

**EXPLORING METHODS TO ADDRESS  
COVID-19 MISINFORMATION ON FACEBOOK**

A Research Paper submitted to the Department of Engineering and Society  
In Partial Fulfillment of the Requirements for the Degree  
Bachelor of Science in Computer Science

By

Zachary White

March 25, 2021

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISOR

Catherine D. Baritaud, Department of Engineering and Society

The cemented presence of social media and the internet in contemporary lives has permanently altered the way in which today's society interacts. Particularly during a pandemic where a lockdown lifestyle is ever-present, a large portion of community, work, and social life has transitioned online for almost everyone on the planet with access. This has put heightened pressure on Facebook's capability of managing its platform as a fair and unbiased place for users to engage with information, which is especially crucial with consequential news. In a survey study by Pew Research Center, almost 90% of participants indicated their awareness of how algorithms have a control of their news feed, with the vast majority also conceding that social media companies have "too much" control over what reaches the top of the timeline (Kirkpatrick 2020). Facebook's specialization in social networking and user engagement creates environments through which misinformation can easily permeate through a network of similar opinions, often called "echo chambers" because of the like-minded viewpoints a user primarily is exposed to (Usher-Layser 2016).

While Facebook has taken measures to introduce fact-checking onto its platform, this has not yet proven enough to combat the effectiveness of the social networks that cultivate these information-based communities (Ardévol-Abreu et al. 2020). Before it turns into an arms race between Facebook and online malicious groups, the current state of the application's role in the spread of this bad information needs to be assessed, and different methods should be explored to address the implications of this "infodemic" (Zaracostas 2020). The technical project is an analysis of current research that exists on how Facebook algorithms and artificial intelligence affect user engagement and the social networks on the platform. This tightly-coupled STS paper discusses how various ideas around addressing Facebook misinformation could have worked differently during this pandemic, and how they can change moving into the future. Actor-

Network Theory (Callon 1986) is used to observe how the platform operates in the flow of news and information, and research reveals that social- and community-based solutions are the crucial component of maintaining a healthy online information environment on Facebook.

### **Algorithmic Personalization**

In social media marketplace competition, the company that is best able to engage the user generally wins the market. If a user likes the media or content more, they will spend more of their time in that application as opposed to less engaging platforms. This is the reason most people visit entertainment applications rather than Wikipedia in their free time; the personalization of news feeds draws more positive feedback from the user.

The easiest way to personalize content for millions of users at the same time is through employing massive amounts of computing resources to collect data and develop predictions for each account. News Feeds are configured every instance by algorithms that predict which content will elicit the most positive response from the user. This “positive” reward is quantified in terms of Likes, Shares, and other direct engagements with the content, including overall time spent on the screen (Celis et. al 2019). The massive amount of the data that Facebook and other social media companies gather and use has brought about the term “surveillance capitalism”.

Facebook has created a billion dollar empire from effectively engaging users in a business model broadly known as Social Networking Marketing (SNM). This describes the platforms ability to leverage their specialty in personalization and networking to better target audiences with relevant advertisements (Liao 2014). As shown in figure 1 below, the same data and algorithms used for content personaliation can predict which ads will be more successful, grouping users by attributes that correlate more strongly with certain types of advertisements. An

account that follows sports, for example, will see more content and advertisements related to sports in hope of keeping a user’s time and money. This role coronates Facebook algorithms as the gateway to all information seen on the News Feed of a given account.

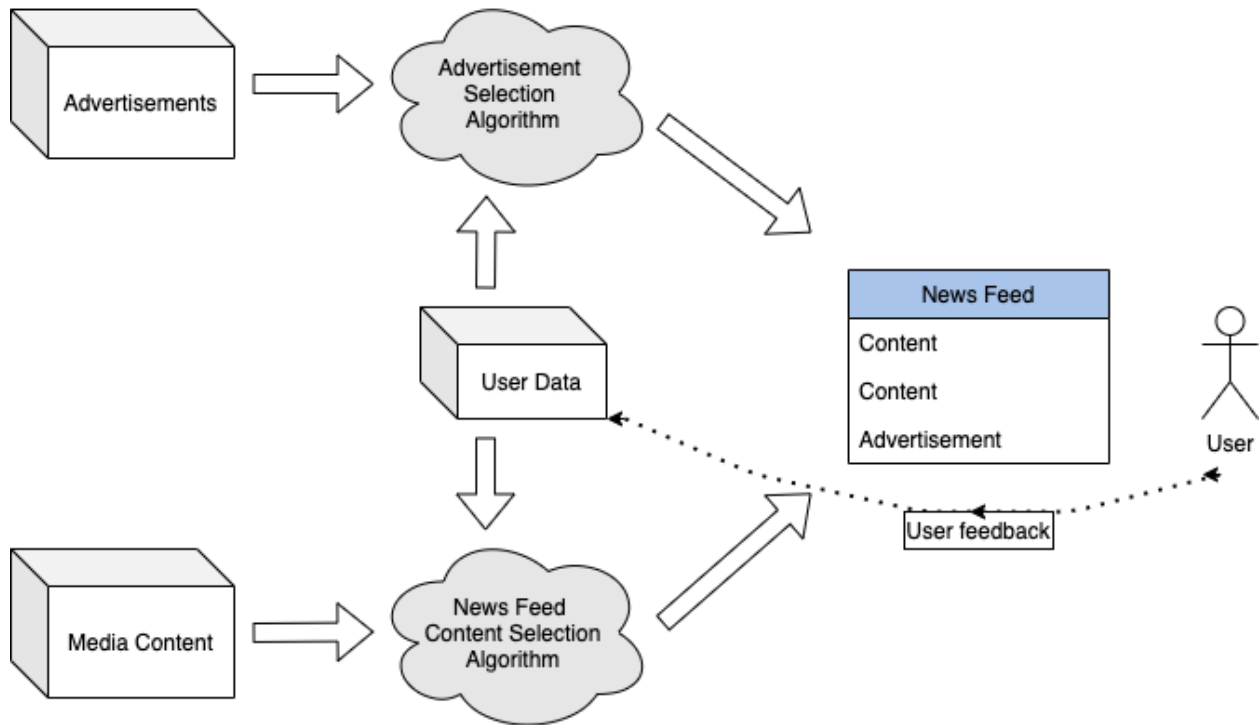


Figure 1: User feedback loop with News Feed: Sustained user feedback loop is how algorithms collect data (Created by White, 2020).

Facebook has publicly encouraged the use of its platform for primarily groups and friends, but *interpretive flexibility* of the platform around the world alters how different cultures use the application. Research has shown how Myanmarese users prioritize news and information in the platform as opposed to solely users from real life (Leong 2020). Facebook’s algorithm does not distinguish between news and real-life connections, and therefore prioritizes these news connections based upon user preference. This user interaction transforms Facebook into primarily a news source for those in Myanmar, and similar scenarios occur within the United

States. Pew Research Center has shown more and more Americans get their news from digital devices as well (Shearer 2021). The most dangerous of Facebook's flexible interpretations are when "echo chambers" and opinion bubbles are prioritized over time in an account's content.

Opinion bubbles are not a phenomenon created by social media; the algorithms that control the applications reflect tendencies of humans over time. Psychology research has shown human preferences to engage with those like-minded and similar to them to exist long before online life (Kunda 1990). Television networks are often known to lean politically left or right and deliver news through a biased lens already. The significant differences with social media technology are the artificial intelligence for personalization and the networks that exist on the platform itself. These two features are the main enablers of the growth of opinion bubbles, and Facebook in particular has been shown to polarize these echo chambers into more extreme views (Kitchens et. al 2020). The growth of polarizing groups online has propagated to real-world consequences, intertwined heavily with the epidemic of misinformation that infiltrates these extremist bubbles. Extreme movements and groups, like the Trump campaign and QAnon conspiracists, have seen amplified success through directly leveraging Facebook's ability to find susceptible users. Myanmar saw a mass genocide and exodus of hundreds of thousands of Rohingya Muslims from their home country because the government successfully spread anti-Muslim propaganda as influencers on the platform. Most recently, the global pandemic opened the door for abundant theories and fake news to spread.

### **COVID-19 misinformation**

When our modern world was brought to its knees by a virus, Facebook and online communities became an even more central part of society than before. The COVID-19 pandemic

initiated a parallel epidemic of its own, spawning online misinformation around the virus in a time where universal truths were less clear than ever. Theories around the origins and behavior of the virus exploded on all political spectrums. Headlines questioned whether the virus actually originated in a lab, whether hydroxychloroquine was an effective treatment, and other provocative ideas (Lewis 2020).

Misinformation can stem from both malicious parties and community misunderstandings. For example, anti-Asian sentiment and violence arose from the ranks of the white supremacist system in America during this pandemic. This is a malicious institution and culture manifesting community-based fake news that targets a minority group, as has been repeated throughout American history. But other sentiment, such as confusion around masks and cures, sprouts from inconsistent messaging from authoritative figures. The FDA was unclear initially on the effectiveness of masks (Lewis 2020), and the Trump administration consistently has confused the population with unfounded theories of consuming hand sanitizer and UV light. Creating a sense of uncertainty and distrust between differently-thinking groups introduces a massive barrier to fighting bias and misinformation, especially when objective facts cannot be agreed upon. Groups like QAnon have capitalized on how polarization can overpower users deemed susceptible by the algorithm, and this particular group has grown significantly since the pandemic began in the United States. More than *two thirds* of Facebook extremist group members joined through the algorithm-controlled group recommendations (Ovide 2021), showing the tools the algorithm gives extremist groups for recruitment and expansion.

Fake news has some notable advantages over equivalent real news in a handful of ways. Fake news can shift stories and headlines to be more engaging to already-polarized users, causing this false content to surface more frequently in News Feeds. This bad information

usually undermines trust in larger institutions that help define a broader truth, in many cases being the government or news, fondly coined “the liberal media”. These adaptations of misinformation content allow it to spread four to six times faster and wider than equivalently true news stories (Vosoughi 2018).

Fake news online appears in various forms, which augments the obstacles in mitigating this information. Facebook content can be images, videos, opinionated posts, external website embeds, and more. One particular instance of “news” existed as an external web archive, titled “5G Syndrome Goes Global” in reference to the theory that COVID-19 was caused by 5G internet infrastructure rollout in early 2020 (Acker & Chalet 2020). This external and ambiguous form allowed it to evade initial automated content moderation implemented by that point by Facebook. Fake news can often overlap with the idea of “free speech” in the United States as well, causing companies and government actors to hesitate when removing and moderating content.

The politicization of COVID-19 caused a divisive approach across the country, which left those most vulnerable to the virus less protected in the system. The communities centering around differing beliefs have become significantly more powerful than the standard of truth and trust that existed between citizens and broader information institutions. The current state of this polarized platform and therefore society requires further analysis and research on ways to best implement a healthier online community.

## **ACTOR-NETWORK THEORY ANALYSIS**

This paper seeks to address the question: “What are ways to mitigate COVID-19 misinformation on Facebook’s platform?” by analyzing how Facebook operates at the center of

social and journalist institutions today. Actor-network theory is appropriate for this scenario because of the large networks and number of actors that are involved in this dynamic. Most of these actors can be generally qualified as either suppliers of information or user groups. As illustrated in figure 2 below, Facebook’s application, engineers, and artificial intelligence are quite literally at the center of these networks, acting as the gateway through which other networks increasingly exist. Therefore, it is Facebook that is the most capable actor to effectively address the health of the networks it manages.

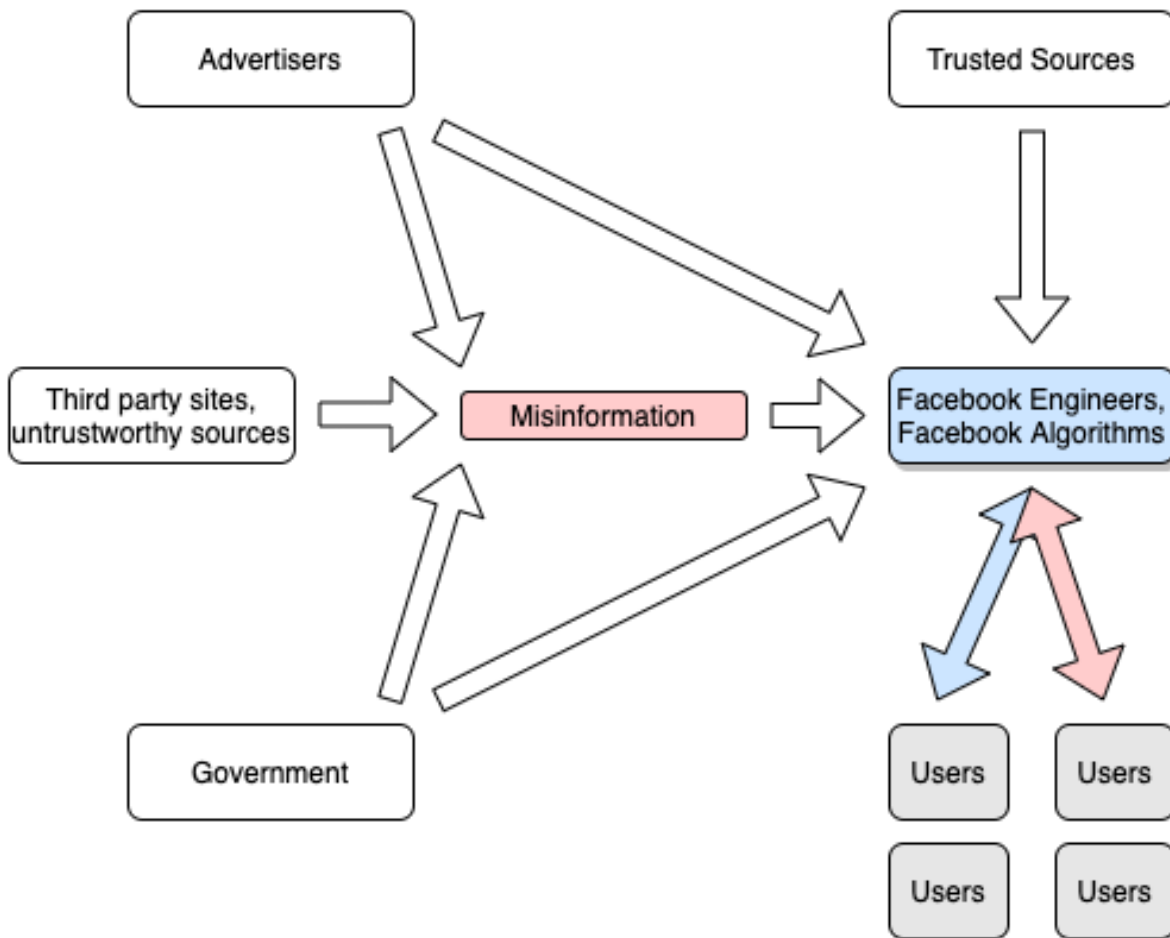


Figure 2: Actor-Network theory analysis of Facebook: Facebook creates distinctly different networks for different user groups (Adapted by White from Callon, 1986).



Figure 2 on page 7 shows how misinformation can originate from various sources, including the institutions that are built around trust, like governments. This widespread origin of bad information makes it difficult to track and eliminate from the source before reaching Facebook's platform. Facebook is the medium through which these networks carry fake news, but this is one of the few shared attributes of the networks between Facebook and different users. This diagram uses colors to represent how users see different information, and how different user groups use Facebook via interpretive flexibility of the personalized News Feed. This causes misinformation to spread more often to *some* but not all user networks, depending on which networks are predicted to be more susceptible.

Algorithms are the key component that enable misinformation to find more relevant networks by using data collection to further differentiate and polarize these user groups. This creates different levels of trust in the relationship between users and the government, the media, and other providers of information. These user groups become communities centered around their own personalized beliefs and content, further ingraining the “divide and conquer” into our society that is often taken by large institutions. The same way that Trump leveraged distrust in a polarizing American system to hijack an entire political party, misinformation permeates better through polarized networks and their inherent distrust of institutions.

## **METHODS TO MITIGATE MISINFORMATION**

The solution to misinformation cannot be one solely centered in computing; artificial intelligence has become a crucial component in modern social and psychological worlds, and this infodemic requires social and psychological solutions alongside application feature changes. The actor-network theory analysis helped to identify polarization of user groups

as a crucial problem in misinformation incubation. If the network that a user engages with has a healthy spread of opinions and connections across opinion groups, this will not create powerful echo chambers, where misinformation is likely to foster undetected. This diversity is a necessary pairing with fact-checking to create online environments with minimal fake news.

### **Fact checking**

Fact-checking functionality on Facebook has proven minimally effective in reducing the spread of misinformation, as these themselves can represent suspicious institutions that user already do not trust. Researchers Alberto Ardévol-Abreu, Patricia Delponti, and Carmen Rodriguez-Wangüemert discuss what factors are involved with fake news spreading throughout contact networks on platforms such as Facebook (Ardévol-Abreu et. al 2020). The authors, researchers funded by organizations through the Universidad de La Laguna in Spain, conducted research to explore how users interact with false information, particularly when it is deemed false by Facebook fact-checking measures.

In designing the online survey, the researchers centered the discussion mostly around misinformation and coronavirus news and tried to measure if certain demographics of user accounts corresponded with likelihood to share false information. This part of the research appeared to mostly conclude that the aspects they chose, including political orientation, age, gender, and more, did not have a significant impact on accounts' chances to share fake news (p. 6). The qualitative studies with the focus groups, however, provided some valuable insight into how users interact with fake news in general.

In the focus groups, the authors hosted spaces where users could interact with each other on the subject of a set of news stories, all deemed at least partially-false by Facebook's

independent fact-checking agency (p. 4). Some significant notes about the sessions indicated that most participants treated the fact-checking feature as an additional asterisk in addition to the article content itself, rather than the gate-keeping presence it is intended to be. The discussion on whether the articles were fake or not centered mostly around the users' own opinions and logic, and most discussion around the fact-checking feature was skepticism rather than trust (p. 6).

This article speaks to an audience of “educators, journalists, fact-checking organizations, and social media companies” in the conclusion with supplementary suggestions to combat misinformation epidemics. The majority of these suggestions urge the large organizations to better engage the public in this process, through education or participation. This is crucial in re-establishing the trust that is needed for societies to exist more peacefully and progressively, especially through turbulent times where accurate information is crucial.

### **Social norms**

Fact-checking online must be accompanied by establishing expectations for social behavior online. Researchers from FIM Research Center in Germany experimented with “the effectiveness of social norms in fighting fake news on social media” (Gimpel et. al 2021). In these experiments, social norms are defined in two ways: “descriptive” of what others do, and “injunctive” of what others *approve* of. In the experimental News Feeds, the descriptive social norms functioned similarly to fact-checking functionality, and these researchers found results similar to Ardévol-Abreu in that fact-checking alone did not have a significant impact on whether users reported misinformation. However, social approval proved to be a significant factor in whether users reported misinformation they engaged with. This aligns with the power of community standards on Facebook that have made it so difficult to mitigate polarization in the

first place. The spread of misinformation can be reduced if more users are effectively encouraged to report misinformation sooner.

Long before the pandemic, researchers for the Association for Psychological Science Lewandowsky et. al discussed extensively the social and psychological issues around individuals receiving misinformation from online. They examine certain problems that cause “backfires” when trying to disprove misinformation, such as information overkill and contrasting world views (Lewandowsky et. al 2012). Analyzing the psychological barriers that reinforce misinformation allows the researchers to make suggestions on how to address misinformation on the individual level. These recommendations include calibrating information toward identities and world view, emphasis on simple and powerful facts, and other psychological approaches.

In short, strengthening the health, diversity, and trust of the Networks on Facebook is the best way to mitigate coronavirus misinformation. This is the most direct way to disrupt opinion bubbles online and barricade misinformation at the News Feed level. Algorithms are well positioned to change to this need, as they control much of what is seen by these networks. Combining this with a consistent truth standard through trusted and active sources are two ways Facebook can systematically strengthen their online world against fake news, and lessen the negative effects of the reward-hungry algorithms that drive personalization.

These standards cannot be solely maintained by external parties; the users and community themselves must be proactive about this online setting. Studies have shown that social approval and social norms can be impactful forces driving users to be aware of and report misinformation; this aligns well with the approval-seeking tendencies that can drive normal life and Facebook activity in general. A sense of community approval and standard is much more powerful than the sole use of external truth-validation.

## REFERENCES

Acker, A., Chaiet, M. (2020). The weaponization of web archives: Data craft and COVID-19 publics. The Harvard Kennedy School (HKS) Misinformation Review.

<https://doi.org/10.37016/mr-2020-41>

Ardévol-Abreu, Alberto; Delponti, Patricia; Rodriguez-Wangüemert, Carmen (2020).

“Intentional or inadvertent fake news sharing? Fact-checking warnings and users’ interaction with social media content”. *Profesional de la información*, v. 29, n. 5, e290507. <http://doi.org/10.3145/epi.2020.sep07>

Callon M. (1986) The sociology of an actor-network: the case of the electric vehicle. In:

Callon M., Law J., Rip A. (eds) *Mapping the Dynamics of Science and Technology*.

Palgrave Macmillan, London. [https://doi.org/10.1007/978-1-349-07408-2\\_2](https://doi.org/10.1007/978-1-349-07408-2_2)

Cho, J., Ahmed, S., Hilbert, M., Liu, B., & Luu, J. (2020). Do search algorithms endanger

democracy? An experimental investigation of algorithm effects on political

polarization. *Journal of Broadcasting & Electronic Media*, 64(2), 150–172.

<https://doi.org/10.1080/08838151.2020.1757365>

Duffy, K. J., & Brown, R. H. (2021). Shouting FIRE! (or worse) on social media: The interplay

of the first amendment and government involvement in efforts to limit or remove

social media content. *Intellectual Property & Technology Law Journal*, 33(4), 3–14.

- Gillespie, T. (2014). The relevance of algorithms. *Media technologies: Essays on communication, materiality, and society*, 167(2014), 167.
- Gimpel, H., Heger, S., Olenberger, C., & Utz, L. (2021). The effectiveness of social norms in fighting fake news on social media. *Journal of Management Information Systems*, 38(1), 196–221. <https://doi.org/10.1080/07421222.2021.1870389>
- Kitchens, B., Johnson, S. L., & Gray, P. (2020). Understanding echo chambers and filter bubbles: The impact of social media on diversification and partisan shifts in news consumption. *MIS Quarterly*, 44(4), 1619–1649.  
<https://doi.org/10.25300/MISQ/2020/16371>
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108 (3), 480 – 498.  
<https://doi.org/10.1037/0033-2909.108.3.480>
- Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. (ACLS Humanities E-Book.) Oxford: Oxford University Press.
- Leong, L. (2020). Domesticating algorithms: An exploratory study of Facebook users in Myanmar. *Information Society*, 36(2), 97–108.  
<https://doi.org/10.1080/01972243.2019.1709930>

- Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3), 106–131. <https://doi.org/10.1177/1529100612451018>
- Lewis, T. (2020). COVID-19 misinformation that won't go away. *Scientific American*, 323(5), 41–43.
- Liao, S.-H., Hsian, P.-Y., & Wu, G.-L. (2014). Mining user knowledge for investigating the Facebook business model: The case of Taiwan users. *Applied Artificial Intelligence*, 28(7), 712–736. <https://doi.org/10.1080/08839514.2014.927695>
- Ovide, S. (2021, February 2). How to fix Facebook groups. *The New York Times*. <https://www.nytimes.com/2021/02/02/technology/how-to-fix-facebook-groups.html>
- Shearer, E. (2021). “More than eight-in-ten Americans get news from digital devices.” “Pew Research Center, Washington, D.C. (January 12, 2021). <https://pewrsr.ch/2MZqns7>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *American Association for the Advancement of Science*, 359(6380), 1146-1151. doi:10.1126/science.aap9559

White, Z. (2020). *User feedback loop with News Feed*. [Figure 1]. *STS Research Paper: Exploring methods to address COVID-19 misinformation on Facebook* (Unpublished undergrad thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

White, Z. (2020). *Actor-Network theory analysis of Facebook*. [Figure 2]. *STS Research Paper: Exploring methods to address COVID-19 misinformation on Facebook* (Unpublished undergrad thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Zaracostas, J. (2020). How to fight an infodemic. *The Lancet*, 395(10225), 676.

[http://doi.org/10.1016/S0140-6736\(20\)30461-X](http://doi.org/10.1016/S0140-6736(20)30461-X)