Radical bubbles on YouTube? Revisiting algorithmic extremism with personalised recommendations
by Mark Ledwich, Anna Zaitsev, and Anton Laukemper

Abstract
Radicalisation via algorithmic recommendations on social media is an ongoing concern. Our prior study, Ledwich and Zaitsev (2020), investigated the flow of recommendations presented to anonymous control users with no prior watch history. This study extends our work on the behaviour of the YouTube recommendation algorithm by introducing personalised recommendations via personas: bots with content preferences and watch history. We have extended our prior dataset to include several thousand YouTube channels via a machine learning algorithm used to identify and classify channel data. Each persona was first shown content that corresponded with their preference. A set of YouTube content was then shown to each persona. The study reveals that YouTube generates moderate filter bubbles for most personas. However, the filter bubble effect is weak for personas who engaged in niche content, such as Conspiracy and QAnon channels. Surprisingly, all political personas, excluding the mainstream media persona, are recommended less videos from the mainstream media content category than an anonymous viewer with no personalisation. The study also shows that personalization has a larger influence on the home page rather than the videos recommended in the Up Next recommendations feed.

Contents
I. Introduction
II. Recent studies of social media radicalisation
III. Analysing YouTube
IV. Findings
V. Discussion
VI. Limitations and conclusions

I. Introduction
Are social media platforms like YouTube facilitating radicalisation by creating filter bubbles or generating rabbit holes of ever more extreme content? A persistent narrative is that YouTube viewers are recommended increasingly extreme content to increase engagement due to algorithms that favour wild claims, hate speech and outrage over a more balanced diet of content (Munger and Phillips, 2022). Those behind the YouTube recommendation algorithm claim this is not the case: the algorithm does not prefer what the platform describes as ‘borderline’ content (Goodrow, 2021).

Our prior study supports the platform’s messaging (Ledwich and Zaitsev, 2020). That study examined the prevalent media and research narratives in 2018 and 2019 (e.g., Roose, 2019; Tufekci, 2018; Ribeiro, et al., 2020) by investigating the behaviour of the recommendation algorithm during November and December 2019 via data flow analysis and concluded that YouTube directed traffic towards mainstream media channels and away from the fringe (Ledwich and Zaitsev, 2020). Our results aligned with the commentary of YouTube representatives, who have consistently attempted to convince users and the media that their algorithm is not a force of radicalisation and does not show users more and more extreme content recommendations (Thompson, 2020; Goodrow, 2021).

Since the publication of our 2020 study, YouTube has taken even more extensive steps in content policing. The platform has taken an aggressive attitude towards specific topics, actively removing videos and channels that go against the company’s policies regarding the COVID-19 pandemic (specifically, vaccination-related and other perceived
conspiracy theories linked to the pandemic); claims of election fraud (YouTube Team, 2020; Criddle, 2021); and the convoluted QAnon conspiracy theory (Hannah, 2021), which is linked to controversial content about both the pandemic and the 2020 U.S. presidential elections. YouTube representatives have typically approached the discourse around the algorithm with carefully worded and strategic comments that do not directly address the filter bubble phenomenon or radicalisation claims (YouTube Team, 2019a; YouTube Team, 2020). However, claims of radicalisation persist. For example, a recent report titled “YouTube Regrets” from the Mozilla Foundation discusses how the algorithm directs people into self-described rabbit holes of misinformation, violence and hate (McCrosky and Geurink, 2021). Thus, we believe the phenomenon requires more research considering the recommendation algorithm’s personalisation aspects.

Personalisation of recommendations is one of the key features that has made YouTube’s recommendation algorithm a success story and the subject of criticism (Kumar, 2007; Markmann and Grimm, 2021). Our 2020 research was conducted using accounts with no watch history that could potentially affect the algorithm’s behaviour. Due to this lack of personalisation, that study presented a generalised overview of the actions of the recommendation algorithm. In this work, we expand on our original study by introducing a level of personalisation into the data collection. We recorded recommendations for ‘persons’ to simulate users interested in different political topics. The personas represent YouTube viewers interested in channels that discuss the political landscape from various specific perspectives and have collected a viewing history from a subset of related channels.

Using our developed personas, this follow-up study examines whether the YouTube recommendation algorithm creates so-called radical bubbles: ‘recommendations that influence the viewers of radical content to watch more similar content than they would otherwise, making it less likely that alternative views are presented’ (Ledwich and Zaitsev, 2020). Radical bubbles are a subset of the filter bubble phenomenon. Filter bubbles are formed when users are provided only with content that corresponds to their current interests, forming an isolated bubble of information that does not present alternative perspectives (Pariser, 2011). The existence of filter bubbles seems a common feature across social media; outside of YouTube, filter bubble phenomena have been observed on platforms such as Facebook, Twitter and Reddit (Kitchens, et al., 2020).

The claim that YouTube would form filter bubbles — and thus potentially radical bubbles — was partially supported in our 2020 study: the algorithm often suggested channels similar to the ‘seed’ channel [1], thus recommending content that was ‘more of the same’. However, our data did not show much evidence for the other three hypotheses we investigated that might support the theory of radical bubble formation. There was no evidence that the algorithm showed a preference for right-wing content, material linked to right-wing radicalisation was not prominently recommended and the algorithm did not generate a significant pipeline of ever more radical recommendations that would lead users to increasingly irrational conspiracy theories or extreme political ideologies (whether left or right).

In this study, where the impact of personalisation is added, we can focus on the formation of radical bubbles. This research investigates whether the YouTube algorithm can be seen as a driving force in creating these ‘radical bubbles’. In other words, we examine whether the YouTube recommendation algorithm encourages users to view only content from within a radical subset of political content or whether users with personalised accounts are suggested a more heterodox recommendations feed (Kitchens, et al., 2020). We find the filter bubble a helpful metaphor, but it is important to clearly state what we mean by this term (Bruns, 2021). In the case of YouTube, the algorithm might attempt to create a filter bubble by suggesting content only from one genre (or a subset of a genre) and filtering out everything that does not fall into this specific niche. For example, the algorithm might direct users towards political content with a religious perspective or sports content solely focused on mountain climbing. Real users are likely to engage with content outside of a very limited niche and thus break away from the algorithmically generated bubble by watching videos from content creators outside the recommendations feed. However, the concern is that a radicalisation pipeline will keep niche content viewers within the filter bubble perpetuated by the algorithm (Tufekci, 2018; Roose, 2019; Faddoul, et al., 2020). To study whether this claim has merit, the personas are a useful tool, as they simulate users who are fixated on their particular political niche.

This paper is structured as follows. The next section reviews recent studies on social- media driven radicalisation published after our first study. This is followed by descriptions of the personas created for the new study and our new data collection method. Next, we describe the recommendations that the algorithm presented to the personas on their home pages and in the recommendations feed. This is followed by a discussion of whether the findings support the filter bubble claims. The paper concludes by detailing the limitations of the study.

II. Recent studies of social media radicalisation

Since the publication of our original study on the flows of the recommendation algorithm (Ledwich and Zaitsev, 2020), there has been a wave of additional studies that have examined YouTube, news consumption on the platform and the recommendation algorithm, either by accessing actual viewership data or simulating personalised data. For example, a recent Pew poll claims that almost 30 percent of adults in the United States use YouTube as one of their news sources (Stocking, et al., 2020). YouTube has retained its role as the nexus for all other social media platforms, where content is hosted, and other platforms direct their users to long-form video content. Even though political content is a relatively niche category on YouTube, several new studies have investigated the platform’s political side (Rieder, et al., 2020).
Radical bubbles on YouTube? Revisiting algorithmic extremism with personalised recommendations

First, a recent study that takes a similar approach to our study researched personalised YouTube account recommendations (Hussein, et al., 2020). That study looked at both YouTube search and recommendation algorithms from the perspective of misinformation spread. The study collected 56,475 videos containing 2,943 unique videos, which they coded as ‘debunking misinformation’, ‘neutral’ and ‘promoting conspiratorial misinformation’, such as 9/11 and moon landing conspiracy theories. This study confirmed that filter bubble phenomena exist within accounts with previous watch history. The study confirmed the trend that we had observed based on anonymous data (Ledwich and Zaitsev, 2020): the algorithm feeds watchers content that takes a similar approach to previously watched content. However, the data suggest that the algorithm seems to have a bias towards more neutral — if not outright debunking — videos when it comes to antivaccination content.

A report by the Anti-Defamation League confirms that users’ watch habits generate a filter bubble effect (Chen, et al., 2022). That research investigated users’ viewership habits engaging in alternative and extremist or white supremacist content. The report was based on data from 2,000 participants acquired from a pool of YouGov America respondents. The study combined poll data describing personal political attitudes with YouTube viewing data collected via a browser extension between July and October 2020. The report found that 18.9 percent of users appeared to follow the suggestions of the recommendation algorithm.

Interestingly, the report also showed that recommendations shown after a viewer watched a video in the ‘extremist’ category were a mix of videos belonging to the categories’ ‘other’ (56.4 percent), ‘alternative’ (14.3 percent) and ‘extremist’ (29.3 percent), disproving a pure radical bubble effect. The drive towards other content was even stronger if the starting point was ‘other’ or ‘alternative’: no extremist videos were recommended when the starting point was ‘other’, and only 2.3 percent of recommended videos were extremist if the starting point was ‘alternative’. These results suggest that the algorithm is not driving users towards more radical content but rather towards a selection of content that might lead the user away from extremist content. While this report yields interesting insights into users’ behaviour, it also has limitations. The study only focused on what the organisation perceived as alternative and extremist or white supremacist channels, thus omitting a large portion of political YouTube from their filter bubble analysis and suffering from a similarly lopsided analysis as prior studies focusing on only one part of the political spectrum (Ribeiro, et al., 2020). Furthermore, the report states that the study’s recommendations data are based on both anonymous recommendations and recommendations based on user profiles. The study could not disentangle these two types of recommendations (Chen, et al., 2022).

Further evidence of a filter bubble effect can be observed from a study whose authors had access to the Nielsen dataset, which collects actual user information via a browser extension (Hosseinmardi, et al., 2021). The dataset consists of views recorded between January 2016 through December 2019. The study used a subset of 309,813 users with at least one recorded YouTube pageview for 21,385,962 watched video pageviews. This means that the dataset obtained by this research team extends to the time before our study’s snapshot from late 2019. The timelines and impressive dataset support the existence of filter bubbles, but the study does not perceive that the recommendations algorithm is the culprit. The authors state: ‘We find no evidence that engagement with far-right content is caused by YouTube recommendations systematically, nor do we find evidence of a pure radical bubble effect. The drive towards other content was even stronger if the starting point was ‘other’ or ‘alternative’: no extremist videos were recommended when the starting point was ‘other’, and only 2.3 percent of recommended videos were extremist if the starting point was ‘alternative’. These results suggest that the algorithm is not driving users towards more radical content but rather towards a selection of content that might lead the user away from extremist content. While this report yields interesting insights into users’ behaviour, it also has limitations. The study only focused on what the organisation perceived as alternative and extremist or white supremacist channels, thus omitting a large portion of political YouTube from their filter bubble analysis and suffering from a similarly lopsided analysis as prior studies focusing on only one part of the political spectrum (Ribeiro, et al., 2020). Furthermore, the report states that the study’s recommendations data are based on both anonymous recommendations and recommendations based on user profiles. The study could not disentangle these two types of recommendations (Chen, et al., 2022).

Finally, a few studies have continued research into anonymous, impersonal data and the YouTube recommendation algorithm. Another study on conspiracy theory content, which uses a dataset with no watch history, similar to that in our original study, confirms that YouTube has been actively suppressing content related to various conspiracy theories (Faddoul, et al., 2020). Conversely, a study on so-called involuntary celibates or ‘incels’ — a niche community on YouTube that is sometimes seen as a subset of men’s rights activists — claims that the recommendation algorithm occasionally favours incel-related videos. The study also applied impersonal recommendations using the YouTube API, and the time frame is, similarly, the latter months of 2019. The study seems to confirm that if a user starts by watching an incel-related video, the recommendation algorithm will suggest more of the same, thus confirming the filter bubble or, in this case, perhaps even the radical bubble phenomenon. However, the study also states: ‘In the absence of personalisation (e.g., for a non-logged-in user), a user casually browsing YouTube videos is unlikely to end up in a region dominated by Incel-related videos’ [2].

III. Analysing YouTube

Our study focuses on the YouTube recommendation algorithm and the direction of recommendations between different groups of political content. Similar to our prior study, the data are limited to English-language YouTube channels directed primarily to audiences in the United States, United Kingdom and other Anglosphere countries. However, the metrics of these channels (i.e., number of channels, channel and video views and engagement) are analysed on a global scale (see Clark and Zaitsev, 2020). Furthermore, the study is limited to channels that primarily discuss politics. Unlike other recent studies, which take a broader look at the recommendations algorithm (e.g., Markmann and
Radical bubbles on YouTube? Revisiting algorithmic extremism with personalised recommendations

Grimme, 2021) or YouTube overall (Rieder, et al., 2020), we are interested in the realm of political YouTube, which includes channels focused on both politics and polarisation in the traditional sense (Pew Research Center, 2021) as well as the politics of the online culture wars (Nagle, 2017).

A. Channel classification and personas

This study builds on the data and classification schema applied in the 2020 study, in which we designed a channel classification that consisted of 17 ‘soft’ tags and three ‘hard’ tags describing content type. The 17 soft tags were then aggregated into 14 soft tags that were applied to compare content to determine algorithmic advantage (Ledwich and Zaitsev, 2020). The schema allowed each channel to be tagged with up to three different soft tags to capture the content category better. The additional hard tags distinguished YouTube content by independent creators from content generated by establishment corporations or government-funded entities (e.g., BBC, MSNBC, ABC). For example, a news channel like MSNBC can be labelled both Partisan Left and Mainstream News, whereas the channel of political commentator Ben Shapiro is classified as Independent YouTuber, Anti-Woke, Partisan Right and Religious Conservative [4].

We have applied a similar set of classifications for the follow-up study with modifications and updates. First, we introduced a new soft tag category labelled QAnon to distinguish channels focusing on the QAnon conspiracy theory from other conspiracy theory content. The QAnon channels were first selected manually from a set of conspiracy theory channels and later expanded with the help of a machine learning algorithm (discussed in detail in Clark and Zaitsev, 2020). The algorithmically discovered QAnon channels were then manually reviewed by the researchers before they were added to the pool of QAnon channels [5].

Furthermore, Social Justice has been renamed Woke, and the Anti-SJW category has been renamed Anti-Woke to keep up with the current vernacular. Similarly, Men’s Rights Activist (MRA) channels are now labelled Manosphere. We also exclude personalised data tags applied in the prior study, such as Educational and State Funded, because these tags do not sufficiently describe user watch preferences.

Next, to study accounts with personalised watch history, we applied a subset of soft tags and created personas based on the content types identified earlier with the soft tags (see Ledwich and Zaitsev, 2020). We generated 14 personas based on the soft tags and channel type classification. To test the radical bubble hypothesis, the premise for each persona was that the persona is mainly interested in a certain type of political content. For example, a Socialist persona would watch content from a channel that we would have previously tagged Socialist. If the radical bubble theory is correct, the Socialist persona would be shown more socialist content than any other type of content due to their personalised watch history.

All of the personas created for this study and their channel viewing preferences are described in Table 1.

<table>
<thead>
<tr>
<th>Persona type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anti-Theist</td>
<td>Watched content from channels created by self-identified atheists who are also actively critical of religion (e.g., Amazing Atheist, Atheism VS Religion).</td>
</tr>
<tr>
<td>Anti-Woke</td>
<td>Watched content from channels with a significant focus on criticising Woke/woke content (e.g., Steven Crowder, Matt Christiansen, Lotus Eaters).</td>
</tr>
<tr>
<td>Conspiracy</td>
<td>Watched content from channels focused on a variety of conspiracy theories (e.g., X22Report).</td>
</tr>
<tr>
<td>Late-Night Talk Shows</td>
<td>Watched content from late-night talk shows (e.g., John Oliver, Trevor Noah, Bill Maher)</td>
</tr>
<tr>
<td>Libertarian</td>
<td>Watched content from channels that focus on individual liberties and are generally sceptical of authority and state power (e.g., John Stossel, Dave Smith, ReasonTV).</td>
</tr>
<tr>
<td>Mainstream News</td>
<td>Watched content primarily from mainstream news channels (e.g., CNN, MSNBC, Fox News), rather than engaging actively with independent YouTube content creators.</td>
</tr>
</tbody>
</table>
Radical bubbles on YouTube? Revisiting algorithmic extremism with personalised recommendations

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manosphere</td>
<td>Watched content with a focus on advocating for men’s rights (e.g., Honey Badger Radio).</td>
</tr>
<tr>
<td>Partisan Left</td>
<td>Watched content from political channels that are exclusively critical of Republicans (e.g., MSNBC, The Young Turks).</td>
</tr>
<tr>
<td>Partisan Right</td>
<td>Watched content from political channels that are exclusively critical of Democrats and support Trump (2019–2021) (e.g., Rebel News, PragerU).</td>
</tr>
<tr>
<td>QAnon</td>
<td>Watched content from channels with a significant focus on the QAnon conspiracy theory (e.g., X22Report).</td>
</tr>
<tr>
<td>Religious Conservative</td>
<td>Watched content from channels focused on promoting Christianity or Judaism in the context of politics and culture (e.g., Matt Walsh, Ben Shapiro).</td>
</tr>
<tr>
<td>Socialist</td>
<td>Watched content from channels critical of capitalism (e.g., Vauh, Philosophy Tube, Contrapoints).</td>
</tr>
<tr>
<td>Woke</td>
<td>Watched content from channels focused on social issues such as identity politics, intersectionality and political correctness (e.g., AJ+, The Young Turks).</td>
</tr>
<tr>
<td>White Identitarian</td>
<td>Watched content from channels espousing the superiority of ‘whites’ (e.g., American Renaissance — NB. now banned).</td>
</tr>
</tbody>
</table>

Because many channels do not fit into only one category, our classification schema allows multiple tags per channel: up to three different soft tags and a mainstream media versus independent creator distinction. This means that the classification attempts to capture both types of content and political leaning, including overlaps between categories. The overlap between the large channel categories is detailed in Appendix A. The overlap between categories is also an important factor when we discuss the filter bubble phenomenon in Section V.

B. Updated dataset

The updated channel classification tags were applied to an expanded dataset that included thousands more channels than in the 2020 study, including 816 manually reviewed channels. Only channels with over 10,000 subscribers or channels with average video viewership of over 10,000 views were included in the first study due to the constraints of manual review. As noted earlier, we expanded the list of channels by applying machine learning for channel identification and classification, which allowed us to find and include more data than would have been possible using manual channel review (Clark and Zaitsev, 2020). The novel algorithmic method of channel inclusion allowed us to find and classify additional smaller channels and novel channel categories that came to brief prominence after the original study was conducted, such as QAnon channels. All channels with a minimum of 20 subscriptions can be included in our data set review (Clark and Zaitsev, 2020). However, there are fewer channel categories — soft tags — included in this study, as detailed in Section III-A.

Figure 1 shows the average daily views and channel type breakdown for the 10,224 channels included in this research. The new dataset better captures the actual number of political channels because the automated channel detection allowed us to find many more channels, including channels with small subscription counts. Manual channel detection using snowball sampling (Lee, et al., 2006) relies on the recommendation algorithm, and the algorithm does not always suggest smaller channels with less user engagement, especially if the platform deems the content unauthoritative or hateful (YouTube Team, 2020; YouTube Team, 2019b).

The new channel discovery based on user subscription data (Clark and Zaitsev, 2020) uncovered, for example, many more Conspiracy and Anti-Woke channels. The new data show that including smaller channels impacts the daily view figures in some, but not all, categories. It should also be noted that some categories remained very small even after an expanded channel search; for example, the White Identitarian category comprised 131 channels with generally low average daily views (under 100,000). For comparison, the 740 videos in the Mainstream News category have an average of almost 100 million daily views. However, the new dataset shows that some more fringe channel types did attain large audiences during the study. For example, 552 channels propagated the QAnon conspiracy theory in the later months of 2020 with 1.9 million views. A large number of smaller channels can together generate more significant viewership figures than a small number of large channels (e.g., 654 Woke channels with 19.7 million daily views versus 1,313 Anti-Woke channels with 21.6 million views).
Figure 1: Daily views and number of channels.

Note: Larger version of Figure 1 available [here](#).
C. Homepage and recommendations feed

The YouTube recommendation algorithm attempts to influence the platform’s users in two ways. First, when a user has watched a video, the recommendation algorithm suggests what to watch next in Up Next recommendations [7] after a watched video. Our prior study focused on this set of recommendations. However, the recommendation algorithm is also present on users’ homepages. The homepage is YouTube’s landing page: the first page that users typically see when they arrive on the platform. The algorithm populates both the homepage and the feed with recommendations based on the user’s profile, including prior watch history, the user’s interactions with content (subscriptions, likes, comments) and new recommended content that is expected to further engage the user with the content on the platform (Zhao, et al., 2019). Both homepage and the Up Next recommendations are illustrated in Figure 3.

**Figure 2:** The experiment procedure.

Note: Larger version of Figure 2 available here.
Radical bubbles on YouTube? Revisiting algorithmic extremism with personalised recommendations

**Figure 3a:** Homepage.

Note: Larger version of Figure 3a available [here](#).

**Figure 3b:** Up Next Recommendations feed.

Note: Larger version of Figure 3b available [here](#).
D. Intracategory and extracategory recommendations

Throughout the rest of the paper, we use two new terms to clarify the types of recommendations the algorithm suggests to the personas. First, when the algorithm attempts to engage a user with content that falls within the persona’s type, this recommendation is labelled an intracategory recommendation. That is, the algorithm will recommend the Anti-Woke persona videos from the Anti-Woke category, the Libertarian persona videos from Libertarian channels, et cetera.

In the second scenario, the algorithm presents the persona videos outside their immediate watch profile. We call this behaviour of the recommendation algorithm extracategory recommendations. All recommendations, political or nonpolitical that are not within the persona category fall within this category of recommendations.

We analysed the intracategory recommendations to see whether the algorithm created radical bubbles that might further radicalise users. The intracategory recommendations will also reveal whether the theory of more benign filter bubbles holds. Conversely, examining extracategory recommendations might reveal other behaviours of the recommendation algorithm.

Figure 4 illustrates intracategory and extracategory recommendations in the Up Next recommendations. In this example, the Anti-Woke persona has watched a video belonging to an Anti-Woke category. The Up Next recommendations show Anti-Woke content (intracategory recommendations), nonpolitical content, and content from other categories (extracategory recommendations).

Figure 4: Intracategorial and extracategorial recommendations.
Note: Larger version of Figure 4 available here.

E. Process for collection of personalised recommendations
The expanded dataset, which we believe captures a good number of YouTube’s political channels, addresses prior studies’ issues related to potential missing data, i.e., our limits on the channel sizes might have omitted a large number of smaller channels, thus potentially distorting the results. The collection of simulated personalised recommendations addresses the issues related to user anonymity, and the impact of personalisation on the recommendations feed.

Fourteen Google accounts were created and assigned a persona to examine the influence of watch history on recommendations. A persona simulates a user with a watch history within one political video category (e.g., Partisan Left, Anti-Woke). Before starting our experiment, each account started from a clean slate. The experiment set-up was to build different ‘bubbled’ watch histories as input for the recommendation system, which collected the recommended videos for each account while the user was watching videos. An anonymous, neutral user who has not created a profile on the platform and who has no personalised and recorded history was used as a control to compare personalised recommendations with recommendations produced for a viewer with no watch history. This allowed us to identify how page and video recommendations are influenced by personalisation.

Figure 2 depicts the entire daily procedure. During the collection, our political dataset was expanded to include over 7,000 channels, and data were collected from thousands of channels. The number of channels and views per persona are summarised in Table 2.

<table>
<thead>
<tr>
<th>Table 2: Channel and view statistics.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channels either watched or tested in the experiment</td>
</tr>
<tr>
<td>Channels in our political dataset at the time of the experiment</td>
</tr>
<tr>
<td>Views by personas to build history</td>
</tr>
<tr>
<td>Views by personas to record recommendations</td>
</tr>
<tr>
<td>Average views by each persona to build history</td>
</tr>
<tr>
<td>Average views by each persona to record recommendations</td>
</tr>
</tbody>
</table>

The following steps were repeated each day from 7 September 2020 through 1 January 2021:

- To build the persona’s history, videos from the last year were chosen randomly using a view-weighted sampling method. For example, the Manosphere persona watched videos from channels with the Manosphere tag. The random selection was weighted by video popularity (i.e., number of views) better to represent the watching behaviour of a typical user. This set was performed for 50 videos initially and five on each subsequent day.

- Each persona’s YouTube homepage was loaded 20 times in rapid succession once a day, starting at 1 AM GMT and the recommended videos were recorded. For each video ‘viewed’, 20 recommendations were recorded. Each persona watched a view-weighted sample of political videos less than seven days old, and the ‘Up Next’ video recommendations were recorded. The personas were viewed on a cloud server container located on the west coast of the United States to simulate an American viewer.

We used a common sample of videos for all personas each day to isolate the impact of watching history. The common sample consists of 100 videos, chosen randomly (proportional to views), published from political channels in the last seven days. They are ‘viewed’ by each persona. The common sample allows us to see what effect personalisation has on recommendations.

- Videos from the common set were not added to the watch history so as not to influence further recommendations. This was done by disabling the watch history for the personas. After each round of common sample viewing, recommendations were enabled again to build up the watch history for the next personalised set.

To label the recommendations gathered in the experiment, we combined two distinct algorithmic classifiers into one. The first classifier attributes labels to a channel based on the other accounts its subscribers also subscribe to. A detailed description of this classifier and its performance can be found in Clark and Zaitsev (2020). To assign the channels, the second classifier uses text features from the channel description, video snippets, captions and comments. This classifier is derived from the architecture describe in Laukemper (2020).
IV. Findings

We believe that the 14 personas with watch history and their anonymous control counterpart with no history can reveal at least an exaggerated version of the personalised recommendations. This also helps us better understand whether the recommendation algorithm attempts to create radical bubbles and potentially leads users down a path of radicalisation. This section presents what types of recommendations the algorithm provides users on their home pages and Up Next recommendations. We also compare the personalised accounts of the personas with the recommendations for a control account with no watch history.

Users are presented with several different types of ways to consume YouTube recommendations. We present our findings in the following order:

A. Up Next recommendations
   a. Up Next recommendations for personas when watching content from the persona category
   b. Up Next recommendations for personas when watching the common sample
B. Homepage recommendations
   a. Recommendations on the homepage for an anonymous account (no personalisation)
   b. Recommendations on the homepage with personalisation (i.e., recommendations for each persona)
C. Comparison of anonymous and personalised recommendations

A. Feed recommendations for personas

First, we investigate how the recommendation algorithm behaves when a persona account engages with content tagged to match the persona type. This corresponds with how the personas were designed to collect watch history: the personas first watch only content tagged with the corresponding category. The persona simulates a user who exclusively consumes a particular type of political content and is indifferent to other types of political channels present on YouTube. For example, our Socialist persona only ever watches channels that produce content from a socialist perspective. This behaviour builds the persona of a watch history that should, in theory, affect future recommendations.

In order to study the radical bubble phenomenon, we sliced the data based on channel type. We analysed which channels were recommended to each persona after they watched a video belonging to a particular channel category. In other words, we discovered what types of recommendations were presented when, for example, the Woke persona, with a history of watching channels with the tag Woke, watched a video from a Woke channel. If the algorithm strives to create a strong filter bubble, we expect the recommendations to feature Woke channels heavily. Conversely, other personas should have fewer recommendations for Woke channels due to their watch history, which indicates a preference for other types of content. The algorithm should present users with more intracategory recommendations and less extracategory Woke content.

To illustrate the recommendation algorithm’s actions, we look at the recommendations data presented to two polar-opposite persona types: the Woke persona and the Anti-Woke persona. First, we present an overview of the recommendations presented to all personas when the seed video, i.e., the first viewed video of each batch of recommended videos, is tagged with a specific tag. Figure 5 shows which types of recommendations are presented to the different personas when the watched content, or seed content, belongs to the category Woke — in other words, the percentage of recommendations in each category that each persona is offered in their recommendations feed if they watch a video tagged as Woke. For example, as shown in the figure, when the Conspiracy persona watches content from a Woke channel, the recommendation algorithm presents them with 32 percent Woke content, 38.7 percent Partisan Left content and so forth.
Furthermore, we can observe from Figure 5 that when each persona watches a video from a channel that promotes Woke content, the algorithm seems not to recommend Woke content as much as content that belongs to a Woke-adjacent category, mainly channels that can be classified as Partisan Left or Late-Night Talk Show. Nonpolitical content is also featured for roughly one-third of all types of personas. The only persona that seems to receive more recommendations from within the tagged category (i.e., intracategory recommendations) is the Woke persona itself (see highlights from Figure 6). Even for that persona, the recommendations surprisingly prefer Partisan Left channels over Woke channels (49.4 percent versus 43.6 percent).
The data suggests that the *Woke* persona does not receive the majority of recommendations as intracategory recommendations. However, recall that the tagging of content is not exclusive, and many channels are tagged with more than one content category classification tag. The prominence of *Partisan Left* recommendations can be explained by the significant overlap between *Partisan Left* and *Woke* (222 channels are tagged as both *Partisan Left* and *Woke*; see Table VI in the Appendix A). This observation might suggest that, within a larger bubble where both *Partisan Left* and *Woke* content are included, the algorithm does not make as clear a distinction between *Partisan Left* and *Woke* content as a human labeler might.

Moreover, we see that the recommendation algorithm applies similar patterns when we study the persona and tag representing the opposite of *Woke*, namely *Anti-Woke*. Figure 7 shows the breakdown of recommendations shown to each persona when the persona watches a video that belongs to the *Anti-Woke* category. The algorithm does feature *Anti-Woke* channels, but it also directs the users towards *Mainstream News*, *Partisan Right* and nonpolitical content such as music videos, popular culture analysis and gaming channels. *Partisan Right* appears to be the counterpart to *Partisan Left*, a favourite of the algorithm that overlaps with *Mainstream News* (see Table 6 in Appendix A).

For the *Anti-Woke* persona, the recommendations are mixed: 44.2 percent of recommendations are for *Anti-Woke* channels, but *Partisan Right* channels are also featured (20.7 percent). *White Identitarian* channels are not featured in the recommendations. For the *White Identitarian* persona, the algorithm suggests *Anti-Woke* content (39.3 percent), nonpolitical content (37.1 percent), *Partisan Right* content (18.5 percent) and *Mainstream News* content (12.4 percent). Figure 8 lists highlighted percentages of the channels discussed in this example.

Does this reinforce the presumption that the algorithm keeps the *Anti-Woke* persona in the *Anti-Woke* bubble? If we compare the *Anti-Woke* persona with the *Woke* persona, we can see that the *Woke* persona does get more recommendations outside the *Woke* bubble, whereas the *Anti-Woke* persona gets more of the same. However, the data also show...
that the Anti-Woke category has more channels and views than the Woke category (1,313 Anti-Woke channels vs 654 Woke channels and 21.5 million daily views vs 19.7 million daily views; see Figure 1). The algorithm prefers popular content that gets engagement in the form of comments and likes (Goodrow, 2021). Anti-Woke content seems to garner more viewer interactions (Hosseinmardi, et al., 2021), which might explain why the algorithm prefers it over other categories.

Figures from Figure 13 in Appendix C present all the different permutations of the recommendation percentages when various types of channel classifications are selected as the source. Notably, for all seed channels, the algorithm favours a particular set of channels: nonpolitical channels, Mainstream News, Partisan Left and Partisan Right. More niche categories are generally not favoured unless the case is for intracategory recommendations.

The final step in the procedure of our persona experiment consisted of showing the personas videos from a common sample. This dataset further illustrates the level of personalisation in the recommendations feed. The theory is that more intracategory recommendations will appear in more personalised feeds even if the seed video is randomly selected from the common sample set, which includes popular videos across all genres. The channel recommendations to all personas following the common set of videos are shown in Figure 9 [8].
Table 9 shows the percentage of each type of recommendation presented to each persona when viewing the common sample set, which consists of the most popular channels in each category. The Mainstream News persona will be used as an example to illustrate what the chart represents.

In Table 9, the diagonal line shows the intracategory recommendations for each persona. The data show that for some personas, intracategory recommendations are prominent. For example, the persona who enjoys Mainstream News content is recommended similar content 38.3 percent of the time. Similarly, the Late-Night Talk Show persona is more likely to be recommended Late-Night Talk Show content and content that leans politically left. This is consistent with the skew of the Late-Night Talk Show content category towards left-wing politics in the channel classification schema, as shown in Table 6, Appendix A.

However, for many personas, intracategory recommendations are not the majority of recommendations. For example, for the Anti-Theist persona, the recommendation algorithm presents various types of channels from across categories, only slightly favouring Anti-Theist channels over other, more niche categories (11.2 percent). The top recommendations for this persona are extracategory: nonpolitical channels (37.6 percent), Mainstream News (25.3 percent) and Partisan Left (21.2 percent).

Furthermore, the most notable exceptions to intracategory recommendation favouritism are the personas engaging with channels that focus on Conspiracy and QAnon content and the White Identitarian persona. These personas are more likely to be recommended content from either nonpolitical channels or the Mainstream News category, including Partisan Right and Partisan Left sources. For the Conspiracy persona, 28.5 percent of recommendations are Mainstream News and 15.8 percent are Partisan Right. For the QAnon persona, 35.3 percent of recommendations are for Mainstream News and 17.2 percent are for Partisan Right, followed closely by 16.2 percent for Partisan Left. The White Identitarian recommendations are 45 percent nonpolitical, 26.2 percent Mainstream News, 13.7 percent Partisan Left and 12.2 percent Partisan Right (see Figure 10). Only 3.2 percent of the recommendations presented to the White Identitarian persona in their recommendations feed are intracategory.
Radical bubbles on YouTube? Revisiting algorithmic extremism with personalised recommendations

Figure 10: Recommendations for Anti-Theist, Conspiracy, Qanon and White Identitarian personas.

Note: Larger version of Figure 10 available [here](#).

**B. Homepage recommendations**

The recommendations feed an important tool for platforms to retain their audiences by suggesting more interesting content. However, the feed recommendations are not the only page where YouTube’s algorithm attempts to influence users, the algorithm also provides recommendations for users on their home pages. With our new data and personas, we can extend our analysis to homepage recommendations, as well.

**Homepage recommendations for anonymous control account:** When there is no personalisation involved — that is when we analyse the behaviour of a control account with no watch history — the homepage consists overwhelmingly of nonpolitical content. Political content categories are almost nonexistent when the algorithm does not personalise homepage recommendations. The breakdown of the homepage content for the anonymous account is shown in Table 3.

<table>
<thead>
<tr>
<th>Account</th>
<th>Anonymous</th>
<th>Anti-Woke</th>
<th>Anti-Theist</th>
<th>Conspiracy</th>
<th>Late Night Show</th>
<th>Libertarian</th>
<th>Manosphere</th>
<th>Mainstream News</th>
<th>Partisan Left</th>
<th>Partisan Right</th>
<th>Religious Conservative</th>
<th>Woke</th>
<th>Socialist</th>
<th>White Identitarian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonpolitical</td>
<td>83%</td>
<td>3%</td>
<td>0%</td>
<td>0%</td>
<td>37%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>2%</td>
</tr>
<tr>
<td>Anti-Woke</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anti-Theist</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conspiracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Late Night Show</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Libertarian</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manosphere</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mainstream News</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partisan Left</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partisan Right</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Religious Conservative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Woke</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Socialist</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White Identitarian</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 3: Home page recommendations for anonymous account.**

<table>
<thead>
<tr>
<th>Content category</th>
<th>Homepage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonpolitical</td>
<td>83%</td>
</tr>
<tr>
<td>Anti-Woke</td>
<td>3%</td>
</tr>
<tr>
<td>Anti-Theist</td>
<td>0%</td>
</tr>
<tr>
<td>Conspiracy</td>
<td>0%</td>
</tr>
<tr>
<td>Late-Night Talk Show</td>
<td>2.8%</td>
</tr>
<tr>
<td>Libertarian</td>
<td>0%</td>
</tr>
<tr>
<td>Manosphere</td>
<td>0%</td>
</tr>
<tr>
<td>Mainstream News</td>
<td>11%</td>
</tr>
</tbody>
</table>
From Table 3, we can identify that nonpolitical content clearly dominates homepage recommendations for an anonymous user who now has a watch history. The second-largest category, *Mainstream News*, does not even come close with its 11 percent share of recommendations, compared with an 83 percent share of nonpolitical content. We cannot describe the exact reason for this breakdown without insider knowledge. However, we can make an educated guess as to why nonpolitical content is so prevalent: a vast majority of YouTube content is nonpolitical. The channels with the most subscribers are music, gaming, comedy or children’s channels (Boyd, 2021); hence, it makes sense that nonpolitical recommendations for anonymous users consist of popular and viral videos, and content garners positive engagement (Zhao, *et al.*, 2019). The percentage given to mainstream news could be explained by YouTube’s strategic decisions to highlight what the company perceives as authoritative sources and to showcase these sources for all users (YouTube Team, 2019b).

**Homepage recommendations for personas**: When personalisation and watch history are included, the homepage recommendations also become more tailored to each persona, as one might expect. However, we can observe that nonpolitical recommendations are still very prominent and even dominant on the personas’ homepage. For example, for the *Partisan Right* persona, 50 percent of recommendations are nonpolitical, and only 38 percent of recommendations are intracategory — that is, the *Partisan Right* persona is shown recommendations for channels that are also classified as *Partisan Right* (see Table 4). In contrast, the *Partisan Left* persona is recommended 41 percent nonpolitical content and 52 percent *Partisan Left* content on their homepage (Table 4).

| Table 4: Home page recommendations for *Partisan Left* and *Partisan Right* persona. |
|---------------------------------|-----------------|-----------------|
| **Content category**            | **Partisan Right** | **Partisan Left** |
| Nonpolitical                    | 50%             | 41%             |
| Anti-Woke                       | 13%             | 1.3%            |
| Anti-Theist                     | 0.3%            | 3.0%            |
| Conspiracy                      | 6%              | 0.1%            |
| Late-Night Talk Show            | 0.2%            | 4.4%            |
| Libertarian                     | 2%              | 0.1%            |
| Manosphere                      | 0.1%            | 0.0%            |
| Mainstream News                 | 8.8%            | 13%             |
| Partisan Left                   | 0.7%            | 52%             |
| Partisan Right                  | 38%             | 0.5%            |
| QAnon                            | 1%              | 0%              |
| Religious Conservative          | 13%             | 0.1%            |
| Woke                             | 0.2%            | 20%             |
For personas engaging with content in more fringe categories, such as *QAnon* or other conspiracies, the homepage recommendations consist of over 60 percent nonpolitical recommendations, with a slight edge given to *Partisan Right* recommendations as the second-largest category (29 percent for the *QAnon* persona and 17 percent for the more traditional *Conspiracy* persona). All homepage recommendation percentages are shown in Appendix B, Table 7.

### C. Personalised versus anonymous recommendations

The final way we can analyse our new dataset, which presents the recommendations suggested for anonymous users and the personas, is to compare the results of these two types. Comparing the anonymous user with the personas will allow us to analyse the degree of personalisation of both the homepage and recommendations feed to assess whether the homepage leans towards intracategory or extracategory recommendations.

The more personalised the recommendations are, the more the persona is directed towards intracategory recommendations rather than other types of content. Figure 11 presents the percentages of recommended content types for the homepages of personalised versus anonymous accounts, and Figure 12 presents the percentages for accounts’ Up Next recommendations.

![Figure 11: Homepage personalisation.](here)
Figure 12: Recommendation feed personalisation.

Note: Larger version of Figure 12 available here.

From these two figures, we can see a difference between the homepage and recommendations feed. The homepage recommendations are very personalised. We can see that the nonpolitical category is featured much less — on average 40 percent less — on the homepage when the account is personalised, and intracategory recommendations are favoured over other content for all personas. The data on anonymous recommendations confirms one aspect of our prior study, which concluded that the YouTube recommendation algorithm prefers mainstream content sources, both left and right, over other types of content (Ledwich and Zaitsev, 2020). The new analysis of anonymous homepage recommendations shows that the recommendation algorithm prefers more mainstream content over niche channels and nonpolitical over political content. This result is in line with the idea that when there is no personalisation involved, the recommendation algorithm prefers popular content with many views and is uncontroversial (i.e., there is no significant like-to-dislike imbalance) (Zhao, et al., 2019; Goodrow, 2021).

However, the personalised Up Next recommendations has a small advantage in nonpolitical content but a sizeable disadvantage when it comes to mainstream news. This view of the data indicates a flaw in our prior study. In the study, we stated that the algorithm favours mainstream media channels and directs users towards left-wing mainstream media (Ledwich and Zaitsev, 2020). The personalisation of the Up Next recommendations shows that nonpolitical channels are favoured over mainstream news in almost all cases, with a few exceptions. The most significant intracategory advantage in the personalisation is given to late-night talk shows, which are presented to the Late-Night Talk Show persona 27 percent more often than for the anonymous account.

V. Discussion

The personalised data seem to support the idea that the algorithm creates at least weak filter bubbles in the case of several political content categories. This aligns with findings from studies applying real user data (Hosseinmardi, et al., 2021; Chen, et al., 2022). However, even when the findings from our collection of personalised recommendations show some support for the filter bubble phenomenon, they do not support the premise that the YouTube recommendation algorithm specifically generates radical bubbles.

Table 5 presents the intracategory feed recommendations for each persona — that is, the percentage of recommendations within the persona’s category when the starting point is also the persona’s designated category type (see Figures 13 to 26 in Appendix C).
### Table 5: Filter bubbles.

<table>
<thead>
<tr>
<th>Persona</th>
<th>Intracategory recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anti-Woke</td>
<td>44.2%</td>
</tr>
<tr>
<td>Anti-Theist</td>
<td>64.1%</td>
</tr>
<tr>
<td>Conspiracy</td>
<td>13.3%</td>
</tr>
<tr>
<td>Late-Night Talk Show</td>
<td>61.7%</td>
</tr>
<tr>
<td>Libertarian</td>
<td>36.9%</td>
</tr>
<tr>
<td>Manosphere</td>
<td>69.6%</td>
</tr>
<tr>
<td>Mainstream News</td>
<td>52.9%</td>
</tr>
<tr>
<td>Partisan Left</td>
<td>62.8%</td>
</tr>
<tr>
<td>Partisan Right</td>
<td>46.0%</td>
</tr>
<tr>
<td>QAnon</td>
<td>3.6%</td>
</tr>
<tr>
<td>Religious Conservative</td>
<td>39.0%</td>
</tr>
<tr>
<td>Socialist</td>
<td>43.0%</td>
</tr>
<tr>
<td>White Identitarian</td>
<td>17.5%</td>
</tr>
<tr>
<td>Woke</td>
<td>43.6%</td>
</tr>
</tbody>
</table>

This summary shows that the YouTube recommendation algorithm might generate a more substantial bubble effect for personas who enjoy Late-Night Talk Show, Mainstream News, Partisan Left or Manosphere channels. Other types of personas are recommended a more diverse set of channels in their recommendation feeds, regardless of their prior watch history [9].

What is notable is the overlap between categories with a strong bubble effect: only Manosphere is an outlier, whereas the other channel types have significant overlap. These results could suggest that YouTube’s recommendation algorithm creators actively discourage the formation of filter bubbles for personas who watch conspiracy, QAnon or other fringe content but are not worried about filter bubbles forming for other types of viewers. This observation seems to be in line with what we saw in our prior study using anonymous data (Ledwich and Zaitsev, 2020), as well as YouTube’s communications on the topic of the direction in which the company is taking the recommendation algorithm with regard to conspiracy theory content (YouTube Team, 2020).

Conspiracy, QAnon, Anti-Woke and White Identitarian channels have been singled out as potential radicalisers and subjected to much scrutiny in media and academic studies, where they have been seen as a gateway to more fringe content (Hosseinmardi, et al., 2021; Ribeiro, et al., 2020; Chen, et al., 2022). To establish the existence of a strong radical bubble phenomenon or alt-right pipeline, wherein more moderate viewers are directed towards extreme content (Ribeiro, et al., 2020; Faddoul, et al., 2020), the recommendations should heavily skew towards Conspiracy, Anti-Woke or even White Identitarian channels. However, neither personalised nor anonymous recommendations favour these channels: QAnon, Conspiracy and White Identitarian channels are not favoured even when the persona is intentionally engaging with the content rather than tumbling down a rabbit hole.

Furthermore, it is very unlikely that personas not already interested in Anti-Woke channels will encounter Anti-Woke content in their recommendations feed. We can see that the Woke persona, the opposite persona to the Anti-Woke persona, does encounter Anti-Woke content 25.1 percent of the time when the seed channel is also Anti-Woke, as shown in Figure 7, but this percentage drops below 10 percent when any other type of channel is selected as the seed channel (see Appendix C; Manosphere is again a slight exception). We can also see from Table 5 that the intracategory recommendation percentage is 43 percent for the Woke persona. The high intracategory percentage means that types of channels other than Woke and Partisan Left are quite unlikely to appear in the recommendations feed (see also Figure 12, Appendix C), generating a bubble of left-wing and mainstream news content around the Woke persona. This bubble is broken only if the persona specifically seeks out content that would not be generally recommended — that is, the persona chooses to find a seed channel that contradicts their political preferences. By favouring intracategory recommendations, the platform lessens or even eliminates algorithmic exposure of personas to many types of content. We can interpret these results as the algorithm’s enforcement of a filter bubble by more
Radical bubbles on YouTube? Revisiting algorithmic extremism with personalised recommendations

About the authors

 VI. Limitations and conclusions

A. Limitations

The most significant limitation of the study is that it relies on simulated users, not actual user data. As already stated, these personas represent very extreme users who are not interested in any content besides their intracategory channels. However, the persona accounts have watch history and ‘preferences’ that resemble those of a real human user. Thus, the algorithm’s behaviour should be the same for both the bot and the human. Furthermore, suppose we are interested in looking at a personality more likely to be entangled in a web of radical content. In that case, a monomaniacal bot might be a decent stand-in.

Moreover, the machine learning algorithm that we applied to expand our dataset is not flawless and might have misclassified content or applied incorrect tags for channels. For example, the algorithm assigned political labels to a cooking channel hosted by a wife of a libertarian comedian, likely due to a large, shared audience. Similarly, a few history channels were manually removed from the set due to their neutral political stance. However, data spot checks and the performance metrics of the algorithm give us confidence in the accuracy of the classification schema (Clark and Zaitsev, 2020).

Finally, one potential limitation might be the length of the viewing history. Our personas viewed content for only three months and thus may not be comparable to many real-life users, who may have years of recorded watch history. However, studies using longitudinal data have not highlighted significant differences in the watch habits of YouTube users over the years (Hosseinmardi, et al., 2021). Furthermore, according to the company’s announcements on what content the platform will prioritise, YouTube has tweaked its recommendation algorithm several times over the last few years (YouTube, 2019: YouTube Team, 2019b; YouTube Team, 2020). Older watch history might not be relevant to the current version of the recommendation algorithm.

B. Conclusions

Our study on personalised recommendations has shown what types of political content the recommendation algorithm will suggest on the personas’ feeds and homepages. Our study is a follow-up to prior work on anonymous recommendation algorithms (Faddoul, et al., 2020; Ledwich and Zaitsev, 2020), and it has extended studies investigating filter bubble studies using bot accounts (e.g., Hussein, et al., 2020). When personalisation is simulated via a set of personas who engage in different types of political content ranging from left to right and mainstream to niche, we can observe that YouTube will generate bubbles for some types of political content but discourages bubbles for others. Furthermore, the study shows that the algorithm also favours popular nonpolitical content, to an extent, especially on the homepage. YouTube’s recommendation algorithm would work poorly if the recommended content had no resemblance to the interests of its users and suggestions were purely randomised. Thus, the existence of some filter bubbles should not be surprising. The filter bubbles that reinforce the political content YouTube has deemed acceptable are a feature, not a bug. Our study shows the inbuilt bias of the recommendation algorithm: mainstream political content is preferred over content by independent creators as it is deemed more authoritative. This trend was already observable in our prior study (Ledwich and Zaitsev, 2020), but in the broader scope of this study, we have also shown significant differences in filter bubble strength.

YouTube has publicly stated that they priorities channels it perceives as belonging to authoritative content creators, including mainstream news outlets (YouTube Team, 2019b). Based on our data, we see no reason to doubt that this is precisely the direction the platform has taken.

About the authors
Radical bubbles on YouTube? Revisiting algorithmic extremism with personalised recommendations

Mark Ledwich is an independent researcher and a software engineer in Brisbane, Australia. E-mail: mark [at] ledwich [dot] com [dot] au

Anna Zaitsev is a postdoctoral scholar and lecturer in the School of Information at the University of California, Berkeley. E-mail: anna [dot] zaitsev [at] berkeley [dot] edu

Anton Laukemper is a data scientist in Berlin, Germany. E-mail: anton [at] laukemper [dot] it

Notes

1. The seed channel is the channel that determines the trajectory of the recommendations feed. When one watches a video from an Anti-Woke category, the recommended videos are determined by a combination of factors, including the seed video and watch history.


4. Shapiro is considered independent due to his affiliation with an alternative conservative news company, Daily Wire, as opposed to a large, established news organisation.

5. After we concluded our data collection, QAnon content, which was already somewhat limited on the platform, has been limited yet further.

6. YouTube allows users to subscribe to their favourite channels. These channels are accessible to users in their Subscription feed, an alternative to the home page.

7. We will use the term Up Next recommendations, adopted from YouTube, to differentiate the recommendations between the home page and the recommendations presented after a watched video.

8. The anonymous persona at the top of the figure has no watch history and operates similarly to the anonymous account applied in our prior study (Ledwich and Zaitsev, 2020).

9. The White Identitarian category has too few channels to make an accurate comparison.

References


Radical bubbles on YouTube? Revisiting algorithmic extremism with personalised recommendations  


Appendix A: Channel category overlap
Due to our classification schema, which allows each channel to be tagged with up to three different content description tags, there is significant overlap between several channel tags. For example, there are 506 channels with the tag Partisan Left and 623 channels with the tag Woke. Of these channels, 222 have been tagged with both labels, creating an overlap of 40 percent, as shown in Table 6. Similarly, there is significant overlap between the Late-Night Talk Show and Partisan Left categories.

Mainstream news channels are also more numerous on the left than the right side of the political aisle. Fox News is the largest right-wing mainstream news channel (66 million daily views) and is an order of magnitude larger than the second-largest right-wing mainstream news channel (the U.K.-focused The Sun, with 8.1 million daily views).

Conversely, Partisan Right channels also overlap with several categories, including Anti-Woke, Conspiracy, QAnon and Religious Conservative. However, it should be noted that there are many smaller Partisan Right channels with small viewership (2,545 channels), whereas there are fewer channels in the Partisan Left category (506). Figure 1 shows number of views for each category.

<table>
<thead>
<tr>
<th>Tag</th>
<th>Anti-Woke</th>
<th>Anti-Theist</th>
<th>Conspiracy</th>
<th>Late-Night Talk Show</th>
<th>Libertarian</th>
<th>Manosphere</th>
<th>Mainstream News</th>
<th>Partisan Left</th>
<th>Partisan Right</th>
<th>QAnon</th>
<th>Religious Conservative</th>
<th>Socialist</th>
<th>White Identitarian</th>
<th>Woke</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Channels</td>
<td>1,118</td>
<td>246</td>
<td>2,538</td>
<td>18</td>
<td>278</td>
<td>116</td>
<td>735</td>
<td>506</td>
<td>2,543</td>
<td>430</td>
<td>94</td>
<td>148</td>
<td>94</td>
<td>632</td>
</tr>
<tr>
<td>Anti-Woke</td>
<td>26</td>
<td>22</td>
<td>27</td>
<td>40</td>
<td>499</td>
<td>2</td>
<td>94</td>
<td>1</td>
<td>420</td>
<td>7</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anti-Theist</td>
<td>26</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conspiracy</td>
<td>22</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1,082</td>
<td>425</td>
<td>460</td>
<td>1</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Late-Night Talk Show</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Libertarian</td>
<td>1</td>
<td>5</td>
<td></td>
<td></td>
<td>48</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manosphere</td>
<td>40</td>
<td>2</td>
<td></td>
<td></td>
<td>1</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mainstream News</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>79</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>Partisan Left</td>
<td>4</td>
<td>1</td>
<td>17</td>
<td></td>
<td>79</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>29</td>
<td>222</td>
<td></td>
</tr>
<tr>
<td>Partisan Right</td>
<td>499</td>
<td>1,082</td>
<td>48</td>
<td>1</td>
<td>30</td>
<td>420</td>
<td>334</td>
<td>53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QAnon</td>
<td>2</td>
<td>425</td>
<td>1</td>
<td></td>
<td>420</td>
<td>7</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Religious Conservative</td>
<td>94</td>
<td>406</td>
<td>1</td>
<td>3</td>
<td>334</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Socialist</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White Identitarian</td>
<td>45</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td>53</td>
<td>1</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Woke</td>
<td>11</td>
<td>18</td>
<td>5</td>
<td></td>
<td>16</td>
<td>222</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>91</td>
</tr>
</tbody>
</table>

Appendix B: Homepage recommendations for personas
Table 7 shows the percentages of which types of channels are showcased on as homepage recommendations for each persona, including nonpolitical recommendations. The diagonal line in the table shows the intracategory recommendation percentages from which we can observe the filter bubble effect on the homepage.

<table>
<thead>
<tr>
<th>Tag</th>
<th>Nonpolitical</th>
<th>Anti-Woke</th>
<th>Anti-Theist</th>
<th>Conspiracy</th>
<th>Late-Night Talk Show</th>
<th>Libertarian</th>
<th>Manosphere</th>
<th>Mainstream News</th>
<th>Partisan Left</th>
<th>Partisan Right</th>
<th>QAnon</th>
<th>Religious Conservative</th>
<th>Socialist</th>
<th>White Identitarian</th>
<th>Woke</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anonymous</td>
<td>81%</td>
<td>2.9</td>
<td>0</td>
<td>0</td>
<td>2.7</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>5.2</td>
<td>3.6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.1</td>
</tr>
<tr>
<td>Anti-Woke</td>
<td>40%</td>
<td>47%</td>
<td>1.1</td>
<td>0.7</td>
<td>0.5</td>
<td>3.1</td>
<td>2</td>
<td>4.4</td>
<td>1</td>
<td>22</td>
<td>0</td>
<td>8.8</td>
<td>0.3</td>
<td>0.2</td>
<td>0.5</td>
</tr>
<tr>
<td>Anti-Theist</td>
<td>37%</td>
<td>12%</td>
<td>42%</td>
<td>0.1</td>
<td>2.7</td>
<td>0.2</td>
<td>0.1</td>
<td>2.8</td>
<td>9.6</td>
<td>4.8</td>
<td>0</td>
<td>0.7</td>
<td>1.3</td>
<td>0</td>
<td>7.1</td>
</tr>
<tr>
<td>Conspiracy</td>
<td>60%</td>
<td>4.3</td>
<td>0</td>
<td>17%</td>
<td>0.2</td>
<td>0.6</td>
<td>0.1</td>
<td>8.6</td>
<td>0.7</td>
<td>17</td>
<td>2.4</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0.4</td>
</tr>
<tr>
<td>Late-Night Talk Show</td>
<td>50%</td>
<td>0.9</td>
<td>0</td>
<td>0</td>
<td>35%</td>
<td>0</td>
<td>0</td>
<td>6.8</td>
<td>44</td>
<td>0.5</td>
<td>0</td>
<td>0.3</td>
<td>0.2</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Libertarian</td>
<td>47%</td>
<td>16</td>
<td>0.4</td>
<td>0.8</td>
<td>0.4</td>
<td>32%</td>
<td>0.3</td>
<td>6.9</td>
<td>1.4</td>
<td>14</td>
<td>0.2</td>
<td>3.3</td>
<td>0.1</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>Manosphere</td>
<td>41%</td>
<td>20</td>
<td>0.1</td>
<td>0.5</td>
<td>0.2</td>
<td>0.3</td>
<td>42%</td>
<td>5</td>
<td>0.7</td>
<td>2.7</td>
<td>0</td>
<td>2.8</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
</tr>
<tr>
<td>Mainstream News</td>
<td>53%</td>
<td>0.8</td>
<td>0</td>
<td>0.1</td>
<td>1</td>
<td>0.4</td>
<td>0</td>
<td>40%</td>
<td>7.4</td>
<td>5.9</td>
<td>0</td>
<td>0.5</td>
<td>0.1</td>
<td>0</td>
<td>2.3</td>
</tr>
<tr>
<td>Partisan Left</td>
<td>39%</td>
<td>1.3</td>
<td>2.9</td>
<td>0.1</td>
<td>4.2</td>
<td>0.1</td>
<td>0</td>
<td>12</td>
<td>50%</td>
<td>0.5</td>
<td>0</td>
<td>0.1</td>
<td>8.5</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>Partisan Right</td>
<td>49%</td>
<td>13</td>
<td>0.3</td>
<td>5.9</td>
<td>0.2</td>
<td>1.9</td>
<td>0.1</td>
<td>8.7</td>
<td>0.7</td>
<td>37%</td>
<td>1</td>
<td>13</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>QAnon</td>
<td>60%</td>
<td>5</td>
<td>0</td>
<td>16</td>
<td>0.3</td>
<td>0.7</td>
<td>0</td>
<td>9.9</td>
<td>1.1</td>
<td>28</td>
<td>13%</td>
<td>6.1</td>
<td>0</td>
<td>0.1</td>
<td>0.3</td>
</tr>
<tr>
<td>Religious Conservative</td>
<td>38%</td>
<td>15</td>
<td>0.3</td>
<td>13</td>
<td>0.2</td>
<td>0.3</td>
<td>0.3</td>
<td>4.7</td>
<td>0.6</td>
<td>37</td>
<td>0.1</td>
<td>42%</td>
<td>0</td>
<td>0</td>
<td>0.4</td>
</tr>
<tr>
<td>Socialist</td>
<td>41%</td>
<td>2.1</td>
<td>0.7</td>
<td>0</td>
<td>1.5</td>
<td>0.3</td>
<td>0</td>
<td>3</td>
<td>20</td>
<td>0.3</td>
<td>0</td>
<td>0.1</td>
<td>41%</td>
<td>0</td>
<td>33</td>
</tr>
<tr>
<td>White Identitarian</td>
<td>62%</td>
<td>19</td>
<td>0.1</td>
<td>1</td>
<td>0.1</td>
<td>1</td>
<td>2.8</td>
<td>4.7</td>
<td>1.2</td>
<td>15</td>
<td>0</td>
<td>2.9</td>
<td>0.4</td>
<td>15%</td>
<td>0.6</td>
</tr>
<tr>
<td>Woke</td>
<td>41%</td>
<td>1.5</td>
<td>1.8</td>
<td>0.4</td>
<td>3</td>
<td>0.1</td>
<td>0.2</td>
<td>7.1</td>
<td>31</td>
<td>0.7</td>
<td>0</td>
<td>0.6</td>
<td>12</td>
<td>0</td>
<td>45%</td>
</tr>
</tbody>
</table>

Appendix C: Recommendation feed recommendations by each category

The figures in this Appendix illustrate all the feed recommendations by each channel category selected as the seed, i.e., first video for further recommendations. For example, Figure 13 shows which recommendations are shown to each persona when the prior video watched is classified as Partisan Left whereas in Figure 14, the watched video is has been classified as Partisan Right and the recommendations are populated according to the watch history where the Partisan Right video has now been added.
Radical bubbles on YouTube? Revisiting algorithmic extremism with personalised recommendations

<table>
<thead>
<tr>
<th>Account</th>
<th>In Tag</th>
<th>Partisan Left</th>
<th>Partisan Right</th>
<th>Religious</th>
<th>Conservative</th>
<th>Leftist</th>
<th>Rightist</th>
<th>Misanthropic</th>
<th>Antifa</th>
<th>Other</th>
<th>Cringeworthy</th>
<th>Parody</th>
<th>Offensive</th>
<th>White Identitarian</th>
<th>Anarchist</th>
<th>Neo-Nazis</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anonymous</td>
<td>0.6%</td>
<td>0.5%</td>
<td>0.9%</td>
<td>2.3%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>2.4%</td>
<td>0.1%</td>
<td>1.9%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>22.2%</td>
<td>19.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Anti-mede</td>
<td>0.7%</td>
<td>0.6%</td>
<td>0.1%</td>
<td>1.7%</td>
<td>0.7%</td>
<td>0.3%</td>
<td>0.2%</td>
<td>25.2%</td>
<td>43.1%</td>
<td>6.6%</td>
<td>1.3%</td>
<td>13.7%</td>
<td>0.6%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>22.6%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Anti-christ</td>
<td>1.4%</td>
<td>0.8%</td>
<td>0.0%</td>
<td>22.3%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>27.8%</td>
<td>56.4%</td>
<td>1.4%</td>
<td>0.1%</td>
<td>20.1%</td>
<td>0.9%</td>
<td>0.0%</td>
<td>24.5%</td>
<td>0.0%</td>
<td>25.0%</td>
</tr>
<tr>
<td>Conspiracy</td>
<td>1.6%</td>
<td>0.5%</td>
<td>1.9%</td>
<td>20.8%</td>
<td>0.3%</td>
<td>0.1%</td>
<td>29.5%</td>
<td>47.4%</td>
<td>7.0%</td>
<td>0.1%</td>
<td>2.8%</td>
<td>16.9%</td>
<td>0.5%</td>
<td>0.0%</td>
<td>20.3%</td>
<td>0.0%</td>
<td>28.6%</td>
</tr>
<tr>
<td>Late night show</td>
<td>0.2%</td>
<td>0.3%</td>
<td>0.0%</td>
<td>43.5%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>15.2%</td>
<td>60.8%</td>
<td>0.5%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>26.6%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Libertarian</td>
<td>2.9%</td>
<td>0.5%</td>
<td>0.3%</td>
<td>19.2%</td>
<td>5.3%</td>
<td>0.1%</td>
<td>27.9%</td>
<td>43.1%</td>
<td>6.6%</td>
<td>0.3%</td>
<td>1.9%</td>
<td>14.7%</td>
<td>0.5%</td>
<td>0.0%</td>
<td>26.8%</td>
<td>0.0%</td>
<td>26.8%</td>
</tr>
<tr>
<td>Misanthropes</td>
<td>4.9%</td>
<td>0.3%</td>
<td>0.0%</td>
<td>19.6%</td>
<td>0.1%</td>
<td>5.6%</td>
<td>83.7%</td>
<td>48.8%</td>
<td>4.8%</td>
<td>0.7%</td>
<td>15.6%</td>
<td>0.5%</td>
<td>0.0%</td>
<td>26.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>26.0%</td>
</tr>
<tr>
<td>Mainstream Media</td>
<td>0.9%</td>
<td>0.6%</td>
<td>0.1%</td>
<td>16.4%</td>
<td>0.3%</td>
<td>0.0%</td>
<td>33.3%</td>
<td>42.5%</td>
<td>4.4%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>14.7%</td>
<td>0.6%</td>
<td>0.0%</td>
<td>81.0%</td>
<td>0.0%</td>
<td>81.0%</td>
</tr>
<tr>
<td>Partisan Left</td>
<td>0.6%</td>
<td>1.7%</td>
<td>0.9%</td>
<td>20.2%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>27.6%</td>
<td>62.0%</td>
<td>2.0%</td>
<td>0.1%</td>
<td>19.8%</td>
<td>2.3%</td>
<td>0.0%</td>
<td>24.0%</td>
<td>0.0%</td>
<td>24.0%</td>
</tr>
<tr>
<td>Partisan Right</td>
<td>2.9%</td>
<td>0.5%</td>
<td>0.6%</td>
<td>19.4%</td>
<td>0.5%</td>
<td>0.2%</td>
<td>83.5%</td>
<td>48.8%</td>
<td>10.9%</td>
<td>0.2%</td>
<td>2.6%</td>
<td>16.4%</td>
<td>0.4%</td>
<td>0.0%</td>
<td>27.0%</td>
<td>0.0%</td>
<td>27.0%</td>
</tr>
<tr>
<td>Catholic</td>
<td>1.6%</td>
<td>0.8%</td>
<td>0.5%</td>
<td>19.4%</td>
<td>0.3%</td>
<td>0.2%</td>
<td>33.1%</td>
<td>53.0%</td>
<td>6.4%</td>
<td>0.8%</td>
<td>0.8%</td>
<td>17.8%</td>
<td>0.3%</td>
<td>0.0%</td>
<td>24.4%</td>
<td>0.0%</td>
<td>24.4%</td>
</tr>
<tr>
<td>Religious Conservative</td>
<td>2.9%</td>
<td>0.6%</td>
<td>2.0%</td>
<td>19.4%</td>
<td>0.2%</td>
<td>0.2%</td>
<td>83.5%</td>
<td>47.6%</td>
<td>13.5%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>16.8%</td>
<td>0.4%</td>
<td>0.0%</td>
<td>28.5%</td>
<td>0.0%</td>
<td>28.5%</td>
</tr>
<tr>
<td>White</td>
<td>0.7%</td>
<td>0.7%</td>
<td>0.0%</td>
<td>19.1%</td>
<td>0.1%</td>
<td>0.0%</td>
<td>35.1%</td>
<td>58.5%</td>
<td>1.7%</td>
<td>0.1%</td>
<td>0.1%</td>
<td>20.9%</td>
<td>0.4%</td>
<td>0.0%</td>
<td>28.5%</td>
<td>0.0%</td>
<td>28.5%</td>
</tr>
<tr>
<td>Socialists</td>
<td>0.8%</td>
<td>0.7%</td>
<td>0.0%</td>
<td>15.9%</td>
<td>0.2%</td>
<td>0.0%</td>
<td>38.7%</td>
<td>57.8%</td>
<td>2.6%</td>
<td>0.1%</td>
<td>0.1%</td>
<td>23.9%</td>
<td>1.3%</td>
<td>0.0%</td>
<td>21.4%</td>
<td>0.0%</td>
<td>21.4%</td>
</tr>
<tr>
<td>White Identitarian</td>
<td>5.5%</td>
<td>0.4%</td>
<td>0.0%</td>
<td>27.7%</td>
<td>0.4%</td>
<td>0.7%</td>
<td>28.7%</td>
<td>48.9%</td>
<td>6.7%</td>
<td>0.7%</td>
<td>0.0%</td>
<td>15.3%</td>
<td>0.7%</td>
<td>0.2%</td>
<td>30.8%</td>
<td>0.0%</td>
<td>30.8%</td>
</tr>
</tbody>
</table>

Figure 13: From Partisan Left.
Note: Larger version of Figure 13 available [here](#).

Figure 14: From Partisan Right.
Note: Larger version of Figure 14 available [here](#).
Radical bubbles on YouTube? Revisiting algorithmic extremism with personalised recommendations

Figure 15: From Mainstream Media.
Note: Larger version of Figure 15 available here.

Figure 16: From Late Night Talk Shows.
Radical bubbles on YouTube? Revisiting algorithmic extremism with personalised recommendations

### Figure 17: From *Woke*.

Note: Larger version of Figure 17 available [here](#).
**Figure 18:** From *Anti-Woke*.
Note: Larger version of Figure 18 available [here](#).

<table>
<thead>
<tr>
<th>Account</th>
<th>Activism</th>
<th>Left Beat</th>
<th>Conspiracy</th>
<th>Left Night Show</th>
<th>Liberalism</th>
<th>roundup</th>
<th>Conservative</th>
<th>Right Wing</th>
<th>Qanon</th>
<th>Religious Extremism</th>
<th>White Nationalism</th>
<th>Neo-Radical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anonymous</td>
<td>11.6%</td>
<td>0.2%</td>
<td>0.4%</td>
<td>6.0%</td>
<td>54.1%</td>
<td>0.1%</td>
<td>13.4%</td>
<td>9.0%</td>
<td>0.0%</td>
<td>0.8%</td>
<td>1.0%</td>
<td>0.1%</td>
</tr>
<tr>
<td>Anti-Woke</td>
<td>19.4%</td>
<td>0.3%</td>
<td>0.2%</td>
<td>6.0%</td>
<td>22.0%</td>
<td>0.9%</td>
<td>10.0%</td>
<td>2.4%</td>
<td>12.8%</td>
<td>3.5%</td>
<td>0.7%</td>
<td>0.2%</td>
</tr>
<tr>
<td>Anti-fieldName</td>
<td>0.7%</td>
<td>15.4%</td>
<td>0.1%</td>
<td>2.3%</td>
<td>21.0%</td>
<td>0.1%</td>
<td>9.7%</td>
<td>8.3%</td>
<td>8.0%</td>
<td>1.3%</td>
<td>2.5%</td>
<td>0.7%</td>
</tr>
<tr>
<td>Conspiracy</td>
<td>9.2%</td>
<td>0.3%</td>
<td>3.3%</td>
<td>6.4%</td>
<td>27.3%</td>
<td>0.1%</td>
<td>13.8%</td>
<td>2.2%</td>
<td>15.6%</td>
<td>9.5%</td>
<td>5.6%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Late Night Show</td>
<td>6.6%</td>
<td>31.5%</td>
<td>6.2%</td>
<td>0.0%</td>
<td>20.0%</td>
<td>0.7%</td>
<td>31.1%</td>
<td>3.0%</td>
<td>3.0%</td>
<td>0.3%</td>
<td>11.4%</td>
<td>39.5%</td>
</tr>
<tr>
<td>Libertarian</td>
<td>14.7%</td>
<td>0.1%</td>
<td>0.5%</td>
<td>6.7%</td>
<td>36.9%</td>
<td>0.1%</td>
<td>9.7%</td>
<td>2.6%</td>
<td>13.4%</td>
<td>9.0%</td>
<td>3.0%</td>
<td>0.7%</td>
</tr>
<tr>
<td>Massosphere</td>
<td>12.1%</td>
<td>0.1%</td>
<td>0.2%</td>
<td>0.3%</td>
<td>25.3%</td>
<td>19.4%</td>
<td>12.3%</td>
<td>1.4%</td>
<td>7.6%</td>
<td>0.9%</td>
<td>1.5%</td>
<td>0.6%</td>
</tr>
<tr>
<td>Mainstream News</td>
<td>7.4%</td>
<td>0.1%</td>
<td>0.1%</td>
<td>1.1%</td>
<td>24.5%</td>
<td>0.2%</td>
<td>26.5%</td>
<td>6.1%</td>
<td>11.3%</td>
<td>1.9%</td>
<td>1.0%</td>
<td>0.2%</td>
</tr>
<tr>
<td>Parties left</td>
<td>5.9%</td>
<td>1.6%</td>
<td>0.0%</td>
<td>2.0%</td>
<td>20.0%</td>
<td>0.1%</td>
<td>15.9%</td>
<td>26.0%</td>
<td>7.1%</td>
<td>0.9%</td>
<td>10.4%</td>
<td>2.5%</td>
</tr>
<tr>
<td>Parties right</td>
<td>12.3%</td>
<td>0.1%</td>
<td>1.5%</td>
<td>0.4%</td>
<td>25.4%</td>
<td>0.3%</td>
<td>13.0%</td>
<td>2.5%</td>
<td>23.6%</td>
<td>9.1%</td>
<td>5.9%</td>
<td>1.3%</td>
</tr>
<tr>
<td>QAnon</td>
<td>0.2%</td>
<td>0.0%</td>
<td>1.6%</td>
<td>0.5%</td>
<td>28.2%</td>
<td>0.2%</td>
<td>17.9%</td>
<td>3.2%</td>
<td>15.3%</td>
<td>1.3%</td>
<td>1.7%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Religious Conserv</td>
<td>10.9%</td>
<td>4.3%</td>
<td>0.4%</td>
<td>23.4%</td>
<td>4.4%</td>
<td>12.9%</td>
<td>2.3%</td>
<td>19.1%</td>
<td>0.1%</td>
<td>11.6%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>White</td>
<td>6.9%</td>
<td>0.6%</td>
<td>0.2%</td>
<td>4.1%</td>
<td>18.7%</td>
<td>0.0%</td>
<td>11.4%</td>
<td>2.0%</td>
<td>6.2%</td>
<td>0.6%</td>
<td>19.9%</td>
<td>5.7%</td>
</tr>
<tr>
<td>Socialist</td>
<td>9.9%</td>
<td>0.4%</td>
<td>0.1%</td>
<td>2.3%</td>
<td>22.7%</td>
<td>0.1%</td>
<td>9.6%</td>
<td>16.2%</td>
<td>6.0%</td>
<td>0.9%</td>
<td>15.3%</td>
<td>16.3%</td>
</tr>
<tr>
<td>White Nationalism</td>
<td>13.7%</td>
<td>0.2%</td>
<td>0.3%</td>
<td>24.1%</td>
<td>0.5%</td>
<td>12.5%</td>
<td>2.6%</td>
<td>11.6%</td>
<td>2.4%</td>
<td>0.0%</td>
<td>0.1%</td>
<td>6.0%</td>
</tr>
</tbody>
</table>

**Figure 19:** From *Libertarian*.
Note: Larger version of Figure 19 available [here](#).
Radical bubbles on YouTube? Revisiting algorithmic extremism with personalised recommendations

Figure 20: From Manosphere.
Note: Larger version of Figure 20 available here.

Figure 21: From Conspiracy.
Note: Larger version of Figure 21 available here.
**Figure 22:** From *QAnon.*

Note: Larger version of Figure 22 available [here](#).

**Figure 23:** From *Socialist.*
Radical bubbles on YouTube? Revisiting algorithmic extremism with personalised recommendations

Note: Larger version of Figure 23 available [here](#).

**Figure 24:** From *White Identitarian*.

Note: Larger version of Figure 24 available [here](#).
Figure 25: From Religious Conservative.
Note: Larger version of Figure 25 available [here](#).

<table>
<thead>
<tr>
<th>Account</th>
<th>Anti-Umla</th>
<th>atchet</th>
<th>Ciao</th>
<th>Catnight</th>
<th>Show</th>
<th>Telenova</th>
<th>Melodrama</th>
<th>NineStar</th>
<th>Te</th>
<th>-Tag</th>
<th>Petronela</th>
<th>OfCom</th>
<th>Religious Conserva</th>
<th>Woke</th>
<th>Sociali</th>
<th>White Nationalist</th>
<th>Neo-Nazi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anonymous</td>
<td>12.7%</td>
<td>6.3%</td>
<td>12.0%</td>
<td>0.7%</td>
<td>0.7%</td>
<td>0.1%</td>
<td>12.4%</td>
<td>4.2%</td>
<td>41.6%</td>
<td>0.4%</td>
<td>42.0%</td>
<td>0.2%</td>
<td>0.2%</td>
<td>0.2%</td>
<td>20.6%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anti-Xenia</td>
<td>22.2%</td>
<td>5.6%</td>
<td>3.7%</td>
<td>0.7%</td>
<td>0.7%</td>
<td>0.1%</td>
<td>13.3%</td>
<td>3.1%</td>
<td>31.6%</td>
<td>0.5%</td>
<td>24.2%</td>
<td>1.0%</td>
<td>0.2%</td>
<td>6.0%</td>
<td>26.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anti-Theist</td>
<td>12.6%</td>
<td>17.9%</td>
<td>3.6%</td>
<td>2.8%</td>
<td>0.6%</td>
<td>0.2%</td>
<td>11.6%</td>
<td>6.8%</td>
<td>23.6%</td>
<td>0.5%</td>
<td>21.7%</td>
<td>2.0%</td>
<td>0.3%</td>
<td>23.4%</td>
<td>23.4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conspiracy</td>
<td>10.4%</td>
<td>0.2%</td>
<td>9.6%</td>
<td>0.7%</td>
<td>0.1%</td>
<td>0.1%</td>
<td>14.2%</td>
<td>2.3%</td>
<td>36.9%</td>
<td>0.6%</td>
<td>32.2%</td>
<td>0.0%</td>
<td>0.1%</td>
<td>20.7%</td>
<td>20.7%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Late Night Show</td>
<td>6.9%</td>
<td>1.0%</td>
<td>2.9%</td>
<td>0.6%</td>
<td>0.6%</td>
<td>9.0%</td>
<td>23.9%</td>
<td>12.7%</td>
<td>0.8%</td>
<td>5.7%</td>
<td>14.7%</td>
<td>1.8%</td>
<td>26.0%</td>
<td></td>
<td>26.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Libertaskan</td>
<td>16.6%</td>
<td>0.2%</td>
<td>8.7%</td>
<td>0.6%</td>
<td>10.7%</td>
<td>0.2%</td>
<td>13.9%</td>
<td>2.8%</td>
<td>23.9%</td>
<td>0.4%</td>
<td>24.3%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>26.0%</td>
<td>26.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manosphere</td>
<td>15.0%</td>
<td>0.1%</td>
<td>3.5%</td>
<td>0.3%</td>
<td>0.0%</td>
<td>11.6%</td>
<td>16.9%</td>
<td>1.5%</td>
<td>29.3%</td>
<td>0.3%</td>
<td>23.8%</td>
<td>0.6%</td>
<td>0.1%</td>
<td>32.6%</td>
<td>32.6%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mainstream News</td>
<td>9.2%</td>
<td>0.1%</td>
<td>3.4%</td>
<td>1.8%</td>
<td>0.6%</td>
<td>0.1%</td>
<td>20.6%</td>
<td>6.7%</td>
<td>23.9%</td>
<td>0.5%</td>
<td>22.7%</td>
<td>1.0%</td>
<td>0.1%</td>
<td>35.5%</td>
<td>35.5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partisan Left</td>
<td>1.5%</td>
<td>0.1%</td>
<td>3.2%</td>
<td>4.0%</td>
<td>0.6%</td>
<td>0.2%</td>
<td>21.2%</td>
<td>27.0%</td>
<td>24.4%</td>
<td>0.2%</td>
<td>18.7%</td>
<td>8.6%</td>
<td>1.7%</td>
<td>20.0%</td>
<td>20.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partisan Right</td>
<td>14.2%</td>
<td>0.1%</td>
<td>5.1%</td>
<td>0.3%</td>
<td>1.5%</td>
<td>0.2%</td>
<td>15.2%</td>
<td>2.6%</td>
<td>42.9%</td>
<td>0.2%</td>
<td>30.0%</td>
<td>0.8%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>34.4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OfCom</td>
<td>10.7%</td>
<td>0.1%</td>
<td>7.6%</td>
<td>0.5%</td>
<td>1.5%</td>
<td>0.2%</td>
<td>17.4%</td>
<td>2.7%</td>
<td>37.0%</td>
<td>0.0%</td>
<td>30.0%</td>
<td>0.8%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>39.3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Religious Conserva</td>
<td>13.5%</td>
<td>0.1%</td>
<td>8.8%</td>
<td>0.5%</td>
<td>0.7%</td>
<td>0.4%</td>
<td>16.4%</td>
<td>2.8%</td>
<td>41.9%</td>
<td>0.2%</td>
<td>36.0%</td>
<td>0.9%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>33.9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Woke</td>
<td>8.7%</td>
<td>1.0%</td>
<td>8.6%</td>
<td>2.0%</td>
<td>0.2%</td>
<td>1.5%</td>
<td>14.1%</td>
<td>36.6%</td>
<td>12.2%</td>
<td>0.1%</td>
<td>23.8%</td>
<td>12.6%</td>
<td>11.8%</td>
<td>0.1%</td>
<td>33.8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Socialist</td>
<td>10.3%</td>
<td>0.5%</td>
<td>3.6%</td>
<td>2.6%</td>
<td>1.8%</td>
<td>0.1%</td>
<td>15.7%</td>
<td>23.0%</td>
<td>25.2%</td>
<td>0.1%</td>
<td>21.8%</td>
<td>12.6%</td>
<td>11.8%</td>
<td>0.1%</td>
<td>33.8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White Nationalist</td>
<td>17.0%</td>
<td>0.2%</td>
<td>3.5%</td>
<td>0.4%</td>
<td>1.5%</td>
<td>0.5%</td>
<td>17.1%</td>
<td>2.2%</td>
<td>29.1%</td>
<td>0.1%</td>
<td>25.0%</td>
<td>0.9%</td>
<td>0.1%</td>
<td>5.0%</td>
<td>40.2%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 26: From Anti-Theist.
Note: Larger version of Figure 26 available [here](#).

Editorial history

Received 5 May 2022; revised 28 November 2022; accepted 4 December 2022.

This paper is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](#).

Radical bubbles on YouTube? Revisiting algorithmic extremism with personalised recommendations
by Mark Ledwich, Anna Zaitsev, and Anton Laukemper.
First Monday, volume 27, number 12 (December 2022).
doi: [https://dx.doi.org/10.5210/fm.v27i12.12552](https://dx.doi.org/10.5210/fm.v27i12.12552)