BROOKINGS

Report

# Echo chambers, rabbit holes, and ideological bias: How YouTube recommends content to real users

Megan A. Brown, Jonathan Nagler, James Bisbee, Angela Lai, and Joshua A. Tucker Thursday, October 13, 2022

Elon Musk's recent effort to buy Twitter along with court fights over social media regulation in Florida and Texas have recharged the public conversation surrounding social media and political bias. Musk and his followers have suggested that Twitter release regular audits of Twitter's algorithm—or that Twitter open source its algorithm—so independent parties can audit it for political bias.

Before Elon Musk entered the fray, however, a growing body of journalistic work and academic scholarship had begun to scrutinize the impact of social media platform algorithms on the type of content people see. On the one hand, most empirical research has found that user behavior, not recommendation algorithms, largely determines what we see online, and two recent studies disputed Musk's claim of anti-conservative bias. On the other hand, disclosures from the Facebook Files last fall suggested that adjustments to Facebook's algorithm amplified angry and polarizing content and may have helped foment the January 6 insurrection. Social media content feeds are crucial to media consumption today. By extension, then, it is critical to understand how the algorithms that generate our feeds shape the information we see.

In a new working paper, we analyze the ideological content promoted by YouTube's recommendation algorithm. Multiple media stories have posited that YouTube's recommendation algorithm leads people to extreme content. Meanwhile, other studies have shown that YouTube, on average, recommends mostly mainstream media content. In our study, which utilizes a new methodological approach that makes it easier for us to isolate the impact of the YouTube recommendation algorithm than previous work, we found that YouTube's recommendation algorithm does *not* lead the vast majority of users down extremist rabbit holes, although it does push users into increasingly narrow ideological ranges of content in what we might call evidence of a (very) mild ideological echo chamber. We also find that, on average, the YouTube recommendation algorithm pulls users slightly to the right of the political spectrum, which we believe is a novel finding. In the remainder of this article, we lay out exactly why such research is important, how we did our research, and how we came to these conclusions.

## YouTube's Recommendation Engine: Why Is It Important?

By many measures, mass polarization is on the rise in the United States. Americans are more willing to condone violence, less open to relationships that cut across party lines, and more prone to partisan motivated reasoning. We've seen two prime examples in the past two years. First, the nation's response to COVID-19: Preventative measures such as mask wearing and vaccination became inextricably linked to partisanship. Even more dramatically, many Republicans claimed that the 2020 U.S. presidential elections (although not the concurrent legislative elections) were riddled with fraud, culminating in the January 6 Capitol attacks, while Democrats largely accepted the results as legitimate.

While few claim that social media actually is the root cause of political polarization, many worry that the affordances of social media are accelerating the more recent rise in political polarization.[1] One prominent concern is that our rapidly evolving information environment has increased the number of ideological news outlets and made it easier for individuals to exist in "echo chambers" where they're rarely confronted with alternative perspectives. Many believe that social media algorithms exacerbate this problem by suggesting content to users that they will enjoy. While this can be harmless or even beneficial in areas like sports or music, in areas such as news content, health content, and others, this type of personalization could lead to harmful societal outcomes, such as siloing individuals into anti-vaccine, extremist, or anti-democratic echo chambers.

In our study, we focus on YouTube. YouTube, started in 2005 and acquired by Google in 2006, has grown to prominence as the internet's archive for video content. Even before Facebook, Twitter, Reddit, or other platforms implemented algorithmically-generated user feeds, YouTube was providing users with recommended videos to watch next. By many measures, YouTube is the largest social media platform in the United States. In 2021, 81% of American adults reported using YouTube, compared to 69% who use Facebook and 23% who use Twitter. YouTube is the second-most visited domain on the internet, just behind Google, part of their parent company. Twenty-two percent of, or roughly 55 million, Americans, also report regularly getting news on YouTube. In addition, YouTube's recommendation algorithm drives around 70% of total views on the platform. Taken together, these statistics suggest that YouTube's recommendation algorithm is vitally important for news consumption. Understanding what content YouTube recommends to users—and the extent to which YouTube recommends various types of harmful content—is both an important and challenging problem to solve.

## Why is studying YouTube's recommendation algorithm difficult?

For starters, the recommendation algorithm is highly personalized, meaning one individual's experience on YouTube can be completely different from another's. While there are documented cases of individuals who are radicalized on YouTube, these cases don't allow us to understand scale, prevalence, or cause, which is key to being able to remedy any adverse effects of the algorithm writ large. In addition, platform algorithms change frequently, and researchers outside the company do not have access to data to conduct audits.

For YouTube, outside researchers are limited to using user watch histories (donated by survey respondents to be used in research) or using web scraping to collect recommendations. Both methods present challenges for understanding the effects of the recommendation system on online consumption. The first, using donated watch histories, does not allow researchers to disentangle user demand for content from the supply the platform provides. That is, we can see what users choose to consume on YouTube. However, what users choose to consume within a YouTube session is a composite of what the platform chooses to show in recommendations, (i.e., the supply of videos) and the user's choices of which videos to actually watch (i.e., user demand, or the user's preference for particular content).[2] If researchers rely on watch histories and find increased consumption of right-wing content among a particular user, it could be a result of YouTube recommending increasing amounts of right-wing content to that user or it could be a result of that user receiving an ideologically diverse set of content but consistently choosing right-wing content. This is the crux of the difficulty in disentangling the effect of the algorithm from the effect of user choice on consumed content.

Alternatively, researchers can collect recommendations produced by an algorithm via web scraping. The way this works is that the researcher automates the process of visiting the YouTube webpage repeatedly and recording what is on it. In this way, the automated program can simulate the experiences of a user visiting YouTube. However, these automated visits to the web sites (we can think of this as a "bot" that watches YouTube videos) do not contain real user histories. No matter how a researcher programs a bot to simulate user behavior on YouTube, these user-agnostic recommendations disconnect the YouTube algorithm from the real user data on which it relies to operate, calling into question the extent to which web-scraped recommendations represent the lived experiences of users on the platform.

In our paper, we overcome the limitations of using watch histories or web scraping by analyzing what YouTube recommended to a sample of real users who participated in our study—that is, we are able to observe the actual videos that were recommended to users based on what YouTube chose to recommend to that person. However, instead of allowing users to choose which video to watch (and thus confounding the impact of the recommendation algorithm with user choice, as is the case when relying on watch histories) we constrain the user's behavior for the duration of our survey by requiring our participants to click on a predetermined recommendation (i.e., always click on the third (or first, or fourth) recommendation). While we could have programmed a bot to do the same thing, we would not have been able to record what YouTube recommended to actual human users with real user histories. With the data collection method we employ, though, we can isolate the impact of the algorithm on which videos are shown to real users. When we combine this with a novel method to estimate the ideology of YouTube videos (described below) we are therefore able to assess the impact of YouTube's recommendation algorithm on the presence of ideological echo chambers, rabbit holes, and ideological bias.

## How We Did Our Research

In Fall 2020, we recruited 527 YouTube users and asked them to install a web browsing plug-in—a piece of software that would allow us to see what appeared in their web browser—to record their YouTube recommendations.[3] Each participant was randomly assigned one of 25 starting videos, consisting of a mix of political content across the ideological spectrum and some non-political content from music, gaming, and sports.[4] Users were then randomly assigned to one of five "traversal rules," which instructed them to always click on a predetermined recommendation by the rank order of the recommendation. That is, a user would be instructed to always click on the first video, or always click on the second video, and so on. Respondents followed their assigned rule for 20 steps, and the browser extension collected the list of recommended videos presented at each traversal step. Thus, for each user, we would collect the set of twenty recommendations they received across twenty traversal steps, allowing us to understand the ideological content they were recommended.

After the survey, we used a novel method involving a machine-learning model (trained on Reddit and YouTube data) to estimate the ideology of each video recommended to users. To do so, we turn to Reddit, which is organized into sub-communities, or "subreddits," for particular interests or beliefs. These subreddits cover a variety of topics, from broad forums like r/politics to discuss political content and r/music to discuss the newest music to more narrow forums like r/dataisbeautiful to show pretty data visualizations and r/backyardchickens to discuss chicken raising. For our method, we focus on subreddits dedicated to political content like the aforementioned r/politics and other subreddits such as r/The_Donald, r/liberal, or r/Conservative. The underlying assumption of our method is that videos shared in these political subreddits are likely on average to be ideologically aligned with that subreddit. For example, a conservative Fox News video would be more likely to appear in r/Conservative than in r/liberal. We can use this information—what videos appeared in what ideological subreddits—to estimate the ideology of YouTube videos.

To that end, we collect all posts with YouTube videos shared on 1,230 political subreddits from December 31, 2011 until June 21, 2021. We remove posts with negative "upvotes" (users can give a post either a thumbs up or a thumbs down, with a negative score indicating that the members of that subreddit did not like the content, which we interpret as an indicating that the content is not aligned with the ideology of the subreddit). We filter the remaining posts for basic popularity metrics to make sure the videos and subreddits being used to train our machine learning model are actually informative for the model.[5] With our final set of videos, we use correspondence analysis, a method commonly used in social sciences, to estimate the ideology of the videos shared on Reddit.

However, in our survey users encountered many videos that never appear on political subreddits. Therefore as a final step, we train a machine learning classifier to predict the ideology of YouTube videos using ideology estimations from the (Reddit-based) correspondence analysis. Using state of the art natural language processing, we trained a BERT model on video titles, tags, and descriptions to predict ideology of videos using the text features for the videos. Using this method, we can estimate ideology for all videos that users encountered in our survey.[6]

## What did we look for?

We use these ideology scores to measure the three concepts noted in the introduction of this article: **ideological echo chambers**, **rabbit holes**, and **system-wide ideological bias**.

**Ideological echo chambers** refer to a distribution of videos recommended to a user that is both ideologically homogeneous and centered on the user's own ideology. For example, we consider a user to be in an ideological echo chamber if they are a conservative user who receives mostly conservative videos recommended from YouTube (and vice versa for liberal users). These users are in an "echo chamber" because they are only exposed to information that is consistent with their own ideology and prior beliefs. Echo chambers, as we define them in our article, are static: They represent the overall distribution of ideologies to which a user is exposed by the algorithm, rather than an evolving process that happens over the course of a traversal. So, if the YouTube recommendation algorithm puts users in echo chambers, we would expect users to see ideologically narrow content centered around their own ideology.

Alternatively, **rabbit holes** capture the process by which a user starts in a rich information environment and ends up in an ideologically extreme echo chamber. While much early social media research explored the prevalence of echo chambers on many platforms, rabbit holes are a specific phenomena related to personalized recommendation systems like YouTube's.[7] The underlying intuition of this hypothesis is that recommendation systems provide a self-reinforcing feedback loop whereby users click on content that they like, and YouTube provides a *more intense* version of that content. In a non-political example, this could look like users watching videos about learning how to start jogging and then receiving recommendations for ultramarathon or triathlon-related videos. In the political context, a user might start on content about the presidential election and land on content espousing Holocaust denial or white supremacy.

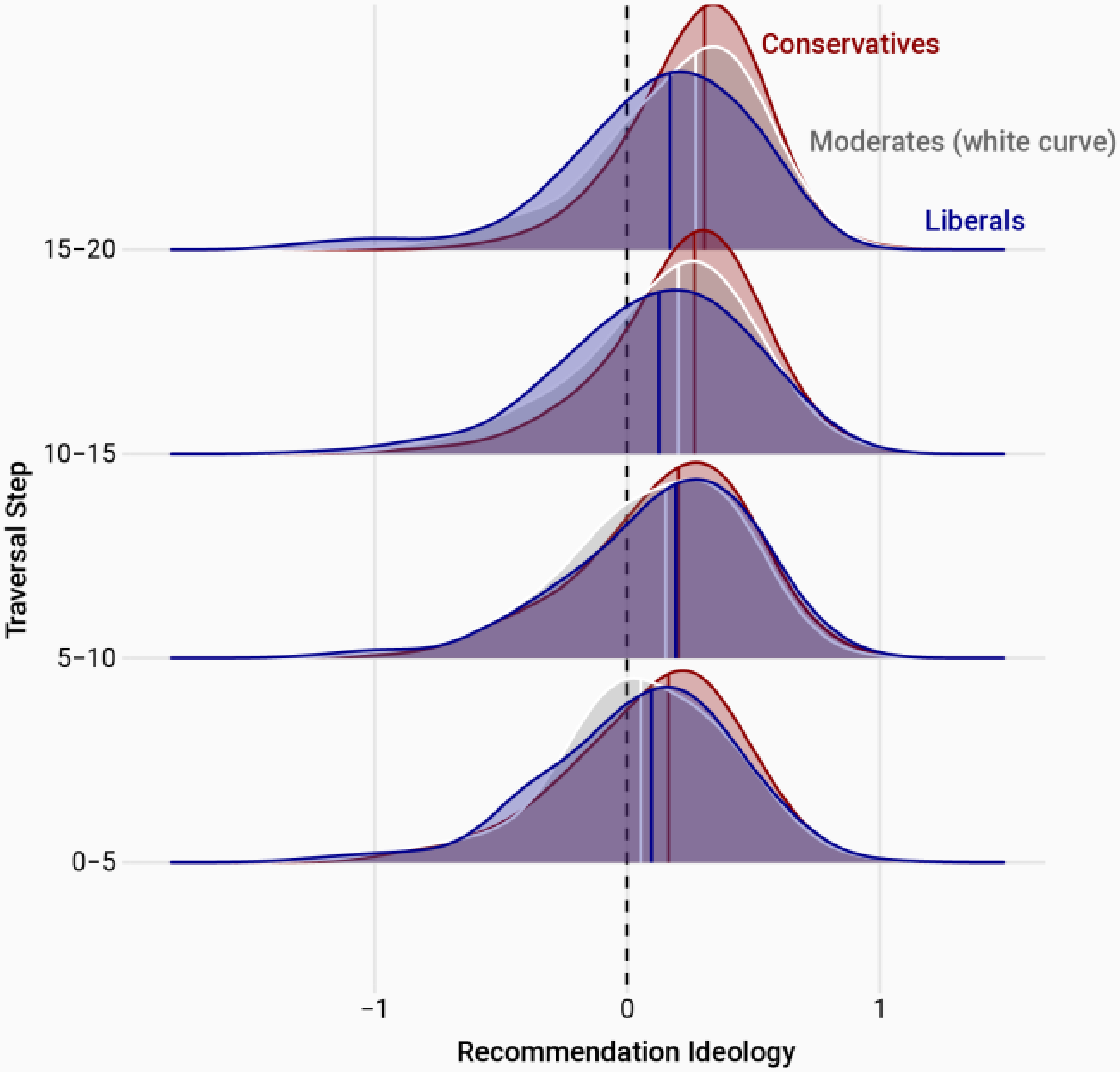
Finally, we look at **system-wide ideological bias**, meaning bias in the overall recommendations of the majority of users. More specifically, system-wide ideological bias refers to the process by which all users—regardless of their own ideology—are pushed in a particular ideological direction. For example, if all users are pushed toward ideologically liberal content, we would consider YouTube's recommendation system to have system-wide ideological bias toward liberal content.

## What we found

FIGURE 1

### YouTube's algorithm pushes users into (mild) echo chambers

Ideological distribution of recommended videos, by traversal step and user ideology



**Source:** NYU Center for Social Media and Politics (data);
Megan A. Brown (figure).
**Notes:** Positive ideology score on the x-axis is more conservative and a negative ideology score is more liberal.

B | Economic Studies
at BROOKINGS

The figure above summarizes our findings from the study and allows us to assess the prevalence—or lack thereof—of echo chambers, rabbit holes, and ideological bias. The figure displays the ideological distribution of recommended videos by traversal step—how many steps they had taken in the recommendation path–(0-5, 5-10, 10-15, 15-20) and a user's self-reported ideology (conservatives = red, moderates = white, and liberals = blue) where a positive ideology score on the x-axis is more conservative and a negative ideology score is more liberal. Therefore, the distributions at the bottom of the figure illustrate the ideological distribution of video recommendations users received during the first five videos watched during the traversal task, while the distributions at top of the figure illustrate the video recommendations users received during the last five videos watched during the traversal task.

If the YouTube recommendation algorithm were fostering echo chambers, we would expect liberals' recommendations to be more left leaning, conservatives' recommendations to be more right-leaning, and little overlap between liberals and conservatives. If the recommendation algorithm was leading users down rabbit holes, we would expect the distribution of video recommendations to shift toward either extreme as the number of traversal steps increased (again, moving up the y-axis). Finally, if the recommendation algorithm has an ideological bias, we would expect the distribution of recommendations to shift uniformly across all users, regardless of their ideology, in one direction or another.

So what can we conclude from the figure above?

**Echo chambers***:* We find that YouTube's algorithm pushes real users into (very) mild ideological echo chambers. As we can see in the figure, by the end of the data collection task (the top part of the figure, traversals 15-20), liberals and conservatives received different distributions of recommendations from each other (we see that the three different color distributions in the top part of the figure do *not* perfectly overlap) in the direction we would expect: Liberals (blue) see slightly more liberal videos and conservatives (red) see slightly more conservative videos. The difference was statistically significant, but very small. Moreover, as the figure clearly illustrates, despite these small differences, there is a great deal of overlap between videos seen by conservatives and liberals at all stages of the traversal process. Furthermore, while the variance of these distributions does decrease across the different traversal steps, meaning that the ideological diversity of recommendations declines, it is only by a minimal amount.

**Rabbit holes**: While we do find evidence of the recommendation algorithm contributing to mild echo chambers, we did not find evidence that the algorithm led most users down extremist rabbit holes. There is little evidence from the figure that the ideological distribution becomes more narrow or increasingly extreme over time.

However, these conclusions are based on average outcomes. Does this mean that extremist rabbit holes don't exist at all? To answer this question, we visually inspected each participant's traversal path, looking for cases where the average ideology grew increasingly narrow (i.e., no liberal videos suggested at all) and increasingly extreme (i.e., more conservative than Fox News). We found that 14 out of 527 (~3%) of our users ended up in rabbit holes, which we defined as recommendations that are more liberal/conservative than 95% of all recommendations, and more narrow than a variance of 0.4 on our scale ranging from -2 to +2. While this is a substantively small proportion of users, due to YouTube's size, a small proportion of users could still amount to non-trivial numbers of individuals occasionally falling into rabbit holes on YouTube. These findings underscore an important point about algorithmic systems and their effect on media consumption: Harmful effects are often concentrated among small numbers of users, and what is true for the platform as a whole can be very different for these sets of users.

**Ideological bias**: Finally, we found that, regardless of the ideology of the study participant, the algorithm pushes all users in a moderately conservative direction. If you look closely at the figure, you will see that all of the curves shift a bit to the right as they move up the y-axis. Although not large, these effects are statistically significant, and somewhat surprising given that we have never previously seen this feature of the recommendation algorithm publicly discussed. Moreover, the magnitude of the conservative ideological bias we identified far outweighs the magnitude of the echo chamber measure. This bias could be a result of two possible states of the world. First, YouTube's library could consist of a normally distributed set of videos centered around moderate content, but the algorithm could choose to only recommend content that is skewed ideologically conservative. Second, the YouTube library could consist of content that skews conservative and the algorithm could recommend videos representative of that underlying distribution. Our study does not allow us to adjudicate between these two causes.

# Discussion

Contrary to popular concern, we do not find evidence that YouTube is leading many users down rabbit holes or into (significant) ideological echo chambers via its recommendation algorithm. While we do not find compelling evidence that these rabbit holes exist at scale, this does not mean that some that the experiences of the small number of individuals who encounter extremist content due to algorithmic recommendations are not consequential, nor does it mean that we shouldn't be worried about the possibility for users to find harmful content online if they go searching for it. However, as we consider ways to make our online information ecosystem safer, it's critical to understand the various facets of the problem.

While our study was designed to test whether the *algorithm* leads users down rabbit holes, into echo chambers, or in a particular ideological direction, these outcomes could still emerge from user choice (recall that the recommendations in our study were collected *without* user choice). So, for example, a well known article by Baskhy et al. shows that Facebook recommended an ideologically diverse array of content but users consistently clicked on ideologically congruent content. In another study of YouTube, Chen et al. found that other platform features—subscriptions and channel features—were the primary path by which users encountered anti-social content.

Furthermore, other platforms, like 4chan, are hotbeds for extremist content. Indeed, if an individual is bound and determined to jump down a rabbit hole online, they can do so fairly easily. What we explore in our work is incidental exposure: that is, users who are perusing content and encounter harmful content by accident, subsequently leading them down a rabbit hole. While recommendation systems may play a small role in this type of incidental exposure, we do not find significant evidence that they drive consumption of harmful content (at least on YouTube). Other studies have found that harmful content is often encountered off-platform via link sharing, driving users to extreme places on YouTube via the internet at large rather than the recommendation algorithm. That's what makes this problem tricky. If it's not just the recommendation engine and instead it's the entire online ecosystem, then how do we fix it?

Our findings are consistent with—and add additional evidence to—a growing body of research showing that YouTube is not consistently pushing harmful or polarizing content to their users but rather that users self-select into viewing the content when offered. Collectively, the research suggests that there is unlikely to be one technological panacea to reducing the consumption of harmful content on YouTube. Instead, we need to be sure we focus both on the amount of harmful content online as well as the (many) paths which users might take to this content. Focusing solely on the role of YouTube's algorithm in advertently luring people to extremist content may make for great headlines, but our research suggests that this alone is not going to get us at the crux of the problem.

---

*The Brookings Institution is a nonprofit organization devoted to independent research and policy solutions. Its mission is to conduct high-quality, independent research and, based on that research, to provide innovative, practical recommendations for policymakers and the public. The conclusions and recommendations of any Brookings publication are solely those of its author(s), and do not reflect the views of the Institution, its management, or its other scholars.*

*Google, YouTube's parent company, provides support to The Brookings Institution. The findings, interpretations, and conclusions in this report are not influenced by any donation. Brookings recognizes that the value it provides is in its absolute commitment to quality, independence, and impact. Activities supported by its donors reflect this commitment.*

## Footnotes

1. [1] Although see Asimovic et al., 2021 for an example of a study that found evidence in Bosnia and Herzegovina that Facebook usage during a period of war remembrance actually reduced levels of ethnic polarization.
2. [2] Of course, another way that Google/YouTube can recommend YouTube videos is through Google Search results. Users can also find YouTube content linked to or embedded in other websites they visit. While this raises interesting questions for future research, our focus was on recommendations from within YouTube.
3. [3] This research was reviewed and approved by NYU Internal Review Board (#IRB-FY2020-4647).
4. [4] Starting videos were selected to be balanced across the ideological spectrum, including five liberal, five conservative, and five moderate starting videos. Additionally, we included nine non-political videos (three sports, three music, and three video gaming) in the starting video set. Respondents were randomly assigned one of these twenty-five videos at the start of the survey. Videos were refreshed throughout the study period to adjust for videos that were deleted or to add newer videos.
5. [5] For more details, see https://csmapnyu.org/research/echo-chambers-rabbit-holes-and-algorithmic-bias-how-youtube-recommends-content-to-real-users.
6. [6] For those interested in more information for how we used Reddit data to estimate the ideology of YouTube videos, including validation measures for the methods described in the text, see our working

paper on "Estimating the Ideology of Political YouTube Videos".

7. 7 This is not to say that other platforms do not have recommendation systems: Facebook and Twitter provide algorithmically-generated newsfeeds; Twitter provides following recommendations, TikTok's experience is fully algorithmically generated.