

／ Social media algorithms are harmless, or are they?

by Marco Bastos

New research published in Science and Nature suggest that Facebook and Instagram are not causing political polarization. But there are limitations in the research design that need to be discussed.



Foto von Glen Carrie auf Unsplash

Are social media algorithms making people more partisan? Perhaps not — at least, not according to some of the peer-reviewed studies, released late July, that resulted from a collaboration between Facebook parent company Meta and outside researchers. One of the studies, appearing in the journal [Science](#), examined the effects of Facebook's and Instagram's content recommendation algorithms around the 2020 U.S. elections. It reportedly found that the algorithms increased user engagement but didn't affect people's existing political attitudes or polarization. [Meta itself](#), and some prominent [press headlines](#), touted such findings as causal evidence against the notion that social media algorithms harm democracy by polarizing political beliefs and the public debate.

There are, however, issues with the research design in some of these studies that merit discussion. In the [aforementioned study](#), researchers recruited participants who were given either the default algorithmically curated feed on Facebook and Instagram or a reverse-chronological content feed. Many [critics](#), including Facebook whistleblowers, have argued that algorithmic recommender systems designed to maximize user engagement create dangerous feedback loops and fuel political polarization. By showing some of the study participants a supposedly more neutral reverse-chronological feed, the researchers set out to compare whether these users' experiences of politically contentious issues differed significantly from those who were subjected to the usual algorithms.

The study expected participants to self-report significant changes to their personal opinions on politics, immigration, COVID-19 lockdowns, healthcare, unemployment, election knowledge, and other hot button issues within the 3-month period when some users were given a respite from the algorithms. The study also assumed that turning the algorithms off would show consequential effects on study participants when it came to consequential behaviors like attending protests or rallies, signing petitions, contributing money to political candidates, or trying to convince people to vote. Perhaps unsurprisingly, the study participants didn't self-report such significant changes in their attitudes and behavior in such a short period, so the researchers concluded that social media algorithms were not [the root cause of phenomena such as increasing political polarization](#).

Survey aside, the research team also had the chance to explore the behavioral data provided by Meta, Facebook and Instagram's parent company. In contrast to observational data like surveys that are based on what one declares, behavioral data are collected by tracking user activity and represent actual behavior. It registers what users did as opposed to what they think they did or what they would tell others that they did. The behavioral data show that the algorithms increased exposure to uncivil content and removed political content, particularly content from moderate friends and ideologically mixed audiences. The algorithms were also found to dramatically increase time spent on Facebook and Instagram. More disturbingly, removing the algorithms decreased exposure to both cross-cutting and like-minded sources, but the effect on the latter is twice as large as the effect on the former (i.e., the algorithms maximize exposure to like-minded content more so than to cross-cutting content). These significant findings in the behavioral data were downplayed by the absence of significant results in the observational, self-reported data collected with surveys.

This research design, in which behavioral data is benchmarked against observational data collected over a short period of time, seems to have been repeated in three out of the four studies released in [Science](#) and [Nature](#). This is strange because a randomized controlled experiment such as A/B testing with Facebook and Instagram users would allow for establishing causation. Yet this set of studies decided instead to rely on survey waves

rolled-out over a short period, which alone can hardly determine causal mechanisms. This is particularly strange because some of the articles imply causation, such as in the titles “Reshares on social media amplify political news but do not detectably affect beliefs or opinions”, and “Like-minded sources on Facebook are prevalent but not polarizing”.

In fairness to the researchers, there may be reasons for this method. The research design was primarily focused on whether the Facebook and Instagram algorithms were associated with the polarization of U.S. audiences only – as such, the researchers acknowledged that their findings can't be generalized to the rest of the world. It also seems that Meta only allowed experiments on participants who volunteered for the study (unlike the A/B testing [the company allows for advertisers, let alone internally for product testing](#)), a limitation that violates the requirements for randomized trial studies. The research team then decided to rely on surveys over a 3-month period to claim causation, despite the conflicting signals from the behavioral data they were given access to by Meta. The decision to focus on polarization, and to rely on observational data to identify changes in attitudes, may have prevented the research team from unpacking the many dimensions of the behavioral data that are only casually reported in the aforementioned study.

Much of the press coverage about the studies highlighted the limitations of this industry-academic partnership, with the project's independent rapporteur cautioning that Meta influenced how researchers evaluated the platforms, and that [independence by permission is not independence at all](#). However, the rapporteur also stated that the academics involved in this project were given ‘control rights’ over the papers and the findings, and thus had the final word on any disagreements with Meta. With the survey being the only component of the study that did not yield significant results, we can only wonder what other research teams would have found if given [comprehensive](#), representative behavioral data from social media platforms.

READ MORE



STORY

Reels of Fortune: Instagram-shaped memories for a bigger reach



OPEN LETTER

DSA must empower public interest research with public data access

／ Marco Bastos



Marco Bastos is an Ad Astra Fellow in the School of Information and Communication Studies at the University College Dublin. He is also Senior Lecturer in Media and Communication in the Department of Media, Culture and Creative Industries at City, University of London. Previous to that he held research positions at the University of California at Davis, Duke University, University of São Paulo, and the Goethe-University of Frankfurt. Marco Bastos is the lead investigator in the Twitter-funded project investigating echo chambers and political polarization in the aftermath of the Brexit vote. His research leverages computational methods and network science to explore the intersection of communication and critical data studies. He is the author of *Spatializing Social Media: Social Networks Online and Offline* (Routledge, 2021) and the forthcoming monograph *Brexit, tweeted*. His research on social media manipulation has been featured in BBC, New York Times, Guardian, Washington Post, Bloomberg, Wired, and BuzzFeed.