PNAS Nexus, 2024, **3**, pgae518
**https://doi.org/10.1093/pnasnexus/pgae518**
Advance access publication 19 November 2024
**Research Report**

# Nudging recommendation algorithms increases news consumption and diversity on YouTube

Xudong Yu [a,1], Muhammad Haroon [b], Ericka Menchen-Trevino [c] and Magdalena Wojcieszak [d,e,*,1]

[a]Department of Communication, University of North Dakota, Grand Forks, USA
[b]Department of Computer Science, University of California, Davis, USA
[c]Independent Researcher, Chicago, USA
[d]Department of Communication, University of California, Davis, USA
[e]Center for Excellence in Social Science, University of Warsaw, Warsaw, Poland
*To whom correspondence should be addressed: Email: mwojcieszak@ucdavis.edu
[1]X.Y. and M.W. contributed equally to this work.
**Edited By** David Rand

## Abstract

Recommendation algorithms profoundly shape users' attention and information consumption on social media platforms. This study introduces a computational intervention aimed at mitigating two key biases in algorithms by influencing the recommendation process. We tackle *interest bias*, or algorithms creating narrow nonnews and entertainment information diets, and *ideological bias*, or algorithms directing the more strongly partisan users to like-minded content. Employing a sock-puppet experiment ($n = 8{,}600$ sock puppets) alongside a month-long randomized experiment involving 2,142 frequent YouTube users, we investigate if nudging the algorithm by playing videos from verified and ideologically balanced news channels in the background increases recommendations to and consumption of news. We additionally test if providing balanced news input to the algorithm promotes diverse and cross-cutting news recommendations and consumption. We find that nudging the algorithm significantly and sustainably increases both recommendations to and consumption of news and also minimizes ideological biases in recommendations and consumption, particularly among conservative users. In fact, recommendations have stronger effects on users' exposure than users' exposure has on subsequent recommendations. In contrast, nudging the users has no observable effects on news consumption. Increased news consumption has no effects on a range of survey outcomes (i.e. political participation, belief accuracy, perceived and affective polarization, and support for democratic norms), adding to the growing evidence of limited attitudinal effects of on-platform exposure. The intervention does not adversely affect user engagement on YouTube, showcasing its potential for real-world implementation. These findings underscore the influence wielded by platform recommender algorithms on users' attention and information exposure.

**Keywords:** social media, news exposure, filter bubbles, recommendation algorithm, computational social science

### Significance Statement

Recommendation algorithms profoundly shape users' attention and information consumption on social media. This project designs and deploys a computational intervention that nudges YouTube's algorithm to increase recommendations to videos from verified and ideologically balanced news channels. Our results from an experiment on 8,600 sock-puppets and a month-long experiment on 2,142 frequent YouTube users show that this intervention increases recommendations and exposure to news and enhances the ideological diversity of users' news diets over time. In contrast, nudging the users themselves has no effects on recommendations or consumption. Increased news consumption has no effects on democratic attitudes (e.g. political participation, belief accuracy, perceived and affective polarization, etc.). These findings underscore the influence of recommender algorithms and add to the recent work finding limited effects of on-platform exposure.

## Introduction

The majority of Americans are active users of social media platforms. As those users navigate the online ecosystem, their attention and information consumption are largely driven by recommendation algorithms, which are designed to maximize engagement by recommending content that aligns with users' inferred interests and biases (1–4). These personalized recommendations drive between 75 and 95% of consumption on platforms (5, 6). In this context, observers (7, 8) and scholars (6, 9–11) worry that algorithmic systems recommend

**Competing Interest:** The authors declare no competing interests.
**Received:** April 9, 2024. **Accepted:** October 29, 2024

like-minded political content and direct users to radical, conspiratorial, or otherwise problematic information. Evidence to support these worries about what we refer to as *ideological bias* in algorithms is mixed at best (6, 9).

This project advances past work by tackling *interest bias* in recommendation algorithms, insufficiently recognized, under-studied, and crucially important for content exposure. Recommendation algorithms are designed to optimize the time the users spend on the platform by learning to recommend the content with which users are likely to engage (2, 3). This feature of all recommendation systems is not on its own problematic. However, because most users go to platforms for entertainment, *not* for verified news and quality public affairs (12–15), and because many engage with low-quality, sensationalist, and click-baity content (16, 17), algorithmic recommendations align with these interests and preferences (1, 18). This, combined with the fact that algorithms actively direct users to entertainment and away from news (19), leads to closed loops of low-quality, nonnews content consumption (4, 13, 20).

Under-consumption of public affairs has crucial democratic consequences. Low-information voters are vulnerable to irrelevant cues in the political environment (21), vote against their personal and group interests (22–24), and are more susceptible to populist, manipulative, and misinformative rhetoric (25). In turn, news exposure leads to more informed citizens (26–30), increases opinion stability and voting in accordance with one's interests (24, 31), decreases beliefs in misinformation (32–34), enhances efficacy, tolerance, and the acceptance of democratic norms (35, 36), and leads to more equitable voting outcomes (37). Therefore, minimizing interest bias in recommendation algorithms and incentivizing greater consumption of verified news among citizens is of importance.[a]

In addition, recommender systems may direct the more politically interested users to congenial, one-sided, and/or hyperpartisan political information, i.e. ideological bias in recommendation algorithms. Although few users inhabit echo chambers (15, 39, 40) or are put in extreme rabbit holes (6, 9), this small subset is consequential. Those users are more strongly partisan, more affectively polarized, more vocal and active in the political arena, and have a disproportionate influence on the democratic process (41–43). We thus also test whether providing *balanced* news input to the algorithm minimizes ideological bias in recommendations and consequently leads users to consume more diverse and cross-cutting news and political content. Because we establish the baseline slant of each user's news diet, from very liberal to very conservative, we can pinpoint daily over-time shifts in the slant of recommendations and exposures resulting from our intervention. In short, our intervention is built to increase news consumption while also minimizing ideological biases in information diets. Our expectation is that nudging the algorithm by altering a user's browsing history increases recommendations to and consumption of news and diverse news and political content.

Furthermore, we nudge individual users by reminding them of the personal and democratic benefits of news exposure. Many citizens avoid news, either actively or unintentionally (44). They may be uninterested in politics, perceive it as irrelevant to their daily lives, or not make the connection between public affairs, governmental policies, and news events and their own and societal interest (38). It follows that making politics more personally and collectively relevant, as well as making it easy for people to access news, could increase its consumption. We propose that nudging citizens to consume news by highlighting the benefits that news brings to the public and to oneself should encourage them to consume more news and political content.

This two-pronged approach—algorithmic and user nudge—is uniquely suited to disentangle two interrelated factors that underlie low news consumption: recommender systems and/or individual disinterest. If low news use is due to interest bias in algorithms (e.g. a sports fan is only recommended sports videos and is effectively secluded from news and public affairs information on platforms (19)), then nudging the algorithm to recommend verified news could increase news consumption simply because more inventory is easily accessible. If, however, people avoid news mostly because they see it as irrelevant (44), reminding users that news benefits them and society could encourage greater news consumption.

In this project, we focus on YouTube, as it is one of the largest platforms, with 1.7 billion unique monthly visitors and 14.3 billion visits per month, more than Facebook, Wikipedia, Amazon, and Instagram. YouTube is also the most popular platform in the United States, used by 81% of the population, and has a steadily growing user base (45).

We rely on two systematic experiments. The first experiment uses 8,600 trained *sock-puppets* to test the effects of various parameters for the algorithmic nudge intervention. We integrate the findings into a Chrome extension we develop, which not only unobtrusively records all recommendations and information consumption on YouTube but also builds in two interventions: an algorithmic nudge and a user nudge (see Supplementary Material, R for details on the extension). We deploy the extension in a longitudinal experiment with a sample of 2,142 frequent YouTube users recruited by YouGov, who were asked to install the extension for 1 month. For the first week, the extension records all the recommendations (both homepage and up-next) and behaviors on the platform. In the next 2 weeks, the extension activates our interventions. In the *algorithmic nudge* treatment, the extension obfuscates user's YouTube watch history by unobtrusively playing videos from verified and balanced news sources, based on expert metrics. In the *user nudge* treatment, the extension changes the YouTube interface to integrate a banner reminding the users of the benefits of news exposure and a link to the YouTube News section (see Supplementary Material, A for the details on the prompt selection and the screenshots of the user interface). There are no changes to participants' YouTube experience in the control.

We examine the effects of our interventions—both during the treatment and over time after the interventions were deactivated—on (i) homepage and up-next recommendations and (ii) the actual consumption of (a) news (using an extensive list of US news organizations, see Supplementary Material, B) and (b) political videos from outside of news channels (determined using our BERT-based classifier, see Supplementary Material, D). We also test shifts in diversity of recommendations and consumption using our validated approach for estimating political slant of individual news and political videos (see (46)) *and also* to extreme, conspiratorial, and otherwise problematic channels from previously compiled lists (11, 47) (see Supplementary Material, E for details). Lastly, we also rely on a posttest survey to test if increased news consumption has effects on self-reported outcomes (i.e. political participation, perceived accuracy of true and false claims, perceived and affective polarization, support for democratic norms, etc.).

We offer four key findings. First, the algorithmic nudge significantly and sustainably increases both recommendations and actual consumption of news on YouTube. In fact, we identify a reinforcing feedback loop between the algorithm and the users, in which recommendations drive news consumption more strongly than news consumption drives subsequent recommendations. Second, the algorithmic nudge also minimizes the ideological congeniality of recommendations and consumption, an effect that is most pronounced among conservatives and that persists after the intervention ends. Third, nudging individual users by changing the YouTube interface to integrate a banner reminding the users of the benefits of news use has no observable effects on the consumption of (diverse) news and political content. Fourth, although the algorithmic nudge boosts the likelihood that users encounter and consume more (diverse and cross-cutting) news and political content, this increased news consumption has no effects on the tested attitudes and beliefs, null effects that do not vary with users' baseline news exposure or partisanship strength. Crucially, the intervention does not decrease the time spent and user engagement with YouTube, which makes it implementable by platforms concerned with engagement.

Before we present the results, we note that the normative foundations of our work are open to critique. Despite the lack of a single standard for good citizenship (48) and growing criticisms of elitism in social science (49, 50), our premise is that citizens should consume verified public affairs information. As aforementioned, awareness of the issues facing the country and one's community sustains a well-functioning democracy (27, 36) and leads to informed populace that is resilient to misinformation (32, 34), votes for parties or candidates that best represent their (often disadvantaged) group interests (22–24), and is able to effectively address various public crises (51). This is not to say that citizens should solely consume public interest content; naturally, such content co-exists with other types of information—including entertainment, silly memes, or partisan media. Yet, some engagement with news produced according to verifiable editorial processes is important. Our approach facilitates this engagement in an online ecosystem where content consumption is currently curated by proprietary black-box algorithms of social media companies.

Additionally, some argue that recommender systems provide individual benefits by offering personally relevant content that users desire. This, however, is akin to saying that if users want ice cream for breakfast and cake for dinner, platforms should reflect users' wishes and serve only ice cream and cake. Nevertheless, to the extent that oatmeal and soup are healthier, *offering* verified public interest content is more responsible, and making content from across the political spectrum easily available supports a diverse diet.[b] In addition, such recommendations do not go against users' desires. Evidence suggests that users want to see more verified, factual, and informative content on platforms (52), and those encouraged to follow news on Instagram and WhatsApp report overwhelmingly positive experiences (34). Currently, there may be a feedback loop where platforms deprioritize news, which decreases exposure and may lead users to lose interest and seek out news less. Our approach can break these loops by putting news in the users' feed and encouraging exposure, thereby increasing future recommendations (4, 20) and users' interest in news and public affairs (53).

## Design and results

### Sock-puppet experiments

To establish the viability of our algorithmic nudge intervention, we first conduct systematic experiments on YouTube using "sock puppets." Sock puppets were implemented as automated web-browsers which mimic the actions of a real user on the platform such as watching videos, seeing, and clicking on recommendations. By controlling the actions of these sock puppets, we could observe how the recommendations change in response to specific actions on the platform and determine the effectiveness of our intervention on recommendations. To maximize ecological validity, the sock puppets replayed the YouTube watch histories of American adults ($n = 1,980$).[c] The act of replaying videos in a given user's watch history allows us to collect and analyze the subsequent recommendations that a particular user might have seen. We refer to this act of replaying videos as *training* the sock puppet.

In total, we trained 8,600 realistic sock puppets from 215 users (with watch history length <100 for scalability; median = 51). These sock puppets were randomly assigned to various algorithmic nudge intervention parameters to determine the optimal parameters to increase news recommendations by playing videos from verified and ideologically balanced news channels in a browser tab that is not active or visible to the user (referred to as "injecting" here).

All intervention videos came from reliable and balanced news organizations, as determined by validated expert metrics from Ad Fontes Media (55). Ad Fontes relies on manually labeled articles, radio, TV, and videos (episodes) from numerous news sources. Each episode is rated by trained human coders and scores are assigned for reliability (from "contains inaccurate/fabricated information" to "original fact reporting") and political bias (from "most extreme left" to "most extreme right"). We selected 39 news outlets categorized as reliable (reliability score >40) and balanced (bias score between −18 and 18), and with corresponding YouTube channels. See Supplementary Material, B for details on the labeling and the outlet selection. Different organizations evaluate news outlets on different and sometimes unclear criteria, yet experts generally agree on the relative placement of news sources (56). We address this in Supplementary Material, B and show that the sources selected from AdFontes rank high in reliability and credibility in rankings from (56), NewsGuard, and Media Bias/Fact Check.

To optimize the subsequent experiment for human participants, we tested the effects of three approaches in this sock puppet study that inject: (i) randomly sampled vs. most popular videos from a news channel ("chronology"), (ii) a fixed number of videos vs. number proportional to the length of the users' history ("proportion"), and (iii) videos at set intervals ("intervals"), as described here.

- **Chronology.** With this parameter, we test whether there is a difference in the number of news recommendations between injecting randomly sampled videos from the most popular or the most recent videos from a news channel. These two sorting mechanisms are provided by YouTube when viewing videos from a channel.
- **Proportion.** With this parameter, we test the minimum number of injections needed to increase news recommendations. We test two approaches: (i) injecting a fixed number of videos and (ii) injecting a number of videos proportional to the length of the user's history, e.g. if the user watched 50 videos, then at 10%, we watch five intervention videos.
- **Intervals.** As opposed to a fixed or proportional number of injections, we test if a recurring injection after a set interval leads to a higher percentage increase in news recommendations.

Of 8,600 sock puppets, 4,300 (50%) were assigned to watch the most popular videos on news channels, and the other 4,300 (50%) were assigned the most recent. Within each, 4,300, 2,150 (25%) sock puppets were assigned to a fixed number of intervention videos, and the other 2,150 (25%) to a percentage number of videos based on the length of the watch history. The fixed number of interventions ranged from 0 to 10 and the percentage of length ranged from 0 to 100% which were also equally distributed amongst the 2,150 sock puppets.

To test if the interventions increase recommendations to news, we rely on a list of a total 1,625 YouTube news channels from two lists: 941 identified YouTube news channels from our curated list of news organizations (57) and 684 channels the lists by Ribeiro et al. (11) and Ledwich et al. (10). Supplementary Material, C presents the details on the list. We calculate the percentage of recommendations to videos from those channels before and after the interventions.

To determine if the interventions increase recommendations to videos about politics outside news channels, we rely on our validated BERT-based neural binary classifier that categorizes YouTube videos as related to politics with high accuracy (0.93, Precision 0.92, Recall 0.91, F1 0.915; see Supplementary Material, D for model training, validation, and performance). In short, our measure of *news* is on the channel level and encompasses all news videos regardless of whether they were about politics or not. Our measure of *political* videos is on the content level and encompasses all videos classified as political and not part of news channels (e.g. those from vloggers, political influencers, and—say—celebrities, in which politics is addressed).

The sock-puppet experiments find the effectiveness of the algorithmic intervention under different hyper-parameter configurations (see Supplementary Material, F for details). Here, we mention the high-level effects and Table 1 details the average increases in news and political recommendations. First, we show that the increase in the percentage of news recommendations is *higher* when injecting randomly selected videos (1.91×) as compared with popular videos (1.61×). The two approaches increase political recommendations equally (1.20×) and (1.24×).[d] Second, injecting number of videos proportional to the number of videos a user previously watched increases news recommendations more pronounced than injecting a fixed number of videos. For both approaches, however, the increase grows steadily as more videos are injected.

Because, unlike sock puppets, human users have an ever-increasing list of watched videos, the proportional approach requires keeping track of users' dynamic watch histories, which presents technical challenges if this approach is to be implemented in an extension for participants. Thus, for the injections to be effective, we need to perform them multiple times as the user continues to use the platform. Our final setup injects one video at a time at regular intervals while the user is active on the platform. These injections at intervals inherently account for the dynamic watch history as they interleave the injections with the user's own watched videos. Doing a recurring injection every 10 min showed an increase in news recommendations of up to 1.49× and political recommendations of up to 1.63× compared with the baseline of no injections.

Thus, our final configuration relies on injections at set intervals with videos randomly sampled from all the videos posted by the news channels on our list in the past 48 h every day. We implemented these parameters into a Chrome extension developed for this project. At a glance, participants installed the extension as part of the study which monitored their YouTube behavior for 4 weeks. The extension collected recommendations, recorded user searches and video watches, and administered the assigned nudges during treatment weeks. All the data were stored with a unique user identifier, which was linked to users' survey responses. The next section further outlines the extension and Supplementary Material, R offers technical details on the extension infrastructure.

## YouTube users experiment

We deployed this extension among human users (who have dynamic watch histories and numerous on- and off-platform behaviors that sock puppets lack) in a month-long experiment on a sample of frequent YouTube users recruited by YouGov. Figure 1 outlines the design. The user experiment was conducted from November 2022 to January 2023. YouGov recruited American adults who visited YouTube more than once a week. Here, 2,142 eligible participants were directed to the Chrome Store, where they were prompted to install our *ResearchTube* extension for 4 weeks. They were then randomly assigned to one of three experimental conditions: algorithmic nudge (35%), user nudge (35%), or control (30%).[e] Following the assignment, participants completed the pretest survey (see Supplementary Material, G for question wording).

For the first week, the extension unobtrusively recorded all platform recommendations (homepage and up-next), exposures, and behaviors to establish a baseline. In weeks 2 and 3, the extension additionally activated our interventions. In the *algorithmic nudge* treatment, the extension obfuscated the user's YouTube watch history by unobtrusively playing/injecting news videos in the background every 10 min when the user's browser was open. For every four videos, one came from left-leaning channels, two from centrist channels, and one from right-leaning channels based on the aforementioned expert metrics from Ad Fontes Media, as detailed in Supplementary Material, B. In the *user nudge* treatment, the extension changed the YouTube interface to integrate a banner on the homepage and under a video a user was watching with a prompt reminding the users of the benefits of news consumption. The extension randomly displayed a prompt from a list of four piloted prompts each time a user opened the homepage or watched a video; see Supplementary Material, A for the prompts. Moreover, a button on the banner would take participants to the YouTube News section, and the YouTube search box suggested "Watch more news" (see Supplementary Material, A for screenshots of the interface). There were no changes to participants' YouTube experience in the control.

At the end of week 3, participants who watched at least seven YouTube videos in weeks 1–3 were invited to complete the posttest measuring various attitudes and beliefs (i.e. political participation,

**Table 1.** Average increase in news and political recommendations as a result of different sock-puppet configurations.

| Number | 0 | 1 | 3 | 5 | 6 | 7 | 8 | 9 | 15 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| News | 0.97× | 1.27× | 1.51× | 1.54× | 1.71× | 1.56× | 1.83× | 2.08× | 2.75× | 2.90× |
| Political | 1.02× | 1.13× | 1.20× | 1.21× | 1.20× | 1.26× | 1.18× | 1.22× | 4.55× | 1.51× |
| Percentage | 0% | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% |
| News | 0.94× | 1.86× | 2.38× | 2.80× | 3.67× | 3.64× | 4.18× | 3.88× | 5.30× | 3.99× |
| Political | 1.01× | 1.18× | 1.18× | 1.26× | 1.32× | 1.33× | 1.52× | 1.55× | 1.58× | 1.55× |

In the top panel, we report the increase as we inject a fixed number of videos. In the bottom panel, we report the increase as we inject a number of videos proportional to the sock puppet's watch history. We see that a higher number of injections yielded more news and political recommendations.

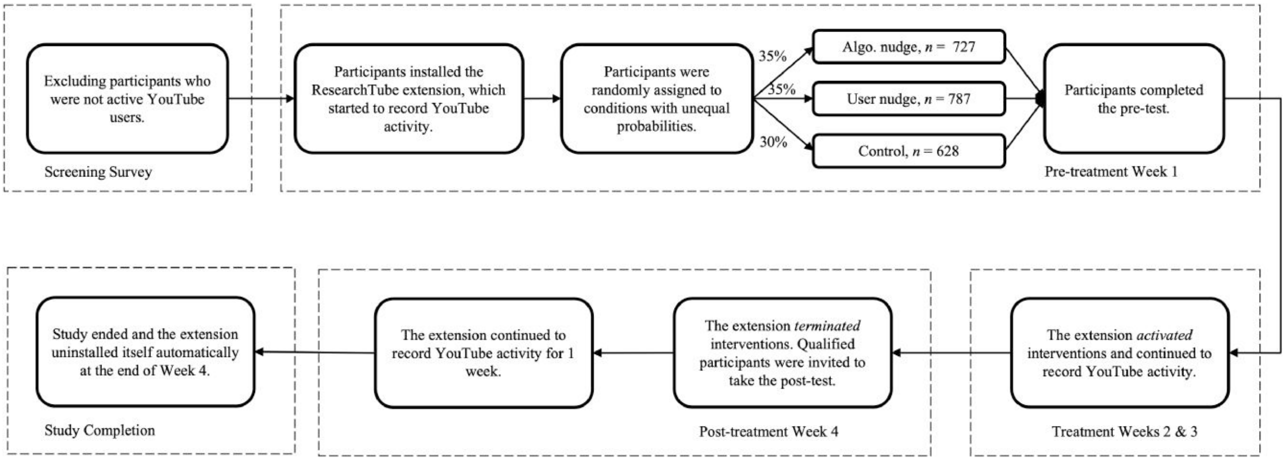


**Fig. 1.** Overview of the research design. Participants were initially screened for their YouTube browsing habits, and then qualified participants were asked to install the ResearchTube browser extension. The extension assigned participants to either treatment or control conditions and then monitored YouTube activities and administered interventions.

perceived accuracy of true and false claims, perceived and affective polarization, support for democratic norms; Supplementary Material, G shows question wording and descriptives). The interventions concluded at the end of week 3. To capture any lasting effects, the extension continued to collect recommendations and on-platform behaviors for an additional full week. After 4 weeks, the extension was automatically uninstalled. The project was approved by the Ethical Board of the European Research Council (ERC) and the University of Amsterdam (ERB number 2022-PCJ-15365, ID 15365), and all participants who took part in this study completed extensive informed consent. Recruitment materials, the text of the privacy policy, and the treatment-specific debriefing materials are presented in Supplementary Material, V.

The final sample includes participants who watched at least seven YouTube videos in weeks 1–3 and at least one video in weeks 2 and 3. This resulted in a total of 1,188 participants, with 430 in the algorithmic nudge treatment condition, 422 in the user nudge treatment condition, and 336 in the control condition. These inclusion criteria ensured that participants were engaging with YouTube and were treated at least once during the treatment weeks. In Table S8, we offer details on the number of participants excluded due to early uninstallation of the extension (e.g. if a participant uninstalled the extension in week 1, they were excluded from the analyses because we lacked their activity data for weeks 2 and 3) and lack of YouTube activity during the study period.[f]

To ensure robustness, we tested alternative inclusion criteria, including participants who (i) watched at least one video during weeks 2 and 3 ($n = 1{,}320$), and (ii) watched at least five YouTube videos over weeks 1–3 and at least one video during weeks 2 and 3 ($n = 1{,}228$). We present the details on the sample and subsample sizes for these different inclusion criteria in Table S7. We reestimated the effects of treatments on YouTube engagement, news watching, and political watching among these samples, finding results consistent with those presented below (see Supplementary Material, W for regression tables and figures).

The mean age of the final sample was 47.7 (SD = 16.5), with 46% males, 70% White, and 66% with a 4-year degree or some college. The majority (60%) identified as Democrats, 25% as Republicans, and 15% as Independents or other. In Supplementary Material, I, we show that randomization to condition was successful.

We note two considerations. First, our sample is not random or representative, and no studies using online samples willing to install browser extensions can claim representativeness. Having a sample of active YouTube users was more important than external validity (hence YouGov included only the panelists who reported using YouTube more than once a week). In Table S6, we show that our initial sample and final sample largely reflect the general YouTube population on basic sociodemographics. Second, it is possible that the rates of uninstallation and/or lack of activity on YouTube may have been a result of the treatment (e.g. if participants were more likely to uninstall the extension or stop using YouTube in the treatment conditions, then excluding those participants introduces bias and prevents causal inference). In Supplementary Material, I, we show that the majority ($n = 188$, 55% of early uninstallers) uninstalled the extension in week 1. Because this was before the experimental treatments were launched, it is not possible that these uninstallations were due to the treatments. Out of those who uninstalled during weeks 2–4, 72% ($n = 109$) were nevertheless included in the analyses because those individuals had the required minimum of YouTube activity. We also show that it is not the case that the uninstallation rates were higher among the treatment groups than among the control. Furthermore, Tables S9 and S10 show that there are no statistically significant differences between treatment conditions and the control condition in early extension uninstallation and lack of YouTube activity during the study period.[g] In other words, there is no systematic bias in participant exclusion across the conditions. In Supplementary Material, I, we additionally show that early uninstallers and those who did not use YouTube during the study period do not significantly differ from the included participants in terms of gender, education, race, and partisanship. We find significant differences on age. We include all the sociodemographic variables in our multivariate models to account for these minor differences. In short, those analyses suggest that attrition bias is minimal or nonexistent, and it is unlikely to influenced the patterns observed.

Overall, we collected data on 5,896,318 recommended and 113,079 watched videos, as well as information on 15,218 YouTube channels the participants browsed and 14,956 search queries they used. Importantly, in Supplementary Material, J we show that neither treatment decreased YouTube usage (number of videos watched, number of days and time spent on the platform). In fact, users in the algorithmic nudge condition had 0.57 *more* active days on YouTube during the treatment weeks (SE=0.12, $P < 0.001$) than the control. This basic finding suggests that it is not the case that seeing news on platforms decreases user engagement and underscores the applicability of our intervention. In sum, the subtle sustained nature of the intervention and our ability to track YouTube recommendations and on-platform behaviors offer a unique opportunity to test the effects of algorithmic and user nudges over time.

### *Effects on recommendations and consumption*

First, we describe the percentage of news and political videos recommended and actually watched by users throughout the 4 weeks (Supplementary Material, U presents the detailed breakdown of percentage changes). As plotted in Figs. 2 and 3, in week 1, the percentage of *news* content watched (average: alg nudge 4.50%, user nudge 4.23%, control 4.22%) and recommendations (average: agl nudge 7.90%, user nudge 8.94%, control 8.26%) was relatively low across all participants. We observe a marked increase in recommendations to (16.42% on average) and viewership of (18.67% on average) news in the algorithmic nudge condition during the treatment weeks. News recommendations and viewership remained under 10% in the user nudge (7.67%, 4.26%) and control (7.69%, 3.59%). In week 4, when the intervention stopped, recommendations continued at 19.02% in the algorithmic nudge condition. Although news viewership declined to an average of 8.71%, it was still about double from the baseline in week 1.[h]

With regard to *political* content from outside news channels, Fig. 3 shows that the percentage of recommended and watched political videos was consistently between 5 and 15% for all participants in week 1 (average recommendations: alg nudge 11.06%, user nudge 10.86%, control 9.56%; average consumption: alg nudge 11.54%, user nudge 9.74%, control 11.59%). There was no noticeable increase in recommendations to or watches of political videos in any of the conditions. These descriptives suggest that the algorithmic nudge—not the user nudge—increased both recommendations to and the actual watching of news but not political content outside of news channels.

We estimated the effects of the interventions, controlling for age, gender, education, race, partisanship, and the baseline number of news videos watched per active day in the pretreatment week. The full models are reported in Supplementary Material, K and plotted in Fig. 4A. Confirming the descriptive plots above, the algorithmic nudge statistically significantly increased news consumption. Participants in the algorithmic nudge condition viewed 1.01 ($P < 0.001$) more news videos per active day during the treatment weeks compared with the control. This effect amounts to 2.59 times the average number of news videos all participants watched per active day in week 1. If sustained for a year, users would consume 369 more news videos. This significant effect persisted even after the intervention terminated: in week 4, participants in the algorithmic nudge condition watched 0.25 ($P = 0.008$) more news videos per active day than the control. In turn, although there were slight increases in news watching in the user nudge condition, these effects were not statistically significant during the treatment weeks ($P = 0.258$) and posttreatment ($P = 0.236$). Also here, the nudges—whether the algorithmic or the user nudge—had no spillover effects on the consumption of political videos (see Supplementary Material, K for regression tables).

We ran parallel models predicting the *percentage* of news videos watched. Figure 4B additionally shows the effects. Participants in the algorithmic nudge condition watched 16% ($P < 0.001$) more news videos during the treatment weeks—nearly a 4-fold increase compared with the average proportion of news videos all participants watched in week 1—and 7% ($P < 0.001$) more than the control in week 4. Again, there were no significant effects of the user nudge ($P = 0.516$ and 0.306, respectively). As before, the models find no effects of the two nudges on the consumption of political content (see Supplementary Material, K). Supplementary Material, K

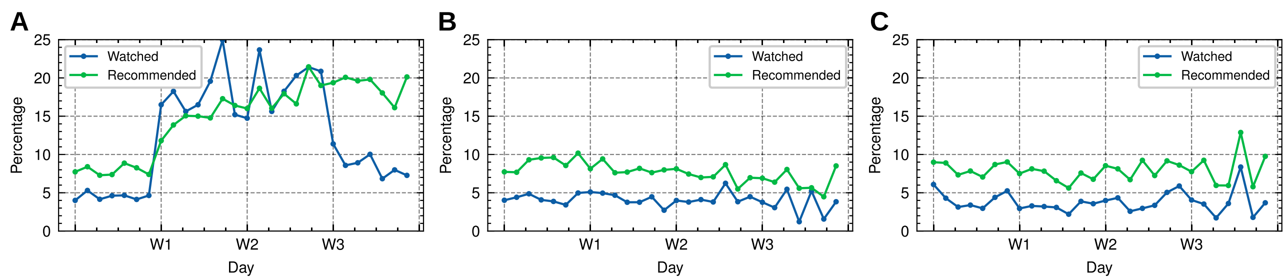


**Fig. 2.** Average daily percentage of *News* videos watched by and recommended to participants in A) algorithmic nudge, B) user nudge, and C) control groups over total watches and recommendations. W1, W2, and W3 refer to the end of weeks 1, 2, and 3, respectively. The treatment ran between W1 and W3.

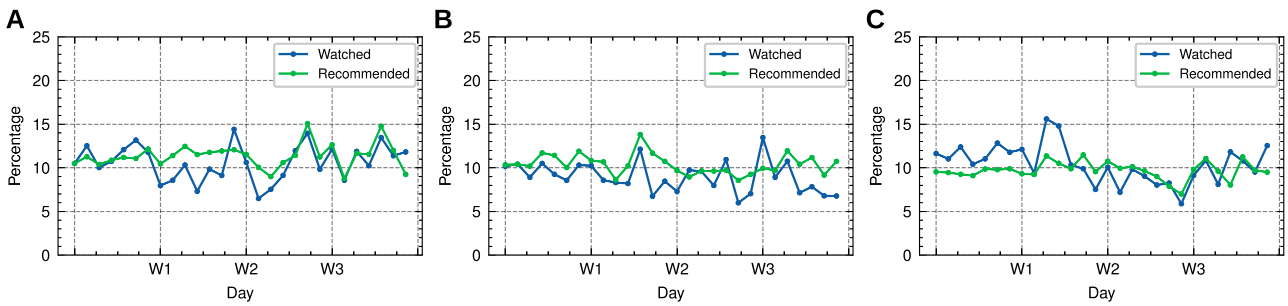


**Fig. 3.** Average daily percentage of *Political* videos watched by and recommended to participants in A) algorithmic nudge, B) user nudge, and C) control groups over total watches and recommendations. W1, W2, and W3 refer to the end of weeks 1, 2, and 3, respectively. The treatment ran between W1 and W3.

additionally shows these models with the algorithmic nudge treatment as the reference, finding that nudging the algorithm was significantly more effective than nudging the users.

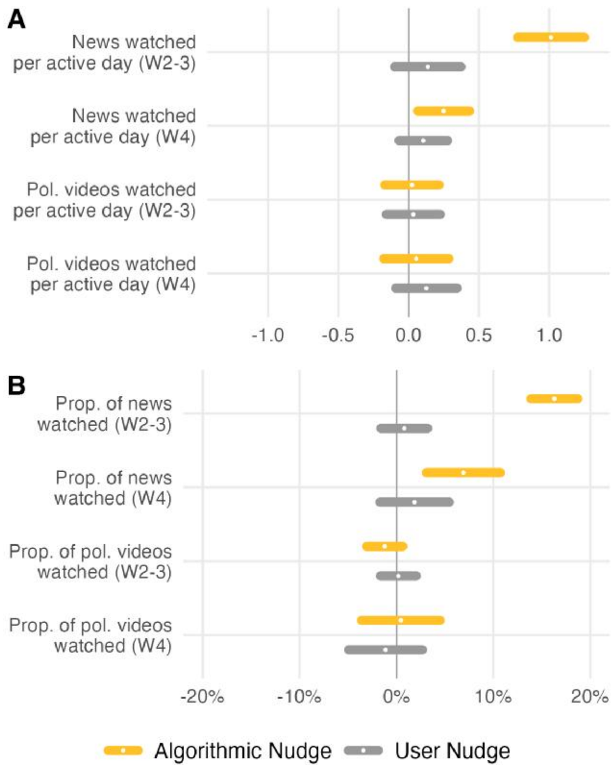


**Fig. 4.** Effects of treatments on the consumption of news videos and political videos. Dots represent coefficients and horizontal bars represent 95% CIs.

*Additional analyses.* We examine the reinforcing feedback loop between recommender system and user behaviors to test if recommendations drive consumption to a greater extent than users' consumption drives recommendations. We ran hierarchical linear regressions predicting the current day's news recommendation from the previous day's news watching, and vice versa (see Supplementary Material, N for full tables).[i] We show that recommendations drive users' on-platform consumption significantly more strongly than the consumption drives subsequent recommendations. A 1% increase in the proportion of news videos recommended led to a 0.16% increase in the proportion of news videos consumed ($P < 0.001$), whereas a 1% increase in news consumption caused a 0.02% increase in news recommendations ($P < 0.001$; $z = 6.02$, $P < 0.001$). Supplementary Material, O details the models and the estimates from 2SLS regressions, which produced similar results.

In addition, in Supplementary Material, M, we present analyses that shed light on the mechanism underpinning the effects of the algorithmic nudge. Also, in Supplementary Material, S, we show the computed transition probabilities between news, political, and other videos. We observe a 1.51× increase in news-to-news transitions and a 1.95× rise in political-to-political video transitions, indicating that algorithmic nudge participants have an increased likelihood to continue viewing news or political videos once started. Lastly, we estimated parallel ordinary least squares (OLS) models testing whether the interventions had any effects on the consumption of videos from radical, conspiratorial, and otherwise problematic channels, such as Alt-right, WhiteIdentitarian, or QAnon (11, 47). We find null effects, detailed in Supplementary Material, L.

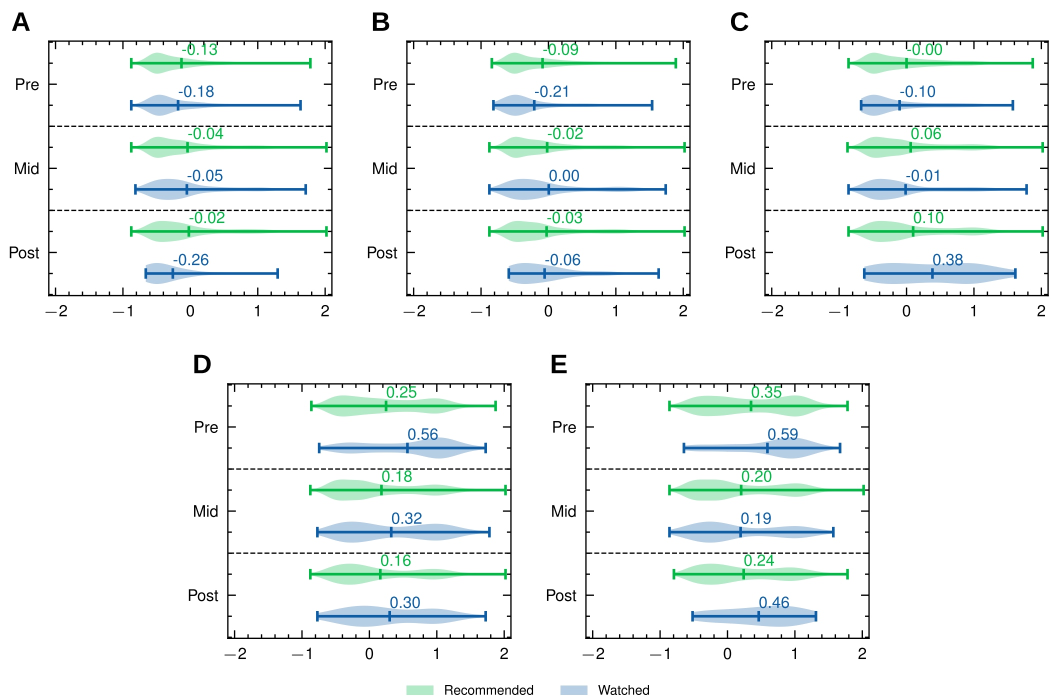


**Fig. 5.** Political slant distribution of the videos recommended and watched, corresponding to the A) Very liberal, B) Liberal, C) Moderate, D) Conservative, and E) Very conservative participants. The three phases of the experiment are specified as pre-intervention (week 1), mid-intervention (weeks 2 and 3), and post-intervention (week 4). The top violin plot in each phase corresponds to the distribution of slant of recommendations and the bottom violin plot corresponds to the distribution of slants of videos watched. The extreme ends represent the interquartile range and the values in-between are the mean slants of the distributions. We see that for the and participants, the recommendations were not as partisan as the videos watched, suggesting that participants self-selected into their ideology rather than recommendations guiding them to do so.

### *Effects on political diversity of recommendations and consumption.*

Did the treatments make users' news diet more diverse and cross-cutting? We estimate the political slant of news and political videos in our dataset by analyzing the audience of each video on Twitter and the political landmarks these audiences follow. The approach, landmarks, score validation, and other details are presented in (46) and the code is available on Github. This approach estimated the slant for 85.2% of all news videos in our data and 72.2% of political videos.

Figure 5 shows the slant distribution of news and political videos recommended to and actually consumed by (i) persons' ideology and (ii) the phase of the intervention.[j] Four clear patterns emerge. First, each user group was recommended news and political videos consistent with users' ideological leaning. This congeniality was especially pronounced for and users (for whom the mean preintervention slant was +0.35 and +0.25, respectively). In contrast, the *Very-liberal* and *Liberal* participants' recommendations were also on the left yet closer to the neutral midpoint (−0.13 and −0.09, respectively).

Second, the algorithmic nudge noticeably and significantly moved the mean slant of the recommended news and political videos toward the center for all groups. This does not necessarily mean that the recommended news and political videos were strictly moderate, but that they were diverse and cross-cutting, thus moving the mean slant toward 0. *Very-liberal* and participants saw the highest shift, with an absolute difference of 0.11 in the mean slant preintervention and postintervention (*Very-liberal* from −0.13 to −0.02; from +0.35 to +0.24). Third, the intervention consequently made the participants' news and political diets *less* congenial. As shown in the middle panel of Fig. 5, the mean slant of the videos watched was closer to 0 during intervention weeks than it was before. These differences between the pretreatment and mid-treatment distributions are statistically significant ($P < 0.05$) for both recommendations and consumption (see Supplementary Material, T).

The last key finding in Fig. 5 regards the videos recommended vs. the videos watched. In almost all cases, the videos the users chose to consume were significantly more congenial than the videos they were recommended (see the second table in Supplementary Material, T). This self-selection to congenial news and political videos was most pronounced for and participants. Preintervention, the average slant of the videos *recommended* to conservatives was 0.25 and 0.35 to very conservative participants, whereas the slant of the videos *watched* was 0.56 and 0.59, respectively. Postintervention, these values were 0.16 and 0.24 (recommended) vs. 0.30 and 0.46 (consumed). A similar—yet less pronounced—pattern of self-selection to congenial content also holds for the *Very-liberal* and *Liberal* users. Yet, caution is needed when interpreting the results from the postintervention week, as attrition resulted in fewer active participants in each ideological group that week.

### *Effects of news consumption on attitudes and beliefs*

Lastly, we test if consumption of news influenced political participation, perceived accuracy of true and false claims, perceived and affective polarization, among other outcomes assessed at the postsurvey at the end of week 3. News exposure was operationalized as the total number of news videos consumed during treatment weeks *and* the difference between the number of news videos consumed per week during treatment weeks and in week 1.[k] As detailed in Supplementary Material, P, OLS regression models found that news exposure had consistent null effects on all survey outcomes. Supplementary Material, P also shows no statistically significant heterogeneous treatment effects on affective polarization and prioritizing partisan ends over democratic means by partisanship strength, ideology, or ideology strength. Additionally, Supplementary Material, Q shows similar null effects of parallel models that did not control for pretest values of the outcomes.

## Discussion

Many observers worry that in their attempt to maximize users' engagement, platform algorithms drive users to one-sided, hyperpartisan, and radical political content. Interest bias in recommendation algorithms, although much more prevalent, is largely overlooked. That is, algorithms create narrow information diets, catering to users' preferences for football, K-pop, outrage, or clickbait, actively redirecting users away from news (19), and thus thwarting users'—especially the marginalized ones'—attention to matters of public interest (4, 20, 58).

This project tackled both issues. We developed and deployed computational interventions aimed at mitigating both interest and ideological biases in YouTube recommendation algorithms. The algorithmic nudge intervention altered a user's YouTube watch history by unobtrusively playing videos from verified and balanced news sources in the background. The user nudge intervention altered the YouTube interface to remind users of the importance and benefits of news consumption. Employing a sock-puppet experiment alongside a month-long randomized experiment on frequent YouTube users, we offer four key findings.

First, nudging YouTube's algorithms by playing news videos in the background increases not only news recommendations but also the actual consumption of news. In fact, we identify a reciprocal loop, whereby recommendations drive exposure to news, which in turn promotes further recommendations, with the first link (recommendations → consumption) being stronger than the second link (consumption → recommendations). In other words, although users' on-platform behavior and content recommendations are intrinsically related, the algorithms seem to matter more to users' information consumption than what the users consume matters to the algorithms, at least in the tested context. This feedback loop amplifies the effects of the algorithmic nudge. Indeed, we observed that a week after the treatment, users in the algorithmic nudge condition were still being recommended and consuming more news than those in the control group. In short, a simple back-end tweak to the algorithms can promote greater engagement with public affairs. We find no effects on recommendations and exposure to political videos from outside news channels (e.g. a vlogger discussing politics or a celebrity endorsing a candidate). This could be because YouTube algorithms pull users away from public affairs (19). While our algorithmic treatment managed to counteract this algorithmic design (after all, we injected news into the users' watch history), it may not have been strong enough to generate a spillover effect on recommendations to political videos.

Second, providing politically balanced news input to the algorithm (recall that our algorithmic intervention was programmed to unobtrusively watch videos from across the political spectrum) leads to more diverse and cross-cutting recommendations and consequently minimizes congeniality in users' political and news exposure. During the treatment, recommendations in the algorithmic nudge condition shifted toward more moderate and cross-cutting, which led to more diverse news consumption. We emphasize, however, that the users were consuming significantly more congenial videos than they were being recommended, particularly the conservative and very-conservative users. Although making news easily available does lead users to consume news, the users exercise agency by self-selecting into specific news that aligns with their ideology, a result also found by Robertson et al. (59) and Hosseinmardi et al. (9). Further work is necessary to tease apart the complex interplay between humans and the algorithms and identify the conditions in which the algorithms vs. people are the primary drivers of information consumption.

Third, nudging users by reminding them that news is beneficial to democracy and personally relevant fails to increase news consumption. This may be due to the limited participants' attention to the nudge, which did not change the behavior that it was theorized and predicted to induce. Only 12.7% participants in the user nudge condition clicked on the *"Watch News"* button (see Supplementary Material, A), and there was no marked increase in search queries for news. In addition, the null effects from the user nudge may be because the participants already reported believing that news consumption is important. In the pretest survey, participants' average agreement with the idea that watching news and politics is crucial for society and themselves was high ($M = 4.05$ out of 5) and there was no significant increase in this self-reported agreement in the user nudge condition compared with those in the algorithmic ($P = 0.959$) and control ($P = 0.060$) conditions.

Fourth, although the algorithmic intervention increased news consumption, this increase had no corresponding effects on a range of survey outcomes, namely political participation, perceived accuracy of false claims, perceived and affective polarization, or support for democratic norms, neither in the aggregate nor across subgroups of participants. This finding adds to the growing evidence that while (algorithmic) interventions can alter users' on-platform exposure, this exposure does not influence attitudes or beliefs (60, 61). We emphasize that this does not mean that promoting news consumption is futile or that the pronounced effects on recommendations and actual news exposure in our algorithmic condition are meaningless. Media effects accumulate over time and our 2-week intervention may have been too short or insufficiently strong for these effects to emerge. It is also possible that increased news exposure may have affected outcomes that we did not measure, such as political efficacy (62), subsequent news seeking (63), or increased awareness of social problems (64). Future studies should include these additional outcomes. We also encourage cross-platform and comparative work testing whether similar results would be detected on other platforms and in other countries.

Our findings carry important implications. Enhancing the accessibility of news on social media not only encourages its consumption but also fosters a reinforcing loop during user-algorithm interactions. This indicates a substantial amount of power wielded by proprietary platform algorithms. Platforms could easily implement an intervention analogous to ours, e.g. upranking and prioritizing verified news content, yet some platforms do just the opposite, redirecting users from news (19) or down-ranking or outright banning such content (65). Although platforms argue that users are not interested in news (66), evidence suggests that users want to see more accurate and educational information (52) and enjoy following news accounts on social media (34). We additionally show that increasing recommendations to news does not decrease users' engagement with YouTube.

That said, we acknowledge that interventions such as ours open many questions. They are built on normative ideals about what citizens should do. Yet, some argue that there should be no interference with users' freedom to produce and consume their desired content (for a review, see (67, 68)). Recent regulations (i.e. the Digital Service Act in the EU or the NetzDG in Germany) and increased attention to platform governance (68, 69), however, recognize that currently, it is the proprietary and commercial algorithms that are shaping online information ecosystem and that regulations reflecting accountability and public interest are needed to guide their development. In addition, such interventions entail decisions about which news organizations or what specific content should be recommended. We selected legacy outlets that were not overly partisan and delivered factual information. As long as other news is not banned or down-ranked, making such news more easily and freely available is not controversial.

Despite these open questions, there are theoretical and normative arguments to be made for encouraging exposure to verified public affairs information on social media platforms. It is also crucial to increase algorithmic transparency and advance measures that build in considerations of societal impacts into decisions as to how content is sorted and displayed to users. Platforms should also revisit their content policies, balancing user engagement and information quality, which are *not* mutually exclusive.

## Notes

[a] We acknowledge that negativity and conflict present in the news may generate stress and anxiety among heavy news users (38). Because most people do not encounter news on platforms (4, 14, 19), our intervention—which does not entail a heavy stream of news—is unlikely to negatively influence users' well-being. We did not measure this outcome at the posttest, so we cannot assess this potential effect.

[b] We emphasize that our algorithmic intervention does not solely "inject" fully moderate news or ban partisan outlets; it simply does not promote strongly partisan news on either side.

[c] Those participants in a large-scale project submitted their online browsing data to researchers. From a total of 129,281,569 web visits, we identified 1,037,392 YouTube video visits by 1,980 participants. The median number of video visits per person was 76 (SD = 1,165). The details on the large-scale project, including extensive informed consent and Institutional Review Board (IRB) approvals, are reported in (54).

[d] We normalize across all other parameters (i.e. number of videos injected) and report the average increase for random and popular video injections.

[e] We assigned participants with unequal probabilities as we expected higher attrition among the treatment groups given that the changes to their platform experience were the greatest.

[f] Supplementary Material, H presents the results of power analysis. The minimum detectable effect for on-platform outcomes (i.e. effects during the treatment weeks) and most survey outcomes is very small ($f^2 < 0.01$) and the effect for the over-time effects in week 4 is small ($f^2 < 0.02$).

[g] We note that the two treatment conditions slightly differed from each other in terms of inactivity without extension deinstallation, such that participants in the algorithmic nudge condition were significantly more likely to be active if they kept the extension installed for the entire 4-week period compared with those in the user nudge condition ($P = 0.006$). This one difference is unlikely to influence our findings, given that it is substantively small and given that we mainly compare the treatment effects with the control.

[h] This drop may be due to the fact that for most participants, the last week of the experiment coincided with Christmas or New Years. Not only is it a slow news cycle but also people may be less inclined to watch news than other content. We tested if news videos were recommended higher up on users' homepages and in up-next

recommendations during the treatment weeks than in week 4, finding no significant differences in the relative placement of news videos.

[i] In addition to the sociodemographic variables mentioned above, the models controlled for the previous day's news recommendation/watching and treatment assignment. The choice of using a daily over a shorter time frame for analysis was dictated by the sparse consumption of news videos relative to nonnews. If a shorter interval was used, the instances with zero news video consumption would have added noise to the data, distorting the findings.

[j] Here, we combine news and political videos; the sparsity of those videos in the data, especially broken down by ideology and the phase of the intervention, would make separate estimates for news and political videos unstable.

[k] The models included age, gender, education, race, partisanship, treatment assignment, pretest values of the outcomes, and news exposure. Perceived accuracy of claims about current events was not measured in the pretest.

## Acknowledgments

The authors are grateful to Rong-Ching (Anna) Chang for the pipeline for video ideology estimation and to Anshuman Chhabra for the development of the political classifier.

## Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

## Funding

The authors gratefully acknowledge the support of the European Research Council, for the Proposal EXPO-756301 (ERC Starting Grant, M.W., PI). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the European Research Council.

## Author Contributions

M.W., M.H., X.Y., and E.M.T. designed the study and study materials; M.H. did sock-puppet experiments and developed the software extension; X.Y., E.M.T., M.W., and M.H. oversaw data collection; X.Y., M.H., and M.W. did data analysis and interpretation; M.W., X.Y., and M.H. wrote the paper; M.W. and X.Y. managed the revision.

## Preprints

This manuscript was posted on a preprint: https://doi.org/10.21203/rs.3.rs-3349905/v1.

## Data Availability

The analytical code and data are made publicly available at https://github.com/ercexpo/Nudging_YT_Algorithms and https://osf.io/t5gwn/, respectively.

## References

1 DeVito MA. 2017. From editors to algorithms: a values-based approach to understanding story selection in the Facebook news feed. *Digit Journal*. 5(6):753–773.

2 Kendall TA, *et al.* 2009. US Patent No. 12/193,705.

3 Kendall T, Zhou D. 2010. Leveraging information in a social network for inferential targeting of advertisements. US Patent App. 12/419,958.

4 Thorson K, Cotter K, Medeiros M, Pak C. 2021. Algorithmic inference, political interest, and exposure to news and politics on facebook. *Inf Commun Soc*. 24(2):183–200.

5 Staff WSJ. 2021. Inside Tiktok's highly secretive algorithm. https://www.wsj.com/video/series/inside-tiktoks-highly-secretive-algorithm/investigation-how-tiktok-algorithm-figures-out-your-deepest-desires/.

6 Chen AY, Nyhan B, Reifler J, Robertson RE, Wilson C. 2023. Subscriptions and external links help drive resentful users to alternative and extremist YouTube channels. *Sci Adv*. 9(35):8080.

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

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

9 Hosseinmardi H, *et al.* 2021. Examining the consumption of radical content on YouTube. *Proc Natl Acad Sci U S A*. 118(32):e2101967118.

10 Ledwich M, Zaitsev A, Laukemper A. 2022. Radical bubbles on YouTube? Revisiting algorithmic extremism with personalised recommendations. *First Monday*. 27(12). doi:10.5210/fm.v27i12.12552

11 Ribeiro MH, Ottoni R, West R, Almeida VAF, Meira Jr W. 2020. Auditing radicalization pathways on YouTube. In: ACM Conference on Fairness, Accountability, and Transparency (FAT*). New York (NY): Association for Computing Machinery.

12 McClain C, Widjaya R, Rivero G, Smith A. 2021. The behaviors and attitudes of U.S. adults on Twitter. https://policycommons.net/artifacts/1894717/the-behaviors-and-attitudes-of-us/2644730/.

13 Wells C, Thorson K. 2017. Combining big data and survey techniques to model effects of political content flows in Facebook. *Soc Sci Comput Rev*. 35(1):33–52.

14 Meta. 2022. Widely viewed content report: what people see on Facebook. https://transparency.fb.com/data/widely-viewed-content-report/#intro.

15 Wojcieszak M, Casas A, Yu X, Nagler J, Tucker JA. 2022. Most users do not follow political elites on Twitter; those who do show overwhelming preferences for ideological congruity. *Sci Adv*. 8(39):eabn9418.

16 Rathje S, Van Bavel JJ, Van Der Linden S. 2021. Out-group animosity drives engagement on social media. *Proc Natl Acad Sci U S A*. 118(26):e2024292118.

17 Yu X, Wojcieszak M, Casas A. 2024. Partisanship on social media: in-party love among American politicians, greater engagement with out-party hate among ordinary users. *Polit Behav*. 46(2):799–824.

18 Rossi WS, Polderman JW, Frasca P. 2021. The closed loop between opinion formation and personalized recommendations. *IEEE Trans Control Netw Syst*. 9(3):1092–1103.

19 Huang S, Yang T. 2024. Auditing entertainment traps on YouTube: how do recommendation algorithms pull users away from news. *Polit Commun*. 41(6):903–920.

20 Thorson K. 2020. Attracting the news: algorithms, platforms, and reframing incidental exposure. *Journalism*. 21(8):1067–1082.

21 Bartels LM. 1996. Uninformed votes: information effects in presidential elections. *Am J Pol Sci*. 40(1):194–230.

22 Lau RR, Andersen DJ, Redlawsk DP. 2008. An exploration of correct voting in recent US presidential elections. *Am J Pol Sci*. 52(2):395–411.

23 Lau RR, Redlawsk DP. 2006. *How voters decide: information processing in election campaigns*. Cambridge University Press.

24 Althaus SL. 1998. Information effects in collective preferences. *Am Pol Sci Rev*. 92(3):545–558.
25 Brennan J. 2017. *Against democracy*. Princeton University Press.
26 Delli Carpini MX, Keeter S. 1996. *What Americans know about politics and why it matters*. Yale University Press.
27 Prior M. 2005. News vs. entertainment: how increasing media choice widens gaps in political knowledge and turnout. *Am J Pol Sci*. 49(3):577–592.
28 Prior M. 2007. *Post-broadcast democracy: how media choice increases inequality in political involvement and polarizes elections*. Cambridge University Press.
29 Barabas J, Jerit J. 2009. Estimating the causal effects of media coverage on policy-specific knowledge. *Am J Pol Sci*. 53(1):73–89.
30 Jerit J, Barabas J, Bolsen T. 2006. Citizens, knowledge, and the information environment. *Am J Pol Sci*. 50(2):266–282.
31 Lupia A, McCubbins MD. 1998. *The democratic dilemma: can citizens learn what they need to know?* Cambridge University Press.
32 Altay S, Nielsen RK, Fletcher R. 2024. News can help! the impact of news media and digital platforms on awareness of and belief in misinformation. *Int J Press/Polit*. 29(2):459–484.
33 Mont'Alverne C, *et al.* 2024. The electoral misinformation nexus: how news consumption, platform use, and trust in news influence belief in electoral misinformation. *Public Opin Q*. 88: 681–707. doi:10.1093/poq/nfae019.
34 Altay S, Hoes E, Wojcieszak M. 2024. News on social media boosts knowledge, belief accuracy, and trust: a field experiment on Instagram and WhatsApp. doi:10.31234/osf.io/hq5ru.
35 Delli Carpini MX. 2000. In search of the informed citizen: what Americans know about politics and why it matters. *Commun Rev*. 4(1):129–164.
36 Delli Carpini MX, Keeter S. 2002. The internet and an informed citizenry. https://repository.upenn.edu/asc_papers/2/.
37 Hayes D, Lawless JL. 2015. As local news goes, so goes citizen engagement: media, knowledge, and participation in US house elections. *J Polit*. 77(2):447–462.
38 Villi M, *et al.* 2022. Taking a break from news: a five-nation study of news avoidance in the digital era. *Digit Journal*. 10(1):148–164.
39 Barberá P, Jost JT, Nagler J, Tucker JA, Bonneau R. 2015. Tweeting from left to right: is online political communication more than an echo chamber? *Psychol Sci*. 26(10):1531–1542.
40 Fletcher R, Robertson CT, Nielsen RK. 2021. How many people live in politically partisan online news echo chambers in different countries? *J Quant Descr Digit Media*. 1:1–56.
41 Abramowitz AI, Saunders KL. 2008. Is polarization a myth? *J Polit*. 70(2):542–555.
42 Barberá P, *et al.* 2019. Who leads? who follows? measuring issue attention and agenda setting by legislators and the mass public using social media data. *Am Pol Sci Rev*. 113(4):883–901.
43 Krupnikov Y, Ryan JB. 2022. *The other divide*. Cambridge University Press.
44 Skovsgaard M, Andersen K. 2020. Conceptualizing news avoidance: towards a shared understanding of different causes and potential solutions. *Journal Stud*. 21(4):459–476.
45 Auxier B, Anderson M. 2021. Social Media Use in 2021. Pew Research. https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/.
46 Wojcieszak M, Chang R-CA, Menchen-Trevino E. 2023. Political content and news are polarized but other content is not in Youtube watch histories. *J Quant Descr Digit Media*. 3:1–63.
47 Ledwich M, Zaitsev A. 2020. Algorithmic extremism: examining YouTube's rabbit hole of radicalization. *First Monday*. 25(3). doi: 10.5210/fm.v25i3.10419
48 Schudson M. 1998. *The good citizen: a history of American civic life*. Free Press.
49 Bucy EP, D'Angelo P. 2004. Democratic realism, neoconservatism, and the normative underpinnings of political communication research. *Mass Commun Soc*. 7(1):3–28.
50 Kreiss D. 2021. "Social media and democracy: the state of the field, prospects for reform," edited by Nathaniel Persily and Joshua A. Tucker. *Int J Press/Polit*. 26(2):505–512.
51 Lazer DMJ, *et al.* 2018. The science of fake news. *Science*. 359(6380): 1094–1096.
52 Rathje S, Robertson C, Brady WJ, Van Bavel JJ. 2023. People think that social media platforms do (but should not) amplify divisive content. *Perspect Psychol Sci*. 19(5):781–795
53 Baum MA, Jamison AS. 2006. The oprah effect: how soft news helps inattentive citizens vote consistently. *J Polit*. 68(4):946–959.
54 Wojcieszak M, *et al.* 2023. Non-news websites expose people to more political content than news websites: evidence from browsing data in three countries. *Polit Commun*. 41:129–151.
55 Ad Fontes Media. 2022. https://adfontesmedia.com/interactive-media-bias-chart/ [accessed 2024 Jul 24].
56 Lin H, *et al.* 2023. High level of correspondence across different news domain quality rating sets. *PNAS Nexus*. 2(9):pgad286.
57 Wojcieszak M, *et al.* 2021. No polarization from partisan news: over-time evidence from trace data. *Int J Press Politics*. 28: 601–626. doi:10.31219/osf.io/hqmuy.
58 Cotter K, Medeiros M, Pak C, Thorson K. 2021. "Reach the right people": the politics of "interests" in Facebook's classification system for ad targeting. *Big Data Soc*. 8(1):2053951721996046.
59 Robertson RE, *et al.* 2023. Users choose to engage with more partisan news than they are exposed to on Google search. *Nature*. 618(7964):342–348.
60 Nyhan B, *et al.* 2023. Like-minded sources on Facebook are prevalent but not polarizing. *Nature*. 620(7972):137–144.
61 Guess AM, *et al.* 2023. Reshares on social media amplify political news but do not detectably affect beliefs or opinions. *Science*. 381(6656):404–408.
62 Moeller J, De Vreese C, Esser F, Kunz R. 2014. Pathway to political participation: the influence of online and offline news media on internal efficacy and turnout of first-time voters. *Am Behav Sci*. 58(5):689–700.
63 Karnowski V, Kümpel AS, Leonhard L, Leiner DJ. 2017. From incidental news exposure to news engagement. How perceptions of the news post and news usage patterns influence engagement with news articles encountered on Facebook. *Comput Human Behav*. 76:42–50.
64 Feezell JT. 2018. Agenda setting through social media: the importance of incidental news exposure and social filtering in the digital era. *Polit Res Q*. 71(2):482–494.
65 Taylor J. 2024. *Meta says Facebook cannot solve media industry's "issues" as it defends ending payments for news in Australia*. The Guardian. https://www.theguardian.com/media/2024/mar/14/meta-facebook-news-media-bargaining-code.
66 Taylor J. 2024. *Meta claims news is not an antidote to misinformation on its platforms*. The Guardian. https://www.theguardian.com/media/article/2024/jul/09/meta-facebook-australia-news-ban-misinformation.
67 Gillespie T. 2018. *Custodians of the internet: platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.
68 Gorwa R. 2019. What is platform governance? *Inf Commun Soc*. 22(6):854–871.
69 Rochefort A. 2020. Regulating social media platforms: a comparative policy analysis. *Commun Law Policy*. 25(2):225–260.