

Modest interventions complement each other in reducing misinformation

Proposals to fight online misinformation range from gently encouraging users to consider the accuracy of information ('nudges') to bans and removing content. Using modelling techniques, we find that these interventions are unlikely to be effective in isolation, but that a combined approach can achieve a significant reduction in the spread of misinformation.

This is a summary of:

Bak-Coleman, J. B. et al. Combining interventions to reduce the spread of viral misinformation. *Nat. Hum. Behav.* <https://doi.org/10.1038/s41562-022-01388-6> (2022)

Published online:

Published online: 23 June 2022

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

The problem

Over the past decade, concern has grown over the spread of misinformation in digital spaces¹. Academics, regulators and technology companies have proposed a wide range of solutions to target misinformation, from reducing its visibility online and encouraging people to share more judiciously to sanctioning spreaders of misinformation through bans and suspensions. None of these solutions is without costs, in terms of time spent identifying and responding to misinformation, reduced platform engagement or limiting user expression. Deciding on an appropriate path requires weighing these costs against potential benefits. Unfortunately, we as researchers do not possess, for instance, Twitter's ability to run experiments at scale and evaluate what will work. Twitter can run these experiments – but without having a theoretical understanding of what might happen, they are experimenting with risks at the scale of elections and pandemics.

The solution

Misinformation is a particularly prevalent challenge for public health officials during a pandemic, who rely heavily on mathematical models of disease spread² to make decisions about which interventions to use, when to use them and how they can be combined. We took a similar approach, building a model of misinformation spread during the 2020 US presidential election. From a large dataset of over 1 billion election-related posts on Twitter, we manually identified specific instances in which false narratives about the election spread rapidly across the platform³. Training our model on this dataset allowed us to simulate alternate worlds in which various interventions had been applied.

The first intervention we examined was perhaps the simplest – removing the related posts. We found this to be effective, particularly if done within the first few hours of its spread (Fig. 1). However, in practice, companies are likely to miss many posts because of the vast amount of content being created every minute. We found similar effects for 'virality circuit breakers', which algorithmically slow the spread of information but do not remove it entirely (Fig. 1) – but again, this intervention has the same limitations as content removal. Next, we investigated nudges, which encourage users to share information more judiciously⁴. We found that nudge interventions can be efficacious but unlikely to reduce misinformation enough on their own (Fig. 1).

To investigate a more drastic approach, we examined banning repeat spreaders of misinformation. We found this approach to be effective (Fig. 1) but that it would require removing many accounts, which remains highly controversial. As none of these approaches appears to be a panacea, we finally investigated what happens if each intervention is modestly used in tandem. We found that a combined approach could reduce misinformation (Fig. 1) without having to catch everything, convince most people to share better or resort to the extreme measure of account removals.

The implications

Our models suggest that we are unlikely to find a single solution for addressing misinformation online. Instead, combined approaches are likely to maximize benefits while minimizing costs. Our results further suggest that companies can reduce the harm caused by misinformation if they are willing to use some or all of these strategies.

One limitation of our study is that we focused on particularly viral misinformation and its spread within a single platform. Our approach cannot answer broader questions about misinformation spread by traditional media outlets or occurring across platforms. These facets of misinformation ecosystems are, for the moment, beyond the reach of mathematical models. However, acting early on rapidly spreading misinformation may have knock-on effects, preventing isolated events from spreading to traditional outlets and consolidating into broader narratives and conspiracy theories.

Our response to misinformation has been largely reactive – often intervening well after the damage is done. Mathematical models of misinformation spread will be an important tool for taking a more proactive, evidence-based approach to managing our digital ecosystems. We hope that our model becomes one of many that companies and regulators can rely on to balance the costs and benefits of such interventions. Progress in this area will critically depend on the willingness of companies to be more transparent and forthcoming about how their platforms and algorithms are designed and managed. If they are willing to open up to researchers and the public, we can start looking ahead and stop playing catch up, in the pursuit to counter misinformation.

Joseph Bak-Coleman and Jevin West
University of Washington Center for an Informed Public, Seattle, WA, USA.

EXPERT OPINION

|| The real significance of this work, in my opinion, is that it offers a data-backed framework by which we can start to compare and evaluate proposed solutions to the problem of fake news. In particular, the project goes all the way from data collection

and representation choices, through model development and simulation, to discussing policies from the points of view of the platforms and users. Such a large scope is rare.” **Lisa Friedland, formerly at Northeastern University, Boston, USA.**

FIGURE

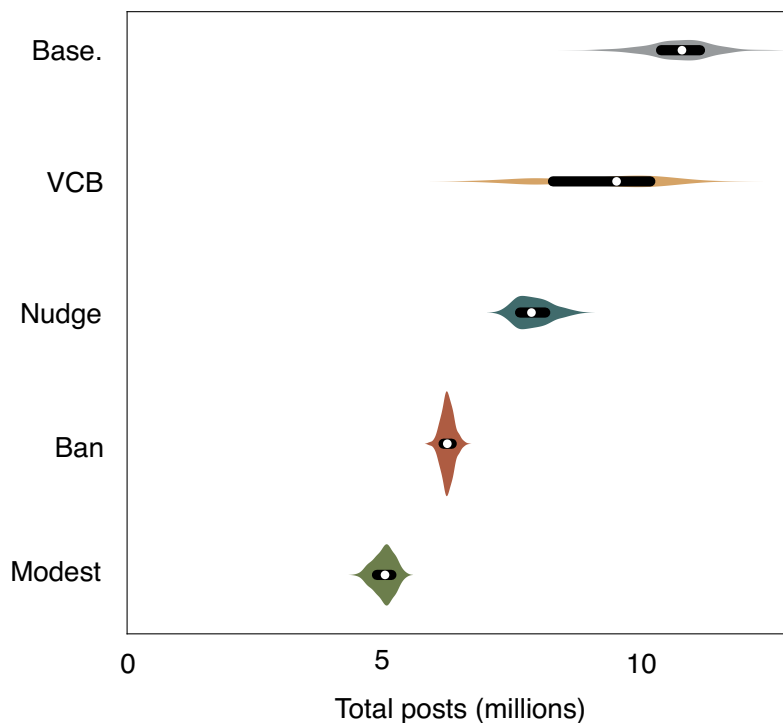


Fig. 1 | Simulations of total misinformation across interventions. Each violin on the plot shows the amount of simulated viral misinformation during the election. Base, baseline case of no interventions; VCB, virality circuit breakers, which algorithmically slow the spread of certain topics; nudge, reminders that encourage better sharing; ban, banning users who repeatedly share misinformation and have large numbers of followers; modest, combining all of these interventions at modest levels of enforcement. © 2022, Bak-Coleman et al., [CCBY 4.0](#).

BEHIND THE PAPER

Like many people in 2020, I spent a considerable amount of time looking at time-series plots of coronavirus case rates and models forecasting future rates if we adopted masking, social distancing and other policies. In the fall of 2020, I joined the Center for an Informed Public and found myself looking at time series of electoral misinformation that had an uncanny resemblance to plots of case counts. Unlike the coronavirus pandemic,

we did not have a rich taxonomy of mathematical models to guide our response. Instead, we watched these small stories in the early fall grow into large narratives culminating in the events at the Capitol Building. I found myself wondering what would have happened in an alternate universe where platforms had intervened in earnest. Through all the analysis and data, this paper is our attempt to gain insight into those alternate realities. **J.B.-C.**

REFERENCES

1. Zarcosata, J. How to fight an infodemic. *Lancet* **395**, 676 (2021).
A review highlighting misinformation challenges during the COVID-19 pandemic.
2. Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359 (2005).
This paper examines the effect of super spreaders on the spread of infectious disease and inspired the model in our paper.
3. Election Integrity Partnership. *The Long Fuse: Misinformation and the 2020 Election* <https://www.eipartnership.net/report> (Election Integrity Partnership, 2020).
A report providing a broader context to misinformation during the 2020 US presidential election and our data gathering process.
4. Pennycook, G. et al. Shifting attention to accuracy can reduce misinformation online. *Nature*. **592**, 590–595 (2021).
This paper describes experimental research proposing the use of nudges to reduce the spread of misinformation online.

FROM THE EDITOR

|| Mathematical modelling by Bak-Coleman et al. provides a rare opportunity to estimate the potential of moderation policies to reduce the spread of misinformation without testing them experimentally and without trying to reconstruct the social network structure. Regulators aiming to curb the spread of misinformation without impeding free speech will find these results illuminating.” **Arunas Radzvilavicius, Associate Editor, Nature Human Behaviour.**