Examining the consumption of radical content on YouTube

Homa Hosseinmardi[a,1], Amir Ghasemian[b,c], Aaron Clauset[d,e,f], Markus Mobius[g], David M. Rothschild[h], and Duncan J. Watts[a,i,j,1]

[a]Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA 19104; [b]Department of Statistics, Harvard University, Cambridge, MA 02138; [c]Department of Statistical Science, Fox School of Business, Temple University, Philadelphia, PA 19122; [d]Department of Computer Science, University of Colorado Boulder, Boulder, CO 80309; [e]BioFrontiers Institute, University of Colorado Boulder, Boulder, CO 80303; [f]Santa Fe Institute, Santa Fe, NM 87501; [g]Microsoft Research New England, Cambridge, MA 02142; [h]Microsoft Research New York, New York, NY 10012; [i]The Annenberg School of Communication, University of Pennsylvania, Philadelphia, PA 19104; and [j]Operations, Information, and Decisions Department, University of Pennsylvania, Philadelphia, PA 19104

Although it is under-studied relative to other social media platforms, YouTube is arguably the largest and most engaging online media consumption platform in the world. Recently, YouTube's scale has fueled concerns that YouTube users are being radicalized via a combination of biased recommendations and ostensibly apolitical "anti-woke" channels, both of which have been claimed to direct attention to radical political content. Here we test this hypothesis using a representative panel of more than 300,000 Americans and their individual-level browsing behavior, on and off YouTube, from January 2016 through December 2019. Using a labeled set of political news channels, we find that news consumption on YouTube is dominated by mainstream and largely centrist sources. Consumers of far-right content, while more engaged than average, represent a small and stable percentage of news consumers. However, consumption of "anti-woke" content, defined in terms of its opposition to progressive intellectual and political agendas, grew steadily in popularity and is correlated with consumption of far-right content off-platform. We find no evidence that engagement with far-right content is caused by YouTube recommendations systematically, nor do we find clear evidence that anti-woke channels serve as a gateway to the far right. Rather, consumption of political content on YouTube appears to reflect individual preferences that extend across the web as a whole.

political radicalization | news diet | user behavior | online platforms

As affective political polarization rises in the United States (1) and trust in traditional sources of authority declines (2, 3), concerns have arisen regarding the presence, prevalence, and impact of false, hyperpartisan, or conspiratorial content on social media platforms. Most research on the potentially polarizing and misleading effects of social media has focused on Facebook and Twitter (4–12), reflecting a common view that these platforms are the most "news-oriented" social media platforms. However, roughly 23 million Americans rely on YouTube as a source of news (13, 14), a number comparable to the corresponding Twitter audience (13, 15), and it is growing in both size and engagement. YouTube news content spans the political spectrum, and includes content producers of all sizes. Recent work (16) has identified a large number of YouTube channels, mostly operated by individuals or small organizations, that promote a collection of "far-right" ideologies (e.g., white identitarian) and conspiracy theories (e.g., QAnon). The popularity of some of these channels, along with salient popular anecdotes, has prompted claims that YouTube's recommendation engine systematically drives users to this content, and effectively radicalizes its users (17–20). For example, it has been reported that, starting from factual videos about the flu vaccine, the recommender system can lead users to antivaccination conspiracy videos (18).

Recent qualitative work (21) has identified a separate collection of channels labeled variously as "reactionary," "anti-woke" (AW), "anti-social justice warriors" (ASJW), "intellectual dark web" (IDW), or simply "antiestablishment." Although these channels do not identify themselves as politically conservative, and often position themselves as nonideological or even liberal "free thinkers," in practice, their positions are largely defined in opposition to progressive social justice movements, especially those concerning identity and race, as well as critiquing institutions such as academia and mainstream media for their "left-wing bias" (21, 22). Concurrently, "anti-woke" rhetoric has increasingly been adopted by mainstream Republican politicians (23), undermining claims that it is intrinsically apolitical. While anti-woke YouTube channels typically do not explicitly endorse far-right ideologies, some channel owners invite guests who are affiliated with the far right onto their shows and allow them to air their views relatively unchallenged, thereby effectively broadcasting and legitimizing far-right ideologies (21). If these channels act as a kind of gateway to the far right, they would constitute a related yet distinct radicalization mechanism from the recommendation system per se (17, 24). Based on these considerations, and recognizing that any label for this loose collection of channels is

Significance

Daily share of news consumption on YouTube, a social media platform with more than 2 billion monthly users, has increased in the last few years. Constructing a large dataset of users' trajectories across the full political spectrum during 2016–2019, we identify several distinct communities of news consumers, including "far-right" and "anti-woke." Far right is small and not increasing in size over the observation period, while anti-woke is growing, and both grow in consumption per user. We find little evidence that the YouTube recommendation algorithm is driving attention to this content. Our results indicate that trends in video-based political news consumption are determined by a complicated combination of user preferences, platform features, and the supply-and-demand dynamics of the broader web.

likely to be inaccurate for at least some members, we refer to them hereafter as anti-woke (AW).

Although reports of various mechanisms driving people to politically radical content have received great attention, quantitative evidence to support them has proven elusive. On a platform with almost 2 billion users (25), it is possible to find examples of almost any type of behavior; hence anecdotes of radicalized individuals (17), however vivid, do not, on their own, indicate systematic problems. Thus, the observation that a particular mechanism (e.g., recommendation systems steering users to extreme content; far-right personalities appearing on anti-woke channels acting as gateways to the far right) might plausibly have a large and measurable effect on audiences does not substitute for measuring the effect. Finally, the few empirical studies (24, 26–29) that have examined the question of YouTube radicalization have reached conflicting conclusions, with some finding evidence for it (24, 26) and others finding the opposite (27, 28). These disagreements may arise from methodological differences that make results difficult to fairly compare—for example, ref. 28 examines potential biases in the recommender by simulating logged-out users, whereas ref. 24 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 far-right content is, in fact, being consumed (vs. produced), how it is changing over time, and to what extent it is being driven by YouTube's own recommendations, spillovers from anti-woke channels, or other entry points.

Here we investigate the consumption of radical political news content on YouTube using a unique dataset 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 4 years from January 2016 to December 2019. To summarize, we present five main findings. 1) Consistent with previous estimates (30), we find that the total consumption of any news-related content on YouTube accounts for 11% of overall consumption and is dominated by mainstream, and generally centrist or left-leaning, sources. 2) The consumption of far-right content is small in terms of both number of viewers and total watch time, where the former decreased slightly and the latter increased slightly over the observation period. 3) In contrast, the consumption of anti-woke content, while also small relative to mainstream or left-leaning content, grew in both numbers of users and total watch time. 4) The pathways by which users reach far-right videos are diverse, and only a fraction can plausibly be attributed to platform recommendations. Within sessions of consecutive video viewership, we find no trend toward more extreme content, either left or right, indicating that consumption of this content is determined more by user preferences than by recommendation. 5) Consumers of anti-woke, right, and far-right content also consume a meaningful amount of far-right content elsewhere online, indicating that, rather than the platform (either the recommendation engine or consumption of anti-woke content) pushing them toward far-right content, it is a complement to their larger news diet.

These results indicate little evidence for the popular claim that YouTube drives users to consume more radical political content, either left or right. Instead, we find strong evidence that, while somewhat unique with its growing and dedicated anti-woke channels, YouTube should otherwise be viewed as part of a larger information ecosystem in which conspiracy theories, misinformation, and hyperpartisan content are widely available, easily discovered, and actively sought out (27, 31).

Methods and Materials

Our data are drawn from Nielsen's nationally representative desktop web panel, spanning January 2016 through December 2019 (*SI 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 1). We quantify the user's attention by the duration of in-focus visit to each video in total minutes (32). 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 its URL, yielding 9,863,964 unique video IDs (*SI Appendix*, section B). To post a video on YouTube, a user must create a channel with a unique name and channel 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.

Video Labeling. Previous studies (21, 24, 27–29) have devoted considerable effort to labeling YouTube channels and videos based on their political content. In order to maintain consistency with the existing research literature, we derived our labels from two of these previous studies which collectively classified over 1,100 YouTube channels. First, ref. 28 classified 816 channels along a traditional left/center/right ideological spectrum as well as a more granular categorization into 18 tags such as socialist, ASJW, religious conservative, white identitarian, and conspiracy. Second, ref. 24 classified 281 channels as belonging to one of IDW, alt-lite, or alt-right, and a set of 68 popular media channels as left, left-center, center, right-center, or right. Because the two sources used slightly different classification schemes, we mapped their labels to a single set of six categories: far left (fL), left (L), center (C), anti-woke (AW), right (R), and far right (fR) (see *SI Appendix*, section C and Table S1 for details). Five of the six categories (fL, L, C, R, and fR) fall along a conventional left–right ideological spectrum. For example, YouTube videos belonging to channels with ideological labels such as "socialist" are considered farther to the left of "left" content and hence were assigned to far left, whereas "alt-right" is a set of ideologies that exemplify extreme right content and so those videos were assigned to far right (24, 27). The "anti-woke" (AW) category, which mostly comprises the labels IDW (24) and ASJW (28), but also a small set of channels labeled "men's rights activists" (MRA) (24, 28), is more easily defined by what it opposes—namely, progressive social justice movements and mainstream left-leaning institutions—than by what coherent political ideology it supports (21). Reflecting its nontraditional composition, we locate it to the right of center but left of right and far right. We note that the ordering of the categories from left to right, while helpful for visualizing results in some cases, is not important to any of main findings. Overall, our data cover 974 channels, corresponding to 523,242 videos (following refs. 24 and 28, all videos published by a channel under study received the channel's label). Details on the number of videos in each category, the assignment of channels, and references used can be found in *SI Appendix*, Tables S1 and S2 and section C.

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 (33). 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 "subreason" for the unavailability. We crawled a uniformly random set of 368,754 videos and extracted these subreasons 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 (*SI Appendix*, Table S3). For channels such as InfoWars, which was terminated for violating YouTube's Community Guidelines (34), none of their previously uploaded videos are available through the YouTube API. Therefore, it is important

Table 1. YouTube data descriptive statistics

Characteristic	Value
Number of unique users	309,813
Number of watched-video pageviews	21,385,962
Number of unique video IDs	9,863,964
Number of unique channel IDs	2,293,760
Number of sessions	8,620,394

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 a model, we first searched for the overlap between our set of unavailable videos and datasets from previous studies (24, 27–29, 33), which had collected the metadata of many now unavailable videos at a time when they still existed on the platform (*SI Appendix*, Table S4). This approach yielded channel IDs for 69,611 of our unavailable videos. We then trained a series of classifiers, which we used to impute 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 by users 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 *SI Appendix*, section D.

For each political channel category, we trained a binary random forest classifier over 96 predictors, which yielded an AUC (Area Under the Curve) of 0.95 for far left, 0.98 for left, 0.95 for center, and 0.97 for anti-woke, right, and far right, on the holdout set. To assign labels, we consider two different thresholds, one with high precision and one with high recall (*SI Appendix*, Tables S5 and S6). For all presented results in the main text, we use a high-precision imputation model, which means fewer videos (with higher confidence) are retrieved from the positive class. 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 *SI Appendix*, section I).

Constructing Sessions. Previous studies have analyzed web browsing dynamics by breaking a sequence of pageviews into subsequences called sessions (35). In this work, we define a YouTube session as a set of near-consecutive YouTube pageviews by a user. Within a YouTube session, a gap less than δ minutes is allowed between a YouTube nonvideo 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 $\delta = 10$ minutes 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 *SI Appendix*, section I).

User Clustering. An individual is considered a news consumer if, over the course of 1 month, they spend a minimum of 1 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 jth entry, ν_{ij}^m, corresponds to the fraction of viewership of user i from channel category j ($j \in \{fL, L, C, AW, R, fR\}$). We then used hierarchical clustering to assign each individual to one of K communities of similar YouTube news diets, with K in the range from 2 to 12. Running 19 different measures of model fit to find the optimal number of communities, six and five had equally the highest number of votes (36), where we set the number of communities to be $K = 6$ to capture all categories. For each of these six clusters, we then identified its centroid, obtained by averaging the normalized monthly viewership vectors of all cluster members (see *SI Appendix*, section E for details). Finally, we labeled each community as $\psi(t)$ ($\psi(t) \in \{fL, L, C, AW, R, fR\}$) according to the predominant content category of its centroid. As a robustness check, we performed similar analysis with more relaxed and more strict definitions of "news consumers" (see *SI Appendix*, section I).

Results

Before we examine the degree to which YouTube's recommendation engine drives users to more extreme forms of political content, we first present a series of analyses to characterize and quantify the overall consumption patterns among different types of content. These patterns allow us to test for several confounding possibilities in the dynamics of news consumption on YouTube, and provide a clear background picture against which to measure systematic deviations caused by recommendations.

These six clusters correspond closely with our six categories of political content (Fig. 1): The centroid or "archetype member" of each cluster devotes more than 90% of their attention to just one content category, with the remaining 10% distributed roughly evenly among the other categories. In fact, the consumption patterns of individual users in each of these clusters aligns strongly with their containing cluster, so much so that more than 70% receive at least 80% of their content from one content type, and more than 95% receive at least 50% of their content from one content type (*SI Appendix*, Fig. S3). This result has two important implications for our analysis. First, it reveals that YouTube users are meaningfully associated with definable "communities" in the sense that consumption preferences are relatively homogeneous within each community and relatively distinct between them. Throughout the remainder of the paper, we will use the terms "community" and "cluster" interchangeably, and will use "category" to refer to the corresponding population of videos. Second, our result demonstrates that the political content categories we used align closely with the actual behavior of users in a parsimonious way (see *SI Appendix*, Figs. S3 and S4 as a robustness check for more strict definitions of news consumers). We present our main results in terms of these communities.

Community Engagement. To check for any overall trends in category preferences, we examine changes in total consumption associated with each of the six communities over the 4-year period of our data, quantified in terms of both population size (Fig. 2*A*) and total time spent watching (Fig. 2*B*) (see *SI Appendix*, Fig. S5 for two other related metrics: pageview counts and session counts; see *SI Appendix*, Tables S8 and S9 for more details on fitted lines.). Consistent with previous work (30), we find that total news consumption accounts for only 11% of total consumption (see *SI Appendix*, section D for definition of "news"). Of this total, the 974 channels for which we have political labels account for roughly one-third (i.e., 3.32% of total watch time). Moreover, the largest community of news consumers—in terms of both population size (0.63%) and watch time (1.65%)—was the "left" mainstream community. Fig. 2*A* shows that the far-right community was the second smallest (after far-left) by population (0.05%) and declined slightly in size after a peak at the end of 2016. In contrast, the anti-woke category started roughly at the same size but grew considerably, overtaking right and almost matching center. Next, Fig. 2*B* shows that monthly watch time for both far-right and anti-woke news grew over our observation period, where the former increased almost twofold, from an average of 0.17% of total consumption in 2016 to 0.30% in 2019 ($t = -3.17$, $P < 0.005$), and the latter grew more rapidly, from an average of 0.31 to 1.02% ($t = -14.08$,
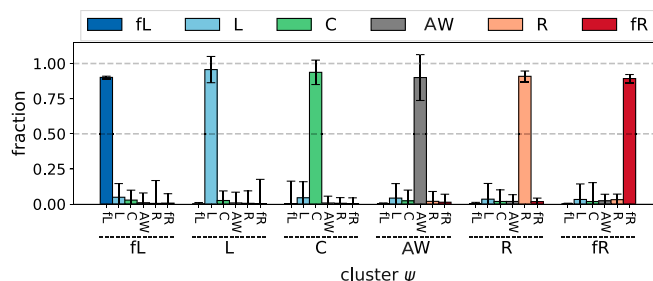

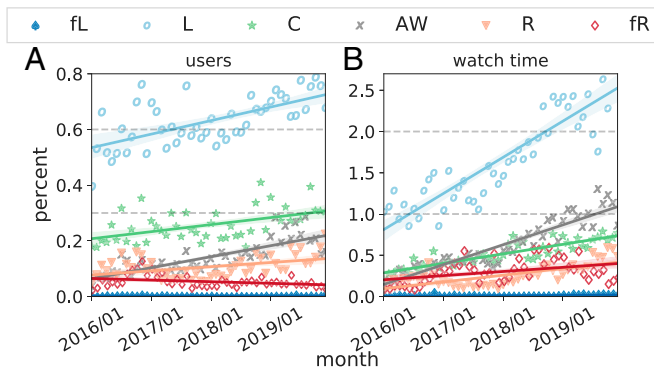**Fig. 1.** The archetypes of news consumption behavior on YouTube for each cluster.

Hosseinmardi et al.
Examining the consumption of radical content on YouTube

PNAS | 3 of 8
https://doi.org/10.1073/pnas.2101967118

Fig. 2. Breakdown of percent of (*A*) users and (*B*) consumption falling into the six political channel categories, per month, January 2016 to December 2019. *A* is the percent of users falling into each community, and *B* presents the percentage of viewership duration from each channel category. Solid lines show the fitted linear models and the shading shows the 95% confidence intervals.

$P < 10^{-5}$). In both cases, watch time grew faster than the overall rate of growth of news consumption (8% of average monthly consumption in 2016 to 11% in 2019 [$t = -13.13$, $P < 10^{-5}$]; *SI Appendix*, Fig. S2 and Table S7), and the anti-woke community ended the period accounting for more watch time than any category except left. We note that these data exhibit strong variations, and hence we caution against extrapolating these trends into the future. Because the far-left community is the smallest by population size (0.002%) and watch time (0.009%) (being barely visible on the scale of Fig. 2), we drop it from our results if the low observation count makes statistical inferences unreliable.

These results indicate that news, in general, and far-right content, in particular, account for only a small portion of consumption on YouTube. As emphasized previously, however, even a small percentage of users or consumption time can translate to large absolute numbers at the scale of YouTube. For example, averaged over the 4-year period, 335,209 Americans consumed far-right content and 764,405 consumed anti-woke content at least once in a given month, where the latter grew steadily in population and watch time. This result is also robust to other choices of consumption metric (e.g., pageviews or session counts), threshold for inclusion in the "news-consuming population," and imputation model (*SI Appendix*, Figs. S5, S13, S16, and 17). To better understand these dynamics, we now investigate individual-level behavior.

Individual Engagement. Complementing the aggregate (community-level) results in Fig. 2, Fig. 3 shows both absolute levels of, and changes over time in, consumption measured at the individual level of videos (Fig. 3*A*) and users (Fig. 3*B*), respectively (see *SI Appendix*, Tables S10 and S11 for more details on fitted lines). Fig. 3 presents evidence of stronger engagement with far-right and anti-woke content than for other categories (*SI Appendix*, Figs. S6–S8 and Tables S12 and S13). The median per-video watch time of far-right videos increased by 50%, from an average of 2 minutes in 2016 to 3 minutes in 2019 ($t = -3.43$, $P < 0.005$), while the per-video watch time of the anti-woke category roughly doubled ($t = -8.55$, $P < 10^{-5}$) over the same 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 and anti-woke communities is up to almost twice the engagement of users in the left, center, and right communities ($P < 10^{-4}$; for all pairwise comparisons of far right and anti-woke with left, center, and right). In other words, while the far-right and anti-woke communities remained relatively small

in size throughout the observation period (with only anti-woke growing), their user engagement grew to exceed that of every other category (see *SI Appendix*, Fig. S14 as a robustness check for more strict definitions of news consumers).

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 side of the political spectrum to the left than the reverse, while individuals in the center are more likely to move left than right. Also, there is more between-community movement from right and far-right than left and center to the anti-woke community (i.e., $P(\psi(t) = AW|\psi(t-1) = fR) = 0.06$, $P(\psi(t) = AW|\psi(t-1) = R) = 0.05$ vs. $P(\psi(t) = AW|\psi(t-1) = C) = 0.03$, $P(\psi(t) = AW|\psi(t-1) = L) = 0.02$ and $P(\psi(t) = AW|\psi(t-1) = fL) = 0.01$), indicating that the anti-woke community gains more audience from right wing than from center and left wing. Also, the most common transition to the far right is from the right, far left, and anti-woke (i.e., $P(\psi(t) = fR|\psi(t-1) = R) = 0.04 > P(\psi(t) = fR|\psi(t-1) = fL) = 0.03 > P(\psi(t) = fR|\psi(t-1) = AW) = 0.02 > P(\psi(t) = fR|\psi(t-1) = C) = 0.01 > P(\psi(t) = fR|\psi(t-1) = L) = 0.00$).

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 (37). To check for this possibility, we define a "burst" of exposure as consumption of at least $M_v^{(k)}$ videos of category k ($k \in \{L, C, R, AW, fR\}$) within a single session, and define a "treatment event" as the first instance in a user's lifetime when they are exposed to such a burst. We dropped the far-left category, as the number of samples were too small for this experiment.

For each content category, we consider three "treatment groups" comprising individuals who are exposed to burst lengths $M_v^{(k)} \in \{2, 3, 4\}$ (*SI Appendix*, Table S14) and compute the difference in their average daily consumption of the same content category k preexposure and postexposure. Finally, we compute the difference in difference between our treatment groups and a "control group" of individuals with maximum $M_v^{(k)} = 1$ video
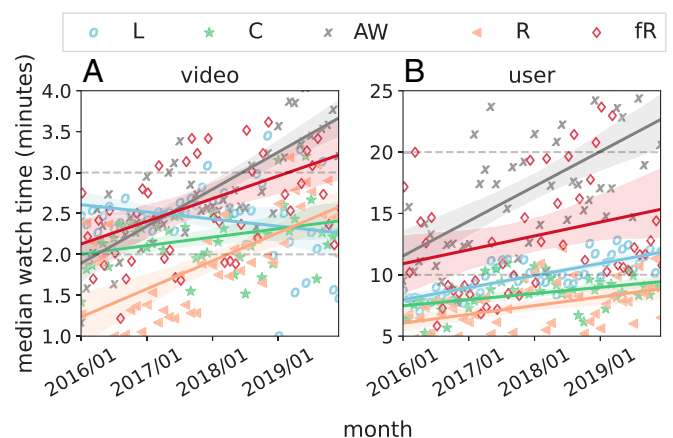

Fig. 3. (*A*) Median monthly video consumption (minutes) across different channel categories, and (*B*) median user consumption (minutes) within each community. Solid lines show the fitted linear models and the shading shows the 95% confidence intervals.
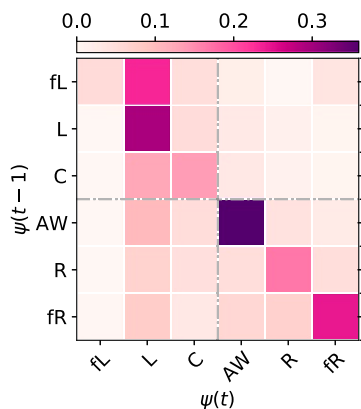
Fig. 4. A heatmap showing the probability that an individual from cluster $\psi(t-1)$ at month $t-1$ will move to cluster $\psi(t)$ at month t. Each month, users may not fall into any of these communities, if they are not among "news consumers" in that particular month.

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 *SI Appendix*, section F for details). We emphasize that these treatments are not randomized and could be endogenous (i.e., individuals are exposed to longer bursts because they already have a higher interest in the content); hence the effects we observe should not be interpreted as causal. As a purely predictive exercise, this analysis reveals 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 remaining content categories. In all cases, increases in burst length from 2 to 4 correspond to higher future consumption relative to the control groups. Individuals exposed to bursts of anti-woke 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 $M_v^{(fR)} = 4$ increases by a gap of almost 30 seconds postexposure, relative to the control group ($\tau = 30.99 \pm 19.50$ with CI = 0.95).

Potential Causes of Radicalization. Summarizing thus far, consumption of far-right and anti-woke content on YouTube—while small relative to politically moderate and nonpolitical content—is stickier and more engaging than other content categories, and, in the case of anti-woke, is increasingly popular. Previous authors have argued that the rise of radical content on YouTube is somehow driven by the platform itself, in particular by its built-in recommendation engine (17, 18). While this hypothesis is plausible, other explanations are too. As large as YouTube is, it is just a part of an even larger information ecosystem that includes the entire web, along with TV and radio. Thus, the growing engagement with 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 and boundaries.

In order to disambiguate between these explanations, we performed three additional analyses. 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 political 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 political 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 1) higher on-platform consumption of radical political content than off-platform, 2) arrival at radical political content dominated by immediately previous video views (thereby implicating the recommender), and 3) increasing frequency of radical political content toward the end of a session, especially a long session. By contrast, the strongest evidence for outside influences would be 1) high correlation between on- and off-platform tastes, 2) arrival dominated by referral from outside websites or search, and 3) 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 six previously identified communities with their consumption of non-YouTube websites classified according to the far-left to far-right web 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 of the web domains accessed by individuals in our dataset, $3,362$ were in the news category. We then used the partisan audience bias score provided by ref. 38 to bucket the news domains into the five political labels—fL, L, C, R, fR—that we used for YouTube channels above (see *SI Appendix*, section C for more details). Although some anti-woke YouTube channel owners may also host external websites, there is no equivalent of the anti-woke category in ref. 38; thus, we do not include it as a category for external websites.

To examine how the consumption pattern of users inside YouTube communities is associated with their web content consumption, Fig. 6 shows the risk ratio $RR_{(i,j)} = P_{ij}/P_j$ for each YouTube community $i \in \{fL, L, C, AW, R, fR\}$, and each web category $j \in \{fL, L, C, R, fR\}$, where P_{ij} is probability of consumption from category j on the web, given users belong to community i on YouTube, and P_j is probability of consuming from category j on the web for random YouTube users. Fig. 6 shows two main results. First, members of the right and far-right communities are more than twice as likely to consume right-leaning content and 3 times as likely to consume far-right content
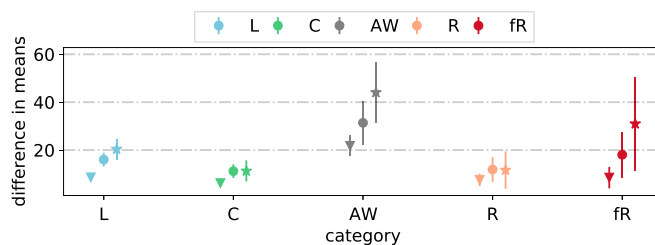

Fig. 5. Difference in means of daily consumption change, in the event of bursty consumption from a specific political category. Individuals are assigned either to bursty consumption group in the event of watching at least M_v^k videos from category k ($k \in \{L, C, AW, R, fR\}$) within a session, or to a 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% CI. 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 AW compared to the other political categories, when controlled for other covariates.
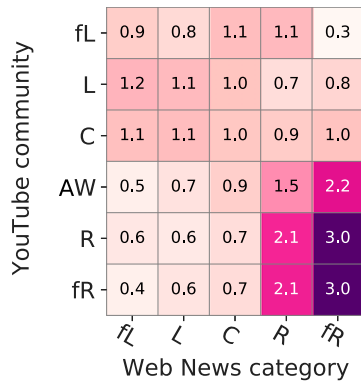
Hosseinmardi et al.
Examining the consumption of radical content on YouTube

PNAS | 5 of 8
https://doi.org/10.1073/pnas.2101967118

Fig. 6. Risk ratio of consumption from category j on the web ($j \in$ {fL, L, C, R, fR}) for users inside each community $i \in$ {fL, L, C, AW, R, fR} on YouTube. Users of far right, right, and AW are more likely than random YouTube users to consume right and far-right content on web.

outside of YouTube compared with an average user. Second, external consumption by members of the anti-woke community is biased in a very similar way to that of right and far-right members (1.5× for right-leaning content and 2.2× for far-right). In other words, even if anti-woke channel owners do not see themselves as associating with right or far-right ideologies, their viewers do. Finally, we note that, while these results do not rule out that recommendations are driving engagement for the heaviest consumers, they are strongly consistent with the explanation that consumption of radical content on YouTube is, at least in part, a reflection of user preferences (27, 31) (see also *SI Appendix*, Fig. S10 for more details on correlations between internal and external consumption).

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 2 shows that, while 36% of far-right videos are preceded by another video, nearly 55% of referrals come from one of the following: external URLs (41%), the YouTube homepage (8%), and search queries (6%) (see *SI Appendix*, Fig. S15 as a robustness check for more strict definitions of news consumers). Moreover, focusing on the subset of videos that are watched immediately after a user visits a news web domain, we find that approximately 50% of far-right/right videos and more than 30% of anti-woke videos are begun after visiting a right or far-right news domain such as foxnews.com, breitbart.com, and infowars.com. 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 bloomberg.com (*SI Appendix*, Figs. S11 and S12).

Session Analysis. Although our data do not reveal which videos are being recommended to a user, if the recommendation algorithm is systematically promoting a certain type of content, we would expect to observe increased viewership of the corresponding category 1) over the course of a session and 2) as session length 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 toward 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 than in shorter ones. Conversely, if we observe no increase in the relative frequency of far-right videos either over the course of a session or with session length, it would be evidence inconsistent with the claim that the recommender is driving users toward radical content.

To test these hypotheses, we assigned each video with political label $k \in$ {fL, L, C, AW, R, fR} an index $i^{(k)} \in \{1, \ldots, M_v\}$ where M_v is the number of videos in the sessions. We then normalized the indices $i_{norm}^{(k)} = (i^{(k)} - 1)/(M_v - 1)$, such that $i_{norm}^{(k)} \in [0, 1]$, meaning that zero indicates the first video and one indicates the final video of a session. Fig. 7A shows mean and SD of the fraction of videos with normalized index $i_{norm}^{(k)}$ for sessions of length $M_v \geq 20$, for each category k, $k \in$ {L, C, AW, R, fR} across different session definitions (*SI Appendix*, Table S16). The fraction of videos from the far-left category is too small to provide clear statistical results (Fig. 7B), and hence we dropped it from Fig. 7A. In all remaining cases, we find a nearly uniform distribution with an entropy deviating only slightly from that of a perfectly uniform distribution (*SI Appendix*, Table S17). For longer sessions, there is a slightly higher density closer to the relative index zero for far-right videos, precisely the opposite of what we would expect if the recommender were responsible (see *SI Appendix*, Figs. S19 and S20 and Table S17 for more details and robustness checks). Complementing the within-session analysis, Fig. 7B shows the average frequency of content categories as a function of session length. All six content categories show overall decreasing frequency, suggesting that longer sessions are increasingly devoted to nonnews content. More specifically, we see no evidence that far-right content is more likely to be consumed in longer sessions—in fact, we observe precisely the opposite.

Discussion

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 (39). On the consumption side,

Table 2. Distribution of the entry points of videos within each category

Category	YouTube homepage	Search	YouTube user/channel	YouTube video	External URLs	Other
fL	9.18	9.7	3.5	27.71	47.77	2.15
L	10.72	10.45	2.72	40.06	33.72	2.33
C	9.36	13.76	1.66	31.96	40.17	3.08
AW	11.98	7.46	3.63	38.62	35.51	2.8
R	7.19	9.12	4.23	37.67	40.19	1.6
fR	7.85	6.36	6.85	35.8	41.29	1.86

Video URLs can start from a YouTube homepage, a search (on/off platform), a YouTube user/channel, another YouTube video, or an external URL. YouTube recommendations (video entry point) have a bigger role for the left (40%) than any other group.
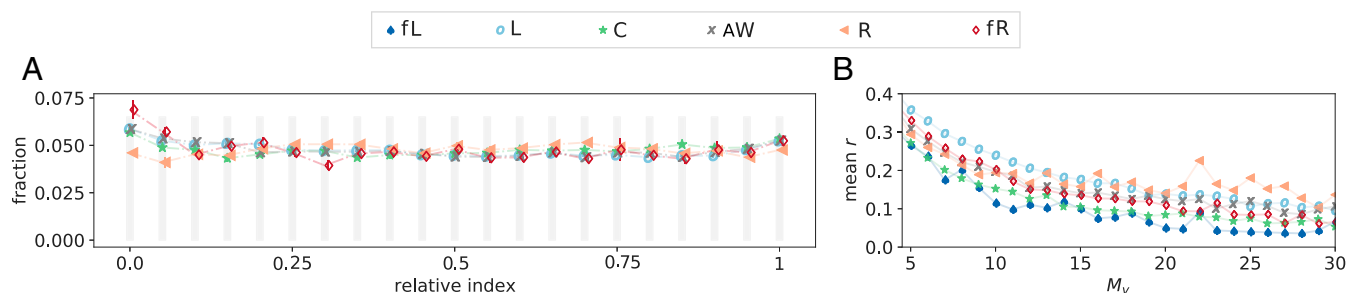
Fig. 7. (*A*) Mean and SD of fractions of videos as a function of normalized relative indices across session definitions for each category k, $k \in \{L, C, AW, R, fR\}$, for sessions with length $M_k \geq 20$. (*B*) Average r, where r is the fraction of videos of category k, $k \in \{fL, L, C, AW, R, fR\}$, for session definition $\delta = 10$ minutes and $\gamma = 60$ minutes. Sessions with length $M_v \geq 30$ (2% of the sessions) are dropped for better visualization.

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 size of online platforms—Facebook and YouTube each have more than 2 billion active users per month—tends to scale up the effect of any flaws in their algorithmic design or content moderation policies.

Here, we evaluated first whether YouTube, via its recommendation algorithm, drives attention to far-right content such as white supremacist ideologies and QAnon conspiracies, and hence is effectively radicalizing its users (24, 26–29, 40). Additionally, we investigate whether and how the large and growing community of anti-woke channels acts as a gateway to far-right content (21). We have investigated 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 videos spanning the political spectrum, over a 4-year time period.

Our results show that a community of users who predominantly consume content produced by far-right channels does exist, and, while larger than the corresponding far-left community, it is small compared with centrist, left-leaning, or right-leaning communities and is not increasing in size over the time period of our study. Moreover, we find that on-platform consumption of far-right content correlates highly with off-platform consumption of similar content, that users are roughly twice as likely to arrive at a far-right video from some source other than a previous YouTube video (e.g., search, an external website, the home page), and that far-right videos are no more likely to be viewed toward the end of sessions or in longer sessions. While none of this evidence can rule out the recommendation system as a cause of traffic to far-right content, it is more consistent with users simply having a preference for the content they consume.

We also find that the anti-woke community, while still small compared with left and centrist communities, is larger than the far right and is growing over time, both in size and engagement. We find evidence that the anti-woke community draws members from the far-right more than from any other political community, and that anti-woke members show an affinity for far-right content off-platform. On the other hand, when they leave, anti-woke members are more likely to move to left, center, and right than far right. Thus, while there do seem to be links between the anti-woke and far-right communities in terms of the content they consume, the hypothesized role of anti-woke channels as a gateway to far right is not supported. Rather, it seems more accurate to describe anti-woke as an increasingly popular—and sticky—category of its own. The implications of this fact are beyond the scope of this paper and left for future work to explore.

Overall, our 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 (27, 31). Although much about the dynamics of this large 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 (31). 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. Notably, 80% of sessions are length one and that 55% of videos are in sessions of no more than four videos. Both the majority of sessions and overall consumption, in other words, reflect the tastes and intentions that users entered with, and they also exhibit in their general web browsing behavior. To the extent that the growing consumption of radical political content is a social problem, our findings suggest that it is a much broader phenomenon than simply the policies and algorithmic properties of a single platform, even one as large as YouTube.

Our analysis comes with important limitations. First, our method of classifying content in terms of channel categories is an imperfect proxy for the content of individual videos. Just because a particular channel produces a substantial amount of far-right content, and hence could be legitimately classified as "far right" in our taxonomy, does not mean that every video promotes far-right ideology or is even political in nature. Future work could adopt a "content-based" classification system that could identify radical content more precisely. Second, 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. Third, our data only include desktop browsing, and hence reflect the behaviors of people who tend to use desktops for web browsing. Using a recently acquired mobile panel that tracks total time spent on YouTube but not detailed in-app usage, we are able to compare the fraction of mobile/desktop panel users who access YouTube at least once per month, and also the median consumption time per mobile/desktop user. As shown in *SI Appendix*, Fig. S21B, roughly 2 times as many mobile users as desktop users visit YouTube at least once a month (40% vs. 20%). Although we cannot rule out that consumption of radical political content is higher on mobile vs. desktop, other recent analysis of mobile device usage found that online news consumption is greater via laptop/desktop browsers (30). Better integration of desktop with mobile consumption presents a major challenge for future work. Fourth, 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 channels. Fifth, because our

Hosseinmardi et al.
Examining the consumption of radical content on YouTube

PNAS | 7 of 8
https://doi.org/10.1073/pnas.2101967118

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. Sixth, 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 in 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 2 thus should be viewed as an upper bound for "external" entrances and a lower bound for "video" entrances. Finally, we emphasize that our analysis is intended to address systematic effects and hence applies only at the population level. It says little about the role of specific YouTube content and the possible radicalization of individuals or small groups of individuals.

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 behavior. Although recent work has shown that television remains by far the dominant source of news for most Americans (30), 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 hyperpartisan, 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.

1. S. Iyengar, Y. Lelkes, M. Levendusky, N. Malhotra, S. J. Westwood, The origins and consequences of affective polarization in the United States. *Annu. Rev. Polit. Sci.* **22**, 129–146 (2019).
2. D. R. Jones, Declining trust in congress: Effects of polarization and consequences for democracy. *Forum* **13**, 375–394 (2015).
3. J. A. Tucker *et al.*, Social media, political polarization, and political disinformation: A review of the scientific literature. SSRN [Preprint] (2018). dx.doi.org/10.2139/ssrn.3144139 (Accessed 11 May 2021).
4. M. D. Conover *et al.*, "Political polarization on twitter" in *The International AAAI Conference on Web and Social Media (ICWSM)* (Association for the Advancement of Artificial Intelligence, 2011), **vol. 133**, pp. 89–96.
5. M. Del Vicario *et al.*, Echo chambers: Emotional contagion and group polarization on Facebook. *Sci. Rep.* **6**, 37825 (2016).
6. P. Grover, A. K. Kar, Y. K. Dwivedi, M. Janssen, Polarization and acculturation in US election 2016 outcomes—Can Twitter analytics predict changes in voting preferences. *Technol. Forecast. Soc. Change* **145**, 438–460 (2019).
7. M. Bossetta, The digital architectures of social media: Comparing political campaigning on Facebook, Twitter, Instagram, and Snapchat in the 2016 US election. *J. Mass Commun. Q.* **95**, 471–496 (2018).
8. M. Alizadeh, J. N. Shapiro, C. Buntain, J. A. Tucker, Content-based features predict social media influence operations. *Sci. Adv.* **6**, eabb5824 (2020).
9. S. Aral, D. Eckles, Protecting elections from social media manipulation. *Science* **365**, 858–861 (2019).
10. D. M. Lazer *et al.*, The science of fake news. *Science* **359**, 1094–1096 (2018).
11. G. Pennycook, D. G. Rand, Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 2521–2526 (2019).
12. J. Roozenbeek, S. Van Der Linden, The fake news game: Actively inoculating against the risk of misinformation. *J. Risk Res.* **22**, 570–580 (2019).
13. T. Konitzer *et al.*, Measuring news consumption with behavioral versus survey data. *SSRN*, [Preprint] (2020). dx.doi.org/10.2139/ssrn.3548690 (Accessed 11 May 2021).
14. A. Schomer, US YouTube advertising 2020. *eMarketer* (2020). https://www.emarketer.com/content/us-youtube-advertising-2020. Accessed 11 May 2021.
15. M. Iqbal, Twitter revenue and usage statistics (2020). *BusinessofApps* (2021). https://www.businessofapps.com/data/twitter-statistics/. Accessed 11 May 2021.
16. S. Clark, A. Zaitsev, Understanding YouTube communities via subscription-based channel embeddings. arXiv [Preprint] (2020). arXiv:2010.09892 (Accessed 11 May 2021).
17. K. Roose, The making of a YouTube radical. *NY Times*, 8 June 2019. https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html. Accessed 11 May 2021.
18. Z. Tufekci, Youtube, the great radicalizer. *NY Times*, 10 March 2017. https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html. Accessed 11 May 2021.
19. J. Bridle, Something is wrong on the internet. *James Bridle* (2017). https://medium.com/@jamesbridle/something-is-wrong-on-the-internet-c39c471271d2. Accessed 11 May 2021.
20. T. Whyman, Why the right is dominating YouTube. *Vice* (2017). https://www.vice.com/en/article/3dy7vb/why-the-right-is-dominating-youtube. Accessed 11 May 2021.
21. R. Lewis, *Alternative Influence: Broadcasting the Reactionary Right on YouTube* (Data and Society Research Institute, 2018).
22. R. Lewis, "This is what the news won't show you": YouTube creators and the reactionary politics of micro-celebrity. *Televis. New Media* **21**, 201–217 (2020).
23. P. Bacon, Why attacking 'cancel culture' and 'woke' people is becoming the GOP's new political strategy. *FiveThirtyEight* (2021). https://fivethirtyeight.com/features/why-attacking-cancel-culture-and-woke-people-is-becoming-the-gops-new-political-strategy/. Accessed 11 May 2021.
24. M. H. Ribeiro, R. Ottoni, R. West, V. A. Almeida, W. Meira, Jr, "Auditing radicalization pathways on YouTube" in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery, 2020), pp. 131–141.
25. M. Mohsin, 10 YouTube stats every marketer should know in 2020. *Oberlo* (2020). https://www.oberlo.com/blog/youtube-statistics. Accessed 11 May 2021.
26. J. Cho, S. Ahmed, M. Hilbert, B. Liu, J. Luu, Do search algorithms endanger democracy? An experimental investigation of algorithm effects on political polarization. *J. Broadcast. Electron. Media* **64**, 150–172 (2020).
27. K. Munger, J. Phillips, Right-wing YouTube: A supply and demand perspective. *Int. J. Press Politics*, 10.1177/1940161220964767 (2020).
28. M. Ledwich, A. Zaitsev, Algorithmic extremism: Examining YouTube's rabbit hole of radicalization. *Clin. Hemorheol. and Microcirc.* **25**, 10.5210/fm.v25i3.10419 (2019).
29. M. Faddoul, G. Chaslot, H. Farid, A longitudinal analysis of YouTube's promotion of conspiracy videos. arXiv [Preprint] (2020). arXiv:2003.03318 (Accessed 11 May 2021).
30. J. Allen, B. Howland, M. Mobius, D. Rothschild, D. J. Watts, Evaluating the fake news problem at the scale of the information ecosystem. *Sci. Adv.* **6**, eaay3539 (2020).
31. T. Wilson, K. Starbird, Cross-platform disinformation campaigns: Lessons learned and next steps. *Harvard Kennedy School Misinform. Rev.* **1**, 10.37016/mr-2020-002 (2020).
32. D. Lazer, Studying human attention on the internet. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 21–22 (2020).
33. S. Wu, M. A. Rizoiu, L. Xie, "Beyond views: Measuring and predicting engagement in online videos" in *Twelfth International AAAI Conference on Web and Social Media*, (Association for the Advancement of Artificial Intelligence, 2018).
34. S. Salinas, YouTube removes Alex Jones' page, following bans from Apple and Facebook. *CNBC*, 6 August 2018. https://www.cnbc.com/2018/08/06/youtube-removes-alex-jones-account-following-earlier-bans.html/. Accessed 11 May 2021.
35. R. Kumar, A. Tomkins, "A characterization of online browsing behavior" in *Proceedings of the 19th International Conference on World Wide Web* (Association for Computing Machinery, 2010), pp. 561–570.
36. M. Charrad, N. Ghazzali, V. Boiteau, A. Niknafs, NbClust: An R package for determining the relevant number of clusters in a data set. *J. Stat. Software* **61**, 1–36 (2014).
37. C. A. Bail *et al.*, Exposure to opposing views on social media can increase political polarization. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 9216–9221 (2018).
38. R. E. Robertson *et al.*, "Auditing partisan audience bias within Google search" in *Proceedings of the ACM on Human-Computer Interaction* (Association for Computing Machinery, 2018), **vol. 2**, pp. 1–22.
39. B. Snyder, Alex Stamos: How do we preserve free speech in the era of fake news? *Stanford Engineering*, 10 January 2019. https://engineering.stanford.edu/magazine/article/alex-stamos-how-do-we-preserve-free-speech-era-fake-news. Accessed 11 May 2021.
40. M. Alfano, A. E. Fard, J. A. Carter, P. Clutton, C. Klein, Technologically scaffolded atypical cognition: The case of youtube's recommender system. *Synthese*, 10.1007/s11229-020-02724-x (2020).
41. H. Hosseinmardi *et al.*, Examining the consumption of radical content on YouTube. Open Science Foundation. https://osf.io/vs4d9/. Deposited 11 July 2021.