Recommending Descent: Understanding How Bias and User Behavior Impact the Political Topography of Recommendations on YouTube

Daniel Covelli
Division of Undergraduate and Interdisciplinary Studies
University of California, Berkeley
Berkeley, United States
danielcovelli@berkeley.com

December 21, 2020

Abstract

This paper attempts to determine the underlying causes of recommendation discrepancies for political content on YouTube. Much of the work on YouTube's watch-next algorithm claims that the platform favors right-leaning political recommendations and consistently leads users down "radicalization pipelines" of extreme content. To measure the validity of these claims, I used three autonomous agents, or bots, to collect and classify the political content of 3750 YouTube videos. Using a random-walk algorithm, these agents traversed YouTube's video recommendation graph guided by conservative, liberal, and positional biases, respectively. This paper finds that left-leaning and right-leaning recommendations occur almost equivalently for users who select for these biases. This paper also finds substantial qualitative differences in structure and content associated with left and right leaning video recommendations. These results suggest that radicalization pipelines are not unique to right-leaning content and that access to these pipelines are largely a user-driven phenomenon.

1 Introduction

Social media has drastically changed the political media landscape. Before social media platforms rose to prominence, political media was primarily dominated by institutional actors like Universities, Newspapers, and Cable News. These gatekeepers had a monopoly on the infrastructure and capital needed to disseminate political information to the larger public. It was therefore entrusted in these legacy actors to screen political information for whether it was appropriate or constructive for public discourse. Web 2.0 and the rise of social media companies changed all of this. YouTube, Facebook, and Twitter, as well as others, have

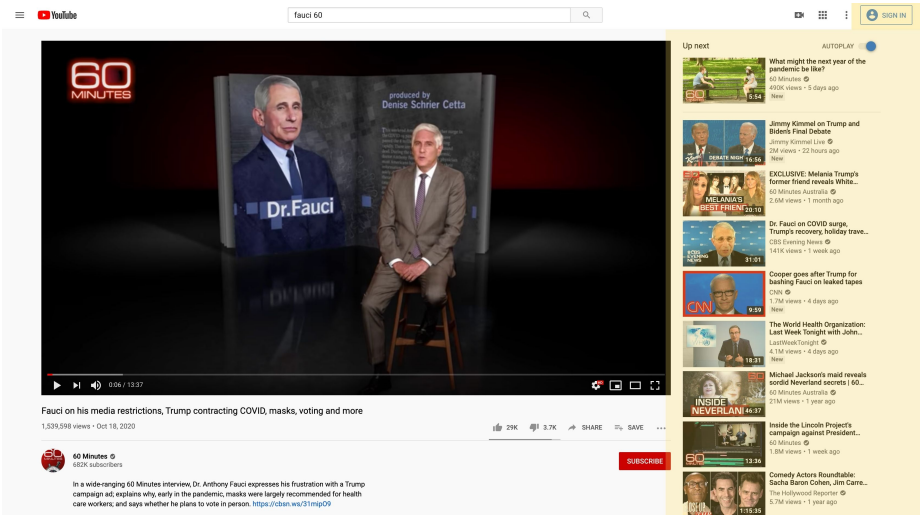


Figure 1: Watch-next algorithm as seen on YouTube's user-interface.

broadened access to the infrastructure needed for disseminating information and opinions to the larger public. Now anyone with a smartphone and a basic understanding of video editing can share information on all of these platforms.

With the barriers to entry in political media greatly decreased, platforms like YouTube have arisen as burgeoning spaces for independent political media producers. However, the democratization in media production has brought with it sharp criticism from political press and academia alike. Some of the most popular and widely echoed critics claim that YouTube and its recommendation system have created "pathways" that lead content consumers to politically extreme content (Lewis 2018; O'Callaghan 2014; Riberio 2019; Roose 2019; Tufekci 2018). Much of the research and thought in this space focuses on right-leaning groups like the Intellectual Dark Web, Alternative Influence Network, and Alt-Right/Light. These researchers argue that YouTube's recommendation system, which is developed primarily to increase viewership, biases for politically extreme content because it is more engaging than politically moderate content. Here, the claim is that YouTube's watch-next algorithm systematically favors recommendations to right-wing content and then leads viewers to the most extreme fringes, including to subjects like white nationalism, Qanon, and 5G conspiracies.

Despite the popularity of these perspectives, there is a sizable push back against the existence of a YouTube radicalization pipeline. Many researchers argue that YouTube has merely given a platform to ideas and beliefs that were always present in society, but were never widely expressed due to the former institutional nature of political media. Here, instead of YouTube and social media being blamed for an increasingly polar media ecosystem, it is an increasing polar user base that drives right-leaning political commentators to prominence (Bakshy

2015; Faddoul 2020; Lewis-Kraus 2020; Munger 2020; Zollo 2018). In addition to these claims, other researchers have suggested that YouTube's watch-next algorithm, and other recommendation systems like it, actually help drive users away from the political fringes by increasing the likelihood of exposing them to political content that they may disagree with or by intentionally recommending them towards more moderate content (Flaxman 2016; Hosanagar 2014; Ledwich 2020).

In order to measure the validity of radicalization pipeline theory and its critiques, I used three autonomous agents, or bots, to collect quantitative and qualitative metadata on roughly 3750 video recommendations offered by YouTube's watch-next algorithm. Each bot was given a politically motivated selection bias for data collection. One bot selected for videos classified as right-leaning, another selected for videos classified as left-leaning, while another consistently picked highly ranked recommendations. Each bot started its data collection from the same set of 25 starting points allowing me to isolate how YouTube's watch-next algorithm reacted to different kinds of political selection bias.

After conducting 75 independent trials, this paper finds that the political composition of recommendations offered by the watch-next algorithm was largely driven by the selection bias of the three bots. Particularly, I found that the probability of receiving content recommendations of the same political denomination as the specified selection bias is equivalent for both the left and right-leaning bots. I also found that when selecting for the highly ranked videos, there was limited political bias in the resulting recommendations. While the proportion of politically biased recommendations is primarily driven by the political bias of the bots themselves, there are substantial qualitative differences in the political topography traversed by each bot. YouTube's right-wing media ecosystem is primarily dominated by Fox News and a wide array of independent Intellectual Dark Web podcasters and theorists. On the other side, left-leaning recommendations seem to be primarily centered on institutional actors like MSNBC, late night talk shows, and "explainer" journalism. Despite the qualitative differences in these spaces, this paper finds limited evidence to suggest that YouTube favors one political content space over the other, casting doubt on the veracity of radicalization pipeline theory and suggesting that the existence of these pipelines is largely a user driven phenomenon.

In order to determine the validity of radicalization pipeline theory, this paper will proceed by reviewing the current literature, discussing the adopted research methodology, and analysing trends in the data.

2 Prior Work

Much of the attention around YouTube's recommendation algorithm and political polarization focuses on a subset of academic and journalistic work centered around right-wing radicalization on the platform. Perhaps one of the most

influential researchers in this space is Zeynap Tufekci. In a 2018 article titled "YouTube, the Great Radicalizer", Tufecki argues that YouTube's watch-next algorithm biases its recommendations towards more extreme content. In the arena of political content, she argues that whether starting from right-wing or left-wing content, recommendations tend to bias towards extreme right-wing content. Other articles seem to echo these concerns, one highlighting the story of a young man indoctrinated by extreme right-wing ideas recommended to him by YouTube's watch-next algorithm (Roose 2018). Both Tufecki and Roose argue that the inflammatory nature of these videos make them more likely to be rewarded by YouTube's algorithm.

Many academics have also corroborated these findings, concluding that YouTube tends to bias its recommendations towards exceedingly extreme right-leaning content. In an analysis of the topography of conservative content on YouTube, researchers describe the unique structure of right-leaning political influencers in a network called the Alternative Information Network (AIN). The researchers suggest that once a content consumer is watching videos in more moderate regions of the AIN, also referred to as the Intellectual Dark Web (IDW), they are likely to be recommended videos from conspiratorial and alt-right regions of YouTube (Lewis 2020). In an analysis of YouTube comments, researchers found that users consistently migrate from milder political content to more extreme political content via content recommendations. They also support the thesis that alt-right content is easily accessible via recommendations from IDW channels (Ribeiro 2020). In a study conducted on English and German YouTube channels, researchers found that content consumers on YouTube who watched content from Extreme Right (ER) YouTube channels were very likely to be recommended more ER content, while being very unlikely to be recommended content that was politically more moderate (O'Callaghan 2014). A study researching the structure of conspiratorial videos on YouTube, provides a compelling explanation for this phenomena. The researchers argue that because conspiratorial videos, including fringe right videos, are novel and provoking they tend to yield higher engagement and subsequently invoke more recommendations (Faddoul 2020). Going forward, this paper will refer to the above body of research as *Radicalization Pipeline Theory* (RPT).

Despite the popularity of radicalization pipeline theory, there is an equally large body of research that walks back or outright rejects many of the claims made by radicalization pipeline theorists. Much of these studies focus on the effects that user behavior and bias have on the kinds of political content recommended to users. One of the most cited studies in this space focuses on content recommendations on Facebook and reveals that users largely drive polarizing recommendations by aggregating in homogeneous communities of interest. Because of selective exposure into approving political groups, content recommendations offered to these users predominately reinforce their preexisting political biases and lead to increasingly self-reinforcing political ecosystems on the platform (Zollo 2018). In an analysis of right-wing content spaces on YouTube, one study argues that fringe content creators seem to be gaining prominence on YouTube because the platform has allowed them to cheaply produce content and more

easily connect with viewers. Here, the paper suggests that groups like the AIN have emerged on YouTube because of existing biases in the larger public that have not been expressed by legacy media outfits (Munger 2020). Other studies have suggested that recommendation systems have the opposite effect of pushing users to the extreme fringes, rather that these systems actually help to bring internet users closer together. In a study focused on online product recommendations, researchers found that because recommendation systems were effective at getting users to browse more often, users were actually more likely to purchase products that other users had bought (Hosanagar 2014). In another study of the browser history of 50,000 online news readers, when exposed to content recommendations within new sites, users were more likely to receive political information that didn't support their preferred political views (Flaxman 2016). Extrapolating these studies to YouTube, it is possible that the watch-next algorithm could similarly be effective at exposing users to political content that they disagree with[1].

The objective of this paper will be to determine the validity of Radicalization Pipeline theory and its criticisms. Particularly, this paper will attempt to understand how YouTube's watch-next algorithm changes as selective bias changes in random-walk scenarios and whether YouTube consistently recommends AIN content over liberal or politically agnostic content.

3 Methods

The goal of this paper is to capture the effect of user behavior on the output of YouTube's recommendation algorithms using autonomous agents as proxies for actual users. These bots were designed to mock the content consumption and browsing behaviors of politically biased users. One bot represents a right wing partisan. This bot search's for videos that have the highest level of right wing bias. Another bot represents a left wing partisan. This bot search's for videos that have the highest level of left wing bias. Another bot will represent a content consumer with no political bias.

It is important to note here that despite how most users interact with YouTube, the bots in this paper did not receive account based recommendations. Rather, the bots accessed recommendations on YouTube as users who were not logged in (Fig. 1). This choice was made because of technical limitations of YouTube's Data API[2], however for the scope of this paper, this approach is sufficient in enabling conclusions to be drawn about recommendations on the platform.

At a high level, the bots can be thought of as agents who walk through YouTube recommendations, making video selection decisions based on a biased heuristic. There are three bots currently in use; one left leaning bot, one right

[1] In another interesting and seemingly contradictory study, researchers found that exposing content consumers to tweets that they politically disagree with actually made them feel more passionate about their own political beliefs (Bail 2020).

[2] https://developers.google.com/youtube/v3/getting-started

leaning bot, and a neutral bot. The random-walk algorithm used by each bot, can be broken down into six steps.

- **Starting Query**: The bots are presented with a list of 'starting' videos. These starting videos are accessed using the YouTube API's query tool[3], which allows search queries and content categories to be submitted in exchange for a number of videos that match the given criteria.
- **Video Categorization**: From this set of starting videos, each video is classified based on its political characteristics using the Recfluence Channel Review Data Set (Ledwich 2020). Videos that contain many conservative leaning Recfluence classifications are given a positive polarization score, videos that contain many liberal classifications are given a negative polarization score. For videos that were not found in the Recfluence data set or for ones that contain apolitical classifications, a zero value polarization score is given.
- **Video Selection**: Given a bots defined political bias, it will pick, from the classified video's, the video that best fulfils its bias. For the right-wing bot, the video with the highest polarization score is selected. For the left-wing bot, the video with the lowest polarization score is selected. For the non-biased bot, the first recommendation is selected irrespective of its given polarization score.
- **Data Collection**: With each video in the array given a Recfluence classification, a polarization score, and with a video selected, the bots then collect all of this information for later analysis.
- **Request Recommendations**: After all important data has been collected, the bot then requests recommendations based on the selected video. YouTube's Data API will then return a new array of ten video recommendations.
- **Repeat**: With these new recommendations, the bots will categorize each video, select the video which best aligns with its bias, collect all of the important data, and request recommendations, repeating the previous four steps 4 times for a total of 5 iterations.

Important design decisions for the previous steps include, the search queries, the classification process, and the kinds of meta-data collected at each iteration

3.1 Starting Queries

As described in the first step, the bots begin their random-walks by submitting a search request in exchange for 10 videos matching the given criteria. Here it is important to discuss how the search requests were derived. Each search request is composed of a query string and a content category. A query string is

[3] https://developers.google.com/youtube/v3/guides/implementation/search

similar to what a user would input into YouTube's search bar, while the content category acts as a filter for the search results. In order to select query strings in a non-arbitrary way, I used Google Trends', Trending Searches[4] feature to select politically relevant query strings. I hand selected trending terms based on recency and my own perception of their political relevance. For instance, recently trending terms that were the subject of current political debate, like *Election Fraud*, were chosen over terms that were trending for other reasons, like *Christmas*. Despite best efforts, this approach is subject to experimental bias and should be formalized for future study[5].

Along with the query string, a content category was also provided. The content category, provided as a `topicId`[6], was statically set to *politics* for all runs of the random-walk algorithm. This helped ensure that all returning search results would be politically relevant even if the query string did not return political videos on its own. After submitting the query string and content category, the bots received a list of 10 relevant videos.

3.2 Classification Procedure

In order for each bot to select a video based on its political content, all videos were mapped to a political classification. The bots achieve this by cross referencing each video and its associated channel with the Recfluence Channel Review Data Set[7]. The data set is composed of ∼6000 hand labeled and machine labeled YouTube channels. The original Recfluence data set was composed of 800 YouTube channels, each hand labeled by a committee who systematically analysed the political content for each channel. Using consensus, the team was able to classify each of the 800 channels into 18 different political classifications (Ledwich 2020). From this original data set, the team has begun using machine learning algorithms to classify channels and has so far classified an additional 5000 channels (Clark 2020). While the accuracy of the machine labeling approach is still undefined, based on preliminary results, the accuracy is likely acceptable for the scope of this paper.

After a bot has cross referenced each video with the Recfluence data set, the bots then pick the video that best aligns with their political bias. In order to do this, each political classification in the Recfluence data set is mapped to a weight, loosely representing the leftness or rightness of a given classification. Right leaning classifications are given positive weights, left leaning classifications are given negative weights, and unbiased classification are given zero weights[8]. These weights were assigned based on the extremity of the given classification. For instance, the `PartisanLeft` classification, associated with late night talk shows and *MSNBC*, was given a weight of `-2.0` while the `PartisanRight` classification, associated with *Fox News*, was given a weight of `2.0`. More extreme political

[4] https://trends.google.com/trends/trendingsearches/daily?geo=US
[5] For a full list of query strings, see Appendix A.
[6] https://developers.google.com/youtube/v3/docs/search/list
[7] https://github.com/markledwich2/Recfluence
[8] For a full list of Reclfuence classifications and their associated weights, see Appendix B.

classifications were given greater weights, for example `SocialJustice` was given a weight of `-3.0`, while `AntiSJW` was given a weight of `3.0`.

However, two concerns immediately arise because on this approach. Firstly, weights were hand assigned to each classification. Secondly, the scale of these weights was largely an arbitrary decision. It should be reemphasised here that the weights should be thought of as a loose numeric representation of the political content of each category. In terms of this research paper, this loose numeric representation is sufficient, however a more formal methodology must be adopted for future work.

After mapping the Recfluence classifications to weights, each video was given a polarization score. In many cases, videos were associated with multiple different classifications in the Recfluence data set. Because of this, many channels were mapped to multiple different weights. In order to compensate for this, the polarization score was created as an average of these weights. With this polarization score, the bots were able to select the best video for their given political selection bias.

3.3 Data Collected

At each step a number of features from each video were captured. Many of the features captured were meta-data supplied by the YouTube Data API. This meta-data included a videos *name*, *channel name*, *video description*, and *data published*[9]. Other features, important for the analysis, were captured using the Recfluence data set. As described above, the Recfluence classifications for each video along with each videos polarization score was collected.

Each video was also given an additional discrete category based on its polarization score. For example, videos with a very large polarization score, greater then `1.5`, were given the general categorization of *Far Right*, videos with moderately positive polarization scores, between `0.5` and `1.5`, were given the general categorization of *Right*, those with a close to zero polarization score, between `0.5` and `-0.5`, *Neutral*, those with moderately negative, between `-0.5` and `-1.5`, *Left*, and with very negative scores, less than `-1.5`, *Far Left*. Again, it should be noted that these general categorizations derived from the polarization score should be thought of as a loose discrete representation of the political content in each video[10].

Data about which videos were and were not selected by the bots was also collected, as well as the video id from which each video was recommended from. This latter field was left empty when videos were returned as a responses from the initially starting query step.

To recap, each bot in this paper was designed to mock the browsing behavior of three different politically motivated YouTube users. The bots use a random-walk algorithm to make politically bias decisions about which videos to select and

[9] https://developers-dot-devsite-v2-prod.appspot.com/youtube/v3/docs/search

[10] For full list of these categorizations with examples, see Appendix C.

request recommendations for. Each bot repeats the steps of the algorithm 5 times during the course of a trial, where each trial is started by a hand-picked query string.

4 Findings

After conducting 75 unique trials and collecting a total of 3750 videos, this paper finds minimal evidence supporting radicalization pipeline theory. This result is broken down into three separate findings:

Left/Right Network Topology Firstly, this paper will look at the general structure of recommendations supplied to the three different bots across their random walks. It finds that right-leaning recommendations are centralized around *Fox News*, and its affiliates, which act like a recommendation beacon for conservative YouTube videos. On the other hand, left-leaning recommendations on YouTube seem to be more decentralized. This suggests that their their may be a radicalization pipeline from Fox News to other, potentially more extreme, conservative channels.

Divergent Recommendation Trends Secondly, this paper looks at general trends in the political content recommendations received by each bot. It finds that in general, the left-leaning bot and the right-leaning bot receive content recommendations with increasingly negative and positive polarization scores as they progress through their random walks. For the neutral bot, there seems to be no substantial skew to either right or left leaning recommendations, as measured by polarization score. This suggests that their does seem to be a radicalization pipeline on the right and the left. It also suggests that users determine whether they are driven to the political extremes or not.

Recommended Category Proportions And finally, this paper looks at the proportion of recommendations to politically left, right, and neutral videos for the three different bots. It finds that the Right Bot receives almost the same proportion of right-leaning video recommendations that the Left Bot receives of left-leaning video recommendations. For the Neutral Bot, there seems to be an equal amount of left-leaning video recommendations and right-leaning video recommendations. For all three of the bots, most recommendations pointed to politically neutral videos. This further suggests that users primarily determine whether they are driven to the political extremes or not.

4.1 Left/Right Network Topology

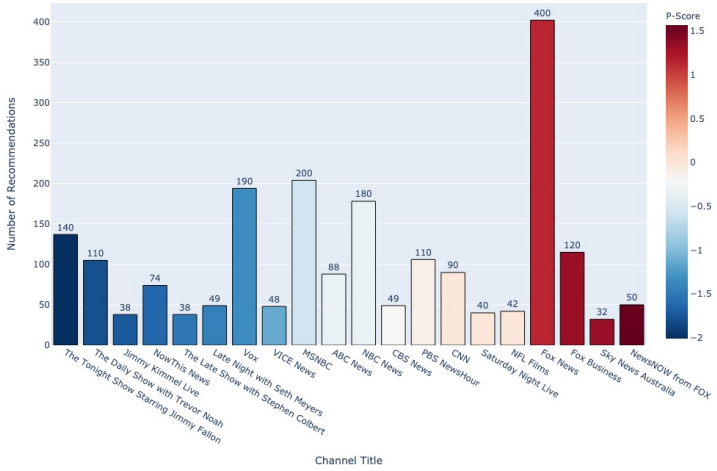
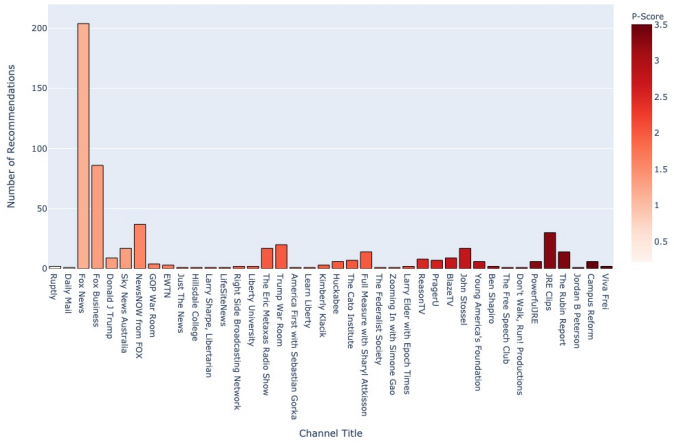


Figure 2: Top 20 most recommended channels and their polarization scores. Each channel is color mapped based on its polarization score (P-score). Numbers above each bar are associated with the y-axis.

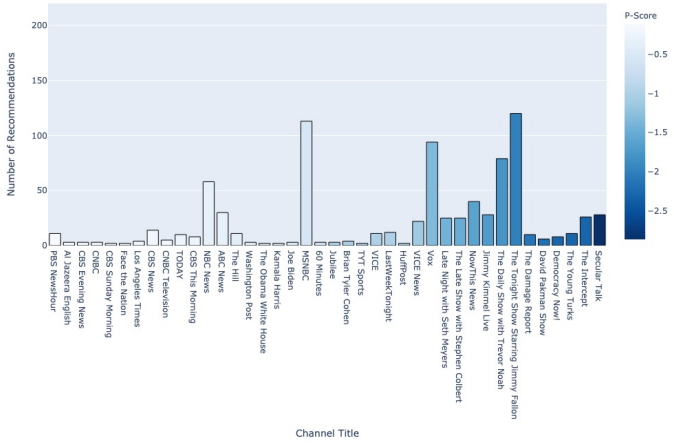
When looking at the most popular channels recommended to each bot, we can see clear differences in the distribution and structure of channel recommendations. Figure 2 shows the top 20 most commonly recommended channels across all trials with each channel color coated by its polarization score. Fox News is the most commonly recommended channel, with about 400 recommendations, while other top channels, like NBC News, MSNBC, and Vox, received around 200 recommendations each. While Fox News is the most commonly recommended channel, left-leaning channels in this list receive a greater number or recommendations in total compared to their right-leaning counterparts. This is even more apparent when looking at channels specifically encountered by the Left and Right Bots.

Figure 3 shows the top 40 channels recommended to the Left and Right bots, respectively. For the Right bot, the vast majority of recommendations it received were to Fox News or Fox Business. The rest of its recommendations are distributed amongst a number of relatively small ideologically conservative channels. A point that should be noted here is that, the Right Bot did not receive a recommendation to channels that are associated with the Extreme Right[11]. Rather, recommendations made to the Right bot, outside of Fox, were to increasingly conservative or pro-Trump channels like Ben Shapiro, BlazeTV, and PragerU. For the Left bot, the vast majority of channel recommendations were to more moderate or mainstream media outlets, like Vox, The Late Show,

[11]Extreme Right in this context referrers to channels that engage in openly anti-semetic, islamophobic, or white nationalistic speech. Term defined in O'Callaghan.



(a) Right Bot



(b) Left Bot

Figure 3: Top 40 channel recommendations for the Right and Left Bots.

and The Young Turks, who received a relatively more distributed share of recommendations compared to the Right Bot.

This analysis shows that the structure of right and left wing recommendations on YouTube vary greatly. In terms of radicalization pipeline theory, it reveals somewhat contradicting evidence. RPT suggests that YouTube consistently draws users away from politically moderate content to more politically extreme content. When looking at the result for the Right bot, the vast majority of recommendations go to Fox News. After selecting a video by Fox, the algorithm mostly continues to recommend more Fox videos. For the Left bot, when it selects a progressive video, the algorithm seems to be serving a more diverse

sample of channels that are ideologically consistent. Determining whether this is an intentional feature of YouTube, to restrict conservative users to more mainstream channels, is beyond the scope of this analysis.

Despite this homogeneity around Fox News, there are a number of more ideologically conservative channels that together formed a substantial share of recommendations to the Right Bot. While these channels do not constitute membership in the Extreme Right, Figure 3 reveals that YouTube has the potential to recommend increasingly more conservative content. Given the addition of more trials, it is possible to expect that YouTube could eventually recommend ideologically extreme channels. Although the effect was not directly observed in this analysis, this ability to access smaller more ideological channels outside of Fox, supports the the theory that YouTube can recommend content in the increasingly more extreme fringes of YouTube conservatism.

4.2 Divergent Recommendation Trends

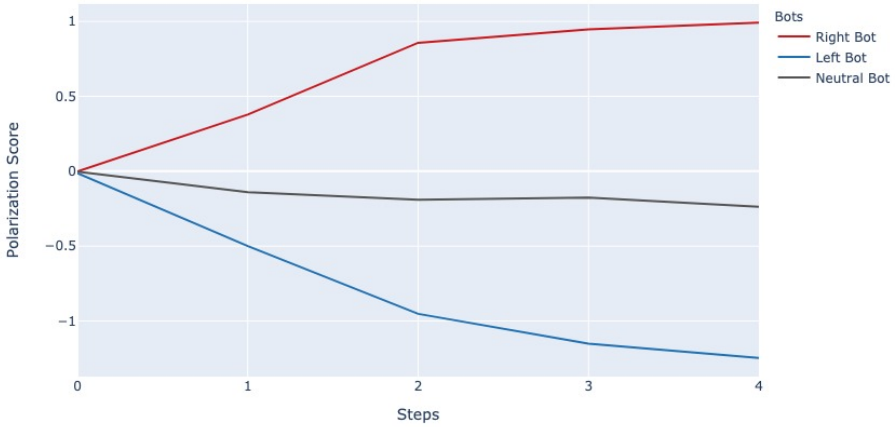


Figure 4: Average polarization score across 5 steps for each group. Positive polarity score is associated with right-leaning content and negative polarity score is associated with left-leaning content.

As the bots traversed YouTube's recommendation graph, the kinds of political content recommended to them varied greatly. Figure 4 represents the average polarization score of content recommended to each bot across 5 steps. As the Right Bot progressed through its steps, the content it encountered was generally more conservative, as measured by polarization score. The same is true for the Left bot, where content recommend to it grew increasingly liberal. For the neutral bot, there seems to be a slight liberal bias in its recommendations but

not significant enough to make affirmative claims of political bias.

The x-axis in Figure 4 captures the step at which the data was collected, while the y-axis captures the mean polarization scores for the data. Its important to point out that these trajectories are the mean of 25 individual trajectories conducted by each bot[12]. At step zero, where each bot enters a search request in exchange for 10 videos, the political content is generally neutral, with a polarization score of effectively `0`. After each bot categorizes, picks, and requests a recommendation from these starting videos, the polarization scores begin to diverge. At step 1, the Right Bot on average receives content with a polarization score of `0.45`, the Neutral Bot receives content with a polarization score of `-0.15`, and the Left Bot receiving content with a polarization score of `-0.5`. As the categorize-pick-recommend steps continue we can see that the bots continue to increase in their differences. By the last step, the Right Bot receives content with an average polarization score of `1`, the Neutral Bot receives content with a score of `-0.2`, and the Left Bot with a score of `-1.25`.

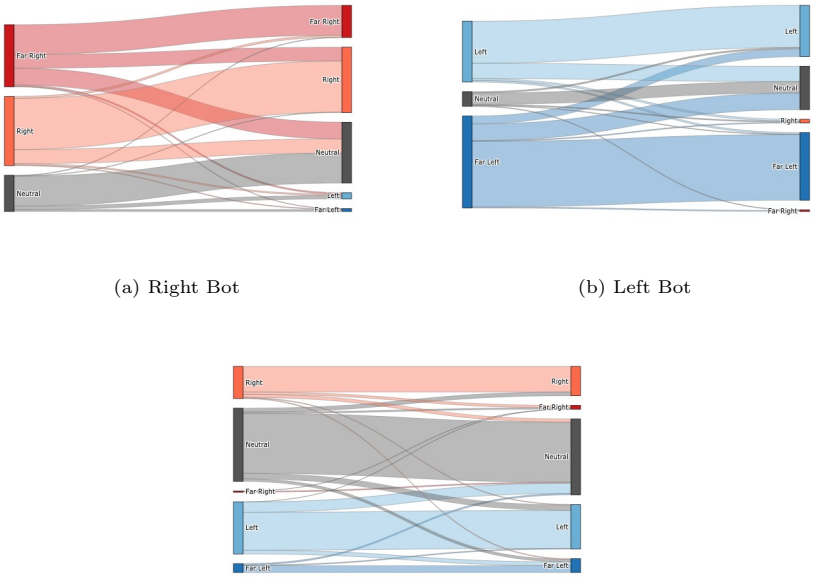
The above analysis contradicts radicalization pipeline theory in two important ways. Firstly, while Figure 4 suggest that YouTube consistently sends users down pipelines of increasingly extreme right-leaning content, unlike in radicalization pipeline theory, these pipelines seems to exist on the left as well. The analysis suggests that users who select for left-learning content, as well as right-leaning content, receive recommendations that increasingly align with their given political bias. Therefore, the analysis suggests that radicalization pipeline theory is limited in its focus on right-leaning content, instead suggesting that the pull towards extremity exists on the left as well as the right.

However, it is important to note that these results may simply reflect the fact that at each step the bots select the video with the *most* amount of favorable bias. Users who generally select for partisan content may not regularly receive increasingly extreme recommendations. It may be the case that the radicalization trajectories seen in Figure 4 merely result from the bots picking videos that are politically extreme. At best, this analysis conveys the plausibility of these pipelines existing for users who are consistently selecting for extremity.

Secondly, the above analysis seems to refute claims that YouTube's watch-next algorithm inherently favors right-leaning recommendations. This can be seen in the behavior of the Neutral Bot. In picking the highest ranked recommendations, videos which the platform thinks users will engage with the most (Zhao 2019), the Neutral Bot encountered negligible political bias across trials. If RPT's assertions were true, we might expect the Neutral Bot to receive recommendations that were at least somewhat biased towards right-leaning videos. However, the above analysis reveals that the Neutral Bot mostly received politically neutral recommendations, suggesting that RPT is flawed in it's assumption that recommendations inherently skew right on YouTube.

[12]For a full set of individual trajectories, see Appendix D.

4.3 Recommended Category Proportions



(a) Right Bot

(b) Left Bot

(c) Neutral Bot

Figure 5: Recommendations between content categories for each group. Thickness of lines represent the number of recommendations from the left category to the right category.

When looking at the proportion of political categories recommended to each bot, we find more evidence that is critical of radicalization pipeline theory and that supports the notion that radicalization pipelines are largely a user-driven phenomenon. Figure 5 presents a category view of recommendations made to the three bots in flow diagram format. These diagrams depict the proportion of recommendations between categories, where the left side of each diagram represent the categories of videos selected, and the right right side represents the subsequent category of recommendations. The height of the each category represents its proportion of total videos. The width of the flows represent the number of videos recommended from category to category. Here the thicker the flows from left to right, the more recommendations there where from the left group to the right group.

These connections reveal important findings about the relationship between selected videos and subsequent recommendations. Figure 5 shows that the Right and Left Bot's only selected videos that were of a favorable political category or were politically neutral, while the Neutral Bot selected videos from all of the different political categories. This result is expected considering at each step the

Table 1: Proportion of recommendations to different categories per group.

	Groups		
Category	Right bot	Left bot	Neutral Bot
Far Right	0.19	0.02	0.03
Right	0.25	0.04	0.16
Neutral	0.47	0.39	0.55
Left	0.06	0.24	0.19
Far Left	0.02	0.31	0.08

Left and Right Bots always pick videos that are most politically favorable, where the neutral bot picks the first recommended video irrespective of classification. From these picked videos, the majority of recommendations were to videos of the same category with mild variation in recommendations to different categories.

Table 1 presents a similar view of the proportion recommendations made to each bot. The columns on the table represent data associated with a given bot, while the rows on the table represent the political categories of recommendations encountered by the bots. For the Right Bot, 19 percent of recommendations were from the Far Right category, 25 percent from the Right category, and 47 percent from the Neutral category. The results for the Left bot are similar, with the majority of recommendations coming from favorable political categories and the neutral category. The Neutral Bot has a roughly even balance of videos that are politically left-leaning and politically right-leaning, with the majority of recommendations being neutral.

These findings help to muddy the narrative established by radicalization pipeline theory in two important ways. Firstly, the analysis reveals that YouTube tends to recommend similar videos, irrespective of political category. Radicalization pipeline theory posits that YouTube tends to favor recommendations to right-leaning videos. Figure 5 reveals that the political composition of recommendations is largely due to the source of the recommendation. If RPT was reflected in this analysis, there would be more recommendations flowing from the Right category to the Far Right category. However, the opposite seems to be true, Far Right videos flow disproportionately towards more politically moderate content compared to other categories. In general, this analysis suggests that the political nature of recommendations is largely influenced by the political nature of previously viewed videos.

Secondly, as revealed in Table 1, users with no political bias tend to receive equally amounts of recommendations from either side of the political spectrum. If radicalization pipeline theory were reflected in this analysis, we might expect the Neutral Bot to have received a higher proportion of right-leaning recommendations. However, it seems that the only way to achieve this one sided bias, is to actively select for it, as seen in the results for the Right Bot. Therefore, the analysis suggests that the political content of YouTube recommendations is largely driven by user selection bias, rather then an innate political bias present

in the YouTube watch-next algorithm.

5 Discussion

Radicalization pipeline theory is the assertion that YouTube's watch-next algorithm skews its recommendations to the right, from which increasingly more extreme content is recommended. This paper finds mixed evidence for these claims. Firstly, this paper does reveal structural and behavioral trends in how YouTube recommends content that appears to support RPT. If a user is selecting for conservative content, it is possible for them to receive increasingly more extreme content recommendations. However, for YouTube users who are not actively searching for increasingly more conservative content, the effect is largely unfounded. Secondly, this paper shows that there is minimal evidence to suggest that their is an inherent quality of recommendations on YouTube that direct them towards the political right. If a user is not actively selecting for political conservative content, this paper finds that the platform will likely not recommend conservative content. Instead of biasing for the political nature of content, YouTube primarily biases its recommendations based on features of the last video watch, often recommending videos from the same channel.

Another important point is that this paper found no recommendations to extreme right content, even amongst trials of the Right Bot. If radicalization pipelines were a repeatable and widespread phenomena on YouTube, we might expect this research to encounter more recommendations to channels containing truly extreme speech. However, this paper does not find that. While this paper finds supplementary evidence supporting a qualified version of radicalization pipeline theory, the evidence proposes a more nuanced and user driven understanding of recommendations on YouTube, ultimately casting doubt on claims made by radicalization pipeline theory.

Its worth here speculating about alternative frameworks for understanding recommendation systems on social media platforms. Instead of thinking of these platforms as directional systems, or of having a particular content specific agenda, I believe the analysis in this paper constitutes a feedback system. Initially, the platform plays a minimal role in instigating a content specific agenda. However, once a user selects what content to consume, the platform is highly optimized to recommend videos that are subject specific to the selection. Most videos will either be from the same channel or about the same subject of the video consumed. From here, any direction, either towards extremity or moderation, is predominantly driven by the user themselves.

While this understanding of YouTube recommendation does diminish claims made by radicalization pipeline theory, the features and outcomes of these algorithms still deserve further inquiry. While the systems themselves may not have a political agenda, they do act as a mirror, continuously feeding consumers what content the platform thinks they may like. Are these the kind of algorithms that would best help inspire a culture of robust debate, intellectual curiosity, or of cordial disagreement that are so important for robust

democracies? It seems that for the time being, these algorithms are doing the opposite. Instead of challenging users with content from different political spheres, these recommendation algorithms are isolating user into reflective ecosystems of decadent recommendation. This paper encourages continued study on the nature of recommendation algorithms and the potential negative effects they may impose on content consumers.

5.1 Limitations

There are several limitations in this paper that reduce its conclusiveness in determining the validity of radicalization pipeline theory. Firstly, this paper uses agents to traverse YouTube's recommendations that do not have profiles. Because most YouTube users access the site via logged-in profiles, this paper can only make general claims about YouTube's watch-next algorithm in a significantly data deprived context. Secondly, the classifications from the Recfluence data set, the polarization's scores, and the discrete categories are limited in several ways. These classifications are not completely accurate in labeling the political content of YouTube channels. Also, the Recfluence data set is not collectively exhaustive, meaning that many channels encountered by the bots were falsely categorized as neutral. Therefore, the categorisations and findings in this paper contain a non-trivial level of error. Despite these limitations, this paper still serves as a good overview of general trends in YouTube's recommendation algorithm.

5.2 Resources and Acknowledgements

This paper was greatly helped by the work of Mark Ledwich, Anna Zaitsev, and Sam Clarke via their Recfluence project and its associated research papers. The data supplied in this paper was collected from the YouTube Data API, with implementation achieved via a Python client. Visualizations in the project were created using the Plotly Python Open Source Graph Library. All code for this project can be viewed at its associated GitHub repository.

References

Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. "Exposure to Opposing Views on Social Media Can Increase Political Polarization." PNAS. Accessed 2020. https://doi.org/10.1073/pnas.1804840115.

Bakshy, E., S. Messing, and L. A. Adamic. "Exposure to Ideologically Diverse News and Opinion on Facebook." Science 348, no. 6239 (2015): 1130–32. https://doi.org/10.1126/science.aaa1160.

Clark, Sam, and Anna Zaitsev. "Understanding YouTube Communities via Subscription-Based Channel Embeddings." arXiv, October 19, 2020. https://arxiv.org/pdf/2010.09892.pdf.

Faddoul, Marc, Guillaume Chaslot, and Hany Farid. "A Longitudinal Analysis of YouTube's Promotion of Conspiracy Videos." arXiv, March 6, 2020. https://arxiv.org/pdf/2003.03318.pdf.

Flaxman, Seth, Sharad Goel, and Justin M. Rao. "Filter Bubbles, Echo Chambers, and Online News Consumption." Public Opinion Quarterly 80, no. S1 (March 22, 2016): 298–320. https://academic.oup.com/poq/article-abstract/80/S1/298/2223402.

Hosanagar, Kartik, Daniel Fleder, Dokyun Lee, and Andreas Buja. "Will the Global Village Fracture Into Tribes? Recommender Systems and Their Effects on Consumer Fragmentation." Management Science 60, no. 4 (2014): 805–23. https://doi.org/10.1287/mnsc.2013.1808.

Ledwich, Mark, and Anna Zaitsev. "Algorithmic Extremism: Examining YouTube's Rabbit Hole of Radicalization." First Monday, 2020. https://doi.org/10.5210/fm.v25i3.10419.

Lewis, Rebecca. "Alternative Influence." Data & Society. Data & Society Research Institute, September 18, 2018. https://datasociety.net/wp-content/uploads/2018/09/DS_Alternative_Influence.pdf.

Lewis-Kraus, Gideon. "Bad Algorithms Didn't Break Democracy." Wired. Conde Nast, January 15, 2020. https://www.wired.com/story/polarization-politics-misinformation-social-media/.

Maia, Marcelo, Jussara Almeida, and Virgilio Almeida. "Identifying User Behavior in Online Social Networks." Social Nets '08: Proceedings of the 1st Workshop on Social Network Systems, April 2008, 1–6. https://dl.acm.org/doi/10.1145/1435497.1435498.

Munger, Kevin, and Joseph Phillips. "Right-Wing YouTube: A Supply and Demand Perspective." The International Journal of Press/Politics, 2020, 194016122096476. https://doi.org/10.1177/1940161220964767.

O'Callaghan, Derek, Derek Greene, Maura Conway, Joe Carthy, and Pádraig Cunningham. "Down the (White) Rabbit Hole: The Extreme Right and Online Recommender Systems." Social Science Computer Review 33, no. 4 (2014): 459–78. https://doi.org/10.1177/0894439314555329.

Ribeiro, Manoel Horta, Raphael Ottoni, Robert West, Virgílio A. F. Almeida, and Wagner Meira. "Auditing Radicalization Pathways on YouTube." Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 2020. https://doi.org/10.1145/3351095.3372879.

Roose, Kevin. "The Making of a YouTube Radical." The New York Times. The New York Times, June 8, 2019. https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html.

Tufekci, Zeynep. "YouTube, the Great Radicalizer." The New York Times. The New York Times, March 10, 2018. https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html.

Zhao, Zhe, Lichan Hong, Li Wei, Jilin Chen, Aniruddh Nath, Shawn Andrews, Aditee Kumthekar, Maheswaran Sathiamoorthy, Xinyang Yi, and Ed Chi. "Recommending What Video to Watch Next: A Multitask Ranking System." ACM Digital Library. Google, Inc, September 2019. https://dl.acm.org/doi/10.1145/3298689.3346997.

Zollo, Fabiana, and Walter Quattrociocchi. "Misinformation Spreading on Facebook." Computational Social Sciences Complex Spreading Phenomena in Social Systems, 2018, 177–96. https://doi.org/10.1007/978-3-319-77332-2_10.

Appendix A Query String List

Full list of query strings used for this paper: Brett Favre, Kyle Rittenhouse, Laura Ingraham, Election Day, GDP, Electoral College, Edward Snowden, Proud Boys, Kirstie Alley, Trump vs Biden, Glen Greenwald, Jon Ossoff, Prop 22, Jo Jorgensen, Democratic Party, Kamala Harris, Kanye West, Lindsey Graham, Republican Party, Joe Biden, Election, Jack Dorsey, Lil Wayne, Moderna, Gavin Newsom

Appendix A: Strings were hand selected from Google Trends', Trending Searches tool based on their recency and political relevance.

Appendix B Classification Table

Recfluence Classification	Example	Weight
WhiteIdentitarian: Identifies with the superiority of western Civilization.	NPI / RADIX, Stefan Molyneux	7
QAnon: Far-right conspiracy theory alleging the existence of a global sex-trafficking ring.	Roseanne Barr, SpirituallyRAW	7
Provocateur: Aims to inflame political discourse and offend liberal sentiments.	MILO, Fleccas Talks	5
Conspiracy: Regularly promotes a variety of conspiracy theories.	X22Report INFO Wars	5
AntiSJW: Significant focus on criticizing `SocialJustice` with a positive view of the marketplace of ideas.	Eric Weinstein, Tim Pool	3
JudeoChristianConservative: Uses religious allegories to inform conservative political discourse.	Jordan B Peterson	3
ReligiousConservative: Takes a strict religious position on social/cultural issues.	The Daily Wire (Ben Shapiro)	3
PartisanRight: Mainly focused on politics and exclusively critical of Democrats.	Fox News, Turning Point	2

Recfluence Classification	Example	Weight
Libertarian: Skeptical of authority and state power (e.g. war, taxes, welfare).	JRE Clips, ReasonTV	2
Mainstream News: Reporting on newly received or noteworthy information.	TODAY, CBS, BBC London	0
Educational: Channels that have a significant focus on educational material	The Aspen Institute, TED	0
StateFunded: Channels funded by a government.	RT America, C-SPAN	0
Politician: The channel is on behalf of a currently running/in-office Politician.	Kamala Harris, Donald J Trump	0
PartisanLeft: Mainly focused on politics and exclusively critical of Republicans.	MSNBC, CNN, VICE News	-1
LateNightTalkShow: Channels with humorous monologues about the news.	Stephen Colbert, Trevor Noah	-2
Black: Black creators focused on cultural/political issues of their community.	Roland S. Martin, Lisa Cabrera	-4
AntiWhiteness: Channels associated who attribute social issues to white people.	Hebrew Israelite, Roland S. Martin	-4
Revolutionary: Calls for revolutionary restructuring of a classless society.	Proletarian TV	-5
Socialist: Suggests that capitalism is the source of most problems in society.	AfroMarxist, Slavoj Zizek	-5
AntiTheist: Self-identified atheist who are also actively critical of religion.	CosmicSkeptic, Matt Dillahunty	-6

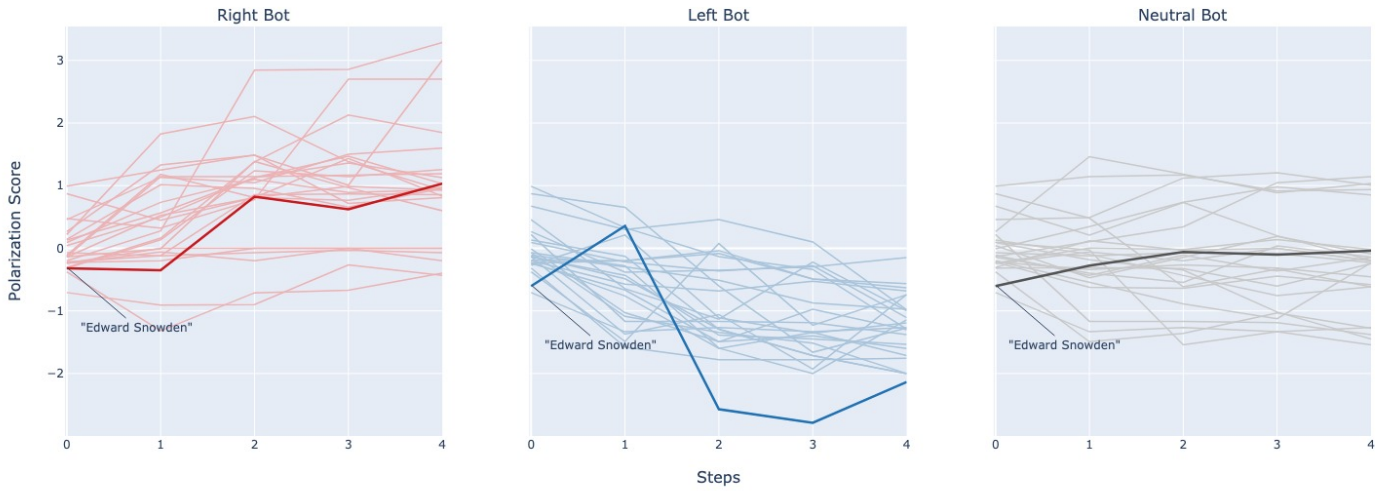
Appendix B: Full list of the political classification and weights used in this paper. Classifications and their definitions are provided by Ledwich and Zaitsev's Recfluence YouTube analysis. Weight's were hand assigned for this paper and act as a loose numeric representation of the left/right-ness of the political classifications, with larger positive weights roughly representing more conservative beliefs and smaller negative weights representing more liberal beliefs.

Appendix C Discrete Categorization Table

Category	P-Score Range	Examples
Far Right	$x \geq 1.5$	Trump War Room, Eric Metaxas, BlazeTV, PragerU, Epoch Times
Right	$1.5 > x \geq 0.5$	Fox News, Fox Business, Donald J Trump, Sky News
Neutral	$0.5 > x \geq -0.5$	NBC News, ABC News, PBS NewsHour CNBC Television, France 24
Left	$-0.5 > x \geq -1.5$	MSNBC, Vox, Vice News, HuffPost, Seth Meyers, The Atlantic
Far Left	$x < -1.5$	Secular Talk, The Intercept, The Young Turks, David Pakman Show

Appendix C: Full list of categories and polarization score (P-Score) ranges used to categorize the political content of videos used for this paper, with examples.

Appendix D Individual Recommendation Trajectories



Appendix C: Polarization score across 5 steps for each of the 25 trials conducted by the three bots. The highlighted trial in each figure represents the trajectory starting from the search query *Edward Snowden*, showing how the three bots generally diverge in polarization score from a common starting point.