focus on how much of the total conversation around misinformation-prone events (such as coronavirus) is dominated by fact-checking.

Findings

Analysis of Twitter data from 15 fact-checking organizations of various sizes and geographic locations and user engagement with fact-checking content from those organizations revealed four key findings, discussed below. First, we examine the increase in activity by fact-checking organizations. Second, we examine the patterns in user engagement with fact-checking content. Third, we examine the heterogeneity in user engagement and associate it with locally salient events. Last, we discuss how much of the change in total engagement is driven by engagement with coronavirus-related content.

Table 1. *List of fact-checking organizations.*

Name	URL	Region/Country	Organization Type	No. of COVID-related Tweets
Africa Check	africacheck.org	Africa	Independent	440
AltNews	altnews.in	India	Independent	272
BOOM Live	boomlive.in	India	Independent	3183
Chequeado	chequeado.com	Argentina	Independent	1638
Correctiv	correctiv.org	Germany	Independent	225
Dogruluk Payi	dogrulukpayi.com	Turkey	Independent	149
EUvsDisinfo	euvsdisinfo.eu	European Union	Independent	69
Falsoo	falsoo.com	Egypt	Independent	3191
Folha de S. Paulo: Agencia Lupa	piaui.folha.uol.com.br/lupa	Brazil	Fact-checking body in larger news network	2280
Liputan6	liputan6.com/cek-fakta	Indonesia	Fact-checking body in larger news network	645
Maldita	maldita.es	Spain	Independent	204
Pesa Check	pesacheck.org	East Africa	Independent	220
Sebernarnya	sebenarnya.my	Malaysia	Independent	223
Teyit	teyit.org	Turkey	Independent	276
Turn Back Hoax	turnbackhoax.id	Indonesia	Independent	492

Note: The number of COVID-related tweets here refers to tweets containing a COVID term, a URL link directing to the fact-checking website's domain and tweeted from the Twitter handles of these fact-checking sites.

Finding 1: Fact-checking organizations increased their debunking efforts in 2020, as evidenced by increased activity on Twitter but there is variation across organizations.

Overall, fact-checking organizations increased their debunking efforts in 2020, as evidenced by increased activity on Twitter, though there is substantial heterogeneity across fact-checking organizations (see Figure 1).[3] Several fact-checking organizations (e.g., EUvsDisinfo, Dogruluk Payi, Sebernarnya, and Teyit)

[3] While looking at debunking efforts, we ignored the issue of multiple iterations of the same fact check impacting the results. This is because our focus was not how many unique stories each fact-checking organization is producing, it was on their overall activity level as seen on Twitter. Given that our focus was on activity (as defined by the number of fact-checking related posts) and engagement (as defined by user interaction with fact-checking), the repetition of fact checks in our dataset is part of the phenomenon being studied.

scaled up their activities in the first months of the pandemic but did not sustain their momentum throughout the year. Others (i.e., Boom Live, Agencia Lupa, and Pesa Check) sustained increased activity throughout 2020 while some (i.e., Africa Check and Alt News) scaled down their fact-checking activities in 2020. The heterogeneity across organizations, as shown in Figure 1, can be attributed to a few factors. Most prominently, organizations may have been posting fact checks on their websites and not posting them on Twitter. For instance, Falsoo's Twitter handle stopped posting fact checks on Twitter in the second half of the year, though their website continued fact-checking activities through October 2020. The decline in fact-checking activity for organizations like Africa Check and AltNews could be a result of the aforementioned reason or reduced funding and, therefore, human resource constraints. Some organizations, such as Sebernarnya and Liputan6, did not engage with fact-checking in 2019. The majority of these websites, however, either maintained similar levels of activity in 2019 and 2020 or scaled up after the pandemic.

Coronavirus-related debunking efforts increased for many fact-checking organizations during the early months of the pandemic (see Figure 2). At times, these increases were accompanied by increases in overall fact-checking efforts, such as for Correctiv, Agencia Lupa, and Sebernarnya. For a few organizations (i.e., Africa Check, AltNews, Chequeado, Boom Live, and EUvsDisinfo) coronavirus-related debunking efforts substituted for other fact-checking efforts, suggested by simultaneous decreases in non-COVID fact checks and increases in COVID fact checks (see Figure 2). The difference between fact-checking activities in 2019 and 2020 does not indicate immediately clear patterns; there is variation across months and websites that cannot be explained solely by the pandemic. This suggests that region-specific salient events may have caused the variance within each year and is supported by the patterns in user engagement, discussed next, as well.

Figure 1. Number of tweets by fact checkers per month. *The number of Tweets by fact-checking organizations on Twitter in 2019 and 2020 are shown as a time series. To be included in the dataset, each Tweet had to include a URL to a fact-checking organization's website and had to come from the official Twitter handle of the fact-checking organization.*

Figure 2. Distribution of COVID-related and unrelated tweets by fact checkers*. The total posts by fact checkers for COVID content includes posts that included a URL to a fact-checking website and at least one of the keywords related to COVID. Posts with non-COVID content included posts that included a URL to a fact-checking website, but no keywords related to COVID.*

Finding 2: User engagement with the fact-checking efforts increased substantially after the pandemic, as evidenced by total engagement with the fact-checking content.

Our dataset shows that the total user engagement—measured as the sum of likes, retweets, quote tweets, replies, and number of tweets by users with a link to a fact-checking website, increased substantially after the pandemic (see Figure 3). A closer look at trends for each fact-checking organization revealed interesting patterns. In most of the cases, the coronavirus pandemic seemed to have engendered a short-term interest in user engagement with fact-checking efforts. We observed a large increase in total engagement for the first few months of the pandemic (February, March, or April), while the engagement in the second half of the year was similar to or less than in 2019. Users in regions where BOOM Live, Correctiv, and Pesa Check operate increased their engagement with fact-checking content throughout 2020. For other regions, such as where Agencia Lupa operates, we only saw a small increase in the first months of the pandemic. According to studies conducted in Turkey, misinformation (Erdoğan & Semerci, 2021) as well as self-reported exposure to misinformation (Tandans Data Science Consulting & Teyit.org, 2020) was especially widespread in the early days of the pandemic when there was little scientific evidence available. Taken together with our findings, this suggests that engagement with fact-checking content may have been the highest when it was needed most.

However, the increased engagement in the initial months of the pandemic did not necessarily translate into an increased engagement with non-coronavirus content (see Figure 4). This suggests we

need further research on why fact checkers cannot sustain user engagement, as well as interventions to sustain the increased engagement with fact-checking content. We explored whether spikes in user engagement were associated with the coronavirus in our next finding.

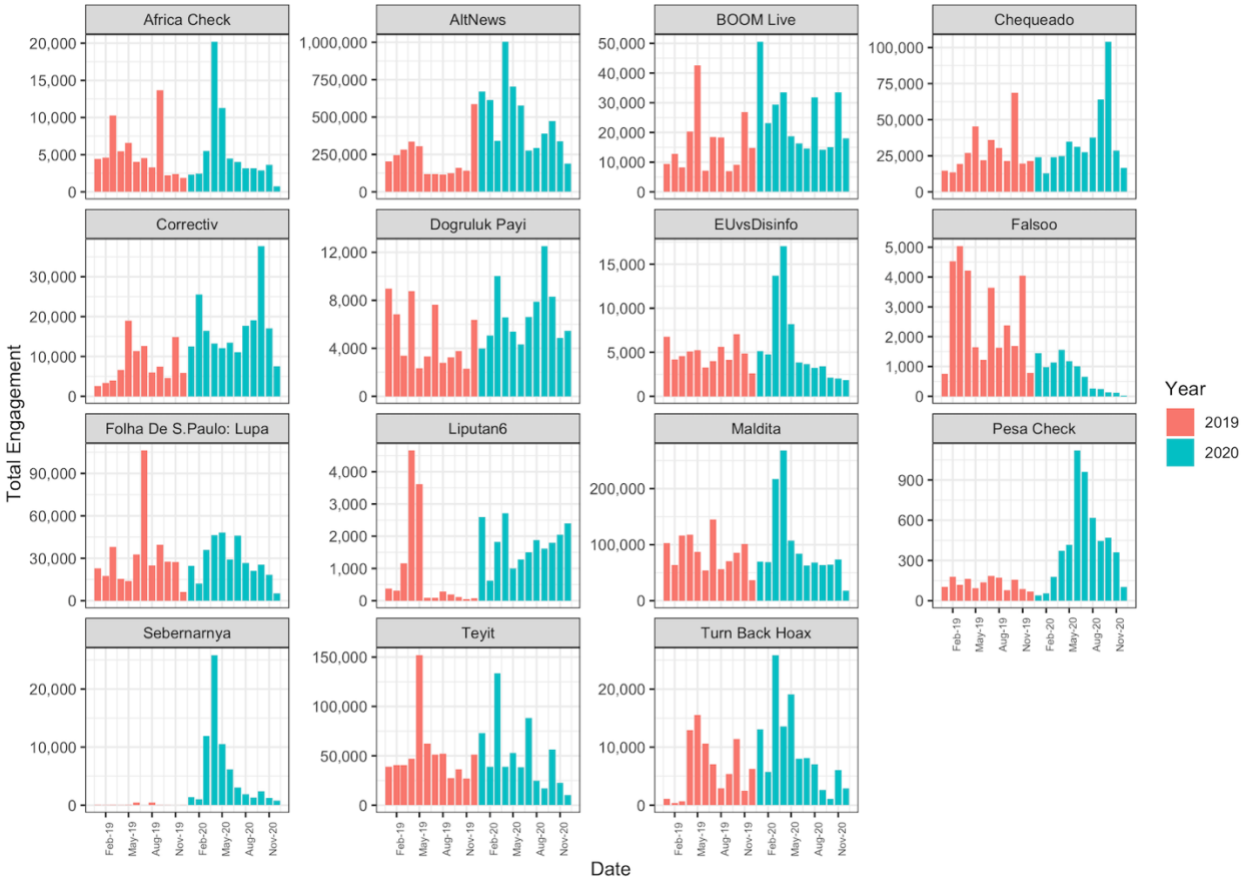

Figure 3. Total engagement by users per month in 2019 and 2020. *We measure user engagement to fact-checking as the sum of likes, retweets, quote tweets, and replies to all fact-checking content as well as the number of posts by users that included a URL to a fact-checking organization's website*.

Finding 3: There is variation in the user engagement with fact-checking content across organizations, which can be explained by locally salient events. The content of fact-checking stories rather than the quantity drives engagement with the fact-checking content.

A closer look at the country-level/regional trends revealed that engagement with fact-checking content increased with politically or socially salient events specific to the region/country or when certain false narratives became widespread without a particular reason. Some of the engagement spikes could also be driven by engagement with coronavirus-related stories but would not have been captured as such. This is because they may have been mislabeled as non-coronavirus-related if they did not contain a coronavirus-related term from our list (see Table A1 in Appendix A) in the body of the tweet. To test the gravity of this limitation, we ran a robustness check, where we took a random sample of 500 tweets from our full dataset that were categorized as COVID and another random sample of 500 tweets that were categorized as non-COVID and manually checked each of them. This revealed a 0% false-positive rate (non-coronavirus tweets labeled as coronavirus tweets) and a 20% false-negative rate (coronavirus tweets labeled as non-coronavirus tweets), suggesting that our methodology captured a majority of the coronavirus-related fact-

checking content.

We took a closer look at some of the regions with increased engagement. In Turkey, we observed a significant increase in engagement with Teyit.org in July. This increase was driven by the debunking of the Wayfair child trafficking conspiracy theory (Spring, 2020), verification of the claim that the Ottoman Empire decriminalized homosexuality in 1858 long before Europe (Yılmaz, 2020), debunking of the claims that the Portrait of Mehmet II (Keskin, 2020) - which was bought and brought back to Turkey by the opposition-held Istanbul Metropolitan Municipality - was fake, and false rumors surrounding the conversion of Hagia Sophia Museum to a mosque (Korkmaz, 2020). Whereas the Wayfair child trafficking conspiracy dominated the online information ecosystem for several days and was a global phenomenon, others were directly related to the politicization of Ottoman history pertaining to the long-standing secular-religious cleavage and identity conflict in the country. Similarly, for Chequeado the spikes are driven by controversies surrounding statements made by president Alberto Fernández (Chequeado, 2020) and the speculation of Fraud in Argentina's PASO elections (Domínguez et al., 2020).

Moreover, the level of total engagement was oftentimes driven by one-off tweets, shared by either fact checkers or by influential Twitter accounts with a high following, rather than by individuals tweeting fact-checking content. For BOOM Live and AltNews, there were no salient events in particular that drove the spikes in user engagement; rather, these spikes were associated with one-off viral tweets. Our findings were confirmed by the patterns in per-post engagement: while total engagement seems to increase in certain months, Figure 5 shows that per-post engagement does not. This metric was obtained by dividing total engagement per month by the total number of posts by general users and fact checkers for the month.[4] In line with our observation that the overall increase in users' engagement is largely driven by one-off viral posts, per-post engagement decreased during the initial months of the pandemic (Figure 5). However, we choose to focus on total engagement as opposed to per-post engagement given our operationalization of the engagement metric.[5]

Lastly, the trends in supply of fact-checking content (see Figures 1 and 2) and the trends in engagement with the fact-checking content (see Figures 3 and 4) are remarkably different from each other. Taken together with our qualitative inquiry mentioned above, this indicates that the content of the fact-checking stories rather than the quantity seems to drive user engagement with debunking efforts. The onset of the pandemic coincided with increased fact-checking activity overall, not just related to the coronavirus, but did not have a sustained or systematic impact on user engagement with fact-checking activity. Our next finding addresses how much of the user engagement was driven by engagement with coronavirus-related content.

[4] At the organization/year/month level.

[5] We have operationalized engagement as the sum of likes, retweets, quote tweets, replies on fact-checking tweets, and the number of posts by users on fact-checking. We choose total engagement over per-post engagement because we aren't necessarily interested in seeing sustained engagement with each fact-checking post; rather, if we can observe increased engagement, even if with only a few posts, we know that more people have been exposed to fact checking and that fact checkers/governments/civil society organizations would need to devise ways to increase engagement across their posts. For instance, a tweet that gets 100 retweets will have a higher average engagement than two tweets that get 100 and 10 retweets each. (100 + 1 / = 101 vs 101 + 11 / 2 = 56). In this situation, the average engagement falls dramatically but total engagement increases (101 vs 112). Moreover, most of the engagement is driven by a few tweets as explained above in the text. Therefore, the decrease in the average engagement and increase in the total engagement may indicate that not only more people were engaged with few engagement-driving tweets but also many individual users with few followers that do not get much engagement from their followers also shared the fact-checking content.

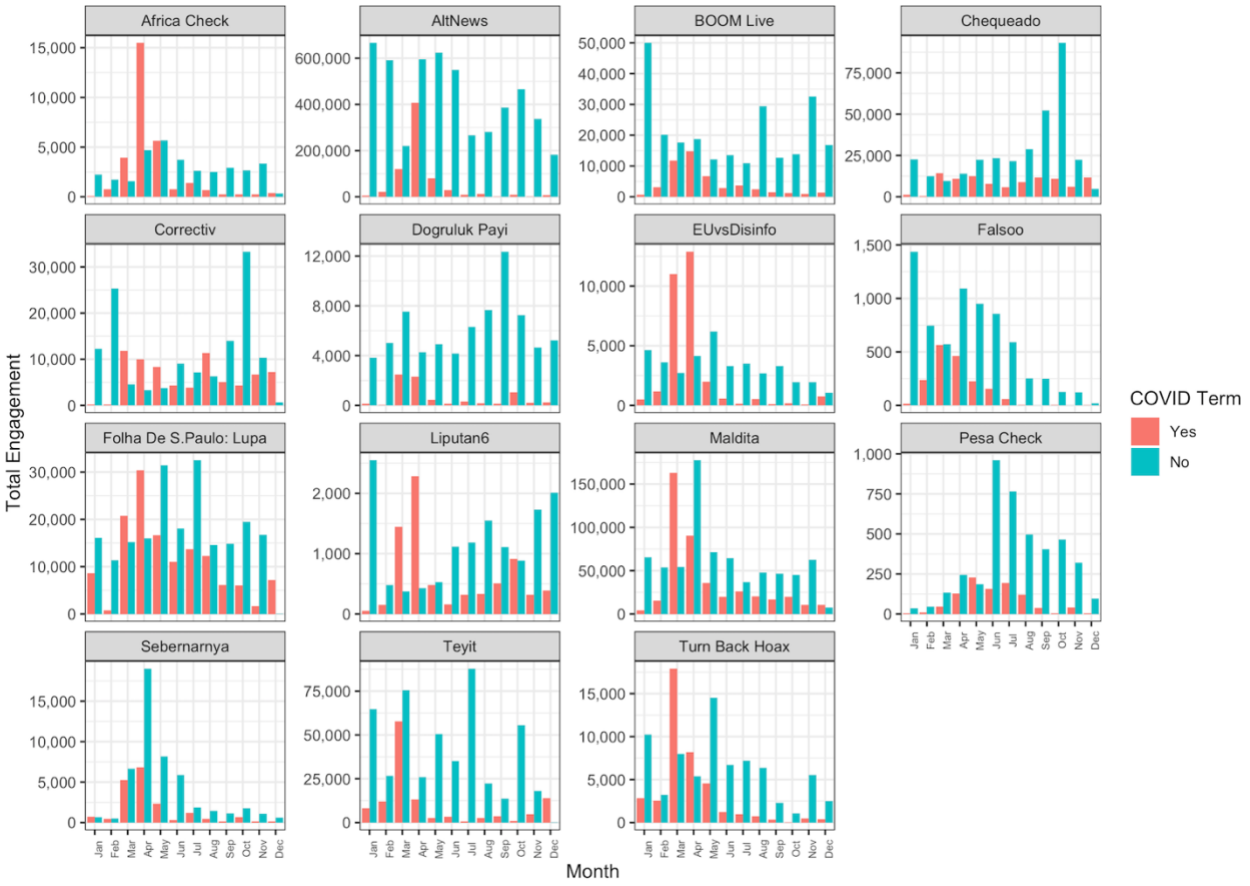

Figure 4. Distribution of the user engagement with coronavirus related and non-coronavirus related content. *The total engagement by users for COVID content included engagement with posts that included a URL to a fact-checking website and at least one of the keywords related to COVID. Engagement with non-COVID content included engagement with posts that included a URL to a fact-checking website, but no keywords related to COVID.*

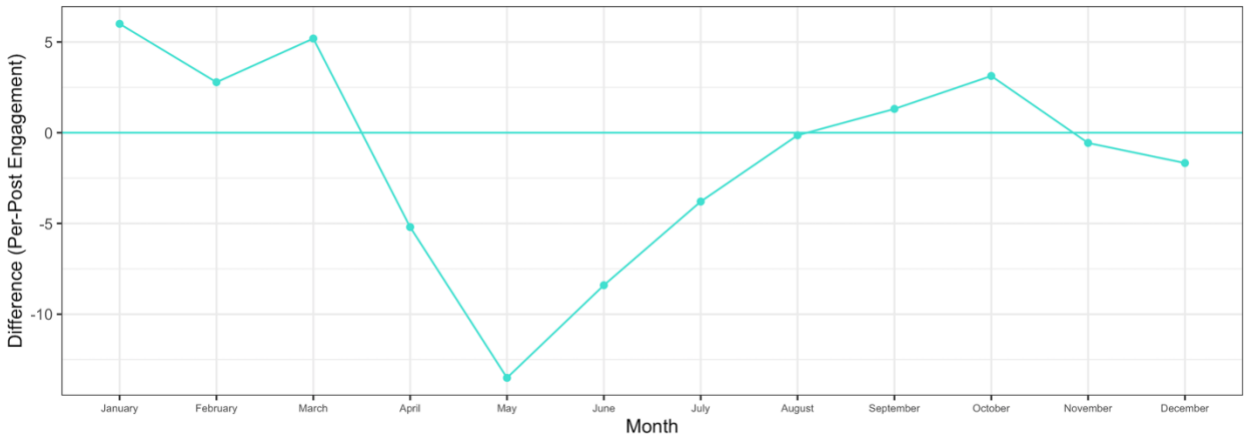

Figure 5. Difference in per-post engagement by users between 2020 and 2019. *Each data point represents the difference between 2020 and 2019 in per-post engagement for each month. Total engagement for a month was calculated by adding the total engagement for all fact-checking organizations. Per-post engagement for a month was calculated by dividing the total engagement for the month by the total number of posts (by users and fact checkers) for the month.*

Finding 4: There is heterogeneity across fact-checking organizations in how much overall engagement with fact-checking content can be explained by engagement with coronavirus-related content.

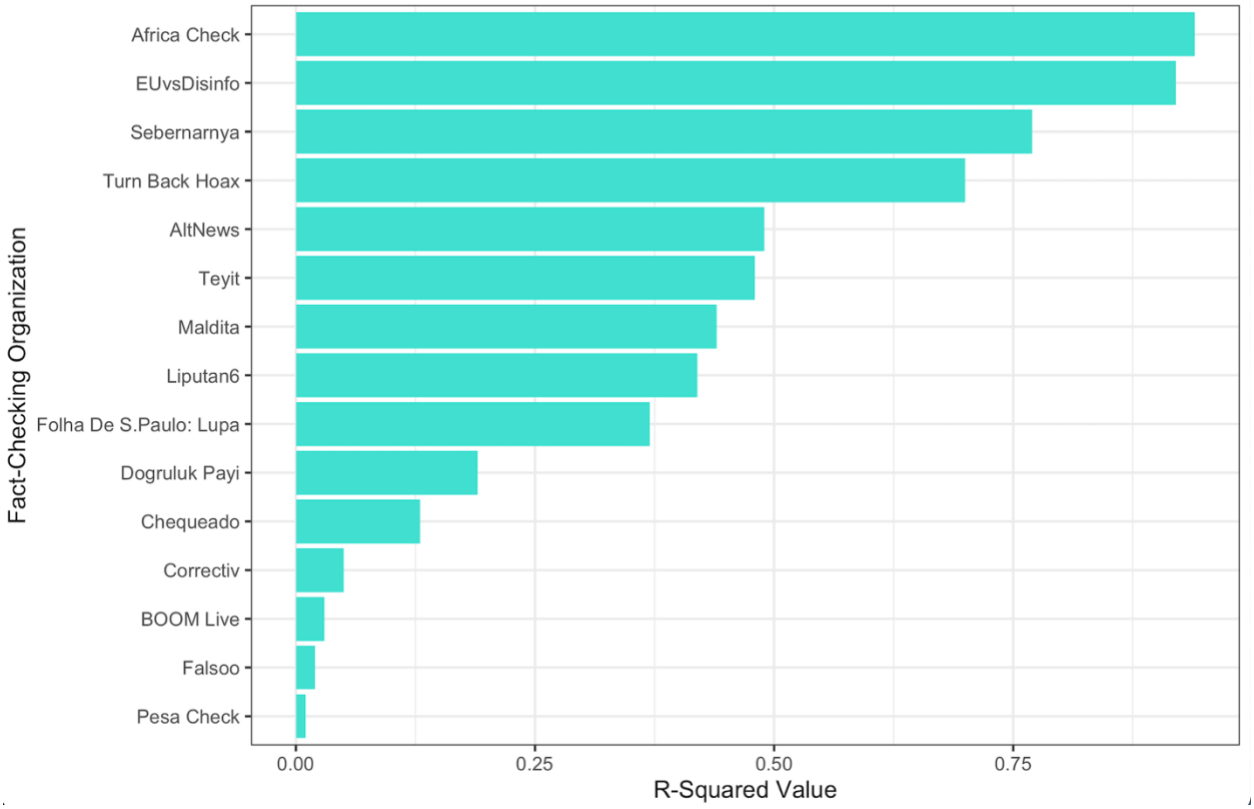

Figure 6. R-Squared Values for Association Between Total Engagement and COVID-Driven Engagement. *R-Squared values show how much of change in total engagement on fact-checking content is predicted by changes in total engagement on COVID-related content.*

To examine how much of the month-to-month variance in fact-checking was associated with COVID-related searches, we regressed the difference between periods in total user engagement on the difference in coronavirus-related engagement. The R^2 value in the regression tells us how much of the overall engagement is associated with engagement with coronavirus-related debunking; that is, a high R^2 value indicates that coronavirus-related debunking is strongly associated with the engagement with fact-checkers' content. Figure 6 suggests the associations between coronavirus-related engagement and overall engagement fall into three broad categories. For some organizations, such as Africa Check and EUvsDisinfo, coronavirus-related content is strongly associated with engagement in 2020. For others, such as Turn Back Hoax and Sebernarnya, coronavirus-related engagement is significantly associated with overall engagement with fact-checking content. For Altnews, Teyit, Maldita, Liputan 6, and Agencia Lupa, coronavirus-related content was associated with 37-49% of the total engagement in 2020 while for others, such as Dogruluk Payi, Pesa Check, Falsoo, BOOM Live, Correctiv, and Chequeado, coronavirus content was not associated with engagement to fact-checking content. Several reasons could explain these differences. First, coronavirus-related content comprised over 40% of the content shared by Africa Check and EUvsDisinfo, the highest in our dataset. Second, Sebernaryna and Turn Back Hoax cover countries where coronavirus was the most salient event of the year, as opposed to Teyit and Dogruluk Payi (which cover Turkey) and Chequeado (which covers Argentina), both countries that had local, politically salient events, such as the conversion of the Hagia Sofia museum in Turkey and the speculation of fraud regarding

Argentina's PASO primary elections respectively.

Methods

Did fact-checking organizations scale up efforts to debunk misinformation in 2020? Did debunking of coronavirus misinformation crowd out or substitute for other debunking efforts of fact-checking organizations? Did Twitter users engage more with fact-checking in 2020 than in previous years? To answer these questions, we used engagement metrics and tweet activity of 15 fact-checking organizations obtained using Twitter's API. We chose the 15 organizations based on two criteria: (1) frequency of the organization's occurrence in our coronavirus misinformation database (Shapiro et al. 2020) and (2) the amount of activity in 2019 and 2020 on its Twitter handle. Each of these organizations operates at either the regional level (e.g., Africa Check) or the local level (e.g., AltNews). Two of the organizations, Agencia Lupa and Liputan 6, are smaller bodies within larger news networks focused on fact-checking.

Using Twitter's research API, we collected data on several public engagement metrics (tweets by users, likes, quote tweets, replies, and retweets) as well as the number of tweets by an organization per month. We collected tweets from two sources: (1) fact-checking organizations' tweets that contained the URL of the fact-checking website for the general query and added coronavirus-related terms (see Table A1 in Appendix A for the full list of terms) for tweets that specifically dealt with coronavirus related misinformation, and (2) users' tweets which contained the URL of a fact-checking organization. We divided tweets of both types into those with and without terms related to debunking coronavirus misinformation. The Twitter activity of fact-checking organizations, or the "supply" of fact-checking activity, was measured through the number of tweets by a fact-checking organization every month. The "demand" for fact-checking was quantified through the total engagement metric, which included likes, retweets, quote tweets, replies, and the number of tweets by users that contained a link to a fact-checking organization. We used a sum of these metrics (as opposed to just likes or retweets) because it allowed us to capture all the ways in which users engage with fact-checking content on Twitter. While it is possible that the same user may have liked and retweeted a single tweet, we were not trying to capture the number of users engaging with tweets; rather, we wanted to look at all possible methods of engagement (and the levels of engagement), and it can be argued that liking and retweeting a tweet indicates greater engagement with a post than only liking or only retweeting. However, to ensure that this total engagement metric was not biased or did not capture falsely inflated user engagement, we ran robustness checks by plotting our figures solely based on either likes or retweets (see Figures B2 and B3 in Appendix B). We found similar patterns in user engagement when looking at only likes and only retweets.

Since some of the fact-checking organizations in our dataset were popular in non-English speaking regions, we augmented our search with region-specific terms for the coronavirus, such as those in Hindi, Arabic, and Turkish, among others. A list of these terms can be found in Table A1 in Appendix A. Overall, we use a set of terms related to the coronavirus in English and translate those into non-English languages as well as use locally relevant terms, as deemed appropriate by our team of local researchers, in non-English languages. We used this full list comprising both English and non-English terms each time we queried COVID tweets on all fact-checking sites. This allowed us to capture tweets with terms that are both globally and locally salient. We selected these keywords based on input from 30 Research Assistants who have analyzed over 5,600 coronavirus-related misinformation narratives from over 80 countries since March 2020.

We summed the engagement metrics to generate a variable for total engagement per organization per month. Using this total engagement variable, we conducted two statistical analyses: (1) a difference-in-differences analysis to gauge changes in the scale of activity and a (2) predictive linear regression to gauge how engagement with coronavirus-related tweets was driving overall engagement. The overall

change in fact-checking activity of fact-checking organizations or user engagement with fact-checking content on Twitter associated with the pandemic can be quantified by means of a difference-in-differences analysis. The idea is to compare the trends of the outcome variable of interest (Twitter engagement with fact-checking posts, in our case) before and after the starting month of the coronavirus pandemic, and between 2020 (the year in which the outbreak took place) and 2019 (a "control" year). As a robustness check, we ran the same analysis by specifying different months as the start of the pandemic, as well as including data points both for the whole year and until June. Website fixed effects were also included in order to eliminate organization-specific factors that might confound our estimates. In the predictive models, we first used differences and employed the change in the coronavirus engagement to predict the change in overall engagement in 2020. We also conducted a difference-in-differences analysis, the results of which can be found in Tables A2, A3, and A4 in Appendix A. These metrics and analyses are an imperfect but useful proxy to compare fact-checking activities before and after the pandemic. However, they are not without their limitations.

Our approach towards capturing tweets about coronavirus (i.e., tweets containing at least a coronavirus-related term and a URL that links to a fact-checking website in our study sample) could miss tweets about coronavirus if the body of the tweet did not consist of at least one coronavirus-related term from our list. Similarly, keyword-based matching could capture noisy data and label it as being about coronavirus. In order to test the gravity of this limitation, we took a random sample of 500 tweets from our full dataset that were categorized as COVID and another random sample of 500 tweets that were categorized as non-COVID and manually checked each of them.

For the sample containing COVID terms, the true positive rate is 100% (i.e., all COVID labeled tweets were accurately categorized as being about COVID and the false-positive rate is 0%). However, 20% of the data in the sample of non-COVID tweets were mislabeled as being not about COVID when they, in fact, were (i.e., the false-negative rate was 20%). Therefore, while we acknowledge that our methodology underestimates the number of tweets about COVID, we have confidence that it captures significant data for most, if not all, websites.

Additionally, although evidence of interaction or engagement with fact-checking content on social media is not proof of changing public opinion on accuracy of information, it still sheds light on the scale of fact-checking activity as well as the trends in user engagement. Aside from running surveys or controlled experiments to see how many people engage with fact-checking or how debunking changes public opinions, the social media engagement metrics are particularly useful in understanding the level of interaction between the suppliers (fact checkers) and the consumers (social media users). We are able to assess how the scale and composition of fact-checking activity changes over time by looking at social media activity of fact-checking organizations.

Lastly, our methodology prevents us from discussing how widespread engagement with fact-checking of coronavirus-related content is relative to all the content associated with coronavirus on Twitter. While this limitation prevents us from commenting on how relevant fact-checking is in the broader context of misinformation-prone events, it creates an opportunity for further empirical work in this direction, using either coronavirus or any other globally/locally salient event as a case study.

Bibliography

Burel, G., Farrell, T., Mensio, M., Khare, P., & Alani, H. (2020). Co-spread of misinformation and fact-checking content during the COVID-19 pandemic. *Social Informatics, 12467*, 28–42. https://doi.org/10.1007/978-3-030-60975-7_3

Chan, M. S., Jones, C. R., Hall Jamieson, K., & 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

Chen, K., Chen, A., Zhang, J., Meng, J., & Shen, C. (2020). Conspiracy and debunking narratives about COVID-19 origins on Chinese social media: How it started and who is to blame. *Harvard Kennedy School (HKS) Misinformation Review, 1*(8). https://doi.org/10.37016/mr-2020-50

Chequeado. (2020, September 4). *Alberto Fernández: "Hoy se consume la misma energía para producción industrial que se consumía el 19 de marzo. Y el consumo es más alto.*" [Alberto Fernández: "Today the same energy is consumed for industrial production that was consumed on March 19 and consumption is higher"]. Chequeado. https://chequeado.com/ultimas-noticias/alberto-fernandez-hoy-se-consume-la-misma-energia-para-produccion-industrial-que-se-consumia-el-19-de-marzo-y-el-consumo-es-mas-alto/

Clayton, K., Blair, S., Busam, J. A., Forstner, S., Glance, J., Green, G., Kawata, A., Kovvuri, A., Martin, J., Morgan, E., Sandhu, M., Sang, R., Scholz-Bright, R., Welch, A. T., Wolff, A. G., Zhou, A., & Nyhan, B. (2020). Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. *Political Behavior*, *42*(4), 1073–1095. https://doi.org/10.1007/s11109-019-09533-0

Domínguez, J. J., Giménez, J., & Di Santi, M. (2020, September 24). *Esteban Bullrich: "Claramente hubo en las PASO un fraude muy, muy grande."* [Esteban Bullrich: "Clearly there was a very, very big fraud in the PASO"]. Chequeado. https://chequeado.com/ultimas-noticias/esteban-bullrich-claramente-hubo-en-las-paso-un-fraude-muy-muy-grande/

Erdoğan, E., Uyan-Semerci P. (2021). *'İnfodemi' İle Etkin Mücadele İçin Bireylerin Yanlış Bilgi Karşısındaki Tutumlarının ve Bu Tutumların Belirleyicilerinin Araştırılması: COVID-19 Örneği Çalışmasının Ön Bulguları* [Investigation of individuals' attitudes towards false information and determinants of these attitudes for effective fighting with 'infodemic': The case of COVID-19 study's preliminary findings report]. Turkuazlab. https://drive.google.com/file/d/1zxTiejG0ex5uqomcnDP4ykgek9eOjVMS/view

Jakesch, M., Koren, M., Evtushenko, A., & Naaman, M. (2018). The role of source, headline and expressive responding in political news evaluation. [Un-reviewed manuscript]. *Social Science Research Network.* https://doi.org/10.2139/ssrn.3306403

Keskin, Ö. H. (2020, July 8). *İBB'nin satın aldığı Fatih tablosunun sahte olduğu iddiası* [The claim that the Fatih painting purchased by IMM is fake]. Teyit. https://teyit.org/ibbnin-satin-aldigi-fatih-tablosu-sahte-oldugu-iddiasi

Korkmaz, B. (2020, July 23). *Yunanistan'da Ayasofya için ulusal yas ilan edildiği iddiası* [The claim that national mourning was declared for Hagia Sophia in Greece]. Teyit. https://teyit.org/yunanistan-ayasofya-icin-ulusal-yas-ilan-etmedi

Nyhan, B., & Reifler, J. (2015). The effect of fact-checking on elites: A field experiment on U.S. state legislators. *American Journal of Political Science*, *59*(3), 628–640. https://doi.org/10.1111/ajps.12162

Oledan, J., Ilhardt, J., Musto, G., & Shapiro, J. N. (2020, June 25). Fact-checking networks fight coronavirus infodemic. *Bulletin of the Atomic Scientists*. https://thebulletin.org/2020/06/fact-checking-networks-fight-coronavirus-infodemic/?utm_source=Newsletter&utm_medium=Email&utm_campaign=Newsletter06252020&utm_content=DisruptiveTechnologies_InfodemicFactcheckingNetworks_06252020

Pennycook, G., Bear, A., Collins, E. T., & Rand, D. G. (2020). The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management Science*, *66*(11), 4944–4957. https://doi.org/10.1287/mnsc.2019.3478

Pennycook, G., & Cannon, T. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology General*, *147*(12), 1865–1880. https://doi.org/10.1037/xge0000465

Robertson, C. T., Mourão, R. R., & Thorson, E. (2020). Who uses fact-checking sites? The impact of demographics, political antecedents, and media use on fact-checking site awareness, attitudes, and behavior. *The International Journal of Press/Politics*, *25*(2), 217–237. https://doi.org/10.1177/1940161219898055

Roozenbeek, J., van der Linden, S., & Nygren, T. (2020). Prebunking interventions based on "inoculation" theory can reduce susceptibility to misinformation across cultures. *Harvard Kennedy School (HKS) Misinformation Review*, *1*(2). https://doi.org/10.37016//mr-2020-008

Siwakoti, S., Yadav, K., Thange, I., Bariletto, N., Zanotti, L., Ghoneim, A., & Shapiro, J. N. (2021). Localized misinformation in a global pandemic: Report on COVID-19 narratives around the world. *Empirical Study of Conflict Project, Princeton University,* 1–68. https://esoc.princeton.edu/publications/localized-misinformation-global-pandemic-report-covid-19-narratives-around-world

Shapiro, J. N., Oledan, J, & Siwakoti, S. (2020). *ESOC COVID-19 misinformation dataset* [Data set]. Empirical Studies of Conflict Project, Princeton University. https://esoc.princeton.edu/publications/esoc-covid-19-misinformation-dataset

Silverman, T., Stewart, C. J., Amanullah, Z., & Birdwell, J. (2016). *The impact of counter-narratives: Insights from a year-long cross-platform pilot study of counter-narrative curation, targeting, evaluation and impact.* Institute for Strategic Dialogue. https://www.isdglobal.org/wp-content/uploads/2016/08/Impact-of-Counter-Narratives_ONLINE_1.pdf

Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research*, *31*(4), 713–724. https://doi.org/10.1086/426605

Spring, M. (2020, July 15). *Wayfair: The false conspiracy about a furniture firm and child trafficking.* BBC News. https://www.bbc.com/news/world-53416247

Yılmaz, M. C. (2020, July 14). *Osmanlı'da eşcinselliğin 1858'de suç olmaktan çıkarıldığı iddiası* [The claim that Ottomans decriminalized homosexuality in 1858]. Teyit. https://teyit.org/osmanlida-escinsellik-1858de-suc-olmaktan-cikarildi

Yilmaz, M. C. (2020). *Information disorder in times of the COVID-19 pandemic: Misinformation, news consumption and fact-checking in Turkey*. Tandans Data Science Consulting & Teyit. https://drive.google.com/file/d/1bD4-cpqvbdcXlYvJf5s0l15aafrbtY2d/view

Walter, N., Brooks, J. J., Saucier, C. J., & Suresh, S. (2020). Evaluating the impact of attempts to correct health misinformation on social media: A meta-analysis. *Health Communication*, 1–9. https://doi.org/10.1080/10410236.2020.1794553

Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2020). Fact-checking: A meta-analysis of what works and for whom. *Political Communication*, *37*(3), 350–375. https://doi.org/10.1080/10584609.2019.1668894

Wintersieck, A. L. (2017). Debating the truth: The impact of fact-checking during electoral debates. *American Politics Research*, *45*(2), 304–331. https://doi.org/10.1177/1532673X16686555

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

Funding
The research team is supported by Microsoft Research.

Competing interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethics
The data for the project were obtained from publicly available sources and via approved access to Twitter's academic research API and thus were exempt from IRB review.

Data Availability
All materials needed to replicate this study are available via the Harvard Dataverse: https://doi.org/10.7910/DVN/YMU2PO
Due to Twitter's privacy policies, we are only able to share tweet IDs and not the raw tweets analyzed for this study.

Appendix A: Tables

Table A1. *List of coronavirus-related terms used in analysis.*

coronavirus, corona, koronavirus, wuhancoronavirus, kungflu, n95, covid-19, corona virus, covid19, sars-cov-2, covd, pandemic, coronapocalypse, chinese virus, chinavirus, cronvirus, كورونا, وباء كورونا, فيروس كورونا, كوفيد, اصاب, حظر تجول, حظر تجوال, لقاح, وباء, كوفيد-١٩ virus, Korona, koronavirüs, aşı, mutasyon, virusi vya corona, कोविड, कोरोना , कोरोना वाइरस, कोभिड, चीनी वाइरस , चिन भाइरस , कोभिड-19

Table A2. *Difference-in-differences analyses – Engagement.*

The estimating equation for the model is:

$$Engagement_{j,i,t} = \alpha + \beta * did_{j,i,t} + \delta * pandemic\_dummy_{j,t} + \gamma * 2020\_dummy_{j,i} + \mathbf{X}'_j + \epsilon_{j,i,t}$$

Here, $Engagement_{j,i,t}$ refers to the total engagement metric (sum of likes, quote tweets, replies, retweets and no. of tweets by users) for website j in year i and month t; $did_{j,i,t}$ is the difference-in-differences indicator for website j, year i, month t (obtained as $2020\_dummy_{j,i} * pandemic\_dummy_{j,t}$); $pandemic\_dummy_{j,t}$ is an indicator equal to 1 for observations on website j in pandemic months (variable) t, 0 for pre-pandemic months; $2020\_dummy_{j,i}$ is an indicator equal to 1 for year 2020, 0 for 2019 observations for website j and year i; $\mathbf{X}'_j$ are website fixed effects and $\epsilon_{j,i,t}$ is the error term.

Table A2: Difference-in-differences, total engagement

	First pandemic month: March		First pandemic month: April	
	All year	Jan-Jun	All year	Jan-Jun
Diff-in-diff	-13252.2	3203.8	-10733.0	10259.4
	(-0.80)	(0.16)	(-0.81)	(0.50)
Pandemic months dummy	4660.1	8471.8	1698.5	6076.7
	(0.61)	(0.65)	(0.26)	(0.50)
2020 dummy	31211.1*	31211.1*	28217.3**	28217.3*
	(2.09)	(2.14)	(2.65)	(2.54)
Observations	360	180	360	180
Website FE	YES	YES	YES	YES

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Difference-in-differences on total engagement (sum of likes, quotes, retweets, replies to, and number of, tweets containing a link to fact-checking organizations' websites). While we detect a significant increase in engagement in 2020 compared to 2019, there is no specific association with the pandemic period.

Table A3. *Difference-in-differences analyses – Posts.*

The estimating equation for the supply-side model is:

$$Posts_{j,i,t} = \alpha + \beta * did_{j,i,t} + \delta * pandemic\_dummy_{j,t} + \gamma * 2020\_dummy_{j,i} + \mathbf{X}'_j + \epsilon_{j,i,t}$$

Here, $Posts_{j,i,t}$ is the total number of posts by website j in year i and month t; $did_{j,i,t}$ is the difference-in-differences indicator for website j, year i, month t (obtained as $2020\_dummy_{j,i} * pandemic\_dummy_{j,t}$); $pandemic\_dummy_{j,t}$ is an indicator equal to 1 for observations on website j in pandemic months (variable) t, 0 for pre-pandemic months; $2020\_dummy_{j,i}$ is an indicator equal to 1 for year 2020, 0 for 2019 observations for website j and year i; $\mathbf{X}'_j$ are website fixed effects and $\epsilon_{j,i,t}$ is the error term.

Table A3: Difference-in-differences, number of posts

	First pandemic month: March		First pandemic month: April	
	All year	Jan-Jun	All year	Jan-Jun
Diff-in-diff	-50.13	-16.72	-64.91	-35.82
	(-0.83)	(-0.27)	(-1.34)	(-0.71)
Pandemic months dummy	83.94	77.58	67.58	64.91
	(1.62)	(1.40)	(1.77)	(1.60)
2020 dummy	115.7*	115.7*	122.5**	122.5**
	(2.07)	(2.00)	(2.96)	(2.89)
Observations	356	180	356	180
Website FE	YES	YES	YES	YES

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Difference-in-differences analysis on the number of tweets by fact-checking organizations' handles. While we detect a significant increase in the number of posts in 2020 compared to 2019, there is no specific association with the pandemic period.

Table A4. *Difference-in-differences analyses - Per-post engagement.*

The estimating equation for the model is:

$$Per\_post\_engagement_{j,i,t} = \alpha + \beta * did_{j,i,t} + \delta * pandemic\_dummy_{j,t} + \gamma * 2020\_dummy_{j,i} + \mathbf{X}'_j + \epsilon_{j,i,t}$$

Here, $Per\_post\_engagement_{j,i,t}$ refers to the per-post engagement metric (sum of likes, quote tweets, replies, retweets and no. of tweets by users, divided by the total no. of tweets by **both users and fact-checking websites' handles**) for website j in year i and month t; $did_{j,i,t}$ is the difference-in-differences indicator for website j, year i, month t (obtained as $2020\_dummy_{j,i} * pandemic\_dummy_{j,t}$); $pandemic\_dummy_{j,t}$ is an indicator equal to 1 for observations on website j in pandemic months (variable) t, 0 for pre-pandemic months; $2020\_dummy_{j,i}$ is an indicator equal to 1 for year 2020, 0 for 2019 observations for website j and year i; $\mathbf{X}'_j$ are website fixed effects and $\epsilon_{j,i,t}$ is the error term.

Table A4: Difference-in-differences, per-post engagement

	First pandemic month: March		First pandemic month: April	
	All year	Jan-Jun	All year	Jan-Jun
Diff-in-diff	-6.811*	-9.872**	-7.944**	-13.69***
	(-2.42)	(-2.80)	(-3.18)	(-3.82)
Pandemic months dummy	2.128	6.383*	2.415	8.586**
	(1.06)	(2.27)	(1.31)	(2.80)
2020 dummy	4.392	4.392	4.658*	4.658*
	(1.81)	(1.72)	(2.36)	(2.31)
Observations	356	180	356	180
Website FE	YES	YES	YES	YES

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Difference-in-differences on per-post engagement (sum of likes, quotes, retweets, replies to, and number of, tweets containing a link to fact-checking organizations' websites, divided by the total no. of tweets by both users and fact-checking organizations' websites' handles containing a link to fact-checking organizations' websites). In all specifications, we detect a statistically significant reduction in average engagement associated with the pandemic months and as compared to the control year (2019).

Appendix B: Figures

Figure B1. Difference in total engagement by users. *The difference in total engagement by users was calculated by subtracting total engagement by users in 2019 from the total engagement by users in 2020.*

Robustness checks for engagement metric

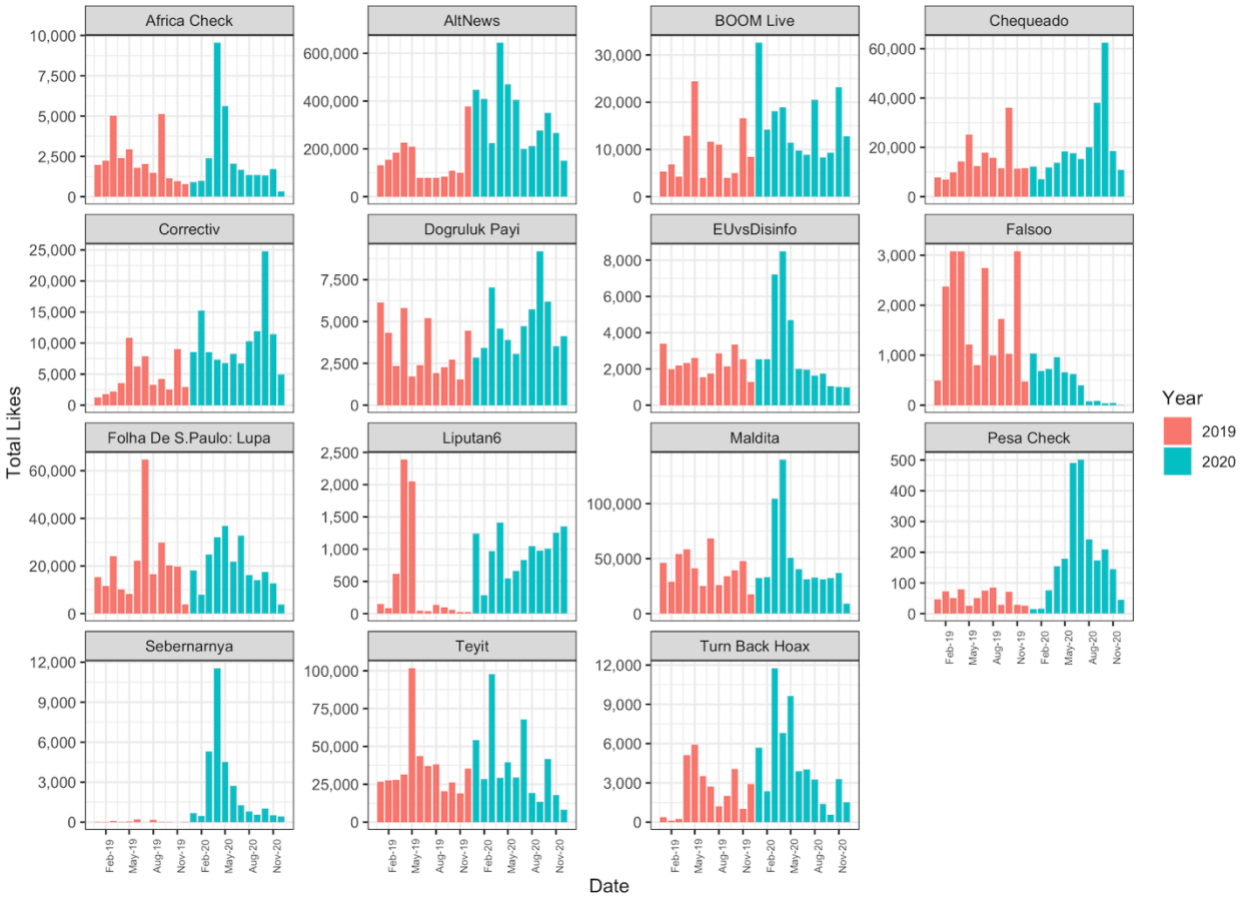

Figure B2. Distribution of userlLikes on fact-checking content. *The user engagement measured as likes on fact-checking content in 2019 and 2020 is shown as a time series.*

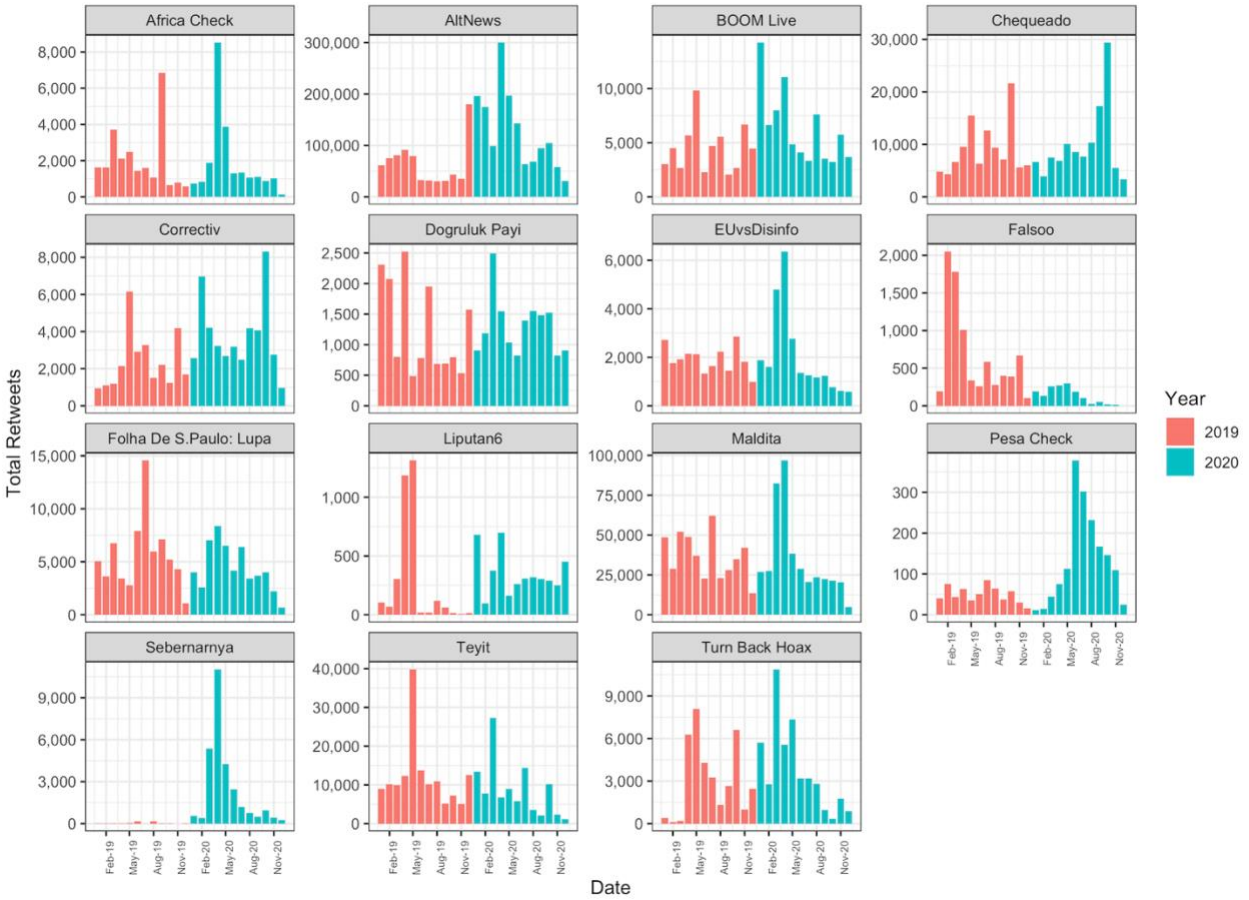

Figure B3. Distribution of user retweets of fact-checking content. *The user engagement measured as retweets of fact-checking content in 2019 and 2020 is shown as a time series.*

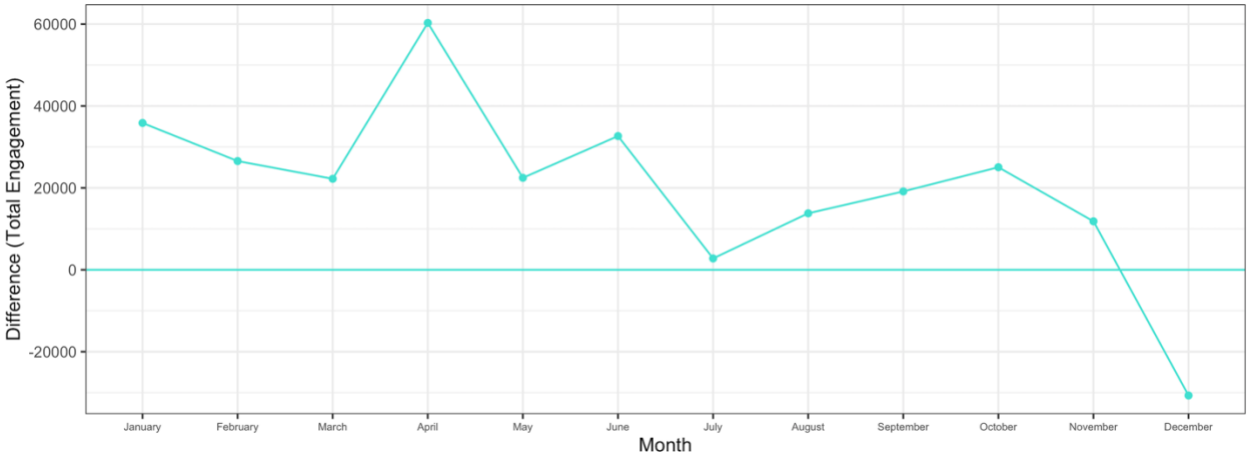

Figure B4. Difference in total engagement by users between 2020 and 2019. *Each data point represents the difference between 2020 and 2019 in total engagement for each month. Total engagement for a month was calculated by adding the total engagement for all fact-checking organizations.*

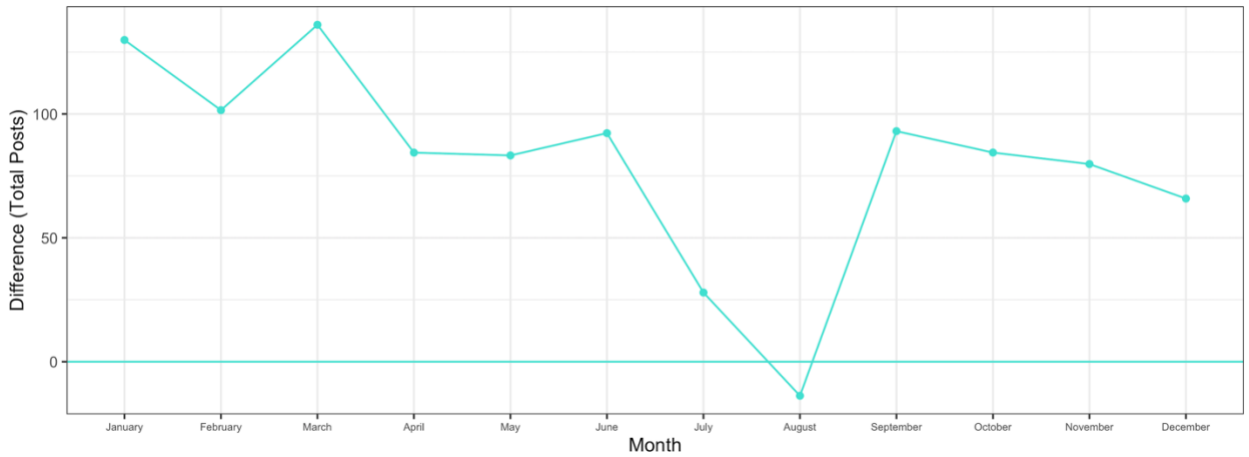

Figure B5. Difference in total posts by fact-checkers between 2020 and 2019. *Each data point represents the difference between 2020 and 2019 in total posts by fact-checkers for each month. Total posts for a month are calculated by adding the total posts by all fact-checking organizations in a month.*