Evaluating the scale, growth, and origins of right-wing echo chambers on YouTube

Homa Hosseinmardi,[1] Amir Ghasemian,[2, 3] Aaron Clauset,[4, 5, 6]
David M. Rothschild,[7] Markus Mobius,[8] and Duncan J. Watts[1]

[1] *University of Pennsylvania, Philadelphia, PA 19104*
[2] *Department of Statistics, Harvard University, Cambridge, MA 02138*
[3] *Department of Statistical Science, Fox School of Business, Temple University, Philadelphia, PA 19122*
[4] *Department of Computer Science, University of Colorado, Boulder, CO 80309*
[5] *BioFrontiers Institute, University of Colorado, Boulder, CO 80303*
[6] *Santa Fe Institute, Santa Fe, NM 87501*
[7] *Microsoft Research New York, 300 Lafayette St, New York, NY 10012*
[8] *Microsoft Research New England, 1 Memorial Dr., Cambridge, MA 02142*

Although it is understudied relative to other social media platforms, YouTube is arguably the largest and most engaging online media consumption platform in the world. Recently, YouTube's outsize influence has sparked concerns that its recommendation algorithm systematically directs users to radical right-wing content. Here we investigate these concerns with large scale longitudinal data of individuals' browsing behavior spanning January 2016 through December 2019. Consistent with previous work, we find that political news content accounts for a relatively small fraction (11%) of consumption on YouTube, and is dominated by mainstream and largely centrist sources. However, we also find evidence for a small but growing "echo chamber" of far-right content consumption. Users in this community show higher engagement and greater "stickiness" than users who consume any other category of content. Moreover, YouTube accounts for an increasing fraction of these users' overall online news consumption. Finally, while the size, intensity, and growth of this echo chamber present real concerns, we find no evidence that they are caused by YouTube recommendations. Rather, consumption of radical content on YouTube appears to reflect broader patterns of news consumption across the web. Our results emphasize the importance of measuring consumption directly rather than inferring it from recommendations.

The internet has fundamentally altered the production and consumption of political news content. On the production side, it has dramatically reduced the barriers to entry for would-be publishers of news, leading to a proliferation of small and often unreliable sources of information. On the consumption side, search and recommendation engines make even marginal actors easily discoverable, allowing them to build large, highly engaged audiences at a low cost. And the sheer scale of online platforms—Facebook and YouTube each have more than 2 billion active users per month–tends to amplify the effect of any flaws in their algorithmic design or content moderation policies. As political polarization rises and trust in traditional sources of authority declines, concerns have naturally arisen regarding the presence and impact of hyper-partisan or conspiratorial content on social media platforms that, while not clearly in violation of platform rules, may nonetheless have corrosive or radicalizing effects on users.

Although research on online misinformation has to date focused mostly on Facebook and Twitter [1–9], roughly 23 million Americans rely on YouTube as source of news [10, 11], a number that is comparable with the corresponding Twitter audience [10, 12], and is growing both in size and engagement. Moreover, recent work [13] has identified a large number of YouTube channels, mostly operated by individual people or small organizations, that produce large amounts of extreme right-wing and–less frequently–left-wing content. The large followings of some of these channels have prompted the hypothesis that YouTube itself is driving traffic to them via its recommendation engine, and is therefore effectively radicalizing its users [14–18]. For example, it has been reported that starting from factual videos about the flu vaccine, the recommender system can drive users toward anti-vaccination conspiracy videos [14, 15].

However, systematic evidence for these claims are elusive. On a platform with almost 2 billion users [19], it is possible to find examples of almost any type of behavior, hence anecdotes do not on their own indicate systematic problems. Meanwhile, the few studies [20–24] that have examined this question have reached conflicting conclusions, with some finding evidence of radicalization [20, 21] and others finding the opposite [22, 23]. These disagreements may arise from methodological differences that make the results hard to compare—for example, Ref. [23] examines potential biases in the recommender by simulating logged-out users, whereas Ref. [20] reconstructs user histories from scraped comments. The disagreement may also reflect limitations in the available data, which is intrinsically ill-suited to measuring either individual or aggregate consumption of different types of content over extended time intervals, such as user sessions or "lifetimes." Absent such data for a large, representative sample of real YouTube users, it is difficult to evaluate how much radical content is in fact being consumed (vs. produced), how it is changing over time, and how it is being encountered (from recommendations vs. other entry points).

Here we introduce a unique data set comprising a large

(N = 309, 813) representative sample of the US population, and their online browsing histories, both on and off the YouTube platform, spanning four years from Jan 2016 to Dec 2019. To summarize, we present four main findings. (i) Total consumption of any news -related content on YouTube accounts for only 11% of overall consumption, similar to previous estimates of news consumption across both web and TV [25], and is dominated by mainstream or moderate sources. (ii) Nonetheless, the fraction of YouTube users strongly engaged with far-right channels has increased over the last four years, where consumers of far-right channels tend to show a more-extreme engagement pattern on YouTube, compared to individuals whose majority of consumption is from channels with other political orientations. (iii) The pathways by which users reach far-right videos are diverse and only a fraction can plausibly be attributed to platform recommendations. (iv) Within sessions of consecutive video viewership, we see no trend toward more extreme content, indicating that consumption of this content is determined more by user preferences than by recommendation. We conclude that while the increasing prevalence and evident appeal of radical content on YouTube may be a real source of concern, the recent focus on the recommendation engine is overly narrow. Rather, YouTube should be viewed as part of a larger information ecosystem in which extreme and misleading content is widely available, easily discovered, and both increasingly and actively sought out [22].

METHODS AND MATERIALS

Our data come mainly from Nielsen's nationally representative desktop web panel, spanning January 2016 through December 2019 (Appendix, section B), which records individuals' visits to specific URLs. We use the subset of $N = 309,813$ panelists who have at least one recorded YouTube pageview. Parsing the recorded URLs, we found a total of 21, 385, 962 watched-video pageviews (Table I). The duration of in-focus visit of each video in total minutes quantifies the user's attention [26]. Duration or time spent is credited to an in-focus page and when a user returns to a tab with previously loaded content, duration is credited immediately. Each YouTube video has a unique identifier embedded in the recorded URL, yielding 9, 863, 0964 unique video IDs (Appendix, section B).

TABLE I. Data statistics.

Number of unique users	309, 813
Number of watched-video pageviews	21, 385, 962
Number of unique video IDs	9, 863, 0964
Number of unique channel IDs	2, 293, 760
Number of sessions	8, 620, 394

To post a video on YouTube, a user must create a channel with a unique name and ID. For all unique video IDs, we used the YouTube API to retrieve the corresponding channel ID, as well as metadata such as the video's category, title, and duration. We then labeled each video based on the political leaning of its channel.

a. Video labeling. Previous studies devoted considerable effort to labeling YouTube channels and videos based on their political leaning. Such labels range from high level (left, center, right) to more granular ideologies (Alt-right, Alt-lite). To construct a unified label set, we used the labeled channels and videos provided by Refs. [20], [22], [23], and [24], and assigned all videos published by each labeled channel to one of five major categories across the political spectrum: far left (fL), left (L), center (C), right (R) and far right (fR) (Appendix, Table IV). For example, YouTube videos belonging to channels with ideological labels such as "Revolutionary Socialist" are considered radical left content whereas ideologies such as "Alt-right" exemplify radical right content. Overall, we compiled a list of 997 highly subscribed channels. This list is not a comprehensive set of all political news channels, but does comprise around 37% of YouTube's total news consumption (Appendix, Fig. 9). Details on the assignment of channels along with the full list of channels we consider can be found in Tables XI-XIII and section I, Appendix.

b. Label imputation. Using the YouTube API, 20.1% of the video IDs had no return from the API (we refer to these as unavailable videos), a problem that previous studies also faced [27]. The YouTube API does not provide any information about the reason for this return value; however, for some of these videos, the YouTube website itself shows a "sub-reason" for the unavailability. We crawled a uniformly random set of 368, 754 videos and extracted these sub-reasons from the source HTML. Stated reasons varied from video privacy settings to video deletion or account termination as a result of violation of YouTube policies (Appendix, Table III). For channels such as "Infowars," which was terminated for violating YouTube's Community Guidelines [28], none of their previously uploaded videos are available through the YouTube API. Therefore, it is important to estimate the fraction of these unavailable videos that will receive each of the political leaning labels, and whether that distribution will affect our findings.

To resolve this ambiguity, we treated it as a missing value problem and imputed missing video labels via a supervised learning approach. To obtain accurate labels for training such model, we first searched for the overlap between our set of unavailable videos and data sets from previous studies [20, 22–24, 27], which had collected the metadata of many now unavailable videos at a time when they still existed on the platform (Appendix, Table VI). This approach yielded channel IDs for 69, 840 of our unavailable videos. We then trained a series of classifiers, which we used to infer the labels of the remaining unavailable videos. For the features of the supervised model, we extracted information surrounding each unavailable video, such as the web partisan score of news domains viewed before and after it, along with the YouTube and political categories of all videos watched in close proximity within the same session. We also exploited a set of user-level features, such as the individual's monthly consumption from different video and web categories during their lifetime in our data. Details on the feature engineering and model selection can be found in Appendix, section D.

For each political channel category we trained a binary random forest classifier over 96 predictors, which yielded an AUC (Area Under Curve) of 0.97 for left, 0.95 for far right, and 0.93 for far left, center, and right on the holdout set. To assign labels, we consider two different thresholds, one with high precision and one with high recall (Appendix, Table IX). For all presented results in the main text, we use a high precision imputation model, which means fewer videos from the positive class are retrieved with higher confidence. Therefore, the results presented here reflect the lower bound of the number of videos in each political category. As a robustness check, we repeat some of our experiments using the upper bound (high recall; see Appendix, section H).

c. Constructing sessions. Previous studies have analysed web browsing dynamics by breaking a sequence of pageviews into sub-sequences called sessions [29]. In this work, we define a YouTube session as a set of near-consecutive YouTube pageviews by a user. Within a session, a gap less than δ minutes is allowed between a YouTube non-video URL and the next YouTube URL, or a gap less than γ minutes is allowed between a YouTube video URL and the next YouTube URL, otherwise the session breaks and a new session will start with the next YouTube URL. External pageviews (all non-YouTube URLs) are allowed within these gaps. For brevity, throughout the rest of the paper, we refer to a YouTube session as simply a session. In the main text we will present the results for sessions created by $\delta = 10$ and $\gamma = 60$ minutes. To check the robustness of our findings to these choices, we repeated the session-level experiments with different values of δ and γ (see Appendix, section H).

d. User clustering. An individual is considered a news consumer if, over the course of one month, they spend a minimum of one minute watch-time on any of the political channels in our labeled set. Each month we characterized every individual who consumed news on YouTube in terms of their normalized monthly viewership vector ν_i^m whose j-th entry, ν_{ij}^m corresponds to the fraction of viewership of user i from channel category j ($j \in$ {fL, L, C, R, fR}). We then used K-means clustering to assign each individual to one of K communities of similar YouTube news diets, with K in the range from 2 to 12. Using the silhouette method [30], we found the optimal number of communities to be $K = 5$ (Appendix, Fig. 12). For each of these five clusters we then identified its centroid obtained by averaging the normalized monthly viewership vectors of all cluster members (see Appendix, section F for details). Finally, we labeled each community as $\psi(t)$ ($\psi(t) \in \{\mathrm{fL}, \mathrm{L}, \mathrm{C}, \mathrm{R}, \mathrm{fR}\}$) according to the predominant content category of its centroid. As a robustness check, we performed similar analysis with nonzero news consumption and at least five minutes news consumption per month as more relaxed and more strict definitions of "news consumers" (see Appendix, Fig. 18).

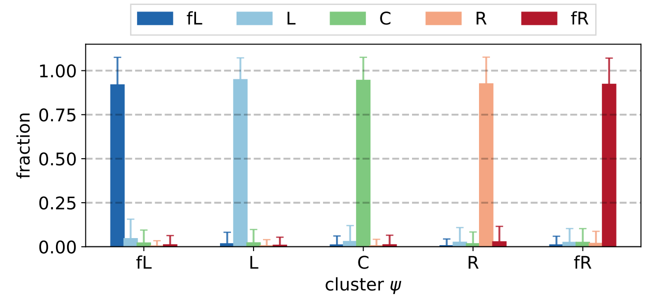

FIG. 1. Archetypes of news consumption behavior on YouTube.

RESULTS

First, we show that these five communities correspond very closely with our identified content categories (Fig. 1): the centroid or "archetype member" of each community devotes roughly 90% of their attention to just one content category with the remaining 10% distributed roughly evenly among the other categories. This result has two important implications for our analysis. First, it reveals that YouTube users are highly segregated into definable communities with preferences that are homogeneous within communities and distinct between them. Because these communities strongly resemble so-called "echo chambers" [31], in which individuals are exposed to ideologically consistent information, we will use these terms interchangeably throughout the remainder of the paper. Second, it demonstrates that the content categories we have identified are able to capture a large fraction of observed behavior in a parsimonious way. We present our main results in terms of these communities.

e. Community engagement. To check for any overall trends toward radical content, we examine changes in total consumption associated with each of the content communities over the four-year period of our data, quantified both in terms of population size (Fig. 2A) and total time spent watching (Fig. 2B) (see Appendix, Fig. 13 for two other related metrics: page view counts and session counts). Consistent with previous work [25], we find that overall news consumption accounted for only a small fraction of overall YouTube activity, whether in terms of population of active news consumers (5.84%) or watch time (3.47%) averaged over the four year period. (These numbers are lower than the 11% which included

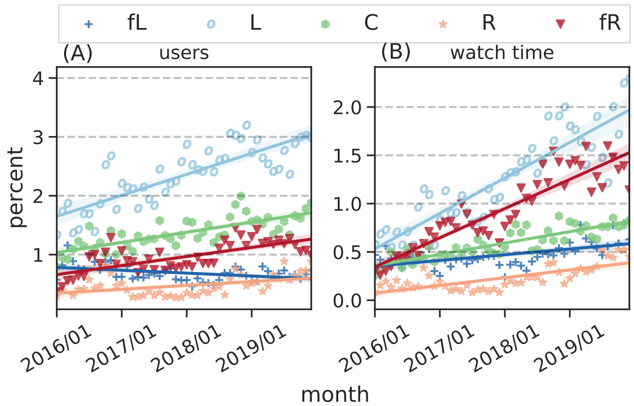

FIG. 2. Breakdown of percent of (A) users, (B) consumption falling into the five political channel categories, per month, from far left to far right, January 2016 to December 2019. Panel (A) is the percent of users falling into each community, and panel (B) presents the percentage of viewership duration from each channel category.

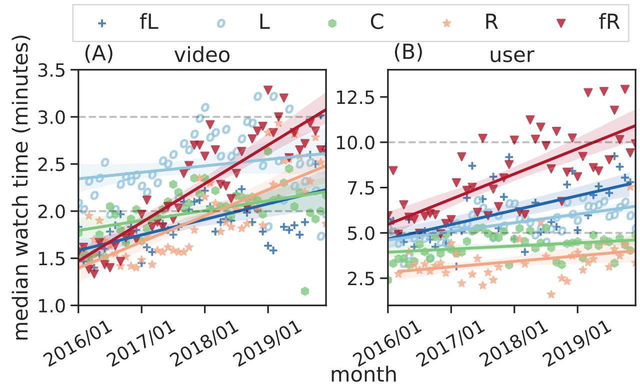

FIG. 3. (A) Median monthly video consumption (minutes) across different channel categories, and (B) median user consumption (minutes) within each community.

all channels labeled by YouTube as "news" whereas these analyses were restricted to the 997 channels for which we had political labels.) Moreover, the largest community of news consumers–again both in terms of population size (2.40%) and watch time (1.25%)–was the "left" mainstream community. However, Fig. 2 also shows that the far-right community was the third largest by population (0.96%) and the second largest by consumption time (0.94%). Moreover, consumption of far-right news on YouTube grew rapidly over the period of the study, increasing three-fold from 0.5% of total consumption in Jan 2016 to 1.5% in Dec 2019, outstripping the rate of overall growth in share of news consumption (8% of total consumption in Jan 2016 to 11% in Dec 2019, Fig. S2). Finally, while Ref. [22] observed a drop in demand–defined as views per video–for "Alt-right" and "Alt-lite" channels beginning in 2017, we find that this drop was a temporary departure from a longer and more sustained upward trend (Fig. 2).

These results indicate that the phenomenon of right-wing radicalization on YouTube, while small in relative terms, nonetheless affects more than 1.85 million Americans monthly, averaged over the four year period, and growing rapidly. (We also note that this result is robust to other choices of consumption metric (e.g. page views or session counts), threshold for inclusion in the "news consuming population," and imputation model (Appendix, Figs. 13, 18, 19 and 20)). To better understand these dynamics we now investigate individual-level behavior.

f. Individual engagement. In contrast with the aggregate (community-level) results, Fig. 3 shows both absolute levels of, and changes over time in, consumption measured at the individual level of videos (Fig. 3A) and people (Fig. 3B), respectively. Fig. 3 presents further evidence of the appeal of far-right content. First, the median per-video watch time of far-right videos has roughly doubled, over the observation period, starting out well below centrist and left-leaning videos but eventually overtaking all of them. Second, the median per-month watch time for individual members of the far-right community also roughly doubled over the same time period, from roughly five minutes per month to more than 10 minutes per month, far exceeding increases in other communities. In other words, while the far-right YouTube community never became the largest news-consuming group, they were the most engaged at the individual level.

Further examining individual behavior, Fig. 4 shows the average probability $P(\psi_j(t)|\psi_i(t-1))$ of an individual member of community ψ_i in month $t-1$ moving to community ψ_j in month t. As indicated by darker shades along the diagonal, the dominant behavior is for community members to remain in their communities from month to month, suggesting that all communities exhibit "stickiness." Moreover, when individuals do switch communities they are more likely to move from the right (R and fR) side of the political spectrum to the left (L and fL) than the reverse, while individuals in the center are more likely to move left than right. Consistent with the engagement results above, however, users in the far-right community exhibit significantly greater stickiness than any other community ($P(\psi(t) = \mathrm{fR}|\psi(t-1) = \mathrm{fR}) = 0.31$, vs. the next largest $P(\psi(t) = \mathrm{L}|\psi(t-1) = \mathrm{L}) = 0.23$). While between-community movement on the left is generally from more extreme to less extreme (i.e. $P(\psi(t) = \mathrm{L}|\psi(t-1) = \mathrm{fL}) = 0.14 > P(\psi(t) = \mathrm{fL}|\psi(t-1) = \mathrm{L}) = 0.04$) on the right side of the partisan spectrum, the most common transition is from less to extreme to more extreme ($P(\psi(t) = \mathrm{fR}|\psi(t-1) = \mathrm{R}) = 0.10 > P(\psi(t) = \mathrm{R}|\psi(t-1) = \mathrm{fR}) = 0.04$).

Finally, the gray horizontal bars in Fig. 4 show the steady states of the transition matrix, representing the fraction of YouTube users who would hypothetically reside in each community in the limit $t \to \infty$ (assuming

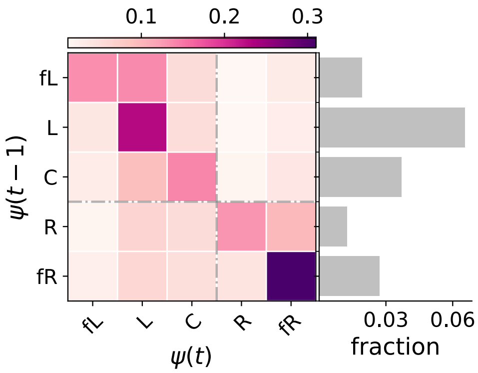

FIG. 4. A heatmap showing the probability that an individual from cluster $\psi(t-1)$ at month $t-1$ will move to cluster $\psi(t)$ at month t. Each month, users may not fall into any of these clusters, if they are not among "news consumers" in that particular month. The bar plot presents the steady states of the transition matrix, as the fraction of YouTube users falling into each cluster, when $t \to \infty$.

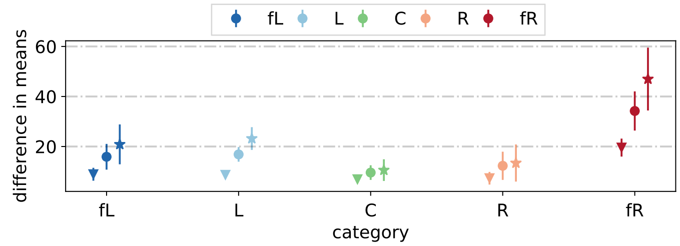

FIG. 5. Difference in means of daily consumption change, in the event of bursty (treatment) consumption from a specific political category. Individuals are assigned either to the bursty consumption group in the event of watching at least M_v^k videos from category k ($k \in \{\mathrm{fL}, \mathrm{L}, \mathrm{C}, \mathrm{R}, \mathrm{fR}\}$) within a session, or to the control group, if none of their sessions has more than one video from the same category with at least M_v^k videos in their lifetime. We run three experiments with different values of M_v^k, where ▼ : $M_v^k = 2$, ● : $M_v^k = 3$, ★ : $M_v^k = 4$. Markers show the difference in means, and the vertical lines present the 95% confidence intervals. The exposure can be driven by user, recommendation, or external sources. Difference in change of daily consumption, after bursty consumption, is almost twice as large for far right compared to the other four political categories, when controlled for other covariates.

constant transition probabilities). In this simulated equilibrium, the left and center communities are projected to remain the largest groups with 6.53% and 3.67% of the population respectively. However, the far-right community is projected to more than double its size by the end of 2019, growing to 2.7% of YouTube users.

g. Concentrated exposure predicts future consumption. Exposure to concentrated "bursts" of radical content may correlate with future consumption more strongly than equivalent exposure to other categories of content [32]. To check for this possibility, we define a "burst" of exposure as consumption of at least $M_v^{(k)}$ videos of category k ($k \in \{\mathrm{fL}, \mathrm{L}, \mathrm{C}, \mathrm{R}, \mathrm{fR}\}$) within a single session, and a "treatment event" as the first instance in a user's lifetime when they are exposed to such a burst. For each content category, we consider three "treatment groups" comprising individuals who are exposed to burst lengths $M_v^{(k)} \in \{2, 3, 4\}$ (Appendix, Table X) and compute the difference in their average daily consumption of same content category k pre- and post-exposure. Finally, we compute the difference in difference between our treatment groups and a "control group" of individuals with maximum $M_v^{(k)} = 1$ video per session, where we use propensity score matching to account for differences in historical web and YouTube consumption rates, demographics (age, gender, race), education, occupation, income, and political leaning (see Appendix, section G for details). We emphasise that these treatments are not randomized and could well be determined endogenously (i.e. individuals are exposed to longer bursts because they already have higher interest in the content), hence the effects we observe should not be interpreted as causal. Nonetheless, as a purely predictive exercise it is interesting to see whether exposure to a fixed-length burst of content at one point in time has different effects at a future point in time across content categories.

Fig. 5 shows the results for $M_v^{(k)} = \{2,\ 3,\ 4\}$ across our five main content categories. In all cases, increases in burst length from 2 to 4 correspond to higher future consumption relative to the control groups. However, individuals exposed to bursts of far-right content show much larger effects than other content categories for all burst lengths, and larger marginal effects for longer vs. shorter bursts. The daily far-right content consumption of individuals exposed to far-right bursts of length $M_v^{(\mathrm{fR})} = 4$, increases by a gap of almost 50 seconds post-exposure, relative to the control group ($\tau = 46.9 \pm 12.5$ with CI=0.95).

h. Potential causes of radicalization. Summarizing thus far, consumption of radical political content on YouTube–while small relative to politically moderate and non-political content–is stickier and more engaging than other content categories, and is increasingly popular. In light of these trends, previous authors have hypothesized that the rise of radical content on YouTube is somehow driven by the platform itself, in particular by its highly influential recommendation engine [14, 15]. While this hypothesis is plausible, other explanations are equally so. As large as YouTube is, it is just a part of an even larger information ecosystem that includes the entire web and TV and radio. Thus, the growing appeal of radical content on YouTube may simply reflect a more general trend driven by a complicated combination of causes, both technological and sociological, that extend beyond the scope of the platform's algorithms.

In order to disambiguate between these alternative explanations, we performed three additional analyses.

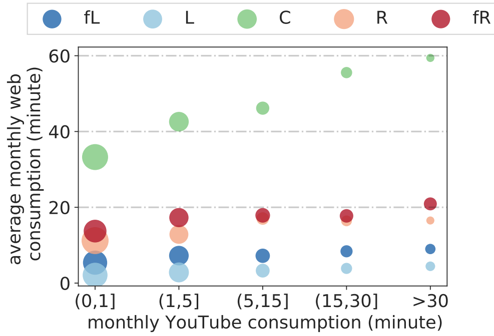

FIG. 6. Average monthly web consumption from category k ($k \in \{\mathrm{fL}, \mathrm{L}, \mathrm{C}, \mathrm{R}, \mathrm{fR}\}$), versus user's YouTube consumption of the same category. Users are assigned to one group according to the majority of their video viewership across categories and binned based on their average monthly YouTube consumption from the same category. Size of circles is proportional to the fraction of users of each category in the corresponding YouTube viewership bin (summation of circles with same color is equal to one). For bin x and category k, y-axis shows the average monthly consumption of users from web news domains of the same category.

First, we examined whether YouTube consumption is aberrant relative to off-platform consumption of similar content. Second, we analyzed the exact pathways by which users encountered radical content on YouTube, thereby placing an upper bound on the fraction of views that could have been caused by the recommender. Finally, we checked whether radical content is more likely to be consumed later in a user session, when the recommendation algorithm has had more opportunities to recommend content. Although none of these analyses on its own can rule out–or in–the causal effect of the recommendation engine, the strongest evidence for such an effect would be (a) higher on-platform vs. off-platform consumption of radical content, (b) arrival at radical content dominated by immediately previous video views (thereby implicating the recommender), and (c) increasing frequency of radical content toward the end of a session, especially a long session. By contrast, the strongest evidence for outside influences would be (a) high correlation between on- and off-platform tastes, (b) arrival dominated by referral from outside websites or search, and (c) no increase in frequency over sessions, even long ones.

On- vs. Off-platform. To check for differences in on- vs. off-platform consumption, we compared the YouTube consumption of members of our five previously identified communities with their consumption of non-YouTube websites classified according to the same content categories. To label websites, we first identified news domains using Nielsen's classification scheme, which distinguishes between themes such as entertainment, travel, finance, etc. Out of all web domains accessed by individuals in our data set, 3362 were in the news category. We then used the partisan audience bias score provided by Ref. [33] to bucket the news domains into equivalent political labels as for the YouTube channels above (see Appendix, section C for more details). Fig. 6 shows that YouTube consumption is correlated with broader web consumption of the same content category, including far left and far right (see Appendix, section E for more details). In the case of far-right content, on-platform consumption exceeds off-platform consumption only for the heaviest YouTube consumers, comprising only 0.56% of the population, who consume more that 30 minutes per month on-platform vs. 20 minutes per month off-platform. These results, while not ruling out the possibility that recommendations are driving engagement for the heaviest consumers, nonetheless show that consumption of radical content extends well beyond YouTube.

Referral mode. Next, we explore how users encounter YouTube content by identifying "referral" pages, which we define as the page visited immediately prior to each YouTube video. We then classify referral pages as belonging to one of six categories (1) the YouTube homepage, (2) search (inside YouTube or external search engines), (3) a YouTube user/channel, (4) another YouTube video, (5) an external (non-YouTube) URL, and (6) other miscellaneous YouTube pages, such as feed, history, etc. Table II shows that while 38% of far-right videos are preceded by another video, nearly 60% of referrals come from one of: the YouTube homepage (10%); search queries (10%); or external URLs (40%). Moreover, focusing on the subset of videos that are watched immediately after a user visits a news web domain (Appendix, Figs. 10 and 11), we find that approximately 50% of far-right/right videos are begun after visiting a news domain such as "breitbart.com," "infowars.com," and "foxnews.com," that are also in the right or far-right categories. In contrast, if the video is from a far-left, left or center channel, it is highly likely (70%) that the external entrance domain belongs to the center news bucket, which indicates domains like "nytimes.com" and "huffingtonpost.com" (Appendix, Figs. 10 and 11).

These results further support the hypothesis that users who consume radical content on YouTube are actively seeking it and do so both on and off the platform.

i. Session analysis. Although our data do not reveal which videos are being recommended to a user, if the recommendation algorithm is systematically promoting extreme content, we would expect to observe increased viewership of far-right videos both over the course of a session and as the session duration increases. For example, if a user who initiates a session by viewing centrist or right-leaning videos is systematically directed toward far-right content, we would expect to observe a relatively higher frequency of far-right videos towards the end of the session. Moreover, because algorithmic recommendations have more opportunities to influence viewing choices as session length increases, we would expect to see higher relative frequency of far-right videos in longer sessions. Conversely, if we observe no increase in the relative frequency of far-right videos either over the course

TABLE II. Distribution of the referral pages of videos within each category. Video URLs can start from a YouTube homepage, a search, a YouTube user/channel, another YouTube video, or an external URL. YouTube videos have a bigger role for left and right video viewership (42%) compare to far left and far right (37%).

category	YouTube homepage	search	YouTube user/ channel	YouTube video	external URLs	other
fL	11.78	11.92	3.96	36.67	32.95	2.73
L	11.35	12.37	2.64	40.64	30.35	2.66
C	9.71	14.83	2.0	34.9	35.2	3.36
R	7.16	10.75	2.39	42.5	34.94	2.26
fR	10.82	9.1	3.72	38.04	35.62	2.69

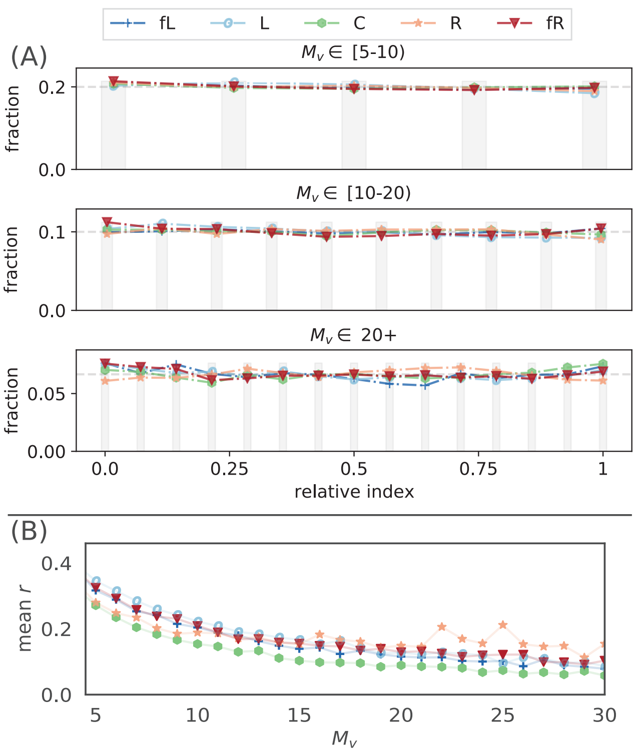

FIG. 7. (A) Distribution of relative index of videos within a session, across channel categories, and grouped by the total number of videos in a session, M_v. Each row presents the distribution of relative index of videos of category k, $k \in$ {fL, L, C, R, fR}, within all sessions with M_v videos, where (1) $M_v \in [5-10)$, (2) $M_v \in [10-20)$, and (3) $M_v \geq 20$. We observe an almost uniform distribution of relative indices, where there is slightly higher density closer to the relative index 1 for center channels within longer sessions, and slightly higher density closer to the relative index 0 for far right, left, and far left. (B) Average r, where r is the fraction of videos of category k. Sessions with length $M_v \geq 30$ (2% of the sessions) are dropped for better visualization.

of a session or with session length, that would undercut arguments to the effect that the recommender is driving users toward radical content vs. simply satisfying their preexisting preference for it.

To test these hypotheses, we assigned each video with political label $k \in$ {fL, L, C, R, fR} an index $i^{(k)} \in \{1,..,M_v\}$ where M_v is the number of videos in the sessions. We then normalized the indices $i^{(k)}_{\text{norm}} = \frac{i^{(k)}-1}{M_v-1}$, such that $i^{(k)}_{\text{norm}} \in [0,1]$, meaning that zero indicates the first video and one indicates the final video of a session. Fig. 7A shows the frequencies of videos in our five political categories as a function of normalized index for sessions of length $5 \leq M_v < 10$ (top panel), $10 \leq M_v < 20$ (middle panel), and $20 \leq M_v$ (bottom panel) respectively. In all cases, we find a nearly uniform distribution with an entropy deviating only slightly from that of a perfectly uniform distribution (Appendix, Table XII). Within longer sessions, there is a slightly higher density closer to the relative index 1 for center channels, and slightly higher density closer to the relative index 0 for far-left, left and far-right channels (see Appendix, Figs. 15, 16 and 17, and Table XII for more details and robustness checks). Fig. 7B shows the average frequency of content categories as a function of session length. All five content categories show decreasing frequency, suggesting that longer sessions are increasingly devoted to non-news content. Regardless, we see no evidence that far-right content is more likely to be consumed either toward the end of session or in longer sessions.

DISCUSSION

While news-related content is a relatively small fraction of all consumption on YouTube, the sheer scale of YouTube makes it a large and growing news audience, representing roughly half of total online news watch time (Appendix, Fig. 14). The potential for this audience to be exposed to extreme and conspiratorial content is substantial, especially if the platform itself is responsible, via its recommendation algorithms [21–24, 34]. Here, we have investigate these possibilities by analyzing the detailed news consumption of more than 300,000 YouTube users who watched more than 20 million videos, with nearly half a million unique videos spanning the political spectrum over a four year time period. Our results show the existence of distinct political news "echo chambers" on YouTube, in which users predominantly consume videos from one political category (Fig. 1). In contrast, online web news consumption does not show such echo chambers [35–37], suggesting that YouTube may play a special role in the consumption of digital news, and in particular, the consumption of far-right content. While the far-right echo chamber is somewhat smaller than the left and centrist communities, it is rapidly growing in population size and watch time. Moreover, its users are more engaged (Fig. 3) and more likely to stay engaged in the future (Fig. 4) than users in other echo chambers, especially when they are exposed to bursts of content (Fig. 5). At the same time, we find little evidence that the YouTube algorithm is responsible for these trends. Rather, we find that consumers of far-right content arrive via multiple "pathways" such as search and external websites, as well as previously watched videos. Nor do we see the rate of consumption of far-right content increase either over the course of a session or with the length of a session (Fig. 7), as would be expected if users were being actively steered to such content via their in-session recommendations.

Together, these findings suggest that YouTube—while clearly an important destination for producers and consumers of political content—is best understood as part of a larger ecosystem for distributing, discovering, and consuming political content [22]. Although much about the dynamics of this ecosystem remains to be understood, it is plausible to think of YouTube as one of many "libraries" of content—albeit an especially large and prominent one—to which search engines and other websites (e.g., Rush Limbaugh's blog or breitbart.com) can direct their users. Once they have arrived at the library, users may continue to browse other similar content, and YouTube presumably exerts some control over these subsequent choices via its recommendations. However 80% of sessions are length one — 55% of videos are in sessions of no more than four videos—and reflect the tastes and intentions that users entered with. To the extent that the growing consumption of radical political content is a social problem, our finding suggest that it is a much broader problem than simply the policies and algorithmic properties of a single platform, even one as large as YouTube.

Our analysis comes with important limitations. First, while our panel-based method has the advantage of measuring consumption directly, it does not allow us to see videos that were recommended but not chosen. Fully reconstructing the decision processes of users would therefore require a combination of panel and platform data. Second, our data only includes desktop browsing, and hence reflects the behaviors of people who tend to use desktops for web browsing. Thus, while our panel is indeed a representative sample of the population, future work would also incorporate mobile consumption. Third, while our sample of videos is large and encompasses most popular channels, we cannot guarantee that all content of interest has been included. Future work would therefore benefit from yet larger and more comprehensive samples, both of videos and people. Fourth, because our sample is retrospective, roughly 20% of videos had been unavailable when we attempted to access their labels using the YouTube API. Although we were able to impute these labels using a classifier trained on a small sample of videos for which labels were available from other sources, it would be preferable in future work to obtain the true labels in closer to real time. Fifth, our method for identifying referral pages does not account for the possibility that users move between multiple "tabs" on their web browsers, all open simultaneously. Moreover viewing credit in the Nielsen panel is only assigned to videos that are playing in the foreground, allowing for the possibility that other videos are playing automatically on background tabs. As a result, some videos that we have attributed to external websites may have in fact been suggested by the recommender in a background tab. The results presented in Table II thus should be viewed as an upper bound for "external" entrances and lower bound for "video" entrances.

In addition to addressing these limitations, we hope that future work will address the broader issue of shifting consumption patterns that are driven by "cord cutting" and other technology-dependent changes in consumer behaviour. Although recent work has shown that television remains by far the dominant source of news for most Americans [25], our results suggest that online video content—on YouTube in particular—is increasingly competing with cable and network news for viewers. If so, and if the "market" for online video news is one in which small, low-quality purveyors of hyper-partisan, conspiratorial, or otherwise misleading content can compete with established brands, the combination of high engagement and large audience size may both fragment and complicate efforts to understand political content consumption and its social impact. While misinformation research has to date focused on text-heavy platforms such as Twitter and Facebook, we suggest that video deserves equal attention.

ACKNOWLEDGMENTS

Acknowledgments: We are grateful to Harmony Labs for engineering and financial support and to Nielsen for access to web panels. The authors also thank Mark Ledwich, Kevin Munger, and Manoel Horta Ribeiro for sharing their data sets with us. Financial support for this research was provided in part by Carnegie Corporation of New York.

[1] M. D. Conover, J. Ratkiewicz, M. R. Francisco, B. Gonçalves, F. Menczer, and A. Flammini, The International AAAI Conference on Web and Social Media (ICWSM) **133**, 89 (2011).
[2] M. Del Vicario, G. Vivaldo, A. Bessi, F. Zollo, A. Scala, G. Caldarelli, and W. Quattrociocchi, Scientific Reports **6**, 37825 (2016).
[3] P. Grover, A. K. Kar, Y. K. Dwivedi, and M. Janssen, Technological Forecasting and Social Change **145**, 438 (2019).
[4] M. Bossetta, Journalism & Mass Communication Quarterly **95**, 471 (2018).
[5] M. Alizadeh, J. N. Shapiro, C. Buntain, and J. A. Tucker, Science Advances **6**, eabb5824 (2020).
[6] S. Aral and D. Eckles, Science **365**, 858 (2019).
[7] D. M. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, *et al.*, Science **359**, 1094 (2018).
[8] G. Pennycook and D. G. Rand, Proceedings of the National Academy of Sciences, USA **116**, 2521 (2019).
[9] J. Roozenbeek and S. Van Der Linden, Journal of Risk Research **22**, 570 (2019).
[10] T. Konitzer, J. Allen, S. Eckman, B. Howland, M. M. Mobius, D. M. Rothschild, and D. Watts, Available at SSRN 3548690 (2020).
[11] Statista, "Number of YouTube viewers in the United States from 2018 to 2022," `https://www.statista.com/statistics/469152/number-youtube-viewers-united-states/` (2019).
[12] Statista, "Number of monthly active Twitter users in the United States from 1st quarter 2010 to 1st quarter 2019," `https://www.statista.com/statistics/274564/monthly-active-twitter-users-in-the-united-states/` (2019).
[13] S. Clark and A. Zaitsev, arXiv preprint arXiv:2010.09892 (2020).
[14] Kevin Roose, "The making of a YouTube radical," `https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html` (2019).
[15] Zeynep Tufekci, "Youtube, the great radicalizer," `https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html` (2017).
[16] James Bridle, "Something is wrong on the internet," `https://medium.com/@jamesbridle/something-is-wrong-on-the-internet-c39c471271d2` (2017).
[17] Tom Whyman, "Why the right is dominating YouTube." `https://www.vice.com/en_us/article/3dy7vb/why-the-right-is-dominating-youtube` (2017).
[18] Tom Whyman, "—," `https://qz.com/706461/facebook-is-predicting-the-end-of-the-written-word/` (2017).
[19] Maryam Mohsin, "10 YouTube stats every marketer should know in 2020," `https://www.oberlo.com/blog/youtube-statistics` (2020).
[20] M. H. Ribeiro, R. Ottoni, R. West, V. A. Almeida, and W. Meira Jr, in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (2020) pp. 131–141.
[21] J. Cho, S. Ahmed, M. Hilbert, B. Liu, and J. Luu, Journal of Broadcasting & Electronic Media , 1 (2020).
[22] K. Munger and J. Phillips, The International Journal of Press/Politics , 1940161220964767 (2020).
[23] M. Ledwich and A. Zaitsev, arXiv preprint arXiv:1912.11211 (2019).
[24] M. Faddoul, G. Chaslot, and H. Farid, arXiv preprint arXiv:2003.03318 (2020).
[25] J. Allen, B. Howland, M. Mobius, D. Rothschild, and D. J. Watts, Science Advances **6**, eaay3539 (2020).
[26] D. Lazer, Proceedings of the National Academy of Sciences, USA **117**, 21 (2020).
[27] S. Wu, M.-A. Rizoiu, and L. Xie, in *Twelfth International AAAI Conference on Web and Social Media* (2018).
[28] Sara Salinas, "YouTube removes Alex Jones' page, following bans from Apple and Facebook," `https://www.cnbc.com/2018/08/06/youtube-removes-alex-jones-account-following-earlier-bans.html/` (2018).
[29] R. Kumar and A. Tomkins, in *Proceedings of the 19th International Conference on World Wide Web* (2010) pp. 561–570.
[30] R. C. De Amorim and C. Hennig, Information Sciences **324**, 126 (2015).
[31] C. R. Sunstein, University of Chicago Law School, John M. Olin Law & Economics Working Paper (1999).
[32] C. A. Bail, L. P. Argyle, T. W. Brown, J. P. Bumpus, H. Chen, M. B. F. Hunzaker, J. Lee, M. Mann, F. Merhout, and A. Volfovsky, Proceedings of the National Academy of Sciences, USA **115**, 9216 (2018), https://www.pnas.org/content/115/37/9216.full.pdf.
[33] R. E. Robertson, S. Jiang, K. Joseph, L. Friedland, D. Lazer, and C. Wilson, Proceedings of the ACM on Human-Computer Interaction **2**, 1 (2018).
[34] M. Alfano, A. E. Fard, J. A. Carter, P. Clutton, and C. Klein, Synthese (2020).
[35] S. Flaxman, S. Goel, and J. M. Rao, Public Opinion Quarterly **80**, 298 (2016).
[36] F. Zuiderveen Borgesius, D. Trilling, J. Möller, B. Bodó, C. H. De Vreese, and N. Helberger, Internet Policy Review. Journal on Internet Regulation **5** (2016).
[37] A. Guess, "(Almost) Everything in moderation: new evidence on Americans' online media diets." (2018).
[38] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, Knowledge-based systems **46**, 109 (2013).
[39] T. T. Nguyen, P.-M. Hui, F. M. Harper, L. Terveen, and J. A. Konstan, in *Proceedings of the 23rd International Conference on World Wide Web* (2014) pp. 677–686.
[40] Mostafa M. El-Bermawy, "Your filter bubble is destroying democracy," `https://www.wired.com/2016/11/filter-bubble-destroying-democracy/` (206).
[41] E. Colleoni, A. Rozza, and A. Arvidsson, Journal of communication **64**, 317 (2014).
[42] P. Barberá, J. T. Jost, J. Nagler, J. A. Tucker, and R. Bonneau, Psychological science **26**, 1531 (2015).
[43] C. Roth, A. Mazières, and T. Menezes, PlosOne (2020).
[44] S. Lall, M. Agarwal, and R. Sivakumar, in *ICC 2020-2020 IEEE International Conference on Communications (ICC)* (IEEE, 2020) pp. 1–7.
[45] R. Ottoni, E. Cunha, G. Magno, P. Bernardina, W. Meira Jr, and V. Almeida, in *Proceedings of the 10th ACM Conference on Web Science* (ACM, 2018) pp. 323–

332.
[46] K. Papadamou, S. Zannettou, J. Blackburn, E. De Cristofaro, G. Stringhini, and M. Sirivianos, arXiv preprint arXiv:2010.11638 (2020).
[47] P. C. Austin, Multivariate behavioral research **46**, 399 (2011).
[48] R. Lewis, Data & Society Research Institute (2018).

Appendix A: Related work

The internet is giving rise to new forms of the spread of knowledge, making it more easily and quickly accessible. It has impacted many aspects of our lives, where individuals increasingly turn to online environments to find information on topics such as education, science and health. With increasing amounts of information in a plethora of social and online media platforms, recommendations have become popular in learning users' interests and directing them to those information or items that best meet their preferences [38]. However, it poses challenges such as narrowing down users' exposure to like-minded people and content, a phenomena known as "filter bubble", "echo chamber" or other terms in the scientific literature [39, 40]. There exists a body of work on these topics on Twitter and Facebook [2, 4, 41, 42], while YouTube has been overlooked. A recent group of studies have examined the role of YouTube's recommendation algorithm in promoting extreme content and user radicalization [21–24, 43]. The majority of these studies rely on API collected data, which has been followed by concerns such as how accurately it reflects real users' engagement on this platform. Ref. [44] is the first effort to provide a longitudinal data set of real users with more than one year of historical data. Yet, the data is comprised of only 243 users, not a large representative sample of the US population to study the political content consumption on YouTube.

In an attempt to find the radicalization pathways on YouTube, Ref. [20] tries to answer the question of whether the YouTube recommendation system is driving users to extreme and far-right contents. Users who left comments for videos shared by each channel are considered as active in that topic. Therefore, the migration of users toward radicalism can be recovered by following users' commenting behavior across these communities. However, besides the fact that only a small fraction of viewers leave a comment for a video [22], a previous study has shown [45] that right-wing channels tend to contain a higher degree of words from "negative" semantic fields, which brings the concern of whether number of comments on a video from a specific category is positively correlated with user's partisan leaning. Finally, this trend can not conclude any causation regarding YouTube radicalizing its users, as YouTube is not the only source of information users are exposed to. While Ref. [20], finds Alt-right videos are only reachable through channel recommendations, looking at the trajectories of users, we observe only around 4% of far-right videos are reached via a channel.

The study closest to our work, Ref. [22], tries to answer why the far right is more successful in receiving viewership on YouTube than the far left. To answer whether it is the YouTube recommendation model's fault, or if other factors can explain it, the authors define view count as the "demand" and number of videos uploaded as the "supply". While they observe the demand (views per video) for Alt-right and Alt-lite channels has decreased since 2017, Ref. [20] pays more attention to comments per view as a notation of strong engagement and concludes an increase in user radicalization.

Augmenting the set of channels introduced in the previous two works to cover highly subscribed YouTube political channels, Ref. [23] studies recommendations made from 816 channels with different ideologies. They record the recommended videos via YouTube API, starting from videos of a set of seed channels. This study came to the opposite conclusion of Ref. [20], that YouTube deradicalizes viewers of extreme content by favoring mainstream and neutral content in their recommendation. One of the main criticisms the authors have received is the strong conclusion made while there was a lack of real users (anonymous accounts with no watch history) in the study of recommended videos/channels.

Similar to Refs. [20] and [23], following the YouTube API recommendations starting from videos posted by a set of 1,080 news and political channels, the authors' focus in Ref. [24] is on the promotion of conspiracy videos. Over a period of more than one year, they observe that after a conspiratorial video is clicked, it is highly likely that the algorithm would recommend another one, yet with a decreasing trend. Consequently, they found existence of the filter-bubble effect for conspiracy content. Ref. [34] also focuses on studying YouTube's recommendation algorithm, concluding that the algorithm appears to have shifted the balance towards promoting longer, more conspiratorial content. Similar to [24], they only look for conspiratorial content, while the study is comparatively smaller, following frequently recommended videos after six selected videos.

Regardless of the disagreements, these methods are problematic due to (i) reliance on logged-out users, (ii) reliance on comments to trace trajectories, (iii) small and unrepresentative sample sizes, and (iv) insufficient history of data. In a recent study on pseudo-scientific videos, Ref. [46] finds that users' watch history affects search results and related recommended videos. Longitudinal characterization of YouTube browsing behavior can shed light on how far away the API data, with no user history, is from real consumption.

Appendix B: Data set

a. Nielsen online panels. Nielsen Company maintains large panels of American households, the members of which agree to have their media habits tracked in exchange for payment. This study makes use of a laptop/desktop web-browsing panel, when a tracking software is installed on the user's web browser(s) on the user's primary laptop or desktop computer. A pageview credit occurs when a content request from the in-focus tab is detected. Duration or time spent is credited to an in-focus page as long as the web session is active (a web session ends when there is more than 30 minutes of inactivity). When a user returns to a tab with previously loaded content, duration is credited immediately. We count any person with at least one record of a YouTube link (whether to a video URL or not) as a user of this platform. The consumption data from Nielsen (Desktop Web) is individual-level and the individuals are a representative sample of the US population. For aggregated statistics derived from this data, we weigh each individual panelist's contribution according to whether their demographics are over or under represented in the panel with respect to the distribution of demographics in the US (Ref. [25] provides more details on the calculation of weights). For aggregated results, we report the weighted sum each month. We also have demographic information of the panelists, age, gender, race, education level, occupation, income and zip code, which will be used in part of the analysis.

TABLE III. Sub-reasons for unavailability of a random set of 368,754 videos. The "other" group includes a wide range of reasons such as nudity, bullying, hate speech, etc.

sub-reason	fraction
unavailable, no sub-reason	0.53
terminated account	0.25
private	0.15
copyright	0.03
other	0.04

b. YouTube API data. Each video on YouTube has a unique 11 character *video ID*, which can be extracted from the logs of YouTube video URLs. A video is published by a channel, where a channel on YouTube is similar to a user on other social platforms and joint with one user account. Sending queries for video IDs, we collected their metadata, such as channel ID, channel category ID, video length, title, etc. using YouTube API. Similar to previous efforts on YouTube platform [27], we also found a subset of video IDs (20%) are not available any more through YouTube API. To have a better understanding of reason behind this unavailability, for a random set of around 18% of these unavailable videos, we crawled their HTMLs and extracted the sub-reason provided by YouTube, e.g. private, or terminated account, Table III.

Appendix C: Defining News Categories

a. YouTube Video Categories. Previous works have a considerable effort on labeling YouTube channels with a variety of ideologies. We borrow the labeled channels from Refs. [23] and [20] and ranked them based on their specific ideologies, Table IV. Next, we bucket them into five political categories, far left (fL), left (L), center (C), right (R) and far right (fR). Combining the set of labeled channels (details in section I), we have 997 channels within the political spectrum of far left to far right, Table IV. By having the channel ID of each video, we assign the label of the corresponding channel to all its published videos. Ref. [24] provides a set of video IDs labeled as "conspiracy", which we assign to the far-right buckets in our set. Aggregating the labels from different sources, we end up with 520,514 videos from these channels, Table V.

b. Online News For news websites, we use the partisan audience bias score for web domains provided by [33], which covers 1263 out of the 3362 original news websites found in our primary data source. These scores are obtained based on proportions of left and right-leaning Twitter users that shared links from each website during a sampling period in 2018 [33]. In contrast to YouTube, news domains are not aggregated by channels, and so bias categorization is done at the level of the individual news website. To bucketize the news domains into equivalent political labels as used for the YouTube, we sort the web domains from lowest (farthest left) to highest (farthest right). Then we define four web domains as the thresholds for cutting the continuous scores into five buckets. On the left, we identify slate.com as the threshold for the far-left bucket, and theguardian.com as the boundary for the left bucket. On the right, we identify foxnews.com as the boundary for right, and breitbart.com as the limit to fall into the far-right bucket, Fig. 8.

TABLE IV. Labeled channels across categories.

channel category	Ref. [23]	Ref. [20]	number of channels	sample channels
far left	social justice/ socialist/ revolutionary socialist/ anti-theist/ anti-white	-	137	The Young Turks
left	partisan left/ left	left/ left-center	132	MSNBC, Vox
center	center	left-center/ center/ right-center	133	ABC News, CBS News, business insider
right	partisan right-right/ libertarian-right	right-center/right	84	Conservative Citizen, Fox News
far right	conspiracy/ IDW/ Alt-right/ MRA/ Alt-light/ religious conservative	Alt-lite/ Alt-right/ IDW/ MRA	511	James Allsup, Black Pigeon Speaks, Red Ice TV

TABLE V. Number of labeled videos over different channel categories. In parenthesis the percentage from the set of videos with metadata from API is provided.

Category	fL	L	C	R	fR
Num. (percent of available) of videos	74,668 (0.76%)	202,462 (2.06%)	94,353 (0.96%)	38,193 (0.39%)	110,838 (1.13%)

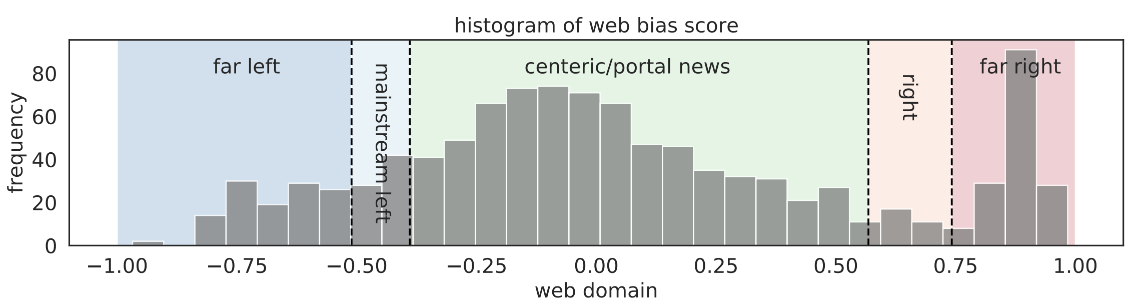

FIG. 8. Histogram of web bias scores for a set 1263 publishers. We rank the domains based on their score and use the scores of “slate.com”, “theguardian.com”, “foxnews.com”, “breitbart.com” domains as cut off points. The vertical lines indicate the threshold for hard labeling of the scores into five buckets from far left into far right.

Appendix D: Missing video imputation

Using the YouTube API, 20.1% of the video IDs had no return from the API. One naive approach would be considering all videos have extreme content. However, previous studies on YouTube have collected video IDs plus a variety of metadata including channel ID, title, close caption, comments, etc., which can help to have a more accurate estimate of these videos. We collected the list of video IDs from these studies and looked for the set of unavailable videos within our set, where we could retrieve the channel ID of $69,840$ unique videos ($74,506$ video IDs), Table VI. Number of retrieved videos for each channel category is as following: far left, 125 (0.18%); left, 1287 (1.84%); center, 749 (1.07%); right, 798 (1.14%); and far right, 3156 (4.52%). Majority of the deleted videos, 63,725 (91.25%), do not belong to any channel of the spectrum of political categories in our set.

We treat these unavailable videos in our data set as missing values, and impute their channel political category. Within retrieved unavailable videos, only around 5% belong to far-right or far-left channels, which shows a naive upper bound of considering all unavailable videos as extreme political content can be misleading. Using the subset of retrieved video labels, we can train a model to learn the patterns of viewership and user characteristics of these videos and infer the labels of the rest via a supervised approach, using ground truth of 7% of our set.

Table VII provides the list of extracted features, at two levels, user-level and session-level. By training a random forest classifier over 96 features, we could infer a probability score for each unavailable video. For each content category k, we design a separate binary classifier of positive class for content of category k, and negative otherwise. We follow the traditional approach for model selection with cross validation on the train set (80%), and evaluating the model on the holdout set (20%). We report the AUC performance metric (Table VIII) which is threshold independent. For each classifier we consider two thresholds, one with higher recall, and one with higher precision and we report the exact number of videos inferred for each class based in these thresholds, Table IX. In the main text, we present the results for a set of videos from Table V, augmented with newly inferred labels with high precision in Table IX. Since we are investigating the existence of echo chambers, we consider a conservative approach and report a lower bound.

To have an understanding of what is the gap between the upper bound and lower bound imputations, we compute the share of news content on YouTube each month for each model, as following:

$$\text{Fraction of news content on YouTube at month } m = \sum_{u=1}^{N_m} \frac{d_{u,m}^{(\text{YouTube news})} \times w_{u,m}}{d_{u,m}^{(\text{YouTube})} \times w_{u,m}},$$

where N_m is the number of users at m-th month , $d_{u,m}^{(\text{YouTube})}$ is the total minutes of YouTube video consumption by u-th user at m-th month. The total minutes of YouTube news consumption by u-th user at m-th month is $d_{u,m}^{(\text{YouTube news})}$. Any video with (i) category ID “politics and news”, or (ii) with label “political news” from [24], or (iii) with a label from far left to far right, is consider under the broad label of “news”. Fig. 9 provides the upper bound and lower bound of share of news content on YouTube.

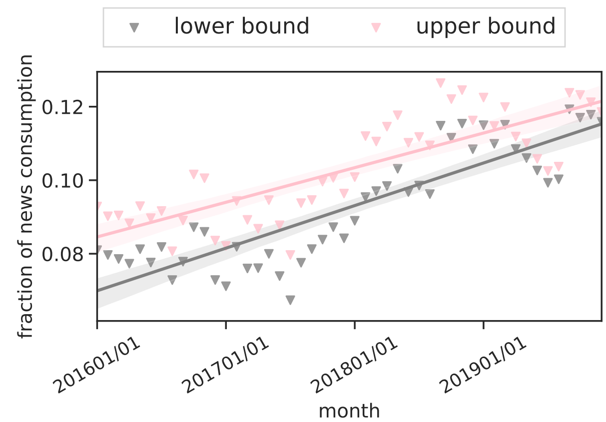

FIG. 9. Share of news content on YouTube.

TABLE VI. Previously collected data sets are used to retrieve the channels who published the videos currently unavailable via YouTube API. Overall, we could retrieve channel IDs of a set of 69,840 unique video IDs.

reference	number of retrieved videos
Fadoul *et al.* [24]	368
Ribeiro *et al.* [20]	3,211
Munger *et al.* [22]	854
Wu *et al.* [27]	18,134
Ledwich *et al.* [23]	51,939

TABLE VII. Detailed list of extracted features for video label imputation, where $k \in \{\mathrm{fL}, \mathrm{L}, \mathrm{C}, \mathrm{R}, \mathrm{fR}\}$, and x is one of the YouTube sub-reasons (Table III).

Feature name	number of features
user's average monthly views from channel category k	5
number of videos from channel category k, within the session	5
user's average monthly web domains visited from news category k	5
number of web domains visited from news category k, within the session	5
user's average monthly views across YouTube categories	19
frequency of the videos across YouTube categories within the session	19
user's average monthly unavailable videos with sub-reason x	15
frequency of the unavailable videos with sub-reason x within the session	15
user's monthly YouTube news consumption	1
YouTube news watch time within the session	1
user's monthly web news consumption	1
web news watch time within the session	1
video duration	1
user's average monthly unavailable videos	1
number of unavailable videos within session	1
frequency of the video across all users	1

TABLE VIII. Area Under Curve (AUC) of the models for each channel category. Left and far right trained models are more accurate in prediction of holdout samples.

Category	AUC
far left	0.93
left	0.97
center	0.93
right	0.93
far right	0.95

TABLE IX. Number (percent) of inferred videos across different channel categories. For assigning binary labels, we consider thresholds for two different settings, when the recall is optimized and when the precision is optimized. Percentages are out of total number of unavailable videos. The average recall of five models with high recall thresholds is 0.86 and their average precision of models with high precision thresholds is 0.65.

Category	fL	L	C	R	fR
Number (percent) of unavailable videos - (<recall>≈0.86)	106,117 (5.32%)	26,322 (1.32%)	41,628 (2.09%)	55,884 (2.8%)	57,066 (2.86%)
Number (percent) of unavailable videos - (<precision>≈0.65)	422 (0.02%)	2,039 (0.11%)	1,057 (0.053%)	1,710 (0.086%)	3,021 (0.15%)

Appendix E: Web and YouTube content correlation

YouTube is part of the whole ecosystem of online media. Users on the web are exposed to a wide range of domains and platforms with different political leanings, and their actions on YouTube are not isolated from external sources. To obtain Fig. 6 in the main text, we cluster our users based on their majority consumption from YouTube channels of the five political categories into five classes ($\psi \in \{\mathrm{fL}, \mathrm{L}, \mathrm{C}, \mathrm{R}, \mathrm{fR}\}$). Within each group, we examine the average web consumption pattern of users from same political category on web. To do so, we bucket the users inside each cluster ψ based on their monthly YouTube consumption of the same content category as their cluster label. Overall, we have 25 groups across five clusters and five buckets. Next, we find the average monthly web consumption of users inside each group, from same content category as their cluster. Fig. 6 shows that users on YouTube are also exposed to content from the same political leaning on web, which necessitates exploration of other contributing factors to fully understand possible pathways to political radical YouTube content.

We also analyze users' news consumption habits on web before switching to YouTube, by finding the subset of YouTube news videos that are begun right after a user visits a news web domain. We look back at the immediate pageview, before a YouTube video, and find all the URL pairs (url_{t-1},url_t) such that url_{t-1} is any external URL in our set and url_t is a YouTube-video URL, with label k. All search URLs, e.g., "search.yahoo.com", are excluded from the set of external URLs. For videos within each category k, we normalize to the total number of external entry points (each row is normalized to have sum one, where we limit the visualization to the unique set of top 15 domains from each video category), Fig. 10.

$$P(\mathrm{url}_{t-1} = x|\mathrm{url}_t = k) = \frac{\text{num. of pairs such that } \mathrm{url}_{t-1} = x \text{ and } \mathrm{url}_t = k}{\sum_x \text{num. of pairs such that } \mathrm{url}_{t-1} = x \text{ and } \mathrm{url}_t = k},$$

where $x \in \{\text{any external domain}\}$, and $k \in \{\mathrm{fL}, \mathrm{L}, \mathrm{C}, \mathrm{R}, \mathrm{fR}\}$. Highly visited pages across all categories include Facebook, Reddit, and Twitter. While it is not identifiable with our data set, it would be interesting to explore whether in fact far-right videos propagate more via email, Twitter or Facebook social media platforms. On the other hand far-left and left YouTube videos have been visited after Reddit, almost twice more compared to right and far right. Domains such as foxnews.com, breitbart.com and infowars.com have been paired more with far-right and right videos, while huffingtonpost.com and nytimes.com are highly visited with left and far-left videos. Even though not a causal effect, this shows exposure of individuals to news sources outside of the YouTube platform is correlated with their next pageview, here a YouTube video URL.

Next, we look at the subset of YouTube news videos that are begun right after a user visits a news web domain. We limit url_{t-1} visits to news domains and collect all pairs of (url_{t-1}=external-news-URL, url_t=YouTube-video-URL), where a "news domain" falls into one of the five buckets of web news political categories, $z \in$ {far left, mainstream left, centric, right, far right} , and a YouTube video should have a label from one of the YouTube channel categories $k \in \{\mathrm{fL}, \mathrm{L}, \mathrm{C}, \mathrm{R}, \mathrm{fR}\}$, Fig. 11.

$$P(\mathrm{url}_{t-1} = z|\mathrm{url}_t = k) = \frac{\text{num. of pairs such that } \mathrm{url}_{t-1} = z \text{ and } \mathrm{url}_t = k}{\sum_z \text{num. of pairs such that } \mathrm{url}_{t-1} = z \text{ and } \mathrm{url}_t = k}$$

Around 50% of far-right/right videos are begun after a news domain from far-right or right categories. In contrast, if the video is from a far-left, left or center channel, it is highly likely (70%) that the external entrance domain belongs to center news bucket. This supports the hypothesis that users are driven towards radical right content through others platforms or domains as well.

	nytimes.com	godlikeproductions.com	drudgereport.com	conservapedia.com	huffingtonpost.com	mail.aol.com	amazon.com	facebook.com	breitbart.com	twitter.com	ebay.com	xvideos.com	apps.facebook.com	docs.google.com	foxnews.com	rumormillnews.com
fL	0.39	0.00	0.02	0.04	0.72	0.49	1.65	12.24	0.06	1.90	0.46	0.55	0.41	0.82	0.09	0.01
L	0.79	0.01	0.18	0.02	0.84	0.49	1.40	12.12	0.13	2.09	0.42	0.30	0.37	0.86	0.31	0.02
C	0.36	0.05	0.32	0.00	0.17	0.66	1.39	10.83	0.05	2.01	0.36	0.22	0.40	1.32	0.27	0.10
R	0.33	0.78	5.63	0.31	0.22	0.72	0.74	14.76	0.89	2.76	0.73	0.14	0.38	0.21	0.93	0.34
fR	0.09	0.11	0.70	0.84	0.06	0.69	1.44	14.31	0.36	3.61	0.65	0.28	0.78	0.44	0.41	0.66

	tumblr.com	netflix.com	news.google.com	en.wikipedia.org	dailykos.com	duckduckgo.com	swagbucks.com	mail.google.com	instagram.com	mg.mail.yahoo.com	mail.yahoo.com	cnn.com	outlook.live.com	reddit.com	infowars.com
fL	1.35	0.98	0.20	2.06	0.28	0.18	0.64	4.68	0.54	0.58	0.48	0.35	0.59	4.96	0.01
L	1.07	0.86	2.59	1.86	0.64	0.30	0.46	3.93	0.55	0.41	0.65	0.59	0.56	4.43	0.06
C	0.50	0.73	1.77	1.56	0.17	0.39	0.38	5.07	0.52	0.83	0.53	0.39	0.60	3.68	0.07
R	0.21	0.13	1.30	0.90	0.05	0.31	0.22	4.80	0.30	0.63	0.72	0.41	0.96	2.47	0.56
fR	0.24	0.59	0.10	2.05	0.05	0.63	0.39	7.75	0.42	0.52	0.55	0.14	0.93	2.56	0.80

FIG. 10. $P(\mathrm{url}_{t-1} = \text{external-URL } x | \mathrm{url}_t = \text{YouTube-video-URL of category } k)$, showing correlation of url_{t-1} and url_t pageviews. In each row, url_t=YouTube-video-URL, with a video from channel category k, where $k \in \{\mathrm{fL}, \mathrm{L}, \mathrm{C}, \mathrm{R}, \mathrm{fR}\}$, and each column shows fraction of external URLs from each domain x = {any external domain}.

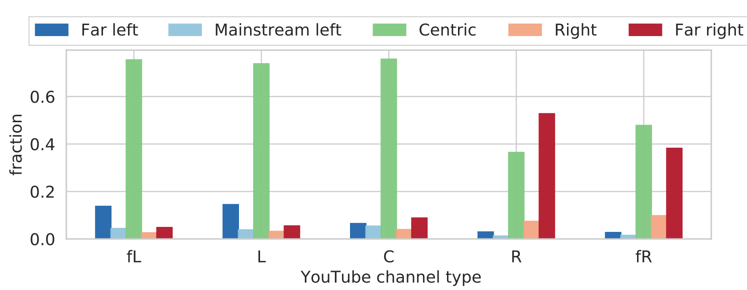

FIG. 11. $P(\mathrm{url}_{t-1}$=external-news-URL of category $z | \mathrm{url}_t$=YouTube-video-URL of category $k)$, where $z \in$ {far left, mainstream left, centric, right, far right} and $k \in \{\mathrm{fL}, \mathrm{L}, \mathrm{C}, \mathrm{R}, \mathrm{fR}\}$.

Appendix F: User clustering

From our set of 997 political news channels on YouTube, we group users into communities of similar consumption patterns. For each user, each month, we create a normalized viewership vector ν_i^m, where ν_{ij}^m is the fraction of viewership of user i from channel category $j \in \{\mathrm{fL}, \mathrm{L}, \mathrm{C}, \mathrm{R}, \mathrm{fR}\}$, during m-th month ($m \in \{1, ..., 48\}$), where the first month is January 2016 and the last month is December 2019. We cluster users' normalized monthly viewership vectors via K-means clustering. We consider a minimum of ζ minute(s) of news viewership per month, and each instance with less than ζ minute(s) of news viewership would belong to a non-news community. We look at silhouette scores for different numbers of clusters, where there is a peak at $K = 5$, Fig. 12. Each community is named according to the political category of predominant viewership share in its archetype, defined as $\psi(t)$ ($\psi(t) \in \{\mathrm{fL}, \mathrm{L}, \mathrm{C}, \mathrm{R}, \mathrm{fR}\}$), where the archetype of community $k = 1, .., K$ is calculated as:

$$c_k = \frac{1}{N^{(\psi=k)}} \sum_{i=1}^{N^{(\psi=k)}} \nu_{ij}^m,$$

where $N^{(\Psi=k)}$ is the total number of users with label k across the four-year period. We found five distinct archetypes, each with a predominant consumption of on average 90% from one category. Consistent with other metrics (Fig. 2), number of views and number sessions within each community is growing, Fig. 13.

For users inside each community, we look at their share of news consumption from YouTube compared to web. For a YouTube news consumer at month m, the share of news consumption on YouTube out of the total news consumption online is:

$$r_{u,v} = \frac{d_{u,m}^{(\text{YouTube news})}}{d_{u,m}^{(\text{online news})}},$$

where $d_{u,m}^{(\text{YouTube news})}$ is the total minutes of YouTube news consumption. Again, any video with (i) category ID "politics and news", or (ii) with label "political news" from [24], or (iii) with a label from far left to far right, is consider under the broad label of "news". The total minutes of online news consumption by u-th user at month m-th month, $d_{u,m}^{(\text{online news})}$, includes consumption from any YouTube news video, or web news domain. Fig. 14 shows the weighted median of $r_{u,v}$'s among the users who fall into the corresponding community each month,. Among all communities, over the last four years, the share of YouTube news consumption of these users has shifted increasingly from 5%-25% to 25%-60%.

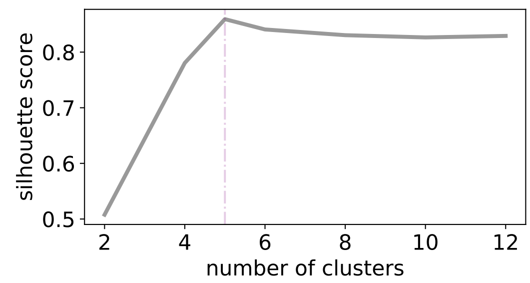

FIG. 12. Silhouette score vs. number of clusters.

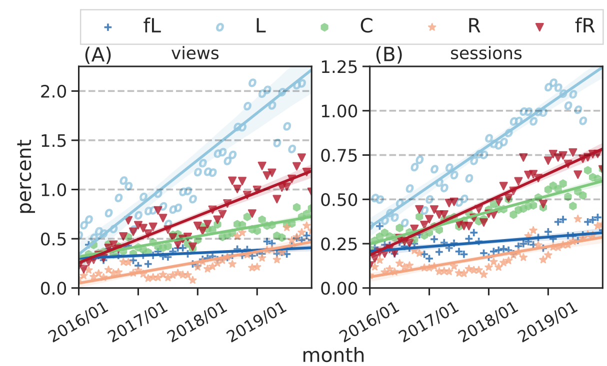

FIG. 13. Breakdown of percent of (A) view, (B) sessions falling into the five different channel categories, per month, from far left to far right, January 2016 to December 2019. Panel (A) is the percent of page views falling into each community, and panel (B) presents the percentage of the sessions (with at least one video from corresponding category) from each channel category.

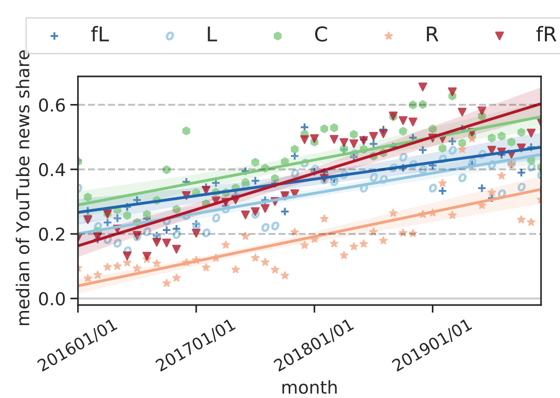

FIG. 14. Median YouTube news consumption share, across users inside each community.

Appendix G: Bursty Exposure

We measure the average daily consumption of the i-th user ρ_i (average seconds per day) pre- and post- event, defined as bursty viewership of videos from category k. Consumption of at least $M_v^{(k)}$ videos from channels of category k within a single session is considered as a proxy for bursty viewership (when $M_v^{(k)} > 1$). The first occurrence of bursty consumption within a user's lifetime is set as the event time. We follow the change in the user's future daily consumption after the event, as $Y_i = \rho_i^{(\text{post})} - \rho_i^{(\text{pre})}$, where superscripts indicate post- and pre- event intervals. Users in the control group have no more than one video from the same category within any of their sessions, with at least $M_v^{(k)}$ non-bursty video consumption from same category in their lifetime. For these users, the first half of their recorded lifetime is considered as pre- event, and the second half is considered as post- event. Users from either group have a minimum of one month of data in the Nielsen panel. As control variables, we utilize a set of observable characteristics including historical web and YouTube consumption rates across categories, and demographics including age, gender, race, education, occupation, income, and political leaning. For each content category, we consider three "treatment groups" comprising individuals who are exposed to burst lengths $M_v^{(k)}$ equal to 2, 3, and 4, as experiments 1, 2 and 3 respectively, Table X. The control groups is the same for all three experiments (users with no bursty viewership).

TABLE X. Individuals are assigned either to bursty consumption (treatment) group in the event of watching at least M_v^k videos from category k ($k \in \{\mathrm{fL}, \mathrm{L}, \mathrm{C}, \mathrm{R}, \mathrm{fR}\}$) within a session, or in control group, if none of their sessions has more than one video from same category with at least M_v^k videos in their lifetime. We run three experiments with different values of M_v^k, where experiment $1:$ $M_v^k = 2$, experiment 2: $M_v^k = 3$, experiment $3:$ $M_v^k = 4$. Number of users in each experiment, plus the break down of treatment and control groups are provided in order.

k	fL	L	C	R	fR
experiment 1	4466 (2065/2041)	10762 (4697/6065)	8064 (3300/4764)	2638 (941/1697)	4180 (1897/2283)
experiment 2	3186 (785/2401)	7982 (1917/6065)	5872 (1108/4764)	1994 (297/197)	3070 (787/2283)
experiment 3	2821 (420/2401)	7096 (1031/6063)	5278 (514/4764)	1809 (112/1697)	2708 (425/2283)

TABLE XI. δ is tolerable gap between non-video YouTube pageview and next YouTube pageview and γ is tolerable gap between a video YouTube pageview and next YouTube pageview.

	$\delta = \gamma =$ 30min	$\delta = \gamma =$ 60min	δ = 30min, γ = 60min	δ = 30min, γ = 120min	δ = 10min, γ = 60min	δ = 10min, γ = 120min
num. of sessions	9,116,367	7,881,304	8,157,087	7,336,722	8,620,394	7,825,521

TABLE XII. δ is tolerable gap between non-video YouTube pageview and next YouTube pageview and γ is tolerable gap between a video YouTube pageview and next YouTube pageview. Normalized video indices across political categories have nearly uniform distribution, with an entropy deviating only slightly from that of a perfectly uniform distribution, regardless of session length and session definition

session length	uniform	δ = 30min	δ = 60min	δ = 30min, γ = 60min	δ = 30min, γ = 120min	δ = 10min, γ = 60min	δ = 10min, γ = 120min
2	0.69	0.69	0.69	0.69	0.69	0.69	0.69
3	1.1	1.1	1.1	1.1	1.1	1.1	1.1
4	1.39	1.39	1.39	1.39	1.39	1.39	1.39
$[5, 10)$	1.61	1.58	1.59	1.58	1.58	1.58	1.58
$[10, 20)$	2.28	2.28	2.28	2.28	2.28	2.28	2.28
20+	2.98	2.98	2.98	2.98	2.98	2.98	2.98

Appendix H: Sensitivity analysis

We tested the sensitivity of our findings to a variety of factors through out the paper, namely: (i) session definition, (ii) news consumer definition, (iii) label imputation, and (iv) matching technique.

1. Sensitivity analysis of session definition

In this section we evaluate the sensitivity of our results to the selected parameters for breaking down the browsing logs into YouTube sessions, Table XI. At the session level, we investigated two possibilities: (i) far-right videos are visited more toward the end of sessions, and (ii) there is a bias across channel categories. If, starting from a random topic over the span of a session, the recommendation algorithm pushes the users via recommendations towards videos with specific political content or towards more extreme ends of the partisan spectrum, we will observe these patterns in the distribution of the order that these videos are visited. In the main text, for $\delta = 10$ and $\gamma = 60$ minutes and high precision imputation, we find no bias on distribution of far-right videos over the session, for sessions with length $M_k \geq 5$. We look at the entropy of sessions, where we find nearly uniform distribution with an entropy deviating only slightly from that of a perfectly uniform distribution, regardless of session length and session definition, Table XII. We are specifically interested in longer sessions, and we look at the mean and standard deviation of the fraction of normalized indices across different definitions of sessions, for high precision (Fig. 15) and high recall (Fig. 16). Both figures show very robust results, where video indices in sessions with high recall video imputation are more uniformly distributed. Fig. 17 also provides replication of Fig. 7A in the main text, again showing a more uniform visits to videos across channel categories compared with the high precision model.

2. Sensitivity of news consumer definition

In the main text, we defined a user as a YouTube "news consumer", with a threshold of at least one minute of total news consumption from any of the 997 labeled channels in our set. Here, we provide a more relaxed and a more strict definition of who may be considered a news consumer on YouTube with non-zero news consumption and at least five minutes news consumption per month respectively. A user who meets this criteria is considered in the clustering model, Fig. 18. In the definition with any non-zero news consumption, it is possible for the individuals with a small duration of news consumption, e.g., have selected a video by mistake while surfing the web, to be considered as news consumers. We consider a threshold for minimum news viewership to reduce this noise.

3. Sensitivity of label imputation

In section D we learn that the fraction of videos belonging to any of the channel categories is very small, and we are facing a highly imbalanced set. For such problems, the best choice of threshold to map the inferred probabilities to binary outcomes depends highly on the application. In the main text we presented the conservative results, with high precision. Here, we look at the upper bound for each category, by changing the threshold to obtain a high recall, even at the cost of low precision. Fig. 19 provides a comparison in the percentage of users inside each community, for no imputation (any unavailable video is considered as no news), high recall and high precision imputations. Similar to Fig. 2 in the main text, Fig. 20 provides the temporal trends across categories for all four choices of consumption metric, where we see our results are robust to the imputation model.

4. Sensitivity of matching

In the main text we present the results with propensity score matching and use the "MatchIt" package with the optimal pair matching method. Here, we utilize an inverse probability of treatment weighting approach [47], using the propensity scores to weight each individuals' outcome to create a synthetic sample in which the distribution of covariates is independent of treatment assignment. Assuming that Y_i denotes the outcome variable measured ($Y_i = \rho_i^{(\text{post})} - \rho_i^{(\text{pre})}$) on the i-th user, we report $\tau = E[Y_i^{(\text{bursty})} - Y_i^{(\text{control})}] = \frac{1}{N_u}\sum_{i=1}^{N_u}\frac{Z_i Y_i}{e_i} - \frac{1}{N_u}\sum_{i=1}^{N_u}\frac{(1-Z_i)Y_i}{(1-e_i)}$, where N_u is the total number of individuals in bursty and control groups. The indicator variable Z_i denotes whether or not the i-th user has bursty viewership and e_i denotes the user's propensity score ($e_i = P(Z_i = 1|\text{covariates})$). The weighting helps to ensure that the differences in an individual's future consumption are not simply explained by the differences in their historical and demographic characteristics. We run this experiment for all political categories, Fig. 21. Across political categories, sensitivity to the severity of exposure is highest for far-right content, and the difference in consumption rate change of bursty viewership and control groups is almost half for other categories compared to far right.

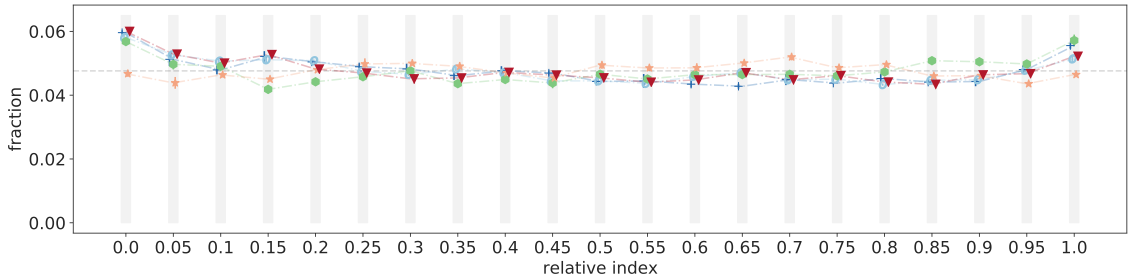

FIG. 15. Mean and standard deviation of relative index fractions across session definitions, for sessions with length $M_k \geq 20$, with high precision imputation.

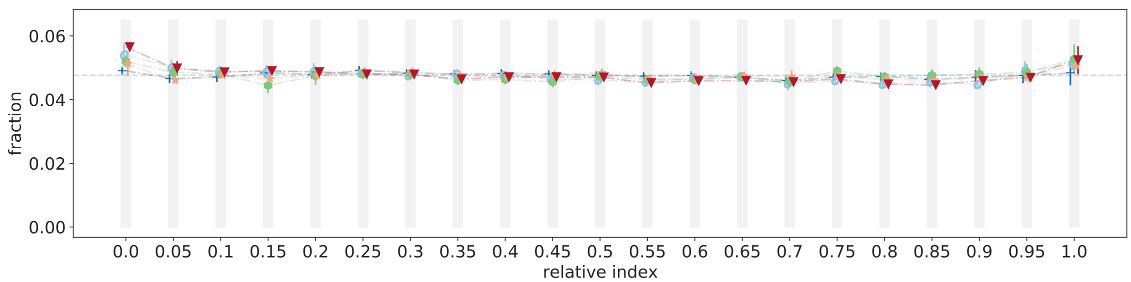

FIG. 16. Mean and standard deviation of relative index fraction across session definitions, for sessions with length $M_k \geq 20$, with high recall imputation.

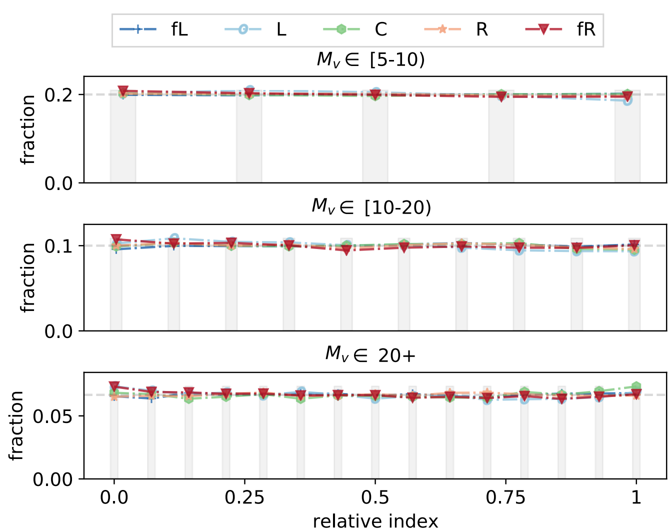

FIG. 17. Distribution of relative index of videos within a session, across channel categories, and grouped by the total number of videos in a session, M_v, with high recall imputation. Each row presents the distribution of relative index of videos of category k, $k \in \{\text{fL, L, C, R, fR}\}$, within all sessions with M_v videos, where (1) $M_v \in [5-10)$, (2) $M_v \in [10-20)$, and (3) $M_v \geq 20$. We observe an almost uniform distribution of relative indices, where there is slightly higher density closer to the relative index 1 for center channels within longer sessions, and slightly higher density closer to the relative index 0 for far right, left and far left.

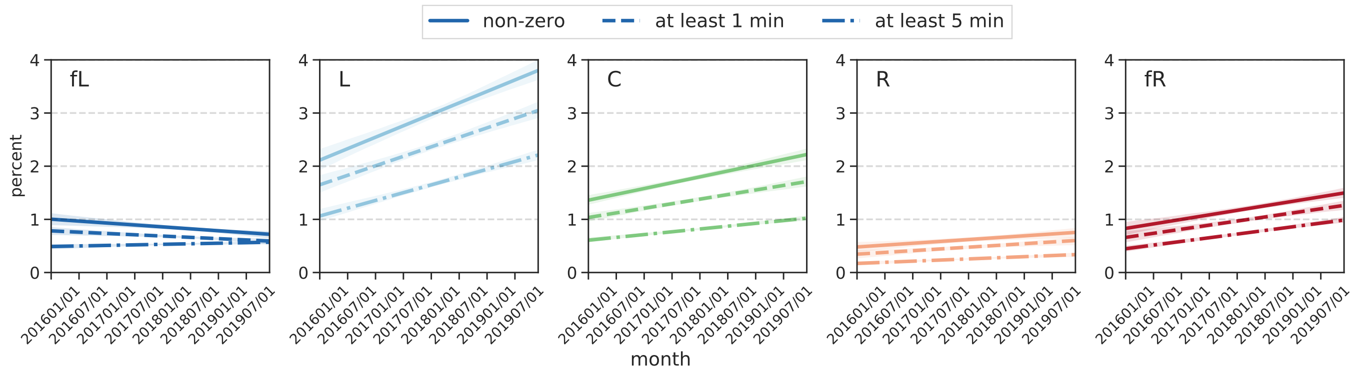

FIG. 18. Percent of users inside each community, from far left to far right, with (i) non zero consumption, (ii) at least one minutes consumption, (iii) at least five minutes consumption per month, from any of the 997 channels in our labeled set.

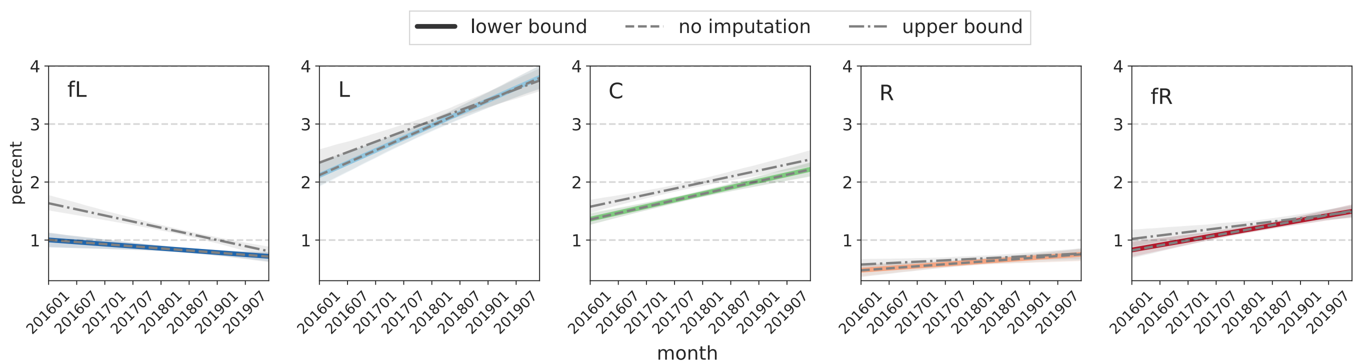

FIG. 19. Percent of users inside each community, from far left to far right, with (i) no imputation, (ii) high precision imputation (lower bound), (iii) and high recall imputation (upper bound). Any user with at least one minute consumption from our set of 997 channels is considered as "news consumer".

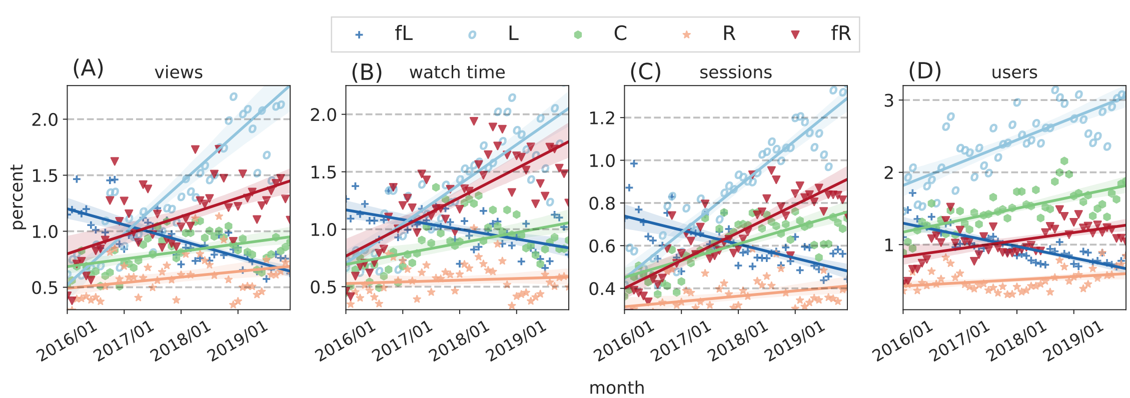

FIG. 20. Breakdown of percent of views (a), consumption (b), sessions (c) and users (d) falling into the five different channel categories, per month, from far left to far right, January 2016 to December 2019, with high recall imputation. Panel (a) presents the percentage of YouTube video pageviews, panel (b) is the percentage of the viewership duration from each channel category, panel (c) is the percent of sessions with at least one video from each bucket, and panel (d) is the percent of users falling into each community.

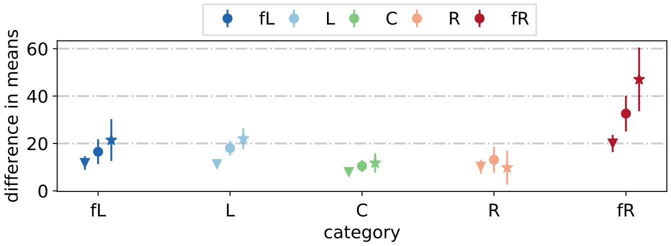

FIG. 21. Difference in means of daily consumption change, in the event of bursty consumption from a specific political category. Individuals are assigned either to bursty consumption group in the event of watching at least M_v^k videos from category k ($k \in \{\mathrm{fL}, \mathrm{L}, \mathrm{C}, \mathrm{R}, \mathrm{fR}\}$) within a session, or in control group, if none of their sessions has more than one video from same category with at least M_v^k videos in their lifetime. We run three experiments with different values of M_v^k, where $\blacktriangledown$: $M_v^k = 2$, $\bullet$: $M_v^k = 3$, $\bigstar$: $M_v^k = 4$. Markers show the difference in means, and the vertical lines present the 95% confidence interval.

Appendix I: Label aggregation

We consider a majority voting between labels provided by Refs. [20], [24], and [23]. If two labels are available, we consider the label toward more extreme end, for example a video with labels right and Center-right will be in the right bucket. If only one label is available, we assign it for that channel directly. Ref. [23] provide high-level label of left/center/right and and a more granular label of channels' ideology. Ref. [24] has video-level labels, only for "Conspiracy" category. Ref. [20] has channels labeled as "Alt-right", "Alt-lite", "Intellectual Dark Web" or "IDW", mostly inspired by Ref. [48].

- far right: any channel labeled as "Alt-lite", "Alt-right", "IDW", and "MRA" (Men's Right Activists) by Ref. [20], or ideology "Conspiracy", "IDW", "Alt-right", "MRA", "Alt-light", "Religious Conservative" by Ref. [23]. Also any channel with label "Right" from Ref. [20], and ideology labels "Conspiracy", "MRA", and "Alt-right" by Ref. [23]. Videos with label Conspiracy from Ref. [24] is labeled as far right video as well.
- right: any channel with labels "right" or "right-center" by Ref. [20], and ideology "Partisan Right" and "right", or ideology "Libertarian" and "right" by Ref. [23].
- center: any channel with label center from both Refs. [20] and [23].
- left: any channel with ideology "Partisan Left", and "left" from Ref. [23], and label "left" or "left-center" from Ref. [20].
- far left: any channel with ideology "Social Justice", "Socialist", "Revolutionary Socialist", "Anti-Theist", and "Anti-white" from Ref. [20].

TABLE XIII. List of channel titles within each channel category.

channel category	channel title
far left (-2)	Xexizy, Thom Avella, AM to DM, Empire Files, ProPublica, The Daily Beast, i−D, Jack Saint, Kristi Winters, Renegade Cut, The Fowler Show, TheThinkingAtheist, Acronym TV, Chomsky's Philosophy, The Point with Ana Kasparian, Socialism Or Barbarism!, Libertarian Socialist Rants, Mexie, Hella Opinions from BuzzFeed News, ATTN:, Tom Nicholas, MEANS TV, Patriot Act, The Atheist Experience, GETchan, Godless Cranium, Hasan Piker, Step Back History, Ezra Klein Show, José, Roland S. Martin, Telltale, Rebel Diaz, Jay Smooth, C0ct0pusPrime, Alternative Facts, azureScapegoat, Randy Rainbow, Sarah Z, ThinkTank, Innuendo Studios, Jason Unruhe, Riley J. Dennis, AronRa, The Damage Report, Cult of Dusty, PoliDice, Chapo Trap House, BadMouseProductions, Some More News, Red Star Videos, TYT's The Conversation, TMM, anarchopac, Creationist Cat, Big Joel, Ana Kasparian, TeleSUR English, Kat Blaque, Claudia Brown, EssenceOfThought, chescaleigh, Mic, Gutian, BuzzFeed News, donoteat01, African Diaspora News Channel, Kevin Logan, Hugo and Jake, Democracy At Work, ContraPoints, Mad Blender, LastWeekTonight, Mother Jones, Rachel Oates, The Young Turks, Rational Disconnect, marinashutup, DemocraticSocialist01, chrisiousity, Angie Speaks, Three Arrows, Rebecca Watson, Thought Slime, let's talk about stuff., Bat'ko, Faraday Speaks, Planned Parenthood, Philosophy Tube, The Serfs, TheeKatsMeoww, Socialist East, Shaun, TVO Docs, hbomberguy, Broadly, NonCompete, Alexandria Ocasio-Cortez, potholer54, ReyRoSho, DarkMatter2525, Cuck Philosophy, VICE, The Majority Report w/ Sam Seder, Milo Stewart, YUGOPNIK, NowThis Her, NowThis World, Mia Mulder, Jeff Holiday, SoulPancake, The Humanist Report, The Michael Brooks Show, Patricia Taxxon, Vaush, The Laura Flanders Show, TYT Nation, NowThis News, VICELAND, RE-EDUCATION, Then & Now, TheFinnishBolshevik, TBA, Hakim, Russell Brand, Quartz, OZY, ACLU, Messy Elliott, Zero Books, lacigreen, AJ+, Peter Coffin, The Intercept, RichardDWolff, truTV, Steve Shives
left (-1)	The Rational National , VICE News , Jesse Dollemore , Maggie Mae Fish , Status Coup , Drunken Peasants , thejuicemedia , Democracy Now! , David Pakman Show , The View , huffington post (huffpost) , vanity fair , Bill Press Video , the verge , cosmopolitan , cbc news , elite daily , lifehacker , Vic Berger , The Jim Jefferies Show , Free Speech TV , slate , Travis &, The New York Times , The Guardian , scoopwhoop , Destiny , Kim Iversen , The Late Show with Stephen Colbert , uproxx , engadget , techcrunch , H. A. Goodman , The New Yorker , Late Night with Seth Meyers , The Grayzone , sky news , Moderate Rebels , yahoo news , The Atlantic , The Jimmy Dore Show , HuffPost , Tony Dortie - 24/7 Eyes , the guardian , hollywood reporter , feminist frequency , Rebel HQ , Jamarl Thomas , san francisco globe , global citizen , Secular Talk , M. Tracey , Full Frontal with Samantha Bee , Lindsay Ellis , Andrew Yang for President 2020 , upworthy , Trevor Noah , Los Angeles Times , Thom Hartmann Program , democracy now , NBC News , Stephanie Miller Show , washington post , CNN , glamour magazine , people magazine , The Nation , Real Time with Bill Maher , The Zero Hour with RJ Eskow , MSNBC , Vox , mashable , (the)atlantic , TLDR News US , rolling stone , merry jane , Washington Post , Newsweek , wired magazine , global news' Jonathan
center (0)	neo , The Agenda with Steve Paikin , consumer reports , King Crocoduck , Bro. Sanchez TV , The Right Opinion , Foundation for Individual Rights in Education , forbes , ABC News , Zarathustra's Serpent, Paulogia , investopedia , goingundergroundRT , Moe Othman , FADE TO BLACK Radio , Rekieta Law , Al Jazeera English , Martymer 81 , RT Documentary , Skavlan , Skeptic , POLITICO , business insider , VeryNearlyViral , Holy Koolaid , Truthstream Media , USA TODAY , Edge of Wonder , Rose of Dawn , South Front , the indian express , Anonymous Official , Steve McRae , CBS News , Earthfiles , PSA Sitch , Jay Myers Documentaries , Wizard Of Cause , TheRabbitHole , Financial Times , PBS NewsHour , Matt Dillahunty , c-span , VisualPolitik EN , friendlyjordies , Corbett Report Extras , Ian R Crane , Charisma on Command , Uncovered , big think , RT America , John Cedars , The Hill , the economist , Genetically Modified Skeptic , TheLivingDinosaur , Chicago Tribune , Maddox , The Big Picture RT , Justin Murphy , In the NOW , Cheddar , The Economist , Associated Press , Philip DeFranco , TED , PBS , Montagraph , Business Insider , Logicked , TheHighersideChats , COAST TO COAST AM OFFICIAL , Kraut , Bunty King , Fortune Magazine , Mouthy Buddha , world economic forum , The Fifth Estate , AFP news agency , Eric Weinstein , financial times , thecrowhouse , Larry King , vocativ , Noel Plum , harvard business review , Rationality Rules , misterdeity , Florida Maquis , today i found out , Think Club , Counter Arguments , C-SPAN , Masaman , Wall Street Journal , Liana K , Autumn Asphodel , Godless Engineer , Joe Rogan University - Fan Channel , WorldsApaRT , Forbes , VOA News , Nuance Bro , TheraminTrees , Media Monarchy , David Wilcock — Divine Cosmos (OFFICIAL) , Your Black World , DarkAntics , relevant magazine , IntelligenceSquared Debates , military.com , SphereBeing Alliance , The Christian Science Monitor , Viced Rhino , Boom Bust , Gary Edwards , Vernaculis , Sputnik , TEDx Talks , recode , mental floss , Gaia , Kalvin Garrah , Ruptly , TheWishmaster , Watching the Hawks RT , A Voice For Men , ADAM FRIENDED , CosmicSkeptic , Jubilee , makeuseof , a16z , The Moore Show , 48 Hours , Big Think

TABLE XIV. List of channel titles within each channel category.

channel category	channel title
right (1)	National Review, gulf news, The Cato Institute, The Liberty Hound, The Common Sense Conservative, Property and Freedom Society, Carey Wedler, ron paul liberty report, American Thought Leaders - The Epoch Times, One America News Network, new york post, Right Wing News!, FOX 10 Phoenix, Anthony Brian Logan, Fox News Insider, Black Conservative Patriot, Judicial Watch, LibertyPen, Right Side Broadcasting Network, The Josh Bernstein Show, Larry Elder Show - The Epoch Times, Sara Gonzales Unfiltered, FreedomToons, Adam Carolla, Fox News, ntd.tv (new tang dynasty), AdamKokesh, judicial watch, GoldSilver (w/ Mike Maloney), Peter Schiff, Gotcha News Network, GOP War Room, The Pholosopher, Mercatus Center, russia insider, Graham Allen, John Stossel, TomWoodsTV, SenatorRandPaul, The Justicar, Conservative Citizen, Epoch News The Epoch Times, Diamond and Silk - The Viewers View, Learn Liberty, Shane Killian, learn liberty, pj media, kytekuttertv, RonPaulLibertyReport, SyrianGirlpartisan, New York Post, OpenMind, Chad Prather, Real Coffee with Scott Adams, national rifle association (nra), Project Veritas, project veritas, Parse The Noise, DECLASSIFIED - The Epoch Times, HIGH ENERGY, That Guy T, HooverInstitution, Gun Owners of America, The Asian Capitalists, Project Veritas Action, Washington Free Beacon, american enterprise institute, Jon Miller, Townhall Media, Conservative Tribune, American Enterprise Institute, NumbersUSA, Newsmax TV, Golden State Times, InstituteForJustice, The Daily Signal, The Federalist Society, Bill Whittle
far right (2)	Kronos, Hannah Wallen, RichieFromBoston, American Pride 2, Irate Bear, Styxhexenhammer666, MR. OBVIOUS, Enlightainment, SkyWatch TV, Stefan Molyneux, Jesse Lee Peterson, Alt Right, SHUTTERSHOT45, Carpe Donktum, Mauritian Struggle, Independent Man, Voice Liberty, Gravitahn, Tomorrow's World Viewpoint, RockingMrE, Dangerfield, 1791, Ben Shapiro, OneTruth4Life, X22Report Spotlight, Nick Fuentes Clips, Destroying the Illusion, And We Know, Airliner World & More, ABitOfBritt, Aydin Paladin, xUnlimitedMagz, Tonka Radio, Awakened Saxon, The Iconoclast, Woz Lee, Shmufkin TV, TheIncredibleSaltMine, ZIEeICoZ, InvincibleNumanist, Fully Sourced, No Bullshit, HowTheWorldWorks, Truth Against The World, eliharman, Englander, Dose of Truth, Halsey News Network, theovonk, Jacob Wohl, Kenn Daily, VikNand, karen straughan, In Pursuit of Truth, Liberty us, Bloggingheads.tv, Black Pigeon Speaks, Borzoi Boskovic, ataxin, JFG Livestreams, AmRenVideos, The Revolutionary Conservative, Scott Adams, Warriors for Christ, David Heavener, Bre Faucheux, Domination Station, LaughingMan0X, Julie Borowski, IRmep Stream, Oof Curator, DAHBOO77, BraveTheWorld, On The Offensive, Censored Movie, Galactic Bubble Productions, Douglas Murray Archive, Way of the World, joerogandotnet, Sandre The Teacher, 50 Shades, theturningpointusa, Dismantle The Matrix, Harmful Opinions, ShortFatOtaku, Bull Brand, Ayn Rand Institute, StevenCrowder, America Uncovered, The Lion, The Heartland Institute, The New Criterion, Mike Nayna, The Pondering Primate, grapjas60, Omegon, Akkad Daily, Asha Logos, PhilosophiCat, Diapers Off, Paul Joseph Watson, Bite-sized Philosophy, Lift the Veil, Breitbart News, Ami Horowitz, Mark Collett, JoyCamp, AltRight Truth, Rebel Media, Rebel News, MILO, Mike Cernovich, Jake 333, Millennial Woes, Faith J Goldy, CRITICAL CONDITION, Felfop Returns, Social Justice Fails, SugarTits, TRS Radio, Brave New World, Actual Justice Warrior, Tommy Robinson, Morrakiu, Jason Goodman, Canadian Association for Equality, Undoomed, StudioBrule, Sinatra-Says, Jason Kessler, Johnny Monoxide, Serp Kerp, People's Veto, Endeavour, Collective Evolution, The James Delingpole Channel, Jean-Francois Gariépy, hOrnsticles3, Vincent James, ArktosOnline, Charlemagne, Corpus Mentis, Liberty Hangout, The Leftovers, Yaron Brook, KristalDGarcia, Coop's Channel, Pat Condell, Cato's Speech, Mister Metokur, The Hateful Gaels, Lindsay Shepherd, Black Pilled, Tragedy and Hope, ybrook, Dissident View, DoctorRandomercam, Leo Stratton, Daily Caller, Toad McKinley, libertydollshouse, Matt Forney, Red Ice TV, Shoe & Skeptic, augustussolinvictus, bmdavll, Red Pill Awareness, Literature Devil, RB-95, Nick Di Paolo, The Golden One, Comics MATTER w/Ya Boi Zack, ThatUmbrellaGuy, sanderson1611, Lionel Nation, Luke Ford, Transliminal, Jordan Peterson Fan Channel, TheArchangel911, Mark Postgate, Libertatem, HydeWars, PragerU, Barbarsossa Rants, Allie Beth Stuckey, Make Cringe Great Again, The Fallen State, ToolTime, Nightmare Fuel, Gavin McInnes, CoalitionForMen, The Great Dolemite, Glenn Beck, RobinHoodUKIP, Andy Warski, News 4 A Change, AIU-Resurrection, Laura Towler, Nicholson1968, Beacom Of Light, tmcleanful, The Rageaholic, Phil Robertson, WisdomTalks, soph, Sanity 4 Sweden, DailyCallerVideo, Baked Alaska 2, CounterCurrentsTV, Reverend Simon Sideways, Raging Golden Eagle, The Andrew Klavan Show, YAFTV, Thescariestmovieever, TylerPreston20, TruNews, ramzpaul, HighImpactVlogs, Armoured Gregory, RFB II, Brother Nathanael, Brittany Sellner, The Daily Truth, 1menaregood1, BlazeTV, Brainlet, FatherlessAmerica, Shoe0nHead, AgatanFoundation, TheQuartering, Essential Truth, TheJohnBirchSociety, MaKaElectric, AmRen Podcasts, Stand Up Europe, DTRHRadioArchives, Traditionalist Youth Network, thkelly67, Alt-Right Tankie

TABLE XV. List of channel titles within each channel category (cont.).

channel category	channel title
far right (2)	Marie Cachet, TRUreporting, Luke Ford Livestreams, Dame Pesos, Twoon, OnlineWipe, Gad Saad, PowerfulJRE, WeAreChange, Conversations with Bill Kristol, Blackstone Intelligence Network, Spy Kitten TV, AllNationsParty, NPI / Radix, Andywarski, Darkstream, TokenLibertarianGirl, X22Report, Rebel Wisdom, Badger Live Streams, Cognitive Thought, Nephtali1981, JerryTheother, Notes For Space Cadets, SomeBlackGuy, No White Guilt, The Daily Wire, TheSchillingShow, "ItsAGundam", Patrick Slattery, BASSFZz, SkidRowRadio, Motte & Bailey, Atlas Society, Blonde in the Belly of the Beast, MillennialWoes, Computing Forever, The Progressive Voice, Timcast, Rebel Edge, Informative, Praise of Folly, Danny 1488, PragerUniversity, Sargon of Akkad, The Academic Agent, Ryan Dawson, Armoured Media, ProductiehuisEU, Baked Alaska, Oppressed Media, The Red Elephants Vincent James, Crysta, successcouncil, JustInformed Talk, HighImpactFlix, Gary Orsum, ReasonTV, TheAtlasSociety, Qronos16, Ben Shapiro Thug Life, jackburton2009, Revenge Of The Cis, CISAus, Pangburn, Sorting Myself Out, AltRight.com, Jreg, Connors Conquest, No Bullshit 2, battleofideas, Shaking My Head Productions, American Joe Show, black chad, SargonofAkkad100, Clash of Ideas, James Allsup, Barbara4u2c, The Glass Blind Spot, Compound Media, The Distributist, A1Cvenom, Jay Dyer, carnellspeaks, Maximilien Robespierre, Katy Khaos, Kelly Day, Jamie Dlux, Jan Kerkoff, Washington Watch, 14 Sacred Words, Patriots Soapbox News Network LIVE 24/7, The Heritage Foundation, Lauren Southern, Operation Freedom, SUIT YouRSELF, Brittany Venti, Fleccas Talks, bloggingheads.tv, Owen Benjamin, LibertarianRealist2, Liberty Machine News, Tim Pool, Hi, I Think Im Real, Semiogogue, Atheist Foundation of Australia Inc, MarkDice, Dont Walk, Run! Productions, Nationalist Media Network, girlwriteswhat, ThePatriotNurse, MillionDollarExtreme2, Vee, Rollo Tomassi, Repzion, JordanPetersonVideos, We the Internet TV, The Propertarian Institute, Hunter Avallone, Mark Dice, Desi-Rae Thinking, America First with Nicholas J Fuentes, Fritz Imperial, Nacionalista Blanco del SoCal, Matt Christiansen, AynRandInstitute, MeaningofLife.tv, Bret Weinstein, Oswald Spengler, THATS THE POINT with Brandon, The Dick Show, Staying Woke, AustralianNeoCon1, Lisa Haven, Liberty Doll NewEuropeANP, jaydyer, 6oodfella, Atheism-is-Unstoppable-2, Matt Jarbo, MG, MBest11x, uzalu, nightmarefuel, Family Policy Institute of Washington, Missing The Mark, Dr. Edward Dutton The Jolly Heretic, John Ward, Traditionalist Worker Party, Augustus Invictus for United States Senate, This is Europa, Candace Owens, Jericho Green, DronetekPolitics, Jordan B Peterson, American Intelligence Media, Georgia Free, AustralianRealist, AVALLONEextra, Laura Loomer, Owen Benjamin Clips, Matthew North, Be Open MInded, TheBritisher, Michelle Malkin, Agatan Foundation, Attila Vinczer, Chris Ray Gun, Jean-François Gariépy, Messages for Men, I,Hypocrite, American Renaissance, Sargon of Akkad Live, JRE Clips, Christopher Anderson, MW Live, Sam Harris, Whitney Dahlin, Tommy Cs SFTP, The Dave Cullen Show, The Rubin Report, corbettreport, The Weekly Sweat, reallygraceful, Michael Knowles, Wild Smile, Real McGoy, Freedom Speaks, WhiteRabbitRadioTV, Tree Of Logic, The Amazing Lucas, The Snark Tank, Brother Nathanael Video, Davie Addison, prayingmedic, OfficialGATG, Bearing, Tipping Point With Liz Wheeler on OAN, The Alternative Hypothesis, Revcon Media, MrAndsn, The Free Speech Club, Armoured Skeptic, The Last Stand, ManOfAllCreation, Alt Hype Streams, The Unplugged Observer, Savage Facts, MichelleRempel, JaclynGlenn, An Ear for Men, Prince of Zimbabwe, Diana Davison, Feminism Sucks, Red Pill Philosophy, Rebel Canada, SJWCentral, Demirep, Nightwave Radio, TheArmenianNation, CoolHardLogic, Count Dankula, deanesmay, Western Man, Anything Goes, Zach Hing, Steve Trueblue, 1791, Turning Point USA, Metokurist Murdochian Archiver, TheBechtloff, ZeroFox Given, patcondell, Peach Balie, Benjamin A Boyce, wikileaksplus, Count Dankula Streams, The Rational Rise, Rekt Idiots, The Thinkery, Ayla Stewart Wife With A Purpose, Tea Clips, BigCatKayla Livestreams, MidWestly, Vincent James of The Red Elephants, American Pride, Rick & Bubba, VertigoPolitix, NPI / RADIX, astrokidnj, NateTalksToYou, andy nowicki, SGTreport, Jordan B Peterson Clips, DailyKenn, Brittany Pettibone, Harrison Hill Smith, Dr. Steve Turley, George Webb, PhilosophyInsights, American Justice, Very Fake News, Proud Conservative, TheDC Shorts, Hard Bastard, Girl Defined, bane666au, HoneyBadgerRadio, ChangeDaChannel, Patrick, Cassiejaye, Lauren Chen, RedPill78, brianoflondon, "Dinesh DSouza", The Reality Calls Show, Mike Buchanan, SJW CRINGE MACHINE, Little Revolution, Blaire White, Trigger Happy Media, Murdoch Murdoch, Slightly Offens*ve, Voxiversity, E;R, TL;DR, Charles Zeiger, The Next News Network, NeatoBurrito Productions, Pragmatic Entertainment, Daisy Cousens, UNITE AMERICA FIRST, ThuleanPerspective, SkepTorr, iambakedalaska, drwarrenfarrell