Investigating Facebook's policy against accounts that repeatedly share misinformation: implementation and impact

Anonymous TTO submission

Abstract

Like many web platforms, Facebook is under pressure to regulate misinformation. According to the company, users that repeatedly share misinformation ('repeat offenders') will have their distribution reduced, but little is known about the implementation or the efficiency of this measure. First, combining data from a fact-checking organization and CrowdTangle, we identified a set of public accounts (groups and pages) that have shared misinformation repeatedly. While we observe a decrease in engagement for pages (median of -43%) after they shared two or more 'false news', such a reduction is not observed for groups. However, we discover that groups have been affected in a different way with a sudden drop in their average engagement per post that occurred around June 9, 2020. No public information was given by Facebook about this sudden decrease. This drop has cut groups' engagement per post in half, but it was compensated by the fact that the overall activity of 'repeat offenders' has doubled between 2019 and 2020. Second, we identified pages that have been warned by Facebook and have shared a screenshot of the 'reduced distribution' notification they have received. We found that their engagement per post following the notification decreased by a modest amount (median of -24%), with some popular pages actually gaining more engagement. Two steps Facebook could take to reduce misinformation is to enforce their 'repeat offenders' policy more strictly on pages, and to start applying it to groups.

1 Introduction

The general public is increasingly getting news related information online, through search engines, social media and video platforms (Mitchell et al., 2016). Hence the spread of misinformation through these platforms has recently received growing attention. Recent studies, along with the political context of January 2021 in the United States, show how the presence of misinformation online can contribute to negative societal consequences. Namely it can fuel false beliefs, such as the idea of a massive voter fraud during the US 2020 presidential election, which may have led to the January 6, 2021 insurrection at the U.S. Capitol (Benkler et al., 2020) and other false stories about presidential candidates (Allcott and Gentzkow, 2017). Misinformation has also confused the public about the reality of climate change (Brulle, 2018; Porter et al., 2019) and stoked skepticism about vaccine safety among the public (Featherstone and Zhang, 2020; Lahouati et al., 2020). In April 2020, a questionnaire from the Reuters Institute found that people in the UK use online sources more often than offline sources when looking for information about the coronavirus. Among social media platforms, Facebook was the most widely used with 24% of the respondents saying they used Facebook to access COVID-19 information in the last seven days (Fletcher et al., 2020). The importance of Facebook in the media landscape is confirmed by Parse.ly's dashboard, which shows that 25% of the visitors of 2500+ media websites are referred by Facebook[1].

Lawmakers and regulators are increasingly pressuring platforms to limit the spread of misinformation. In the US, the House of Representatives organized hearings and convened representatives of the main platforms to testify on how they are being weaponized to spread "misinformation and conspiracy theories online" (Donovan et al., 2020). In Europe, the European Commission has established a 'Code of Practice on Disinformation'[2] that enjoins platforms to voluntarily

[1] https://www.parse.ly/resources/data-studies/referrer-dashboard, accessed on 2021-07-08.

[2] https://ec.europa.eu/digital-single-market/en/code-

comply with a set of commitments (Heldt, 2019). However, there is little data available and few established processes to monitor the implementation of these measures and quantify their actual impact. Here we propose a methodology to monitor Facebook's implementation of its policy to reduce the visibility of accounts repeatedly spreading misinformation. We chose to focus on Facebook as it is the biggest social media platform with more than two billion users worldwide.

Facebook announced a three-part policy to address 'misleading or harmful content': they claim to *remove* harmful information, *reduce* the spread of misinformation and *inform* people with additional context[3]. Facebook has developed the most extensive third-party fact-checking program with dozens of partner institutions to assist the company in this endeavour[4]. Facebook informs page or group owners when published posts on their pages or groups are marked as misinformation, inviting them to correct the posts. Facebook also states that the virality of the posts marked as 'False' or 'Partly False' will be reduced.

Facebook's *reduce* policy is not only applied to individual posts, but also to organizations and communities that often publish posts containing misinformation, as indicated by this statement in their publishers' help center[5]:

> *Pages and websites that repeatedly share misinformation rated False or Altered will have some restrictions, including having their distribution reduced.*

Facebook ranks each post by assigning to it a relevancy score, where a high score leads to a high likelihood of the post to appear on a user's newsfeed. Doing so, Facebook can make a post or a whole account less visible by decreasing the relevancy score of its content; this is precisely the *reduce* measure[6].

So far Facebook has not provided data showing how their *reduce* policy is implemented, which would allow researchers to quantify its impact on the spread of misinformation. To the best of our knowledge, the impact of the *reduce* policy has not yet been audited directly. Hence the present research article departs from articles studying the overall levels of misinformation on platforms (Allcott et al., 2019; Kornbluh et al., 2020; Resnick et al., 2018), by focusing on monitoring a specific policy against misinformation.

We used CrowdTangle, a public insights tool owned and operated by Facebook, to access Facebook data (Team, 2021). CrowdTangle exclusively tracks public content, and provides access to engagement metrics (such as the number of likes, shares and comments), but not to the reach (number of views) of content[7]. We first investigated how Facebook enforces its 'reduce' policy by combining data from one of Facebook's fact-checking partners identifying URLs sharing misinformation and tracking engagement metrics of the Facebook accounts that repeatedly share such misinformation. We further investigated the effects of Facebook's policy on engagement metrics of a set of Facebook pages claiming to be under reduced distribution.

2 Investigating the 'reduce' policy on Facebook accounts repeatedly sharing misinformation

To investigate the effect of fact-checking on Facebook accounts that repeatedly share misinformation, we first analyzed the engagement per post received by these accounts. One would expect this metric to decline if the accounts' posts become less visible in Facebook's feed.

2.1 Methods

We used data from Science Feedback, which is part of Facebook's third-party fact-checking program[8]. Science Feedback is a fact-checking organization, where academics review the credibility of science-related claims and articles. We obtained from Science Feedback a list of 4,000+ URLs reviewed by its team. We relied on the 2,452 URLs marked as 'False', which we refer to as 'false news links', excluding the URLs marked

practice-disinformation.

[3] https://about.fb.com/news/2018/05/inside-feed-reduce-remove-inform/

[4] https://www.facebook.com/business/help/341102040382165

[5] https://www.facebook.com/business/help/2593586717571940, https://www.facebook.com/business/help/297022994952764

[6] https://about.fb.com/news/2018/05/inside-feed-reduce-remove-inform/

[7] https://help.crowdtangle.com/en/articles/3192685-citing-crowdtangle-data, https://help.crowdtangle.com/en/articles/4558716-understanding-and-citing-crowdtangle-data

[8] https://sciencefeedback.co/science-feedback-partnering-with-facebook-in-fight-against-misinformation/

as 'Partly False', 'Missing Context', 'False headlines' or 'True', as well as the URLs marked as 'False' but 'corrected' by the publisher, because these labels do not contribute to the 'repeat offender' status according to Facebook's guidelines. The list of 'false news links' was obtained on January 4, 2021 and cover links flagged in 2019 and 2020.

Using the '/links' endpoint from the CrowdTangle API, we collected the public Facebook groups and pages that shared at least one false news link between January 1, 2019 and December 31, 2020. Due to the API limitations, if a URL was shared in more than 1000 posts, we collected only the 1000 posts that received the highest number of interactions[9]. We focused on the accounts that spread misinformation the most often, choosing a threshold of 24 different false news links shared over the past two years.

The corresponding 307 Facebook accounts (289 Facebook groups and 18 Facebook pages) are referred to as 'repeat offenders accounts'. All the posts they published between January 1, 2019 and December 31, 2020 were collected using the '/posts' endpoint. We calculated the engagement per post by summing the number of comments, shares and reactions (such as 'like', 'love', 'favorite', 'haha', 'wow', 'sad' and 'angry' reactions) that each post has received.

'Repeat offenders' accounts are supposed to have their distribution reduced, according to Facebook's official communication, but the precise rule Facebook uses to classify an account as 'repeat offenders' is not specified. An undisclosed source obtained by a journalist indicated that "The company operates on a 'strike' basis, meaning a page can post inaccurate information and receive a one-strike warning before the platform takes action. Two strikes in 90 days places an account into 'repeat offender' status".[10]

Based on this 'two strikes in 90 days' rule and the list of strike dates known by Science Feedback, we inferred periods during which each account must have been under repeat offender status. If a post shares a misinformation link which was previously fact-checked as 'False', we used the date of the post as the strike date. However, if an account shares a link, which later gets fact-checked as 'False', then the fact-check date was used as the strike date. A repeat offender period is defined as any given time in which an account shared two or more 'false news links' over the past 90 days (see Figure 1 for an example).

2.2 Results

Figure 1 displays the engagement metrics for one 'repeat offender' group named 'Australian Climate Sceptics Group'. The known strike dates appear as red lines at the bottom and the inferred 'repeat offender' periods are shaded in red. The average engagement per post varies throughout the past two years, but does not appear to be related with the shift between 'repeat offender' and 'no strike' periods (see Figure 1). We compared the average engagement metrics between the 'repeat offender' and the 'no strike' periods, expecting a decrease in engagement during the 'repeat offender' periods. However we observe a 61% increase in engagement.

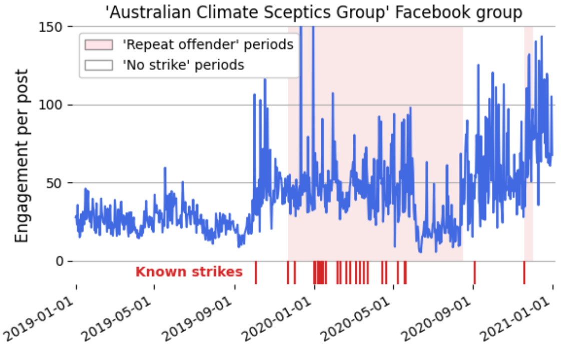


Figure 1: Average engagement (the sum of comments, shares, likes, ...) per post for the 'Australian Climate Sceptics Group' Facebook group for each day in 2019 and 2020. Each red line at the bottom represents the date of a known strike for this group, and the areas shaded in red represent the 'repeat offender' periods as defined by the 'two strikes in 90 days' rule.

To provide a general overview, we calculate the percentage change between the 'repeat offender' and the 'no strike' periods for each of the 256 Facebook accounts that have published at least one post during each period (see Figure 2).[11] The average percentage change is 7%, and the median −6%. A Wilcoxon test shows that the values are

[9] https://github.com/CrowdTangle/API/wiki/Links

[10] https://www.nbcnews.com/tech/tech-news/sensitive-claims-bias-facebook-relaxed-misinformation-rules-conservative-pages-n1236182

[11] The percentage changes were calculated on the periods between January 1, 2019 and June 8, 2020. Because of the drop in engagement described further, the second semester of 2020 was excluded for its vastly diminished and not representative engagement level (see Figure 3).

not significantly different from zero (W = 16051, p-value = 0.74).

When we consider groups and pages separately, the percentage changes are different for the two. The median percentage change for Facebook groups is -3% (not significantly different from zero), while the median for Facebook pages is -43%. A Wilcoxon test applied only to the Facebook pages' percentage changes, shows they are significantly different from zero (W = 21, p-value = 0.0034).

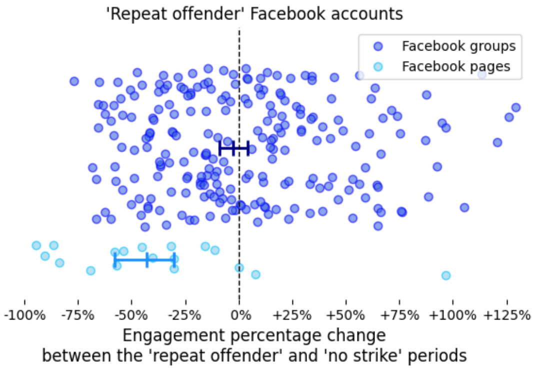


Figure 2: Percentage changes between the average engagement per post during the 'repeat offender' periods and the 'no strike' periods. Each deep blue dot represents a Facebook group, and each light blue dot a Facebook page. The bars show the median and its 90% confidence interval. Confidence intervals are estimated using a bootstrap method.

To see whether the strikes would otherwise influence the repeat offenders accounts' engagement over time, we analyzed the total amount of engagement received by all the posts published by each of the 307 repeat offenders accounts for each day of the 2019-2020 period (Figure 3). This metric, representing the total engagement generated by these accounts on Facebook (top panel), can be decomposed as the number of posts published each day (middle panel) times the average number of engagement per post (bottom panel).

The total engagement per day is stable from January to September 2019, however we observe a rise from September 2019 to June 2020. This rise is explained by the increase in activity of the misinformation accounts (with a doubling of the number of posts per day) while the engagement per post remained rather constant. Around June 9, 2020, the total engagement metrics have massively dropped. This decrease is entirely explained by a corresponding drop in engagement per post (Figure 3).

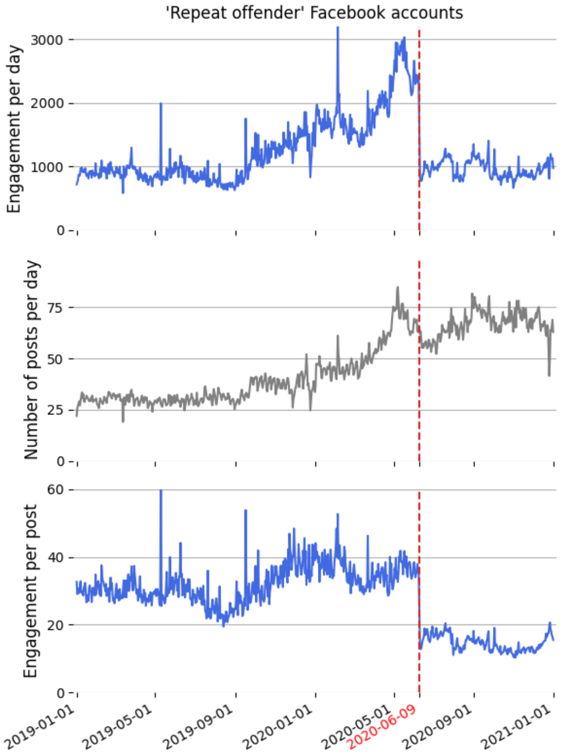


Figure 3: Metrics aggregated over the 307 Facebook accounts that repeatedly shared false news links. **(Top panel)** Engagement per day. **(Middle panel)** Number of posts per day. **(Bottom panel)** Average engagement per post. The dotted line marks the date of June 9, 2020, when a sudden drop in engagement is observed.

To further quantify this 'June drop', we calculated the percentage change in engagement for each account between 30 days before June 9, 2020 and 30 days after (Figure 4). The average percentage change is -21%, and the median -43%. Most of the accounts (219 out of 289) experienced a decrease in engagement[12], and a Wilcoxon test indicates that these percentage changes are significantly different from zero (W = 9012, p-value = 4.6×10^{-17}).

It appears that the Facebook pages are not affected by this decrease, with a median percentage change of -5%, while the groups have a median percentage change of -45%. When tested separately, the Facebook pages' percentage changes are not significantly different from zero (W = 73, p-value = 0.61).

To verify whether this drop was specific to this set of groups, we compared these dynamics to

[12] A decrease in engagement on June 9, 2020 can be seen for the 'Australian Climate Sceptics Group' in Figure 1 (the percentage change was -60% for this example).

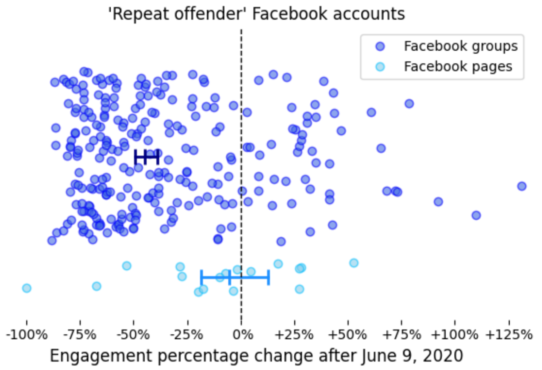


Figure 4: Percentage changes between the average engagement per post 30 days before June 9, 2020 and 30 days after. Each deep blue dot represents a Facebook group, and each light blue dot a Facebook page. The bars show the median and its 90% confidence interval.

those of a control set of accounts consisting of Facebook pages and groups associated with established news outlets that did not publish misinformation. No such drop in total or per post engagement metrics was observed around June 9, 2020.

We can only explain such a massive change by a modification in how Facebook's algorithm promoted the content from these groups starting on June 9, 2020. For 'repeat offenders' pages, we did observe a relationship between the strike dates and a decrease in engagement. For 'repeat offenders' groups, we observed no such link. Hence it seems that Facebook only took action against these groups with the June one-shot measure.

One limitation of the results described in this section is that we obtained the links labelled as 'False' from only one fact-checking organization (Science Feedback), while Facebook partners with over 60 fact-checking organizations[13]. The true 'repeat offender' periods could thus be longer than the ones inferred, potentially changing the magnitude of the 'reduce' effect.

3 Investigating the 'reduce' policy on self-declared 'repeat offenders'

In the present section, we used a different methodology to collect accounts for which we are sure that they are under 'repeat offender' status.

3.1 Methods

We noticed that two popular pages ('Mark Levin' and '100 Percent FED Up') have publicly shared a message claiming to be placed under 'repeat offender' status with a screenshot as a piece of evidence. To gather a list of such self-declared repeat offenders, we searched on CrowdTangle for posts published since January 1, 2020 with the following keywords:

- 'reduced distribution' AND ('restricted' OR 'censored' OR 'silenced')
- 'Your page has reduced distribution'

For this we used the '/posts/search' endpoint of the API on November 25, 2020.

We manually opened the resulting posts, and kept the ones which met the following criteria (see Figure 5 top panel for an example):

- The post should include a screenshot of the Facebook notification.
- In the screenshot, the Facebook notification should say: 'Your page has reduced distribution and other restrictions because of repeatedly sharing of false news.'
- In the screenshot, the name of the page should be visible.

Doing so, we obtained a list of 94 pages. We found only Facebook pages in this case, and no groups. A search using the terms 'Your group has reduced distribution' did not yield any result.

To verify whether Facebook applied any restriction to these pages, we collected all the posts that these 94 pages have published between January 1, 2019 and December 31, 2020 from the CrowdTangle API using the '/posts' endpoint. The collection was run on January 11, 2021. We were only able to collect data from 83 of these pages, as 11 were deleted from the CrowdTangle database since our search in November 2020. This highlights an important issue when studying misinformation trends on Facebook: some data disappears as accounts are deleted or changed to 'private'.

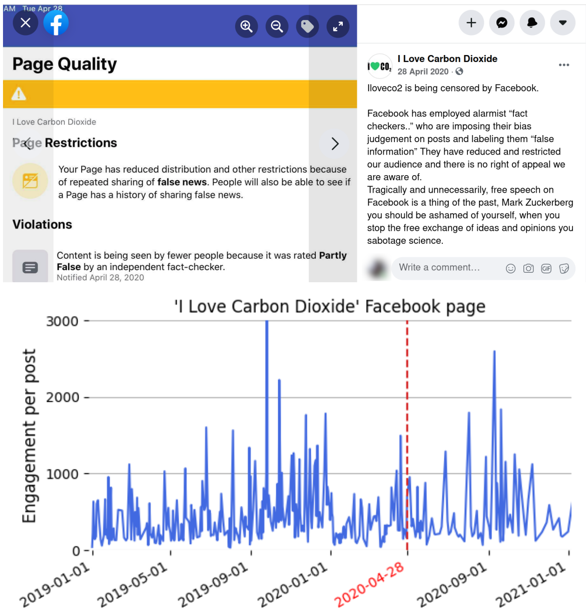
The date of the last notification of was used as the inferred start date of reduced distribution, when it appeared in the screenshot. When it was not visible, we used the date of the post as the inferred start date of reduced distribution.

3.2 Results

Figure 5 shows a screenshot of the Facebook notification shared by the 'I Love Carbon Dioxide'

[13] https://about.fb.com/news/2020/04/covid-19-misinfo-update/

page on April 28, 2020, and the average engagement per post of that page over the past two years. The engagement does not appear to be reduced after April 28, 2020. When we compare the engagement during a 30-day period before and after this date, the percentage change is 2%, indicating that the engagement is not affected by the 'repeat offender' status.


AM Tue Apr 28

Page Quality

I Love Carbon Dioxide

Page Restrictions

Your Page has reduced distribution and other restrictions because of repeated sharing of **false news**. People will also be able to see if a Page has a history of sharing false news.

Violations

Content is being seen by fewer people because it was rated **Partly False** by an independent fact-checker.
Notified April 28, 2020

I Love Carbon Dioxide
28 April 2020 ·

Iloveco2 is being censored by Facebook.

Facebook has employed alarmist "fact checkers.." who are imposing their bias judgement on posts and labeling them "false information" They have reduced and restricted our audience and there is no right of appeal we are aware of.
Tragically and unnecessarily, free speech on Facebook is a thing of the past, Mark Zuckerberg you should be ashamed of yourself, when you stop the free exchange of ideas and opinions you sabotage science.

Write a comment...

Figure 5: **(Top panel)** Screenshot of a post from the 'I Love Carbon Dioxide' Facebook page sharing a reduced distribution notification from Facebook. **(Bottom panel)** Average engagement per post for the "I Love Carbon Dioxide" page for each day in 2019 and 2020, with the reduced distribution start date shown in red.

To provide a general overview, we calculate the percentage change in engagement during a 30-day period before and after the reduced distribution start date for each of the 82 Facebook pages that published at least one post during each period (see Figure 6). The average percentage change is -16%, the median is -24%, and a Wilcoxon test reveals that the percentage changes are significantly different from zero (W = 911, p-value = 0.00026). We can thus suggest that the 'reduced distribution' status is associated with a modest decrease in engagement.

However, there is a large heterogeneity across the different Facebook pages. The engagement of some popular pages have actually increased following the notification, such as the 'Tucker Carlson Tonight' page with a 38% increase (from 104k to 143k interactions per post).

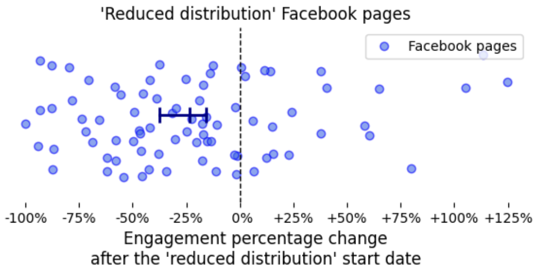


Figure 6: Percentage changes between the average engagement per post 30 days before the 'reduced distribution' start date and 30 days after. Each dot represents a Facebook page. The bars show the median and its 90% confidence interval.

Finally, we verify whether an important drop in engagement also occurred in June 2020 for this set of Facebook pages. When we compare the engagement metrics before and after June 9, 2020, the percentage changes are not significantly different from zero (W = 1093, p-value = 0.055), and the median percentage change is 3%. This confirms that Facebook pages have most likely not been affected by the *reduce* measure implemented on June 9, 2020 and evidenced in the previous section.

4 Discussion

Facebook, the most widely used social media platform in the world, has announced a series of measures to curb the spread of misinformation, notably by reducing the visibility of 'repeat offenders', which are accounts that repeatedly share false information. However, the effects of the platforms' diverse policies to tackle misinformation remains understudied (Pasquetto et al., 2020). The present research article aims to contribute to filling this knowledge gap by verifying the application and measuring the consequences of Facebook's 'reduce' policy on the targeted accounts' engagement metrics.

As a first step, we investigated the reach of 307 Facebook accounts (mainly groups) having repeatedly shared misinformation using a fact-checker's dataset. Sharing of two false news links over a three-month period is supposed to be penalized by a reduced visibility of the account's content. We did observe a significant decrease (median of -43%) in the engagement per posts published by pages under a presumptive repeat offender status. However, we find no evidence that this policy is leading to a significant reduction in engagement for Facebook groups.

As a second step, we identified 83 Facebook pages that shared a notification from Facebook announcing that their account was under reduced distribution. The pages' engagement metrics were significantly lower after the date of the notification (median of -24%), suggesting that the 'reduced distribution' measure was indeed applied to the pages. We noted that no group was found when searching for accounts sharing a reduced distribution notification, which confirms that the 'repeat offender' policy is applied only to Facebook pages, and not to groups.

By analyzing the time series of the repeat offenders' engagement over the past two years, we also discovered a sudden drop affecting the groups around June 9, 2020. For many groups, the decrease was quite drastic (up to 70% - 80%), with a median drop in engagement of 45%. The 18 Facebook pages from the first sample, as well as the 83 pages from the second sample, were not concerned by this decrease. This 'June drop' does not correspond to any official communication by Facebook on that matter. It indicates that the company has very likely taken internal decisions that heavily impact the organic reach of repeat offenders' groups, in ways that differ from its stated policy against repeat offenders pages. More transparency from Facebook would be needed to understand the nature and origin of this change. It would also bring clarity on how rules aimed at limiting the spread of misinformation are being enforced.

Facebook pages and groups have different aims: pages are supposed to be for official communication, while groups should foster interactions between users[14]. Pages are thus always public, while groups can be public or private. Pages' posts can also be monetized and promoted. Despite these differences, both pages and groups are being used to spread false news. In the interest of curbing the spread of misinformation, applying its 'repeat offender' policy to groups as well as to pages would help Facebook decrease the amount of misinformation in their users' feeds. It is also not clear why only repeat offender Facebook groups, and not pages, were reduced in June 2020. Studies have highlighted that misinformation persists at high levels on Facebook and other platforms (Kornbluh et al., 2020; Resnick et al., 2018). In the context of the COVID-19 pandemic, concerns rose about the amount of misinformation spreading on Facebook, and its potential harm on users (Johnson et al., 2020). It is possible that such concerns have driven Facebook to apply a 'quick fix' to decrease the engagement of posts shared in groups spreading misinformation and compensate for the absence of a repeat offender policy. One should note that since the overall activity in these misinformation groups doubled between September 2019 and June 2020, the 'June drop' has only succeeded in bringing the overall engagement level back to its early 2019 values (see Figure 3 top panel).

Online misinformation can be a threat to societies around the world, and the role of the platforms in its regulation has been the subject of intense debate over the past few years (Rogers, 2020; De Gregorio and Stremlau, 2020). As a consequence, researchers (Mena, 2020; Yaqub et al., 2020) and journalists[15] have begun to monitor the actions platforms take against misinformation and their efficiency. Given the facts that 1) fact-checking typically takes more time to perform than false news to go viral, that 2) accounts that have shared misinformation in the past tend to keep sharing misinformation and that 3) a small number of accounts is responsible for a large proportion of the misinformation being shared (at least regarding COVID-19[16]), acting against 'repeat offenders' is one of the most effective policy platforms can take to protect their users against manipulation.

We emphasize the need for further research to thoroughly verify and shed light on the platforms' actions against misinformation. Notably, while our results provide information on the relative drop in engagement per post resulting from Facebook's repeat offenders policy, more research is needed to quantify the impact of such policies on the overall prevalence of misinformation in users' feeds.

References

Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2):211–36.

[14] https://www.facebook.com/help/337881706729661/

[15] https://www.economist.com/graphic-detail/2020/09/10/facebook-offers-a-distorted-view-of-american-news, https://www.nytimes.com/2020/11/24/technology/facebook-election-misinformation.html

[16] https://www.counterhate.com/disinformationdozen

Hunt Allcott, Matthew Gentzkow, and Chuan Yu. 2019. Trends in the diffusion of misinformation on social media. *Research & Politics*, 6(2):2053168019848554.

Yochai Benkler, Casey Tilton, Bruce Etling, Hal Roberts, Justin Clark, Robert Faris, Jonas Kaiser, and Carolyn Schmitt. 2020. Mail-in voter fraud: Anatomy of a disinformation campaign. *Available at SSRN*.

R Brulle. 2018. 30 years ago global warming became front-page news–and both republicans and democrats took it seriously. *The Conversation*.

Giovanni De Gregorio and Nicole Stremlau. 2020. Internet shutdowns in africa— internet shutdowns and the limits of law. *International Journal of Communication*, 14:20.

J. Donovan, N. Jankowicz, C. Otis, and M. Smith. 2020. House intelligence committee open virtual hearing: "misinformation, conspiracy theories, and 'infodemics': Stopping the spread online". https://intelligence.house.gov/news/documentsingle.aspx?DocumentID=1092.

Jieyu Ding Featherstone and Jingwen Zhang. 2020. Feeling angry: the effects of vaccine misinformation and refutational messages on negative emotions and vaccination attitude. *Journal of Health Communication*, 25(9):692–702.

Richard Fletcher, Antonis Kalogeropoulos, Felix M Simon, and Rasmus Kleis Nielsen. 2020. Information inequality in the uk coronavirus communications crisis. *Reuters Institute for the Study of Journalism*.

Amélie Heldt. 2019. Let's meet halfway: Sharing new responsibilities in a digital age. *Journal of Information Policy*, 9:336–369.

Neil F Johnson, Nicolas Velásquez, Nicholas Johnson Restrepo, Rhys Leahy, Nicholas Gabriel, Sara El Oud, Minzhang Zheng, Pedro Manrique, Stefan Wuchty, and Yonatan Lupu. 2020. The online competition between pro-and anti-vaccination views. *Nature*, 582(7811):230–233.

K Kornbluh, A Goldstein, and E Weiner. 2020. New study by digital new deal finds engagement with deceptive outlets higher on facebook today than run-up to 2016 election. gmf the german marshall fund of the united states. viitattu 16.12. 2020.

Marin Lahouati, Antoine De Coucy, Jean Sarlangue, and Charles Cazanave. 2020. Spread of vaccine hesitancy in france: What about youtube™? *Vaccine*, 38(36):5779–5782.

Paul Mena. 2020. Cleaning up social media: The effect of warning labels on likelihood of sharing false news on facebook. *Policy & internet*, 12(2):165–183.

Amy Mitchell, Jeffrey Gottfried, Michael Barthel, and Elisa Shearer. 2016. The modern news consumer: News attitudes and practices in the digital era. *Pew Research Center*.

Irene V Pasquetto, Briony Swire-Thompson, Michelle A Amazeen, Fabrício Benevenuto, Nadia M Brashier, Robert M Bond, Lia C Bozarth, Ceren Budak, Ullrich KH Ecker, Lisa K Fazio, et al. 2020. Tackling misinformation: What researchers could do with social media data. *The Harvard Kennedy School Misinformation Review*.

Ethan Porter, Thomas J Wood, and Babak Bahador. 2019. Can presidential misinformation on climate change be corrected? evidence from internet and phone experiments. *Research & Politics*, 6(3):2053168019864784.

Paul Resnick, Aviv Ovadya, and Garlin Gilchrist. 2018. Iffy quotient: A platform health metric for misinformation. *Center for Social Media Responsibility*, 17.

Richard Rogers. 2020. Deplatforming: Following extreme internet celebrities to telegram and alternative social media. *European Journal of Communication*, 35(3):213–229.

CrowdTangle Team. 2021. Crowdtangle. *Facebook, Menlo Park, California, United States*.

Waheeb Yaqub, Otari Kakhidze, Morgan L Brockman, Nasir Memon, and Sameer Patil. 2020. Effects of credibility indicators on social media news sharing intent. In *Proceedings of the 2020 chi conference on human factors in computing systems*, pages 1–14.