COVID-RELATED MISINFORMATION ON YOUTUBE

The Spread of Misinformation Videos on Social Media and the Effectiveness of Platform Policies

Aleksi Knuutila, Aliaksandr Herasimenka, Hubert Au, Jonathan Bright, Rasmus Nielsen, Philip N. Howard

SUMMARY

For this memo, we identified all Covid-related videos which circulated on social media, but which YouTube eventually removed because they contained false information. Between October, 2019 and June, 2020 there were 8,105 such videos - less than 1% of all YouTube videos about the coronavirus. We find that:

- It took YouTube on average 41 days to remove videos containing false information, based on a subset of videos for which this data was available.
- Surprisingly, Covid-related misinformation videos do not find their audience through YouTube itself, but largely by being shared on Facebook.
- Facebook placed warning labels about false information only on 55 videos, less than 1% of the misinformation videos shared on the platform.
- Misinformation videos were shared almost 20 million times on social media, which is more than the shares gathered by the five largest English-language news sources on YouTube combined (CNN, ABC News, BBC, Fox News and Al Jazeera)

INTRODUCTION

Many false claims related to the coronavirus (Covid-19) pandemic circulate on the internet. These include claims about ineffective or sometimes dangerous cures, such as consuming large amounts of vitamin C or exposing one's body to heat. Misinformation also commonly includes misleading narratives about the actions of governments as well as international bodies like the WHO or the UN, and the origins of the virus as well as accusations that minorities have contributed to the spread of the virus.[1]

Misinformation and conspiratorial claims related to the coronavirus are a problem for stemming the pandemic. Public health depends upon people having accurate knowledge about the severity of the problem, how they can avoid infection and what treatments can help them. Studies have shown that believing in conspiracy theories makes people less likely to participate in behaviors that protect their health, such as obtaining vaccinations.[2]

While all large social media platforms can host misinformation, research suggests that YouTube has played a particularly important role as a source for misinformation related to the Coronavirus pandemic. According to a survey conducted in the United Kingdom, YouTube was the information source that was most strongly associated with belief in conspiracy theories. Out of those who believed that 5G networks caused Covid-19 symptoms, 60% stated that much of their knowledge about the virus came from YouTube.[3]

YouTube has changed its moderation policies and practices to contain the spread of misinformation through the platform. In April 2020, YouTube's Chief Executive Susan Wojcicki stated that the company was increasing its efforts to remove "medically unsubstantiated" videos as well as claims linking the pandemic to 5G.[4] Wojcicki stated that the company would temporarily rely more heavily on automated detection, instead of its workforce, for the identification of such videos.[5] YouTube's COVID-19 Medical Misinformation Policy, published in May 2020, states that "YouTube doesn't allow content about COVID-19 that poses a serious risk of egregious harm", and names the World Health Organization and local health authorities as the authoritative source of information.[6] YouTube has also stated that it will place content that it deems "high quality" or originating from authoritative sources more prominently into its search results and recommendations, to effectively reduce the visibility of potential misinformation, even when it is not outright removed.[7]

According to YouTube's "Community Guidelines" enforcement report, the company does remove videos that break its rules on a large scale. Between January and March 2020, over 6 million videos were removed because a violation of the Community Guidelines. Most of these through automatic flagging of the videos, and according to YouTube's report 75% of these videos were removed before they received 10 views.[8] Information on how many times the remaining 25% of videos were viewed is not published. A significant number of videos with misinformation about the coronavirus do reach a large audience before being removed.

In this memo, we use the categorizations made by YouTube itself to examine misinformation. We identified Covid-related videos that had been removed by YouTube due to breaking the Community Guidelines but that had been shared publicly on social media before being removed. There are several reasons why these videos might have been removed. The Community Guidelines prohibit publishing false or harmful information about the coronavirus, but also for instance nudity or harassment. We refer to these videos as Covid-related misinformation, and rely on YouTube's judgments and its community guidelines for validation of labelling. since it is reasonable to assume that false or misleading content is the dominant reason for why the videos we examined were removed.

DATA AND METHOD

YouTube publishes only aggregated information about the videos that break its Community Guidelines and that are removed. It is, however, possible to gather information about individual removed videos from various public data sources. These data sources have their limitations, but the resulting dataset is a relatively comprehensive sample of videos that circulated on social media and then were removed by YouTube because they contained false information.

We identified Covid-related videos by looking for posts on Facebook, Reddit and Twitter that link to YouTube and that match Covid-related keywords. For Twitter, we used an open access dataset that that covered the period from October, 2019 to the end of April, 2020.[9] This dataset was based on a set 268 Covid-related keywords. We simplified and updated this list of keywords to a total of 71 keywords (see Data Supplement). We used the CrowdTangle service to search for posts on Reddit and Facebook between the 1st of October 2019 and the 30th of June 2020. CrowdTangle is a database that contains public groups and pages from Facebook and Reddit. CrowdTangle does not include Facebook in its entirety, but includes public Facebook pages and groups of a significant size. CrowdTangle contains almost all Facebook pages and groups with over 100,000 followers or likes, and over half of those with 30,000.[10]

This search resulted in a list of 1,091,876 distinct videos. We then followed the YouTube link to each video, and where the videos were no longer available we recorded the reason that the YouTube site gave for the video having been removed. With this method, we identified 8,105 Covid-related videos that YouTube had removed because they breached is Community Guidelines.

For these 8,105 videos, we recovered additional information and metadata from other sources, since YouTube itself only published the reason for their removal. Firstly, we recovered the titles and part of the description for all the videos that have been posted to Facebook. The posts on Facebook displayed the original titles and the first 157 characters of the video's description, which we could read by programmatically retrieving every Facebook post. Additionally, we recorded whether the Facebook posts that linked to misinformation videos contained a "warning label" stating that the link contained potential misinformation, based on third-party fact checking. We also queried the Facebook Graph API to get the total number of shares, comments and reactions that the videos had received across the entire platform, including posts to individual profiles and closed groups. The data collection was undertaken in July, 2020.

Lastly, we recovered metadata about the videos from the archive.org's "WayBack Machine", a service that archives the older versions of webpages. Copies of the deleted YouTube pages were accessible through the WayBack Machine's API in 935 cases. For these videos, we could access the view counts, channel subscriber counts, full descriptions of the videos as well as the video's creation date. In 420 cases, we were also able to approximate how long the video had been visible, by noting the date at which the WayBack Machine had archived the first copy of video's page that stated its removal. It is important to note that there is a bias in the variables that we collected from the WayBack Machine, as the service is more likely to collect and save copies of videos that are widely shared via social media.

FINDINGS

Misinformation videos and their reach

Table 1 describes the 10 videos that received the most reactions on social media over this period. Three videos have exceptionally large audiences. Firstly *Plandemic*, a documentary claiming that the coronavirus was constructed in a laboratory to create profit from vaccinations. Two other popular videos feature physicians Dr. Erickson and Dr. Buttar, outsiders in the medical community who question the dangers of the virus and the benefits of face masks. Fact checkers have found several unfounded and harmful claims in these videos.[11] These three videos alone gathered 40% of the shares, comments and reactions given to all of the misinformation videos in our dataset.

Many popular misinformation videos feature individuals who claim to speak from positions of authority, such as doctors or whistleblowers, who claim that they are being silenced by people in positions of power. Another common technique used is to recontextualise parts of other videos, for instance by editing clips from Bill Gates' TED talks in a way that is misleading. The videos frequently cover a range of themes, including questionable motivations of public authorities, the low

Table 1: The 10 Misinformation Videos with the Highest Share Count on Facebook, Twitter and Reddit

YouTube video title	Shares (millions)	Comments (millions)	Reactions (millions)
Plandemic Documentary: The Hidden Agenda Behind Covid-19	3.15	9.94	8.82
EXCLUSIVE: Dr. Buttar BLASTS Gates, Fauci, Faked Pandemic Numbers EXPOSED As Economy Collapses	1.88	4.12	4.63
Dr. Erickson COVID-19 Briefing	1.14	1.14	1.52
Frank Hahnel A real Doctor telling it like it is. It's all Fake.	0.24	0.50	0.42
Corona virus:Doctor exposes covid-19/watch before it gets deleted again	0.19	0.25	0.22
COVER UP Fauci Approved Chloroquine, Hydroxychloroquine 15 Years Ago to Cure Coronaviruses	0.13	0.13	0.27
COVID-19: This is a WARNING to all people - What they're not telling us	0.12	0.09	0.10
5G Bill Signed Into Law While Everyone Is Distracted By Coronavirus	0.12	0.14	0.22
Otra útil sugerencia en la lucha contra esta pandemia llamada Coronavirus	0.12	0.02	0.06
Tại Vũ Hán, một gia tộc hơn 70 người không ai bị nhiễm Covic 19 bởi cách làm đơn giản này	0.11	0.02	0.22

Source: Authors' calculations based on collected data, based on 8,105 videos.
Notes: Numbers have been aggregated where an individual video has been reuploaded several times.

mortality rate of the virus and treatments without scientific foundation.

Some of the most popular misinformation videos have been uploaded and reuploaded to YouTube numerous times, in part as a response to YouTube removing these videos. For instance, we found 89 copies of the Plandemic video. For the videos in Table 1, we have searched our dataset for duplicates of the same video and aggregated the numbers.

For a subset of all misinformation videos, we could retrieve data on how many times they were viewed and how long they were visible on YouTube. Misinformation videos were on average visible on YouTube platform for 41 days before they were removed, based on the 420 videos for which this information was available. The misinformation videos gathered on average 149,825 views before they were removed, based on 935 videos.

The audience for misinformation videos is only a small share of the audience for all corona-related YouTube videos. In absolute terms, however, their reach is substantial. The misinformation videos in our dataset gathered 20 million shares and 71 million reactions and comments on Facebook, Twitter and Reddit. As a point of comparison, we examined all the YouTube videos published by CNN, ABC News, BBC, Fox News and Al Jazeera. These are the five English-language news broadcasters with the highest subscriber count on YouTube. During the study period, from the beginning of October 2019, to the end of June 2020, the videos from these five YouTube channels gathered 15 million shares and 42 reactions and comments. Hence, the misinformation videos as a whole had a significantly higher reach on social media than the YouTube videos by five large US-based news broadcasters.

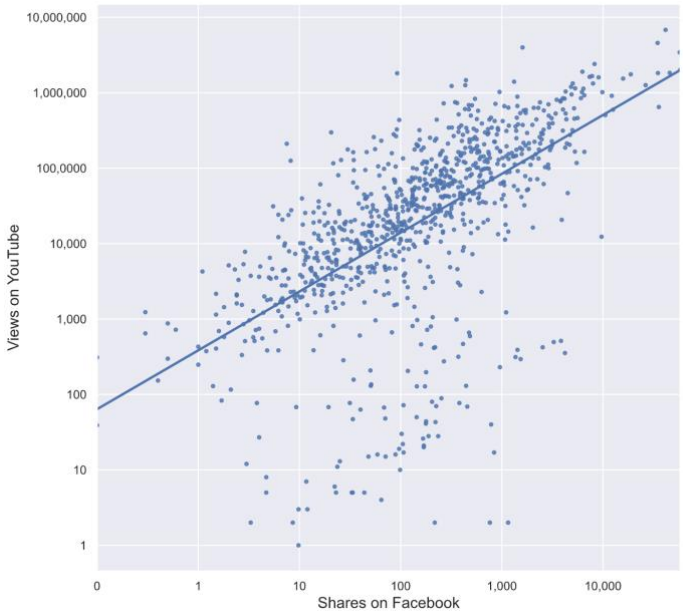
The spread of misinformation videos on social media

The dataset shows that there are a significant number of misinformation videos that gather a large audience on YouTube. How do these videos find their audience?

The data suggests that Facebook is the most significant channel through which the misinformation videos spread. Firstly, the number of shares and reactions to the misinformation videos on Facebook were significantly higher than on other platforms. On Facebook, the misinformation videos were shared on average 2,427 times and gathered on average 8,771 comments and reactions. In comparison, the videos were posted on Twitter and retweeted on average 63 times. All of the videos in our dataset combined were posted on Reddit only 955 times.

The number of shares and reactions that a misinformation video received on Facebook were also much more strongly correlated to the video's count of views on YouTube than the shares and reactions on other platforms. In particular, the number of shares that a misinformation video received on Facebook predicted how many views the video would receive on YouTube. The correlation factor between these two variables was 0.70. This means that it is rare for a Covid-related misinformation video to have many views unless it is frequently shared on Facebook. For every share on

Figure 1. The Relationship Between Facebook Shares and Views on YouTube



Source: Author's calculations based on collected data,

Facebook, on an average a video received 15 views. This relationship is visualized in Figure 1. The number of reactions or comments on Facebook also had a positive correlation with the video's view count, but the connection was less strong (see Data Supplement for more details).

The subscriber count of YouTube channels, in contrast, had a much weaker connection with the view count, with a correlation factor of 0.12. The average channel in the dataset had 192,253 subscribers. However, having a large number of subscribers has not systematically led to videos being seen on many occasions. This fact might be linked — though this study could not confirm it — to YouTube's efforts to deamplify or limit the visibility of videos on the platform that it deems "poor quality".[12] It is possible that YouTube is less likely to present videos from sources that tend to contain misinformation to users even when they have subscribed to the channels in question. Such policies for limiting the visibility of content on one platform are, however, not sufficient for containing misinformation if they find their audience through other channels. Through links from Facebook, millions of users discover and share misinformation videos on YouTube.

One of Facebook's policies to limit the spread of misinformation has been to build a network of independent fact-checkers to evaluate content shared on the platform. Where these fact-checkers judge a link to contain false information, users will see a warning label attached to the post. Additionally, the visibility of the post will be reduced.[13] Out of the 8,105 misinformation videos, only 55 had warning labels attached to them when we undertook our data collection; less than 1% of all the misinformation videos. This suggests that Facebook's network of independent fact-checkers do not focus on YouTube videos in their work or have sufficient reach to cover misinformation on YouTube.

The Facebook communities spreading misinformation

Covid-related misinformation videos have been posted on hundreds of thousands of personal profiles, pages and groups. With the data sources that are available, it is possible to examine only public Facebook pages and groups of a significant size that have shared links to these videos. Our analysis in this section focuses on this subset of pages and groups. Facebook offers both pages and groups as a way of broadcasting information and hosting discussions. In this section, we will refer to both pages and groups as "communities" for the sake of simplicity.

We found 32,607 distinct Facebook communities that had posted links to misinformation videos. Moreover, we found that a small number of these have been much more influential than the others. Just 250 communities have provided the same amount of visibility, in terms of shares, to misinformation than all the remaining communities combined. The posts linking to misinformation videos on these communities have in total been shared 1.2 million times. Out of these 250 communities, 91 operated predominantly in the English language. We briefly categorize and describe these 91 communities to examine the kinds of individuals and communities that participate in the dissemination of misinformation.

Our examination led to the identification of six distinct types of communities. The two most common types of communities are conspiracy-related communities (27 out of 91) and political communities (25 out of 91). We classified under "conspiracy-related groups" the groups that were focused on the following theories and issues: Q-Anon, Flat Earth and Chemtrails communities. We also included communities focused on opposing vaccinations and 5G networks, where these seemed to share unfounded claims related to their harms or aims. Almost all of these conspiracy-related communities existed before the pandemic but have now become conduits for the spread of misinformation about the Coronavirus. The conspiracy-related communities posted very large numbers of misinformation videos, on average 119 distinct videos per community. The conspiracy-related communities frequently had relatively few followers —on average 51,006 followers— but were still influential because they shared misinformation frequently.

We classified as political communities the communities that were focused on supporting or opposing a specific cause, politician or policy. The communities in our dataset largely represented conservative or Republican political views in the United States or Canada. Though

Table 2. The Facebook Communities that Generated Most Shares for Misinformation Videos

Name of community	Subscribers	Misinformation videos posted	Shares generated	Reactions generated
One America News Network	676,602	4	12,958	26,392
Collective Action Against Bill Gates. We Wont Be Vaccinated!!	167,839	576	11,004	21,104
Behold Israel	1,043,193	1	8,403	13,006
OFFICIAL Q / QANON / r / ra / rawgr / Q+ / Q+ +++	191,174	797	8,020	20,454
Chemtrails Global Skywatch	179,632	265	7,048	15,460
An0maly	880,551	1	6,267	23,075
The Official Rush Limbaugh Facebook Group	60,474	175	4,310	7,425
Rush Limbaugh	2,397,093	1	4,290	18,712
Dr. Ruediger Dahlke	177,593	6	4,256	12,873
TB Joshua Ministries	3,809,174	1	4,182	35,784

Source: Authors' calculations based on collected data

some groups supported individual politicians ("Trump / Pence AGAIN in 2020"), they were mostly organised in opposition to liberal politicians ("Recall Gavin Newsom") or in opposition particular policies, such as policies to contain the Coronavirus ("Reopen Alabama"). The political communities also included anti-establishment movements with wider agendas (for instance "Yellow Vests Canada").

Political commentators and news networks made up 18 out of the 91 communities. These were almost exclusively US-based, and were mostly Facebook communities that broadcast material produced by individual journalists or commentators. We also included in this category communities of supporters of political commentators. For instance, the members of the official Facebook group of Rush Limbaugh supporters shared links to Covid-related misinformation videos to that group on 179 occasions.

Finally, 7 out of 91 communities were religious in nature, and in particular associated with North American evangelical Christianity. Among these were pages of individual churches, pastors, and religious communities ("I Love Being A Christian"). An equal number of communities (7 out of 91) were focused on alternative medicine. The majority of these were related to individual practitioners in alternative medicine, but also included communities discussing particular treatments, such as the Fabunan anti-viral injection. A small number of communities were related to musicians and athletes (4 out of 91). Three communities did not fit the categories we developed. The Data Supplement to this memo lists all the 250 communities that were influential in spreading misinformation videos.

In this section, we have examined public pages and groups on Facebook. Overall, these make up a small portion of the total reach of the misinformation videos. The roughly 4.2 million shares, reactions and comments these videos have acquired on public pages only amounts to less than 5% of the 91 million shares, reactions and comments they gathered on Facebook. However, public pages may be an important conduit from which the videos are disseminated elsewhere, shared on individual profiles or other social media platforms.

CONCLUSIONS

This study has presented evidence about the nature and scale of misinformation on YouTube. It shows that there were thousands of misinformation videos related to the coronavirus that, on average, gathered hundreds of thousands of views. Despite YouTube's investment in the automated removal of videos that breach its Community Guidelines, many of these videos remained visible on the platform for weeks before being removed. The results in this memo suggest that there are flaws in the platforms' policies for containing Covid-related misinformation, which have significantly contributed to the spread of false information during the pandemic.

The interaction and links between platforms helps to explain how the videos found their audience. Even though YouTube has amplified credible sources for users who search for information about the pandemic on its own platform, our data shows that much of the traffic to misinformation videos has stemmed from other platforms; in particular Facebook. This suggests that strategies of deamplifying harmful content, in particular for platforms such as YouTube that host content commonly shared on other mediums, does not work in isolation. A strategy of reducing the visibility of misinformation on one platform will not be successful if the same content is shared tens of thousands of times on another platform. Our study also suggests that the reach of Facebook's network of third-party fact checking organizations is insufficient, given the amount of misinformation on YouTube. For the misinformation videos we identified, it was a rare exception that they

had been flagged as containing false information by Facebook's fact checkers.

Some researchers and advocates have suggested that banning or deamplifying communities and organizations that systematically spread false information from social media platforms could potentially contain the spread of misinformation.[14] Our analysis does show that a small number of Facebook communities, some focused on cultivating belief in theories such as the dangers of vaccination or 5G, increased the visibility of misinformation videos significantly. However, while the downgrading of the visibility of these communities could reduce the spread of misinformation, we suggest that placing an emphasis on communities will most likely have only a limited impact. We have found that the types of groups sharing misinformation videos are diverse, spanning, for instance, religious groups and musicians. Moreover, we have found that the great majority of shares on Facebook originate from outside of public communities, and from what are probably millions of individuals sharing misinformation videos from their own accounts. As such, we suggest that it is wrong to characterize misinformation as a problem that is caused by particular types of communities or malicious actors.

This study was necessary in part because the information published by platforms was insufficient for understanding the nature or scale of misinformation. YouTube has, among other platforms, began to periodically publish a transparency report on their own efforts on moderating content, including flagging or removing misinformation.[15] Their presentation of the data, however, reveals little about how many times videos with misinformation or other harmful content were viewed. For the data and reports that are released by platforms to be genuinely useful, they should be comprehensive enough to make apparent potential failures in the platform's moderation policies.

REFERENCES

[1] Brennen, S., Simon, F., Howard, P., and Nielsen, R.K. "Types, Sources, and Claims of COVID-19 Misinformation", Reuters Institute for the Study of Journalism, 2020. https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation

[2] Dunn, A.G., Surian, D., Leask, J., Dey, A., Mandl, K.D., and Coiera, E. "Mapping Information Exposure on Social Media to Explain Differences in HPV Vaccine Coverage in the United States", *Vaccine*, 2017.

[3] Allington, D., Duffy, B. Wessely, S., Dhavan, N., and Rubin, J. "Health-protective Behaviour, Social Media Usage and Conspiracy Belief During the COVID-19 Public Health Emergency", *Psychological Medicine*, 2020.

[4] Cellan-Jones, R. "YouTube Bans 'Medically Unsubstantiated' Content", *BBC News*, Apr. 22, 2020. https://www.bbc.co.uk/news/technology-52388586

[5] YouTube, "Protecting Our Extended Workforce and the Community", *YouTube Creator Blog*, 2020. https://youtube-creators.googleblog.com/2020/03/protecting-our-extended-workforce-and.html

[6] Google, "COVID-19 Medical Misinformation Policy - YouTube Help". Google, 2020. https://support.google.com/youtube/answer/9891785?hl=en

[7] Google, "How Does YouTube Combat Misinformation?; YouTube Scam and Impersonation Policies - How YouTube Works*"*. Youtube, 2020. https://www.youtube.com/howyoutubeworks/our-commitments/fighting-misinformation/

[8] Google, "YouTube Community Guidelines Enforcement – Google Transparency Report". Google, 2020. https://transparencyreport.google.com/youtube-policy/removals

[9] Dimitrov, D., *et al.*, "TweetsCOV19 -- A Knowledge Base of Semantically Annotated Tweets about the COVID-19 Pandemic', *arXiv*, 2020. http://arxiv.org/abs/2006.14492

[10] CrowdTangle, "What Data is CrowdTangle Tracking?". CrowdTangle, 2020. http://help.crowdtangle.com/en/articles/1140930-what-data-is-crowdtangle-tracking

[11] ACEP & AAEM, "ACEP-AAEM Joint Statement on Physician Misinformation". ACEP & AAEM, 2020. https://www.acep.org/corona/COVID-19-alert/covid-19-articles/acep-aaem-joint-statement-on-physician-misinformation/

[12] Marchal, N., Au, H., and Howard, P.N. "Coronavirus News and Information on YouTube". Computational Propaganda Project, Oxford Internet Institute, 2020.

[13] Rosen, G., and Lyons, T., "Remove, Reduce, Inform: New Steps to Manage Problematic Content", Facebook, 2019. https://about.fb.com/news/2019/04/remove-reduce-inform-new-steps/

[14] Avaaz, "Facebook's Algorithm: A Major Threat to Public Health", Avaaz, 2020. https://secure.avaaz.org/campaign/en/facebook_threat_health/

[15] Google, 'Google's Updated Transparency Report Highlights Its Efforts To Maintain A Safe YouTube Community', Google, 2020. https://www.digitalinformationworld.com/2019/11/google-transparency-report-youtube-community-guidelines-enforcement.html

ACKNOWLEGMENTS

The authors gratefully acknowledge the support of the European Research Council for the project "Computational Propaganda: Investigating the Impact of Algorithms and Bots on Political Discourse in Europe", Proposal 648311, 2015–2020, Philip N. Howard, Principal Investigator. Project activities were approved by the University of Oxford's Research Ethics Committee, CUREC OII C1A 15-044. We are also grateful to the Adessium, Luminate, Ford Foundations and the Oxford Martin Programme on Misinformation, Science, and Media for supporting this work. Any opinions, findings,

conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the University of Oxford or our funders.

ABOUT THE PROJECT

The Computational Propaganda Project (COMPROP), which is based at the Oxford Internet Institute, University of Oxford, involves an interdisciplinary team of social and information scientists researching how political actors manipulate public opinion over social networks. This work includes analyzing how the interaction of algorithms, automation, politics, and social media amplifies or represses political content, disinformation, hate speech, and junk news. Data memos integrate important trends identified during analyses of current events with basic data visualizations, and although they reflect methodological experience and considered analysis, they have not been peer reviewed. Working papers present deeper analysis and extended arguments that have been collegially reviewed and engage with public issues. COMPROP's articles, book chapters, and books are significant manuscripts that have been through peer review and formally published.