Election Fraud, YouTube, and Public Perception of the Legitimacy of President Biden

James Bisbee, Megan A. Brown, Angela Lai, Richard Bonneau, Joshua A. Tucker, Jonathan Nagler

Abstract. Skepticism about the outcome of the 2020 presidential election in the United States led to a historic attack on the Capitol on January 6th, 2021 and represents one of the greatest challenges to America's democratic institutions in over a century. Narratives of fraud and conspiracy theories proliferated over the fall of 2020, finding fertile ground across online social networks, although little is known about the extent and drivers of this spread. In this article, we show that a convenience sample of 361 YouTube users who were more skeptical of the election's legitimacy were more likely to be recommended content that featured narratives about the legitimacy of the election in November and December of 2020. Our findings underscore the tension between an "effective" recommendation system that provides users with the content they want, and a dangerous mechanism by which misinformation, disinformation, and conspiracies can find their way to those most likely to believe them.

1 Introduction

The 2020 U.S. presidential election left the voting public divided, with Biden voters overwhelmingly believing that the outcome was legitimate, and Trump voters expressing much more skepticism (**zilinsky_nagler_tucker_2021**). These patterns are not unprecedented – supporters of the losing candidate have always been more skeptical of an election's results (**pennycook2021examining**). However, Trump supporters held a set of beliefs about election fraud, and a range of detailed theories about its causes, that was unprecedented. Alarmingly, these beliefs were enough to motivate skeptics to attend a #StopTheSteal rally organized by President Trump in Washington DC on January 6th, 2021 and – for a subset of these individuals – to commit violent criminal acts in storming the US Capitol as the election results were being certified.

Why did the United States, a country whose tradition of peaceful power transitions has been a point of national pride, falter so dramatically in the winter of 2020? There are myriad explanations, ranging from increasing mass polarization (**barber2015causes**) to declining cross-cutting identities (**mason2018uncivil**) to the pernicious consequences of conspiracy theories and misinformation online (**allcott2017social**; **allcott2019trends**; **aslett2022news**). Of particular concern among the popular press is the role played by online recommendation algorithms, which are thought to contribute to echo chambers, filter bubbles, and radicalization (**tufekci2018youtube**; **weill2018youtube**; **nicas_2018**). Yet there is little evidence to support this claim in scholarly

work which primarily examines whether partisans are exposed to different streams of information (**barbera2015tweeting**; **guess2018avoiding**; **bakshy2015exposure**; **ledwich2020algorithmic**). To the extent that such online echo chambers even exist in the first place (**guess2018avoiding**), they are thought to be the product of intentional human behaviors, not recommendation algorithms (**chen_and_nyhanadl2021**).

In this article, we focus on a specific type of content – YouTube videos about fraud in the 2020 U.S. presidential election – to test whether online recommendation systems potentially contributed to a polarized information environment in which content about Trump's claims were disproportionately suggested to participants who were most likely to believe them. We find this to be the case, providing a cautionary brake a growing consensus that recommendation algorithms have little influence on what users of social media platforms consume (**guess2018avoiding**; **chen_and_nyhanadl2021**; **hosseinmardi2020evaluating**).

To make this case, we gathered data on the YouTube videos recommended to a convenience sample of 361 American YouTube users between October 29th and December 8th of 2020, and show that those participants most skeptical about the legitimacy of the election were recommended disproportionately more fraud-related content than non-skeptic participants. However, our results suggest that the overall prevalence of these types of videos was low, with the most skeptical participants being suggested only 8 more videos than the least skeptical participants, out of approximately 410 total recommendations being shown to each participant in our study. Furthermore, stance analysis suggests that the majority of this difference in recommended content was comprised of a mixture of videos that endorsed Trump's false claims of election fraud, and those that reported neutrally his claims. Yet among those participants who were specifically concerned that fraudulent ballots were being counted, and invalid ballots were being discarded – both of which were key dimensions of Trump's claims – videos endorsing Trump's false claims were significantly more prevalent in their recommendations.

Taken together, our findings address the role played by automated recommendation systems for an increasingly polarized electorate. While an effective algorithm that can cater content on a user-specific basis is not inherently problematic, our analysis isolates a context in which a small number of participants were shown a small number of videos that may have reinforced their concerns about a stolen election. And while we have no reason to believe that our specific participants went on to participate in the January 6th, 2021 rally, we argue our findings suggest that people most skeptical of the election results were more likely to be shown videos on the topic of fraud, which may have increased or maintained their interest.

2 Context and Theory

YouTube was the most popular social media platform among U.S. adults in 2020, and second only to Facebook as the primary source of news among the same (**pew2019social**). Videos on the platform can be searched for directly, shared via URL across other parts of the internet, or recommended on a user-by-user basis via (potentially several) bespoke algorithms.[1] Among these sources, the recommendation algorithm's influence is preponderant, responsible for approximately 70% of what users actually watch (**solsman_2018**).

While the algorithm's inner workings are a trade secret, its core purpose is to maximize user engagement. As discussed by Google engineers in **covington2016deep**, "engagement" is defined as clicks per impression, meaning that the algorithm is intended to recommend videos on which a given user is most likely to click. Although such a tailored system can be valuable in many contexts, we posit that it lead to a subset of Americans who were already skeptical about the legitimacy

---

1. It is not clear whether the recommendation algorithm that suggests videos on the sidebar of a given YouTube page is the same as the algorithm that suggests content on a user's homepage, although we assume that their core purpose is similar enough for the theory that follows.

of the 2020 presidential election to be shown content that reinforced their concerns, potentially exacerbating the polarized information environment in which the January 6th insurrection took place.

Our claim connects with a broad and growing literature on "echo chambers" – information environments in which an individual only sees content that reaffirms her pre-existing beliefs. These information environments (also referred to as "filter bubbles") may occur due to user behaviors (i.e., only clicking on links from ideologically congruent sources), social networks (i.e., conservative users only friending / following other conservative users), or platform algorithms (i.e., YouTube only suggesting conservative content to conservative users) (**barbera2015tweeting**; **garrett2009echo**; **guess2018avoiding**). Despite the popularity of the concept, a recent review finds mixed evidence regarding the prevalence of online echo chambers (**barbera2020social**). Moreover, research on social media algorithms specifically has, if anything, suggested that any echo chambers that do exist are more likely to be driven by user choices than by underlying algorithms (**bakshy2015exposure**; **ribeiro2020auditing**; **ledwich2020algorithmic**).

Nevertheless, the theoretical intuition for why a recommendation algorithm might nudge users into an echo chamber follows straightforwardly from a basic understanding of how the algorithms are designed. YouTube's algorithm uses a combination of user metadata (i.e., demographic characteristics, AdSense metrics, etc.), their watch histories (the record of every video they have watched or engaged with in the past), and account-level information such as subscriptions, to predict which videos users will be most interested in.[2] In practice, YouTube engineers treat this as an extreme multiclass classification challenge in which the target of interest is accurately predicting the video watched at time t (w_t), conditional on high-dimensional embeddings for the user-context ($\mathbf{X}, C$), and high-dimensional embeddings for the video j (v_j), which are learned via a deep neural network (**covington2016deep**). With some slight changes to notation for tractability with our theory, YouTube's optimization function as of 2016 is represented as

$$P(w_t = j | \mathbf{X}, C) = \frac{e^{v_j x}}{\sum_{k \in V} e^{v_k x}} \tag{1}$$

where $k \in V$ is the universe of content on YouTube.

2.1 Theory

Why might such an algorithm recommend videos about election fraud conspiracies to users most likely to believe these stories? We start from the observation that these metadata and watch histories are only proxies for the underlying concepts of interest: user i's utility for watching a specific video j, or $U_i(v_j)$. In line with existing work (**barbera2013birds**; **eady2019media**), we use a random utility model to formalize $U_i(v_j)$ as a function of the user's ideal point α_i and the content of the video a_j. Formally,

$$U_i(v_j) = -||\alpha_i - a_j||^2 \tag{2}$$

In applied work, α_i and a_j are typically formulated as a dimension-reduced measure of ideology. In our setting, we conceive of α_i as the user's prior belief about whatever topic v_j is about, and argue – consistent with a body of research spanning sociology, psychology, economics, and political science – that information which challenges existing beliefs is more costly than information which doesn't (see nickerson1998confirmation,levy2015correlation,little2022motivated for reviews and contributions in the fields of psychology, economics, and political science). When the content of a video a_j is further from the user's ideal point α, the user will experience lower utility and therefore be less likely to watch the video. Substantively, this would mean that videos disputing election fraud would provide less utility to users who believe in Trump's narrative of a stolen election. This leads to our core hypothesis **H1**: *Participants more skeptical about the legitimacy of the 2020*

---

2. AdSense is a Google tool that allows web publishers to monetize their content. AdSense works by matching ads with website content and visitors.

presidential election will be recommended more content supporting the view that the election was fraudulent.[3]

Typically, researchers would control for user-specific covariates in order to isolate the association between the recommendations provided by the algorithm and user skepticism about the legitimacy of the election. In a standard causal DAG, user-specific covariates might be considered a source of omitted variable bias. For example, if the association between skepticism and recommendations does not hold after controlling for user partisanship, the researcher would conclude that partisanship is causally prior to both election skepticism and recommendations and that failing to condition the regression on partisanship produces a spurious association between the user skepticism and the algorithm.

However, based on what we know about YouTube's recommendation algorithm, these covariates are a core component of the personalized data used by the algorithm's ability to infer the user's α_i parameter. More specifically, YouTube's recommendation algorithm learns about α_i from $\mathbf{X}_i$ and C; and about a_j from v_j. As such, controlling for these covariates in a standard setting effectively holds constant the source of variation required by the algorithm to infer user preferences for recommendations. More generally, our research question is interested in whether real users were differentially suggested content containing election misinformation in the fall of 2020 based on their predisposition to believe such content. A null result from a specification that controls for user covariates might accurately reflect the inability of the recommendation algorithm to differentially suggest misinformation to two Republicans, only one of which is more skeptical of the election's legitimacy, but such a finding obscures the substantively important fact that the more skeptical participants in our study were disproportionately more likely to be recommended misinformation.

As such, our main results do not control for user covariates in order to present the overall picture of how real users experienced YouTube's recommendations in the fall of 2020. However, in SI section 4.3 we include a series of robustness checks that add basic demographic controls, including age, gender, ethnicity, educational attainment, and partisanship, finding that our results are attenuated but persist. Importantly, we also control for the content of participants' watch histories among the subset who provided us with this information, again finding that the significant positive association persists (SI Section 4.4). Taken together, our analysis isolates an independent effect of the algorithm that targeted users who were skeptical about the election with recommendations for videos about election fraud.

3 Data Collection Strategy

We are fundamentally interested in testing whether YouTube's recommendation algorithm suggested videos about election fraud to users who were most skeptical about the election's legitimacy. We operationalize these quantities of interest with topic models of video metadata to identify which videos are about election fraud (the outcome variable), and ask our respondents a battery of questions about their belief in election fraud narratives (the main explanatory variable). However, before describing these measures in detail, we start with a discussion of the steps taken to measure the independent effect of the recommendation algorithm.

Isolating the effect of a recommendation algorithm is – perhaps surprisingly – a complicated statistical challenge (**wagner_measuring_2021**). On the one hand, relying on observational data

---

3. Our application of the spatial model to the choice of which recommendation to click is necessarily simplistic. Alternative considerations may and likely do exist. For example, users might obtain perverse utility from grotesque or extreme content even if it is far from their prior α_i. Or there might be non-content related characteristics of the video (i.e., the thumbnail or clickbait title) which are unevenly distributed across different types of videos. While interesting extensions to the core intuition, these are beyond the scope of our analysis. We argue that our framework, while overly simplistic, is nevertheless useful for guiding the intuition about why an effective recommendation algorithm might suggest prior-confirming content to users.

(i.e., user watch histories) confounds the suggestions of the recommendation algorithm with the preferences of actual users. On the other hand, using automated methods of data collection such as YouTube's API or bespoke web scraping programs removes the personalized data that are essential to capturing an ecologically valid snapshot of the algorithm's influence on content consumed on the platform by actual human users of the platform – arguably our key population of interest.[4]

To address these concerns, we fielded a survey experiment in the fall of 2020 in which we carefully controlled the behavior of real YouTube users while they were on the platform. We build on the prior literature in the social sciences for auditing algorithmic systems for bias. Originating in offline contexts, these audits identified discrimination in areas such as jobs, housing, mortgaging, loan lending, or credit card financing (**auditing_discrimination**). In the digital age, these biases have persisted in digital personalization algorithms, where studies have shown racial and gender bias in online ads and job recommendation systems (**gender_discrimination_in_ads**; **racial_discrimination_in_ads**). With the rise of social media and concern about echo chambers and filter bubbles, researchers have adapted traditional auditing methodology to online contexts to identify political bias in Google search results, Twitter search results, Twitter's news feed ranking system, and more (**robertson_google**; **hannak_web_search**; **klimansilver_websearch**; **kulshrestha_twittersearch**; **twitter_bias**). Using similar methodology, we recruited participants by advertising on Facebook between September and December of 2020 and restricting our sample to respondents who lived in the United States and had a YouTube account. Participants were asked to install a temporary browser extension that downloaded the list of recommended videos they were shown (**ji_youtube_plugin**), and then were asked to navigate through 20 videos by clicking on one of the recommendations as proscribed by the treatment arm they were in at each step, which we refer to as a "traversal".[5] Crucially, we required that they complete this task while logged in to their YouTube accounts, ensuring we are capturing what real users experienced in real time.[6]

To isolate the independent effect of the algorithm on what was recommended to our participants, we experimentally manipulated our participants' experience on YouTube in two ways. First, we randomly assigned them to begin their time on the platform on one of a series of pre-selected videos (their "seed" video). Forcing them to start on a randomly assigned seed ensures that any patterns we observe are not partially driven by the participant's preference at the initial point of data collection. The list of seed videos was selected to include fifteen political videos across the ideological spectrum and nine videos from nonpolitical categories. To keep these current and to replace any that might be taken down or set to private, we updated the list weekly. In all our regressions, we control for seed video fixed effects to isolate the differences in recommendations that were shown to participants who started on the same seed video but differed in their concerns about the election's legitimacy. For a list of the seed videos, see the SI section 10.

Second, we randomly assigned them a pre-determined recommendation to click on at each of the 20 steps in their traversal, which we refer to as a "traversal rule". For example, a respondent who was randomly assigned to a traversal rule of only clicking on the second recommendation would always click on the second in the list of recommendations at each step, regardless of the

---

4. There are recent innovations in the use of web scraping that incorporate researcher-determined "cookies" that include simulated personalized data (**haroon2022youtube**). However, it is not yet clear the degree to which these methods accurately approximate the personalized data associated with a real user's account, nor yet what relying on un-personalized data mean for the validity of the measures gathered (**narayanan_2019**; **ledwich2020algorithmic**).

5. Our plugin was neither a bot nor a tool built using the YouTube API, but rather copied the raw HTML of a YouTube page, capturing exactly what the user saw, similar to the tool described in (**chen_and_nyhanadl2021**). We then parsed the raw code to identify the list of recommended videos, which we saved in a .json file that was associated with a unique ID assigned to our participant. Typically, this procedure gathered about 20 recommendations for each page, although more would appear if the user had a higher resolution monitor, was zoomed out, or scrolled down. More details can be found in SI Section 1.

6. We don't control for subscriptions in our analysis, which means we are unable to disentangle a story in which YouTube's algorithm puts its finger on the scale of content, versus one in which election skeptic users subscribe to a certain set of channels that started to produce content about election fraud in the fall of 2020. Nevertheless, our results are robust to controlling for user watch histories among the subset of participants who provided these data. A more detailed discussion can be found in SI Section 4.4.

content of that video. As above, this manipulation holds constant the possibility that skeptical participants might click on a particular recommendation about election fraud, which would then lead them to subsequent recommendations on the same topic based on their revealed preferences. Put another way, these restrictions on participant behavior ensure that any correlations between recommendations and participant skepticism are the product of the recommendation algorithm, with our respondents acting as "confederates" in the traditional audit design terminology.[7]

We are able to precisely identify participant compliance with these instructions in the data we collect via the plugin. Specifically, we can confirm that each participant clicked on their randomly assigned seed video link by checking whether the first video they start on is the same as the one provided in the link. In addition, we can confirm that they followed their assigned traversal rule by checking whether the video they clicked on was in the correct position of the list of recommendations on the preceding page. In some cases, a user assigned to a traversal rule of 2 would click on the third recommendation on a few pages. The richness of our scraped data ensures that we are able to further confirm that this was not user error, but rather due to the second recommendation being a YouTube movie or an advertisement, both of which we explicitly instructed users to avoid while completing the traversal task.[8] Overall, there were few instances of systematic non-compliance (i.e., where the user would not wait the required amount of time for the plugin to work, or would click seemingly at random, or would terminate the traversal task after only a few steps), and we removed these participants from our analysis.

In addition to this traversal task, participants also filled out a short survey that collected basic demographic information, as well as their opinions on a number of political topics.[9] Those who responded to our survey between October 29th and December 8th were asked a battery of questions about the 2020 U.S. presidential election between Donald Trump and Joe Biden, several of which captured respondents' skepticism about the validity of the election and concerns regarding possible fraud, which we treat as the main explanatory variables of interest.

In order to maximize participant attention during the traversal task, we asked them to complete this more complicated procedure first before giving them the short survey. As such, an alternative interpretation of our data might be that participants' skepticism about the election's legitimacy was causally affected by the videos they were recommended during the data collection procedure, since we asked these questions after they had completed the traversal task. We argue that this interpretation is unlikely for the simple reason that participants did not spend a sufficient amount of time on any video to be plausibly influenced by the appearance of fraud-endorsing videos in their list of recommendations, or by spending more time watching fraud-endorsing videos. We test this alternative interpretation in our Supporting Information, Section 5, showing that 1) there is no relationship between watch time and fraud content, 2) predicting beliefs as a function of fraud content watched interacted with time watched is a null (or negative) relationship, and 3) our results are robust to dropping respondents who spent longer than a certain threshold of average time on recommended content.

Depending on the specific date of participation and the measure of skepticism used, our final sample ranges between 354 Americans (when we use questions about the election's outcome, which were asked from November 4th) and 361 (when we use measures of concern about election fraud, which were asked from October 29th). These numbers drop to 331 respondents for post-election analysis,

---

7. It is clear that the randomly assigned seed video and the recommendations a user clicks on according to their randomly assigned traversal rule both influence the recommendations shown at any given traversal step. For example, the recommendations on a video about sports include many suggestions for other videos about sports. However, we argue that it is unlikely these randomized components alter the *personalized* component of the recommendations, which is our core quantity of interest. To isolate this component of the algorithm, we include seed video and traversal rule fixed effects in in our main analysis to confirm that our results hold when comparing two users with differing levels of skepticism who started on the same seed video and followed the same traversal rule.

8. Clicking on ads would take them away from YouTube, while the recommendations on a YouTube movie are comprised entirely of other YouTube movies, representing a very specific type of echo chamber that we were not interested in for this study.

9. A detailed description of the sequencing of the survey is included in the Supporting Information, Section 1.

and 338 respondents for concern analysis, after removing non-compliant participants with errors in their traversal data. Importantly, our convenience sample is neither nationally representative nor even representative of YouTube users (descriptive statistics of our respondents are summarized in the SI, section 1). The majority of our participants are Democrats who are better educated and younger than the American population, and disproportionately took our survey on desktop or laptop computers running some variant of the Windows operating system.[10] This skew in our convenience sample is particularly important given our substantive interest in measuring whether YouTube's algorithm recommended content about election fraud to users most likely to believe it, a group we assume is more conservative and Republican leaning. As such, we emphasize that our findings contribute evidence of a systematic *pattern* of behavior in the recommendation algorithm, but shouldn't be used to inform the *extent* of misinformation on the platform in the fall of 2020.[11] Nevertheless, by experimentally instructing the process by which real users were exposed to recommended content, we trade limited external validity for rich internal validity to collect novel data on the recommendation algorithm of one of the most popular online social networks during a crucial period of American political history.

3.1   Measuring Skepticism

Our explanatory variable of interest is participants' pre-existing belief in election fraud, which we theorize drives their utility when consuming content (parameter α_i in Equation 2). To capture this concept, we asked respondents who participated between October 29th and December 8th a battery of questions about the 2020 presidential election, including a number of questions about election fraud. These included questions about their concern for fraudulent ballots being counted, valid ballots not being counted, non-U.S. citizens voting, and interference by foreign governments. Respondents were asked to indicate their level of concern on a 0 to 100 scale, where zero indicates no concern and 100 indicates extreme concern. We refer to this group of questions as the *fraud battery*.[12]

In addition, we asked those respondents who participated after November 4th to express their opinion about who won the election, whether this outcome was legitimate, and whether Trump should concede or contest the result.[13] These were recorded as yes/no binaries. We refer to this group as the illegitimacy battery, capturing a shared belief that the post-election results were unjust.

Our analyses look at each predictor in isolation, as well as indices constructed by binarizing the continuous measures and rescaling such that positive values indicate skepticism / concern and negative values indicate confidence. The combined indices are created by summing across these -1/+1 binaries. A detailed description of our data are included in the Supporting Information.

Without access to YouTube's trade secret algorithm, we can't confidently claim that the recommendation system infers a user's appetite for election fraud content using their past watch histories, their

10. The computer differences are partially due to our requirement that participants log into their YouTube account on a Chrome-based browser and not via a mobile device. We discuss these requirements and their implications for generalizability in more detail in SI section 1.

11. Given that our respondents were predominantly better educated and liberal than the general public and we know that conservatives were more likely to think the election results were fraudulent, one might view our aggregate findings as a likely lower bound on the empirical relationship that exists in the population.

12. Foreign interference is not typically associated with election fraud in the literature. We choose to categorize it as fraud in the context of the 2020 US presidential election because concerns about fraudulent ballot machines built by SmartMatic and Dominion were couched in fears that, because these companies were partially or wholly foreign-owned, they were anti-Trump. Or, conversely, there was persistent concern among Democrats and those on the left that Russia intervened in the 2016 election to help Trump and might do so again in 2020. The detailed wording of our questions are included in the SI.

13. The survey made the legitimacy question contingent on the "who won" question, meaning that those who responded "we don't yet know" were not shown the legitimacy question. However, we failed to make the concede / contest question similarly contingent, meaning that we have several respondents who indicated in the initial days after the election that "we don't yet know", but who were also forced to indicate whether they thought Trump should concede or contest. We drop these observations from our main analyses.

demographic data, or some combination of both. For the purposes of our contribution, we treat the algorithm as the black box that it is, and instead simply ask whether it will disproportionately recommend election fraud content to those users who are more skeptical of the election's legitimacy to begin with.

3.2   Identifying Fraud Content

Our outcome measure of interest is the number of videos about election fraud recommended to each user, which we theorize are chosen by YouTube's algorithm to minimize the distance between the user's skepticism (parameter α_i in Equation 2) and the content of the recommendation (parameter a_j in Equation 2). We use unsupervised topic models of the video metadata (title, description, and video tags) to characterize the content of the recommended videos shown to our respondents.[14] Specifically, we estimate topic models via Latent Dirichelet Allocation (LDA) (**blei2003latent**), which uses the co-location of terms (words) within documents (video metadata) to estimate the topic(s) of conversation in each video. A crucial hyperparameter that must be set by the researcher is the choice of the total number of topics, k. Our main results set the total number of topics to 150, although we provide extensive robustness checks in our Supporting Information, Section 4.2 to demonstrate the durability of our conclusions to different choices of k.

Of primary interest to us are the topic-word probabilities $\phi_{w,k}$ that express the probability of word w occurring in topic k, and the topic-video probabilities $\theta_{k,v}$ that express the probability of topic k occurring in video v. We calculate a fraud score for each topic as the sum of the topic-word probabilities relating fraud keyword w to topic k, $\phi_{w,k}$.[15] Figure 1 summarizes the highest-scoring fraud topic (topic #108). The top-10 highest scoring terms are indicated in grey with the x-axis indicating the $phi_{w,k}$ value for each term. The total number of fraud keywords are indicated in red, with the legend indicating the $\phi_{w,k}$ value for each in parentheses. As illustrated, the highest scoring fraud topic has 37 total keywords appear among the related terms, with a net $\phi_{w,k}$ of almost 4%.

Looking at the most relevant terms can help us understand *what* the topic is, but not the *stance* of videos on the topic. By "stance" we mean the perspective embodied by the content, which might neutrally cover a news story about Trump's attempts to overturn the election results, might dispute or disparage Trump's claims, or might endorse and support Trump's claims. Substantively, we are more concerned with content that neutrally reports on, or explicitly endorses election misinformation.[16] Figure 2 summarizes the top 50 videos associated with topic #108 based on the topic-video probabilities, $\theta_{k,v}$. As illustrated, the videos most strongly associated with the fraud topic #108 are either from the White House itself, or by NewsNOW, an affiliate of the Fox News network that is heavily biased toward Trump. The video titles strongly suggest that topic #108 represents content that promotes Donald Trump's narrative of a stolen election. Furthermore, among the minority of video titles that don't explicitly suggest an endorsement of Trump's claims, the titles nevertheless suggest neutral reporting on the topic.

However, we caution that LDA results are a mixture of topics across documents (or in our case, videos). While the highest scoring videos indeed appear to almost universally endorse Trump's claims, this doesn't mean that content which scores lower on this continuous measure adopts the same stance. As an additional test, we manually labeled 6,006 videos that scored above zero for

---

14. We also run a similar analysis of video transcripts, presented in our Supporting Information Section 4.2, and find substantively similar results. Because only roughly 20% of YouTube videos have transcripts, we decided to use the textually sparse meta data which we have for over 85% of the recommendations.

15. The full list of fraud terms was curated by the authors and is included in the SI, along with a discussion of what they capture and the regular expressions used to search for them in the text data of the recommendations.

16. As discussed below, and in SI Section 2, we also manually labeled a subset of these videos for stance.
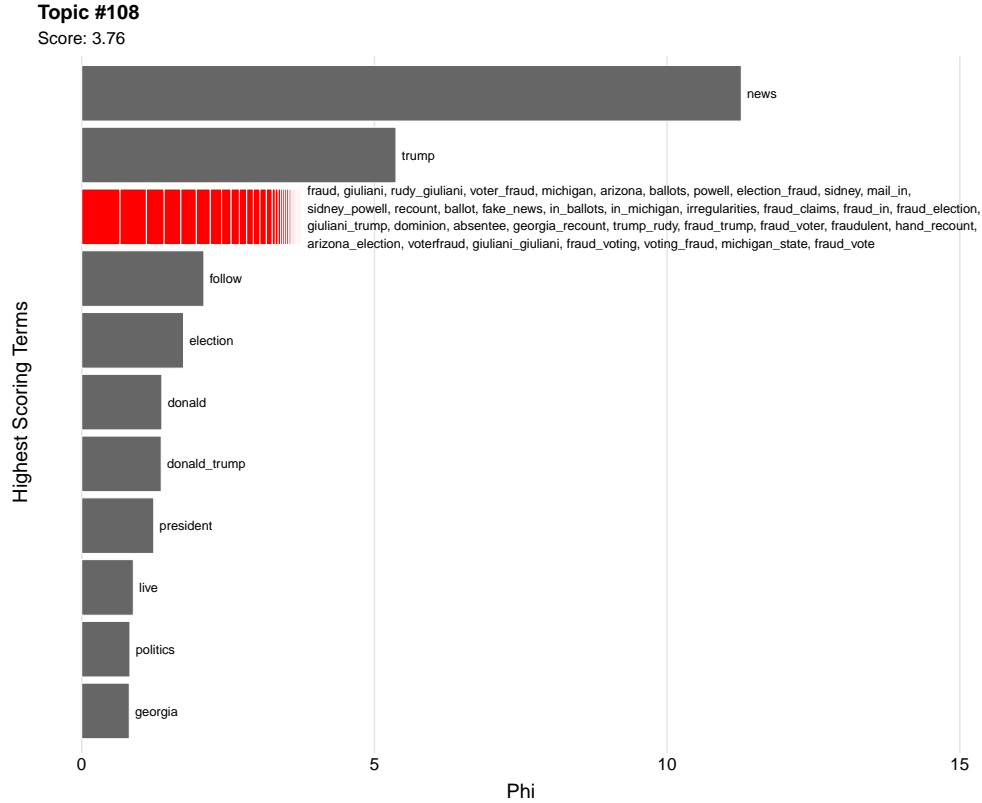
Figure 1: Top 10 terms associated with the highest scoring fraud topic (top panel, #108). X-axis indicate the $\phi_{w,k}$ values for each term, with the fraud keywords given in red.

our scorer across all choices of k.[17] We assigned each video one of four labels: (1) endorsing Trump's false claims of election fraud, (2) refuting Trump's claims, (3) neutrally reporting on these claims, and (4) not about election fraud.[18] As illustrated in Figure 3, videos that were labeled as endorsing or neutrally reporting on the false claims of fraud scored significantly higher on the topic #108 score. However, of the 6,006 human labeled videos, only 147 were coded as endorsing Trump's claims, while 380 were labeled as refuting these claims (208 were labeled as neutrally reporting on Trump's claims, while the remainder were labeled irrelevant). Aggregating these by their associated value of θ indicates that topic #108 consists of 82 videos endorsing Trump's claims, 84 that cover it neutrally, and 123 that refute these claims, out of a total of 1,312 videos.[19]

3.3   Methods

The raw data is indexed by participant-traversal step-recommended video, meaning that each row records one of (roughly) 20 recommendations suggested to a user at a given traversal step. For ease of description, we collapse this data to the user and calculate the proportion of videos about election fraud as the average of the document-topic probabilities $\theta_{k,v}$ for a given fraud topic

---

17. We conducted this task in 2021, leading to the possibility that we are missing many of the most extreme endorsements of Trump's false claims due to YouTube's crackdown on election misinformation on December 8th, 2020 (https://blog.youtube/news-and-events/supporting-the-2020-us-election/). Since we were only able to hand label videos that were still actively hosted on the platform, it is likely that we are missing a substantial amount of content that was recommended to our participants.

18. 500 of these videos were labeled by two human coders, with an intercoder reliability of 0.89. The remaining were coded by a single human.

19. As illustrated in the marginal histogram of Figure 3, the bulk of the coded videos were unlikely to be about topic #108, which is why so many were labeled by our human coders as not about the election.
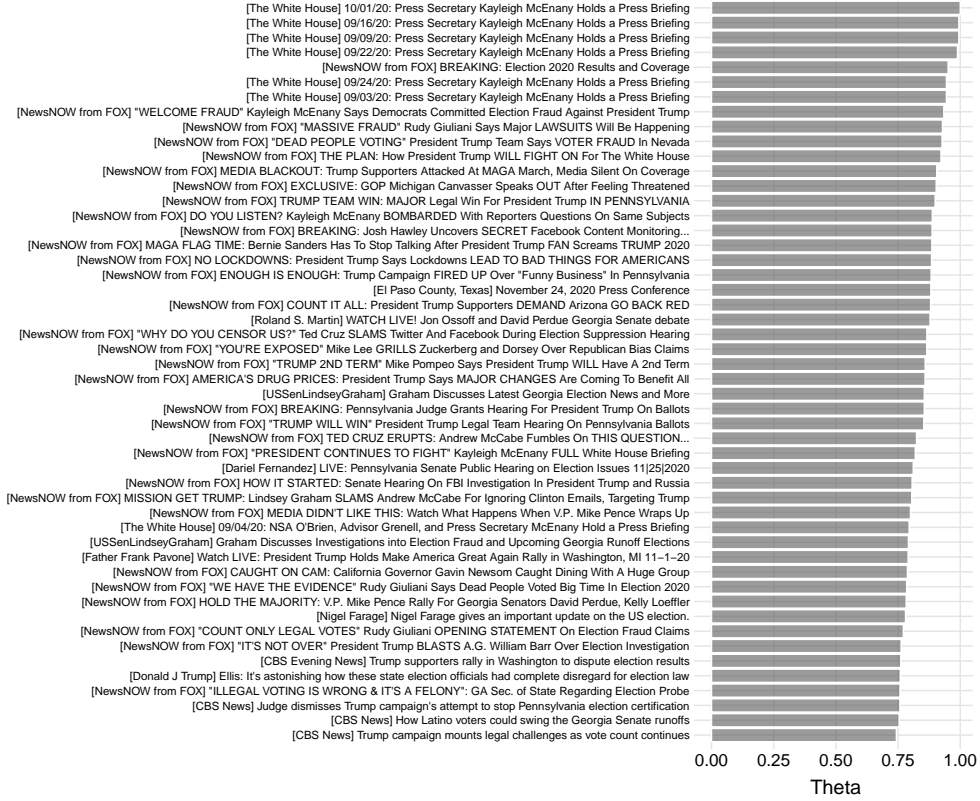
Figure 2: Top 50 videos associated with the highest scoring fraud topic #108. X-axis indicate the $\theta_{k,v}$ values for each term.

k which, in our main analysis, is the aforementioned Topic #108. We estimate the correlation between the proportion of fraud-related content in a user's recommendations and their skepticism about the legitimacy of the presidential election.[20] We predict the proportion of fraud content $y_{i,t}$ recommended to user i surveyed at week w as a function of i's beliefs about the legitimacy of the election ($Skep_i$), controlling for week fixed effects (δ_w), seed video fixed effects (α_s), and traversal rule fixed effects (γ_r). Formally:

$$y_{i,w} = \beta_1 Skep_i + \delta_w + \alpha_s + \gamma_r + \varepsilon_{i,w} \tag{3}$$

If the algorithm does recommend videos about election fraud disproportionately to participants who are skeptical about the legitimacy of the election, we would expect β_1 to be positive.[21]

---

20. We can also run the same analysis on the user-traversal step aggregated data, or on the raw user-traversal step-recommendation indexed data. In the latter two analysis, we cluster the standard errors at the respondent level. The results are almost identical in all settings.

21. There is an alternative interpretation of $\beta_1 > 0$: since the opinion questions were asked after participants completed the traversal task, it may capture the causal effect of exposure to fraud-related content on user beliefs. We argue that this interpretation is unlikely given that (1) our respondents rarely stayed on any video long enough for the "dose" of the fraud narrative to have an effect, and (2) our results examine recommendations, not watched videos, and it is hard to imagine a small thumbnail and truncated video title containing enough information to influence respondents' beliefs. We test the reverse causality concern in SI Section 5, finding no evidence to support this interpretation.
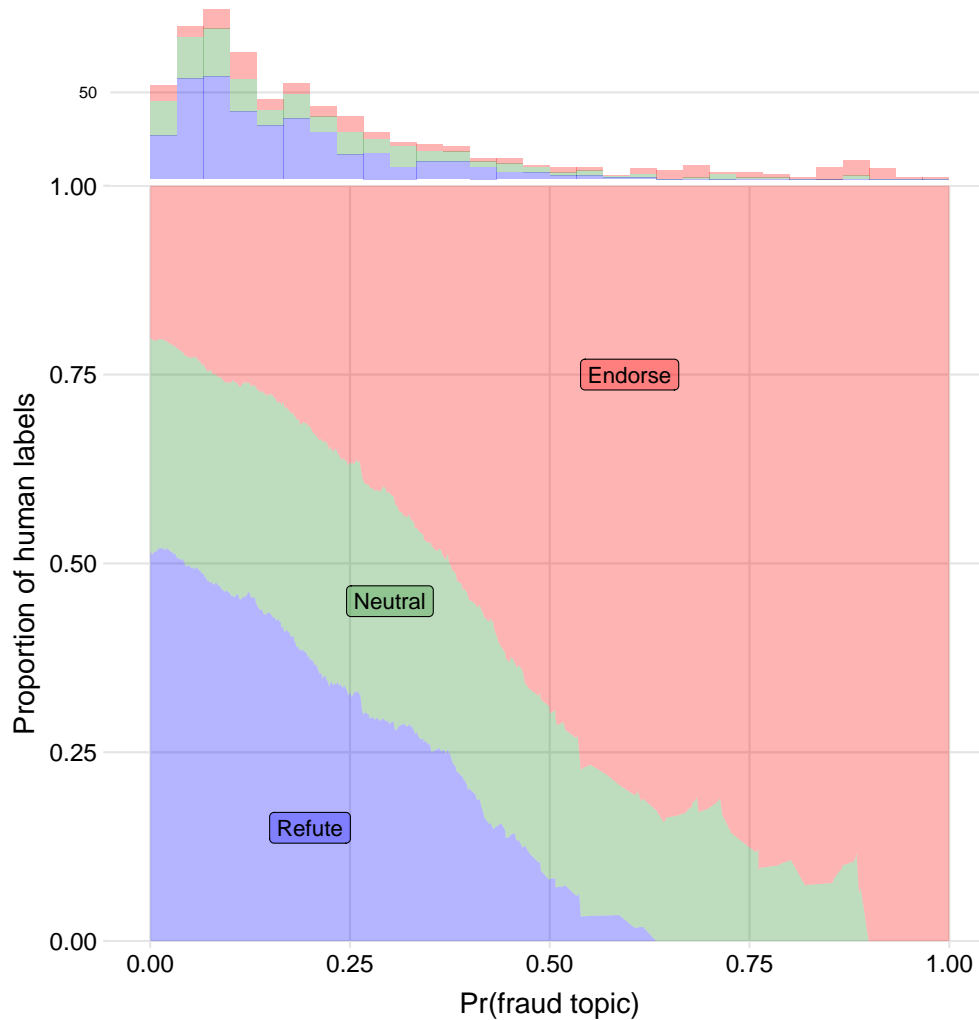
Figure 3: The y-axis indicates the share of human labels for videos that refute Trump's claims (in blue), those that report on Trump's claims neutrally (green), and those that endorse Trump's claims (red). The x-axis indicates the θ value for topic #108. The total number of videos falling into each category is indicated with the histogram across the top of the plot.

4   Results

Were election skeptics recommended more content about election fraud in the fall of 2020? Our main results predict the total proportion of recommendations pertaining to election fraud as a function of the combined skepticism index, based on the 331 respondents who responded to all questions about the election. We then disaggregate by specific belief, predicting the total proportion of recommendations pertaining to election fraud as a function of each skepticism dimension in turn. These disaggregated estimates are based on either 331 respondents (for those questions asked after November 4th) or 338 respondents (for the questions about concern asked from October 29th). We adjust for multiple comparisons across predictors using the Holm-Bonferroni step down procedure (**holm1979simple**), which we refer to as the family-wise error rate or FWER.

Figure 4 plots the β_1 coefficients from Equation 6 that capture the relationship between skepticism and recommendations. Circles indicate the estimated relationship (x-axes) between exposure to fraud-related videos (columns) and respondent agreement with different dimensions of skepticism

about the election integrity (y-axis). Two standard errors are indicated by the horizontal bars. Estimates that are significant at the naive 95% threshold are indicated in thin black borders with hollow points, and those that are significant at the FWER-corrected 95% threshold are indicated in thick black borders and solid points.
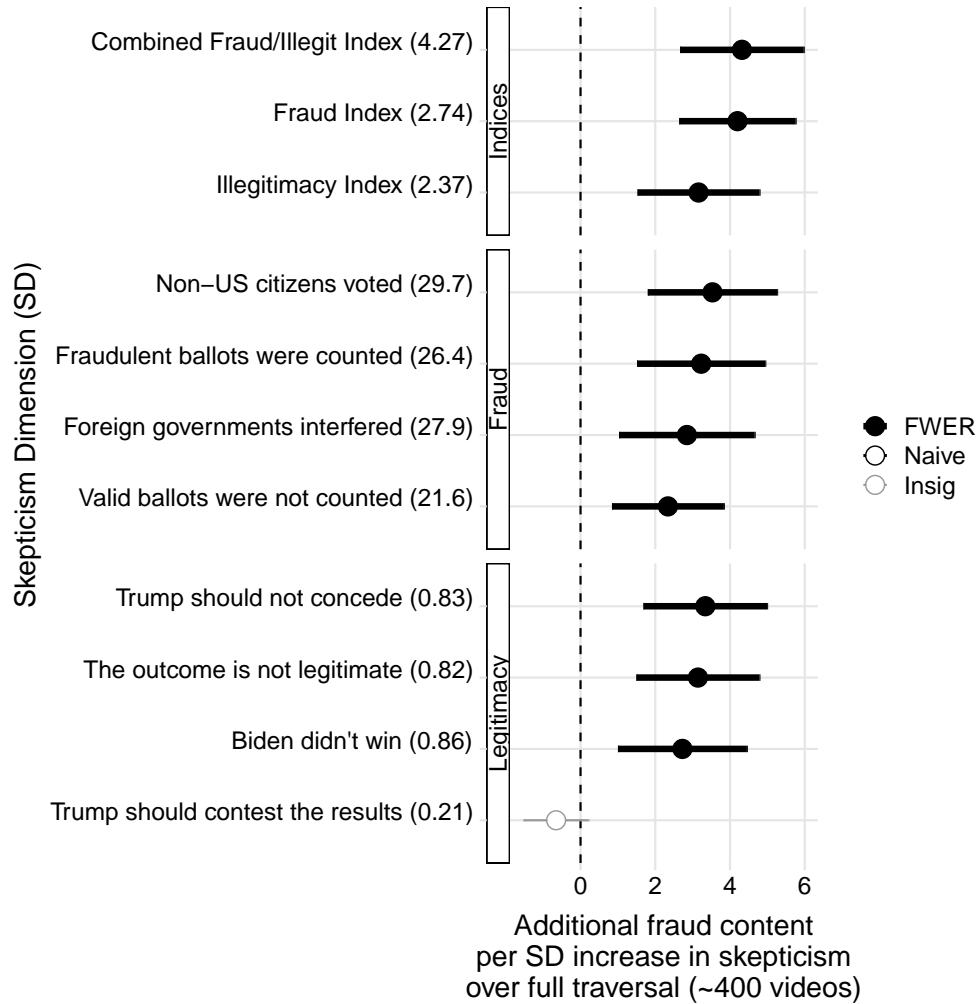

Figure 4: Regression coefficients (β_1) between skepticism over the 2020 presidential election (rows) and exposure to fraud-related content on YouTube. Gray points and bars are not statistically significant at the 95% confidence level, thin black bars with hollow points are significant at the 95% threshold, and thick black bars with solid points are significant at the 95% threshold after adjusting for the family-wise error rate (FWER) using the Holm-Bonferroni step down procedure (**holm1979simple**).

Each row captures an estimate from a separate regression linking the prevalence of recommendations with a different measure of election skepticism, starting with the combined index and then disaggregating to subsets based on opinions about the legitimacy of the outcome, opinions about different types of fraud, and then disaggregating further to the constituent opinions.[22] The first three rows present the combined indices that add up the four measures of fraud ("Fraud Index"), the four measures of legitimacy ("Illegitimacy Index"), and all questions combined ("Combined Fraud/Illegit Index"). As indicated by the thick black borders on the fraud index and the combined

---

22. We also run combined regressions where groups of skepticism measures are estimated jointly, summarized in the Supporting Information.

index, there is a strong positive correlation between holding skeptical beliefs about the election's legitimacy and being recommended fraud-related content on YouTube. These standardized coefficients can be interpreted in terms of standard deviations – a one-standard deviation increase in combined skepticism (roughly 4.27 units on the index ranging from -10 to +10) is associated with approximately 4 additional videos over the course of the traversal task which gathered, on average, 410 recommendations in total for each respondent.[23]

As we disaggregate the skepticism measures to the separate indices for fraud-related concerns and general legitimacy concerns, we see that the relationship is strongest for the combined index, while the illegitimacy index is the least strongly associated. Disaggregating further reveals that the question about whether Trump should contest the results exhibit no statistically significant relationships, and the point estimates are very close to zero.[24] Conversely, the strongest overall patterns are found for the fraud-related questions, specifically the concerns that non-US citizens voted, foreign governments interfered, and fraudulent ballots were being counted.

In sum, videos about the topic that contained the most fraud-related keywords are recommended disproportionately more to participants who expressed skepticism about the legitimacy of the election. As discussed above, this topic is associated with not only the general coverage of the election, but is specifically about the fraud concerns espoused by Donald Trump. Furthermore, videos that score highly on this topic are disproportionately those that either endorse Trump's claims or report on them neutrally.

To evaluate the substantive magnitude of these results, we turn to a descriptive analysis of the data, plotting the total number of videos about election fraud against each participant's self-reported concern about fraudulent ballots (Figure 5). As illustrated, moving from the least to the most concerned participants corresponds to an increase of roughly 8 additional videos being recommended over the course of our survey. On the one hand, this is a small proportion relative to the total number of recommendations that participants were shown over all topics during their traversal (approximately 410 in total). On the other hand, this constitutes an increase from roughly 4 videos to roughly 12, or a 200% increase.

Finally, of these additional videos, what can we say about their stance? Figure 6 plots the proportion of human labeled recommendations that are about the election which were labeled as (1) endorsing Trump's claims, (2) neutrally reporting on Trump's claims, or (3) refuting Trump's claims. We bin respondents into the bottom 10%, the middle 80%, and the top 10% of concern that fraudulent ballots were being counted, revealing two important patterns.[25] First, there is clear evidence of positive associations between concern about election fraud and recommendations that endorse or neutrally report on Trump's claims, and a negative association with content that explicitly refutes Trump's claims. Second, across all groups, the proportion of refuting content is nevertheless greater than content that endorses or neutrally reports. We underscore that these labels were assigned after YouTube purged much of the election misinformation on their platform on December 8th, 2020, meaning that these proportions are a lower bound on the true state of the world experienced by our respondents in the fall of 2020.[26]

---

23. The number of recommendations we were able to measure at a given page are a function of the participants' computer and browser, with elements like monitor resolution and browser zoom influencing the total list of recommendations captured by our plugin. The standard number of recommendations shown before the user starts to scroll down is typically 20 per page, but higher resolution browsers can yield more.

24. These patterns reflect the limitations of relying on a convenience sample gathered using Facebook advertisements. As discussed in the Supporting Information, the majority of our respondents on Democrats and liberal. It is possible that our results would be even stronger were we able to field a nationally representative sample in the fall of 2020.

25. We report these thresholds as they clearly create three distinct groups, the bottom 10% and top 10% being far apart on the measure of interest. However, we test alternative cutpoints (such as 20-60-20 and 33-33-33), and found our results to be robust to this choice.

26. We test whether the videos suggested to more skeptical users were more likely to be taken down after the December 8th crackdown in the SI, Section 6, finding no evidence of a systematic association. However, missingness is a noisy proxy for videos that promoted misinformation and were taken down.
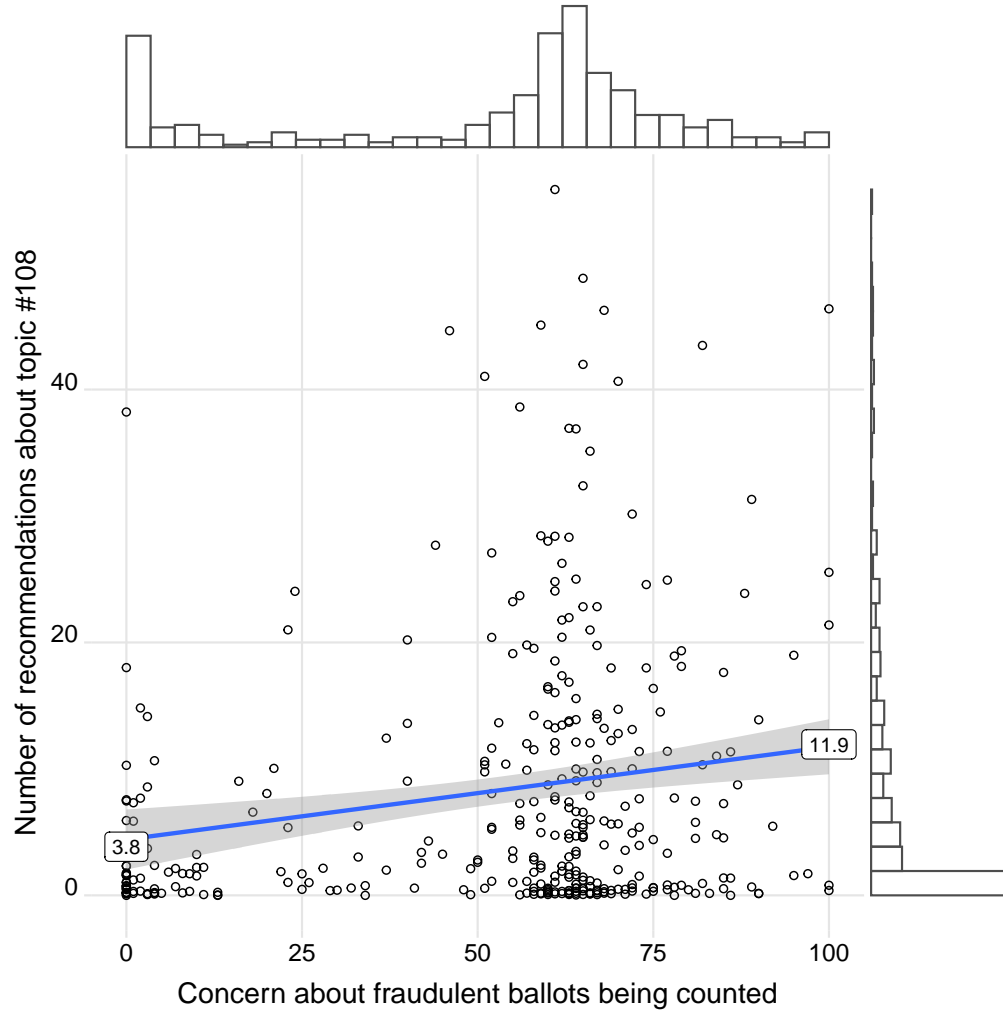
Figure 5: Expected number of videos about topic #108 (y-axis) shown to each user (points) as a function of how concerned the participants were about fraudulent ballots being counted (x-axis).

5 Conclusion

The 2020 presidential election in the United States represented the greatest threat to the country's democratic institutions in over a century. What role did online social media environments play in fanning the flames of distrust? In this article, we document a systematic association between skepticism about the legitimacy of the election and exposure to election fraud-related content on YouTube. By gathering these data over the course of the fall of 2020, and by assigning participants to click through 20 recommended videos at random from a randomly assigned starting video, we are able to capture an ecologically valid snapshot of the role played by YouTube's recommendation algorithm in feeding fraud-related content to participants most likely to believe that the election was marred by fraud.

We show that user skepticism about the legitimacy of the election is positively associated with the number of videos about election fraud that were recommended by the algorithm. Substantively, we show that the least skeptical participants were recommended roughly 4 videos about election fraud on average, while the most skeptical participants were recommended roughly 12 – an increase of roughly 200%. However, we caution against an extreme interpretation of our findings that YouTube
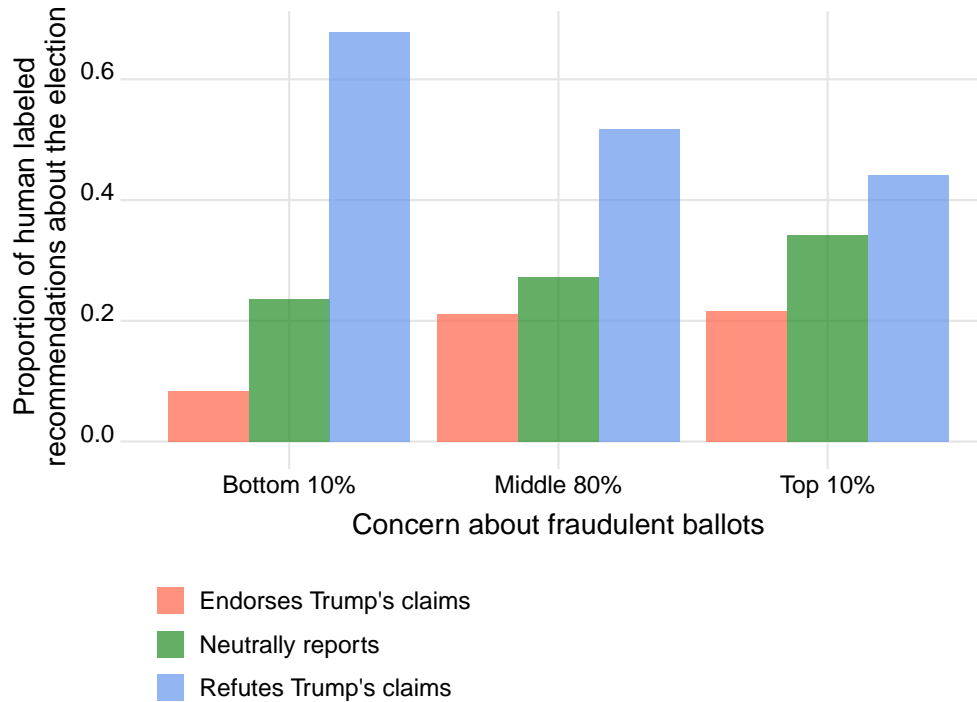
Figure 6: Proportion of all human labeled recommendations that were about the election (y-axis) by human label (endorsing Trump's claims indicated in red, neutrally reporting on Trump's claims indicated in green, and refuting Trump's claims indicated in blue), by degree of user concern about fraudulent ballots being counted (x-axis). Human labeled data was generated in 2021, after YouTube removed thousands of videos containing election fraud on December 8th, 2020, meaning that these proportions are a lower bound.

pushed skeptics over the edge with a deluge of misinformation about Trump's claims in the fall of 2020. While the positive associations between participant skepticism and fraud-related content are statistically significant, they amount to a difference of, on average, roughly 8 more videos out of roughly 410 total recommendations shown to the most extreme skeptics versus the least. In addition, these recommendations range from content that refutes Trump's claims of election fraud, to content that neutrally reports on Trump's claims, to content that explicitly endorses these claims.[27]

Nevertheless, we interpret our findings as evidence of the possible role played by the YouTube in contributing to distrust in America's democratic institutions among those most inclined to be distrustful. Furthermore, we emphasize that our calculations of the total effect are likely a lower bound for two reasons. First, our convenience sample is disproportionately comprised of younger, liberal respondents with at least a college degree, meaning that the group of greatest substantive interest – conservative, less educated Trump supporters – is underrepresented in the data which can bias our estimate of the total effect downward. Second, our human labeled data were gathered in the spring of 2021, which is after YouTube cracked down on election misinformation content on the platform and removed roughly 8,000 channels. Since we were only able to hand label videos that were still actively hosted on the platform, it is likely that we are missing some content that was recommended to our participants. However, missingness only affects 2% of our recommendations,

---

27. In the Supporting Information Section 2, we re-analyze our data, replacing the LDA topic score with a binary indicator for each of the 3 human labels we assigned to videos with non-zero fraud scores. The strongest associations between skepticism and recommendations are for the endorsing category, followed by the neutral category and then the refuting category, although the coefficients are not significantly different from each other.

and is not positively associated with skepticism, as we demonstrate in Section 6 of the Supporting Information.

We acknowledge that these results are consistent with the core goal of an effective recommendation algorithm: namely to suggest content that users would be interested in. In many, if not most, of the domains of the content hosted on YouTube, suggesting videos that users are most likely to enjoy is a harmless if not beneficial quality of the recommendation algorithm. But in the context of misinformation or conspiracy theories such as Trump's claims about election fraud, the recommendation algorithm's precision can lead to potentially socially harmful outcomes, as we demonstrate in this paper.[28]

We also acknowledge that our singular focus on isolating the real world impact of the recommendation algorithm ignores several important dimensions of the broader topic of content consumed on social media platforms. A particularly instructive question is to ask what a counterfactual world would look like in which there was no recommendation algorithm. We expect that users would still be watching socially harmful misinformation about the election in a manner correlated with their broader skepticism about the election's legitimacy. This differential consumption would be driven by a combination of supply (i.e., certain content creators would learn that such content increases ad revenue or donations and would produce more of it) and demand (i.e., users would still seek out this content, or be recommended it not by an algorithm but by their similarly-minded peers), as described in **munger2019supply**. And as discussed above, while our findings are likely a lower bound on the independent influence of the recommendation algorithm, they are nevertheless quite small, paling in comparison to the demand-driven effects documented in **chen_and_nyhanadl2021**, and – more generally – to the finding in **muise2022quantifying** that partisan news segregation is orders of magnitude more pronounced on cable TV compared to online social media.

Nevertheless, our results provide an important caveat to a growing body of research that finds little evidence of algorithms contributing to echo chambers, and highlights the need for issue-specific analyses where the goal of suggesting user-specific desired content can have pernicious consequences for society.

---

28. For the policy-minded reader, a natural question would be whether YouTube should focus on tweaking the algorithm to avoid recommending socially harmful content, or should instead focus on curating the library of content to remove offensive videos. We provide a descriptive investigation of YouTube's efforts to remove socially harmful content in SI Section 9, finding that the spread of videos about election fraud that were linked to posts on Twitter dropped off precipitously after YouTube started removing election misinformation on December 8th.

Authors

James Bisbee is an Assistant Professor of Political Science and Data Science at Vanderbilt University. He is the corresponding author and can be reached at jhb362@nyu.edu.

Megan A. Brown is a senior research engineer and data scientist at NYU's Center for Social Media and Politics.

Angela Lai is a PhD student in the Center for Data Science at NYU and a Graduate Research Associate at NYU's Center for Social Media and Politics.

Joshua A. Tucker is Professor of Politics, affiliated Professor of Russian and Slavic Studies, and affiliated Professor of Data Science at New York University. He is the Director of NYU's Jordan Center for Advanced Study of Russia, and a co-Director of the NYU Center for Social Media and Politics (CSMAP).

Richard Bonneau is Professor of Biology and Computer Science, and co-Director of the NYU Center for Social Media and Politics

Jonathan Nagler is Professor of Politics and affiliated faculty at the Center of Data Science at New York University. He is a co-Director of the NYU Center for Social Media and Politics. Nagler is a past president of the Society for Political Methodology, as well as an Inaugural Fellow of the Society for Political Methodology

Acknowledgements

Data Availability Statement

Replication files are available at: **I assume these will be available on dataverse as part of publication?**

The study preanalysis plan is registered with: Not Applicable

Funding Statement

Ethical Standards

This study was deemed except by New York University's IRB, #IRB-FY2020-4647.

Keywords

Appendices

Appendix A: Survey Details

We surveyed 709 individuals between October 2nd and December 9th, 2020 as part of a larger research project. Data collection proceeded in two waves. The first ran over the first two weeks of October while the second ran from October 29th through December 9th. Respondents were randomly assigned to one of six traversal rules, five corresponding to always clicking on one of the first five recommended videos, and one in which they clicked on the video they were most interested in. For the purposes of this survey, we subset our attention to only those who were assigned to one of the random traversal conditions, and to only those in the second wave of the survey during which we asked questions pertaining to election fraud. Our 361 total respondents participated as depicted in Figure 7.
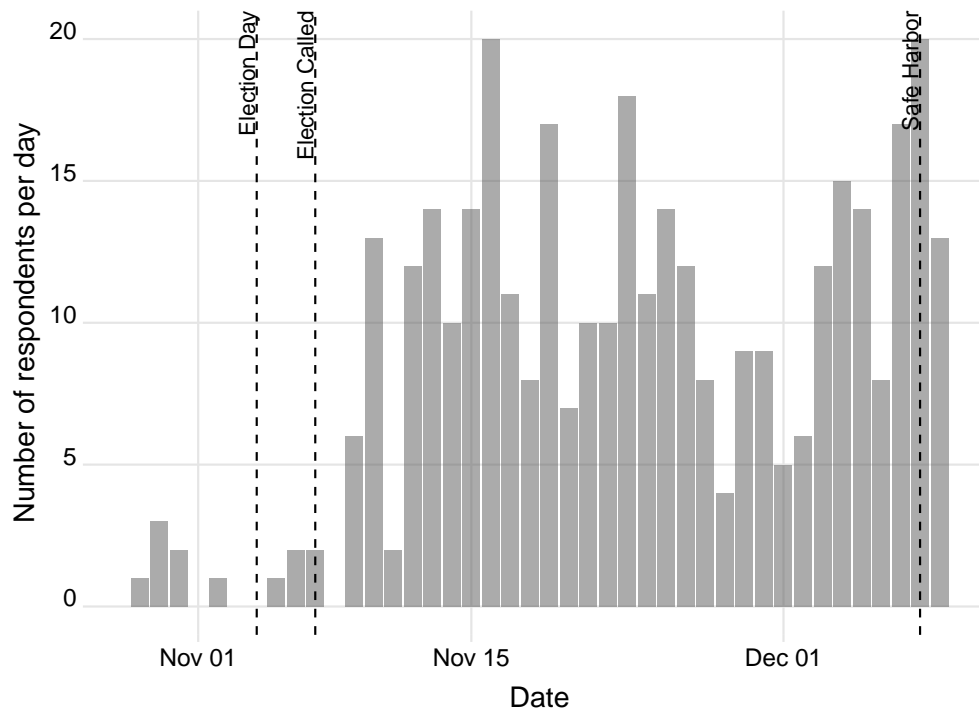
Figure 7: Total number of respondents by date. Vertical dashed lines indicate election day (November 3rd, 2020), when the election was called for Biden by the Associated Press, Fox News, and others (November 7th, 2020), and the "Safe Harbor" deadline at which the states certified their final tallies (December 8th, 2020).

We advertised on Facebook, using the platform's targeted ads feature to recruit only Chrome browser users on desktop computers living in the United States. Our focus on these technical dimensions was to ensure compatibility with our bespoke plugin that we made available on the Chrome app store. We describe the plugin in greater detail below, and summarize the distribution of our respondents by platform in Figure 8. As illustrated, the vast majority of our respondents accessed the survey via Chrome browsers using Windows operating systems. The two exceptions were using Edge and Opera browsers, both of which are based on the Chrome architecture. To the extent that these patterns diverge from the population at large[29], we posit that this is likely due to

---

29. As of June 2021, Windows comprised roughly 61% of the American desktop market, followed by Mac OS at 28%. https://gs.statcounter.com/os-market-share/desktop/united-states-of-america

the popularity of Apple's bespoke Safari browser. Since our plugin only worked on Chrome-based browsers, it is likely that many potential respondents were unable to participate due to their use of Safari. To the extent that these differences might limit generalizability, we argue that they shouldn't do so more than the demographic profile of our convenience sample.
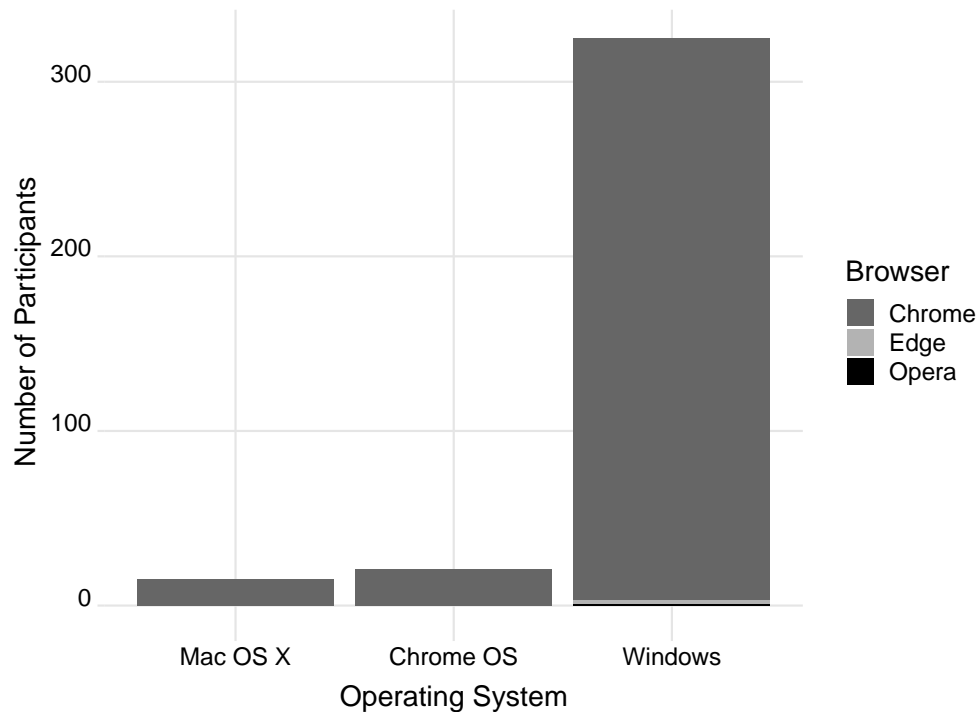

Figure 8: Total number of respondents by operating system and internet browser.

Our recruitment campaign evolved over the course of our study. For the second wave, we relied on fliers depicted in Figure 9 to attract interest. Participants were rewarded via three channels. First, by completing the basic survey and traversal task, they were compensated with $5 via a gift card of their choice. Second, they were given the opportunity to receive an additional $5 in return for uploading their YouTube watch history. Finally, they were entered in a raffle to win a $500 gift card of their choosing, with odds equal to approximately 0.00167%.

Upon clicking on the ad, Facebook users would be taken to the survey hosted on Qualtrics. An automatic IP test would confirm that they were located in the United States and not disguising their location using a VPN, as per the techniques discussed in **winter2019simplified**. After passing this check, along with a Qualtrics-based confirmation that they were accessing the survey via a Chrome browser, they were taken to the consent form for the survey, approved by NYU's IRB (FY2020-4647). Upon confirming their consent, they were then taken to the instructions for how to complete the traversal task, reproduced in Figure 10 below. We wanted to avoid users getting trapped in two types of recommendations: "Mixes" and "Movies". In pilot tests of the survey, we discovered that a YouTube "Mix" would replace the default recommendations with a curated list of videos all on the same theme, such as a sequence of videos related to a musician or sports team. Similarly, clicking on a YouTube "movie" gave the viewer the opportunity to watch a trailer for a movie that they could then spend additional money to rent. On these types of videos, every recommendation provided was for another movie in the YouTube library. Both of these cases, while potentially interesting in their own right, were beyond the scope of our substantive interest in an ecologically valid audit of the platform.

After reading the instructions for how to complete a traversal, we then took the users to a separate

Figure 9: Recruiting advertisement posted on Facebook.

instruction page that described how to install and log in to the plugin we created for this task. These instructions are depicted below in Figure 11, and included a tutorial video to ensure the installation went smoothly. The vast majority of respondents were able to follow these instructions without issue. A small minority of respondents experienced technical difficulties that our researchers were able to assist them with. In every case, these difficulties were due to user error, and not bugs in the extension.

Finally, users were taken to the traversal task completion page in which they were given the link to their randomly-assigned seed video as well as their randomly-assigned traversal rule, as depicted in Figure 12. Upon completion of the 20th traversal, the plugin displayed a completion code that the respondent was instructed to enter into the Qualtrics survey, allowing us to link their survey responses with their traversal data. After copying this code from the plugin pop-up window, the extension would self-uninstall.

After completing the traversal task, respondents would then be given the opportunity to upload their YouTube watch history for an additional $5 reward. If they indicated interest in doing so, our survey gave detailed instructions to guide them through how to access their watch history and upload the zipped file to the survey, as depicted in Figure 13.

In the following section, we will ask you to install our plugin ⊛ and navigate to a starting YouTube video that we will provide.

You will then be asked to click on the **first video** in the list of recommended videos. At this video, *you must wait five seconds* for the plugin to download the list of recommended videos you are shown.

You will be asked to repeat this process **20 times**, each time clicking on the first video in the list of recommended videos, and waiting five seconds for the plugin to download the recommendations. A counter on the plugin icon will keep track of how many you have done ⊛.

There are two types of recommended videos that should be ignored: YouTube **Mixes** and YouTube **Movies**.

- **Mixes** start with the word "Mix" and feature this overlay on the thumbnail: ▪️.
- **Movies** are hosted by "YouTube Movies" and have the text "Buy or Rent" in green: Buy or Rent .

If the first video is either a "Mix" or a "Movie", please click on the first video that is NOT a "Mix" or a "Movie". An example is given in this image.
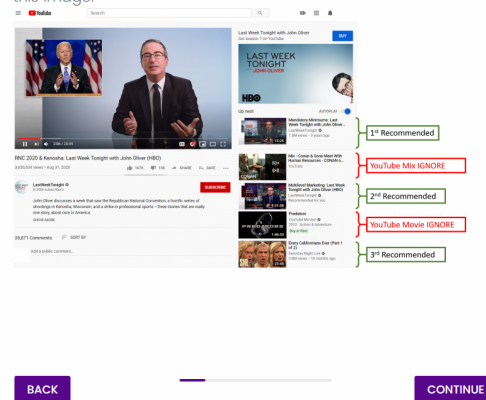

BACK   CONTINUE

Figure 10: Traversal task instructions.
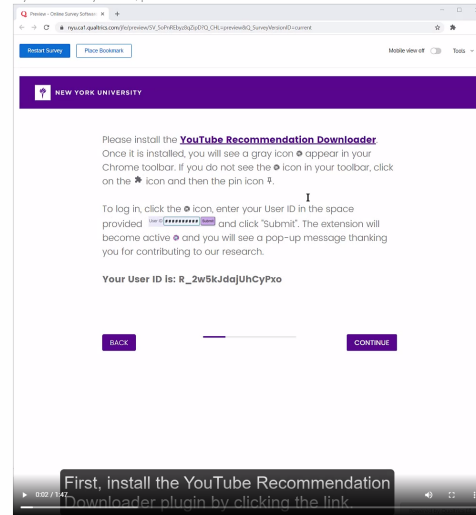
Figure 11: Extension installation instructions.

Figure 12: Traversal task random assignment.

Log in to the account you typically use to watch YouTube, and then navigate to https://takeout.google.com/settings/takeout/custom/youtube.

1. Click the "All YouTube data included" button, and then **uncheck** every box except "history".
2. Click "Next step" and then click the blue "Create export" button with the default settings.
3. You will see a warning from Google that the process can take "hours or even days" but if you only selected "history" it shouldn't be more than a few minutes.
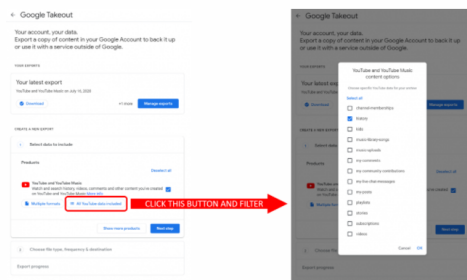
Within a few minutes, you will receive an email from noreply@google.com, which will allow you to download your data. Click "Download your files," which will automatically download the .zip file of your data. Unless you otherwise specify, your data will download to the "Downloads" folder on your computer and will start with "takeout."

Please upload your data here. NOTE: We automatically extract only the anonymized history data from the zip file and destroy everything else. This process ensures that there is **no chance your anonymity will be compromised**, even if you unintentionally downloaded additional information from Google.

Drop files or click here to upload
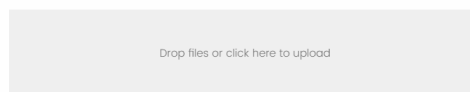
Figure 13: Watch history upload page.

Appendix A.1   Summary Statistics

Finally, the respondents were asked to complete a short survey about how they use YouTube, some basic demographic information, and a handful of questions pertaining to the 2020 presidential election. The questions on the 2020 presidential election are reproduced in Figure 14. Summary statistics for these outcomes are presented in Figure 15.

Figure 14: Question wording for the concern dimension.

In addition to these continuous outcome measures, we also asked respondents a series of questions about the factual outcome of the election, including who won, whether the outcome is legitimate, and whether Donald Trump should concede the election or contest the results in court. These questions are reproduced in Figure 16 Summary statistics for these questions are displayed in Figure 17. As illustrated, the majority of respondents indicated that the outcome was not yet known. Among those who indicated that the outcome of the election was known, the majority responded that they believed Biden won, that Trump should concede, and that the outcome was legitimate. This is likely due to the timing of these responses, which were gathered over a period during which the outcome was still uncertain, even after the election was officially called on November 7th, 2020. As illustrated in Figure 18, uncertainty over the outcome was observed throughout our study period, reflecting the prevailing uncertainty due to Trump's refusal to concede.

Our combined indices for beliefs about fraud, legitimacy, and overall skepticism were generated by re-scaling these outcome measures such that +1 indicated skepticism, -1 indicated trust, and we
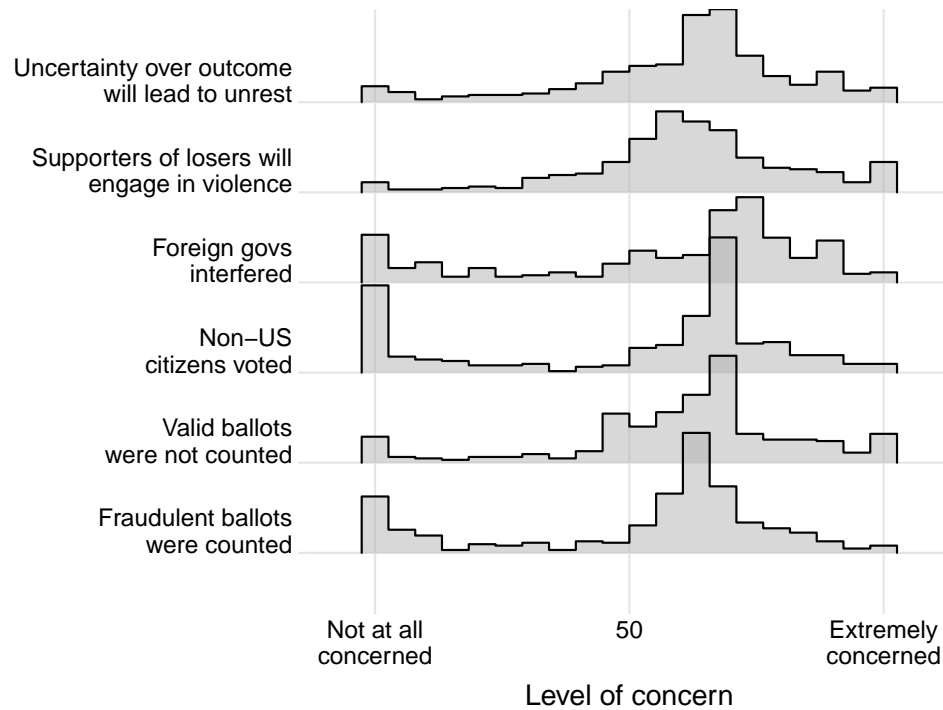
Figure 15: Distribution of concern over election.

divided the continuous measures of concern at 50. The indices were then the simple sum of these rescaled measures, summary statistics of which are presented in Figure 19.

Our convenience sample was gathered using targeted advertisements on Facebook. As such, we should not expect to have balance on demographic covariates. Indeed, as illustrated in Figure 20, our sample skews liberal, Democrat, male, and young.

Q61   ✶

When do you expect to know who the winner of the presidential election will be?

○ Before the "safe harbor" deadline (December 8th)

○ Before the formal electoral college vote (December 14th)

○ Before the certification deadline of the vote (January 6th)

○ Before inauguration day (January 20th)

○ The winner is already known.

Q67

▾ Display this question

If When do you expect to know who the winner of the presidential election will be? The winner is already known. Is Selected

Who is the winner?

○ Donald Trump

○ Joe Biden

○ Someone else

Q74

▾ Display this question

If When do you expect to know who the winner of the presidential election will be? The winner is already known. Is Selected

Is this outcome legitimate?

○ Yes

○ No

-- Page Break --

☐ Q75   iQ   ···

▾ Display this question

If Is this outcome legitimate? No Is Selected

Please indicate whether the following statements are true or not.

	Definitely true	Probably true	Probably false	Definitely false
Mail ballots are being manipulated to favor Joe Biden	○	○	○	○
Illegal immigrants voted fraudulently	○	○	○	○
Russia is attempting to influence the election in favor of President Trump	○	○	○	○
Many people were illegally prevented from voting this year	○	○	○	○
We will never know the true outcome of the election	○	○	○	○

Figure 16: Question wording for the legitimacy dimension.
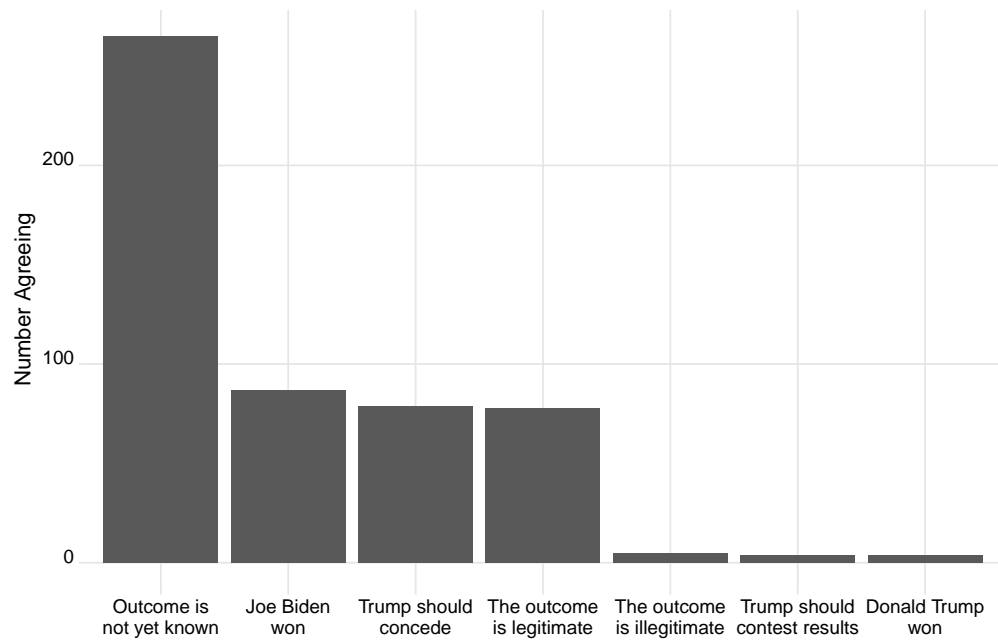
Figure 17: Agreement with statements about the election outcome.
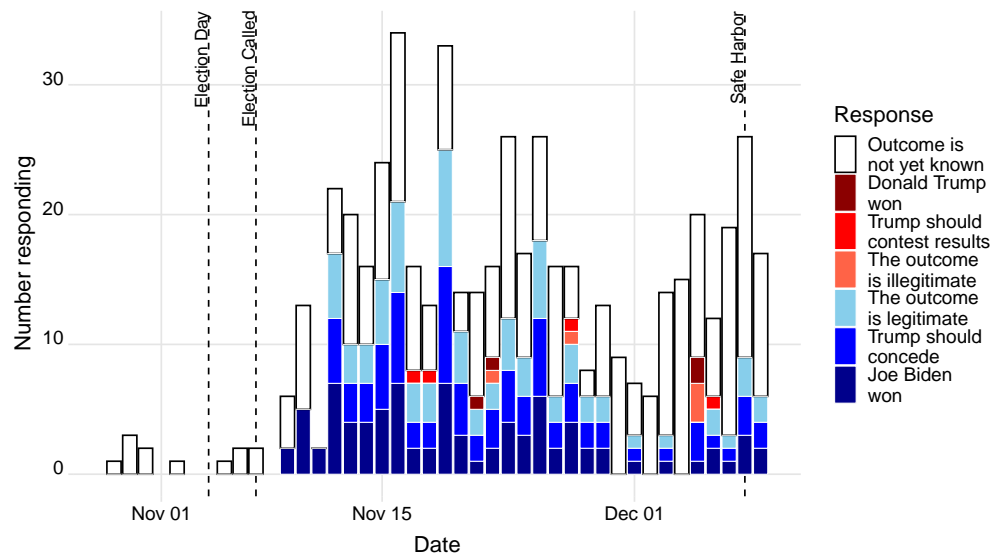
Figure 18: Over-time distribution of responses.
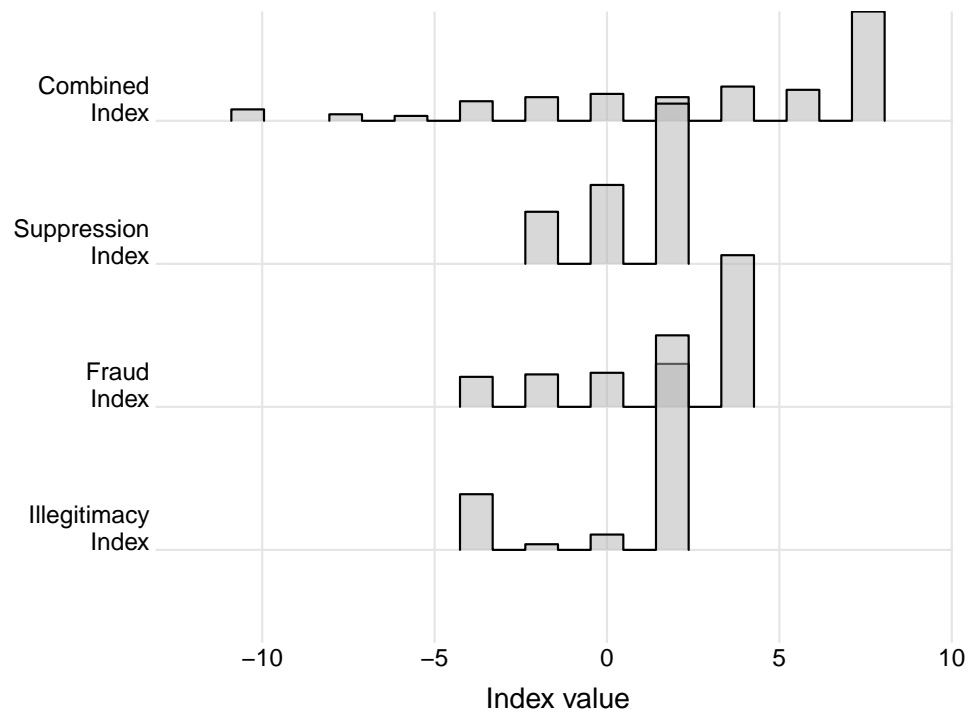
Figure 19: Descriptive stats of the indices.

	Democrat	Independent	Republican	
Zoomer	34	16	5	Age
Millennial	168	41	45	
Gen X	33	8	5	
Boomer	5		1	
Graduate Degree	50	8	18	Education
College Degree	159	41	34	
Some college	21	9	1	
High school grad	10	5	3	
< High school		2		
Asian	31	10	7	Ethnicity
Black	32	7	4	
Native American	25	12	5	
Some other race	8	5	2	
White	144	31	38	
Male	156	35	41	Gender
Female	81	27	15	
Other	3	3		
Conservative	91	11	34	Ideology
Moderate	11	15	11	
Liberal	138	39	11	
More than $80k	173	37	42	Income
$40k to $80k	35	11	7	
Less than $40k	32	17	7	

Figure 20: Number of respondents by self-reported partisanship (columns) broken out by different demographic groups (rows).

Appendix B: Human Labeled Validation

Our main results use topic models combined with a bespoke scorer to calculate the proportion of videos watched that promoted Trump's misinformation about election fraud. As presented in the validation of this approach, we are confident in our ability to identify videos about the topic of election fraud writ large, and are reassured by descriptive evidence that those videos which scored highest were substantively those that endorsed the misinformation. However, a topic model is not the only solution to identifying which videos were about election fraud in our data.

We also had humans label a subset of the recommendations that received a non-zero score in the topic models across every permutation of the LDA model (with and without lemmatization, predicted using metadata or transcripts, and different choices for k, amounting to 6,006 total videos). We asked the coders to watch enough of each video to select one of four labels with a reasonable amount of confidence in their choice. The full text of the instructions is reproduced below.

> For each video, please watch enough to confidently assign one of the following four labels. (Typically, this can be accomplished by watching approximately the first 30 seconds of substantive content, and then by skipping to later in the video to confirm.) All labels are answers to the following question:

> "Does this video endorse Donald Trump's claim that the 2020 U.S. presidential election was illegitimate due to widespread fraud, either through ineligible absentee ballots being cast, illegal voting, rigged voting machines, or miscounting of results?"

> If a video appears to both endorse and refute Trump's claims, please do your best to determine whether it is more endorsing or more refuting. For example, a video that acknowledges some irregularities but concludes that these were not sufficient to overturn the election result should be labeled as refuting Trump's claim.

Label	Description
Yes, it supports Trump's claims that the outcome is illegitimate.	Select if the video or part of the video takes a position in support of Trump's election fraud narrative (i.e., that voting machines were rigged, that Biden didn't win, that Trump did win, that the election was stolen, that invalid votes were being counted, that valid votes were not being counted)
No, it refutes Trump's claims that the outcome is illegitimate.	Select if the video or part of the video takes a position contradicting Trump's election fraud narrative (i.e., that voting machines were rigged, that Biden didn't win, that Trump did win, that the election was stolen, that invalid votes were being counted, that valid votes were not being counted)
Neither, it reports on the issue without supporting or refuting Trump's claims that the outcome is illegitimate.	Select if the video or part of the video is about Trump's election fraud narrative, without taking a stance either way.
Neither, it is not about the legitimacy of the 2020 presidential election.	Select if no part of the video is about election fraud in the 2020 US presidential election. This includes any videos that are *not* about the election in addition to videos that are about the election but do not mention election fraud. Also select if the video does not load or is in a language other than English.
Comment Field (optional)	Include an open-ended text field for worker comments + observations

Table 1: Labels assigned to 6,006 videos by human coders.

There were two human coders assigned to 500 of the videos, while the remainder were labeled by a single coder. The two coders agreed on 475 out of 500 videos, with a Cohen's κ of 0.89. Of the 25 disagreements, we manually reviewed 24 of the videos and selected the most appropriate label.[30] The list is given in Table 2, revealing that all discrepancies are between either neutrality and stance, or irrelevance. We also manually reviewed the human labels on the top 100 videos with the highest value of Pr(topic #108).

Table 2: List of disagreements in coding

Channel	Video	Label 1	Label 2	Author
ABC News	Wisconsin Attorney General Josh Kaul gives updates on the 2020 election	Irrel	*Refute*	Refute
ABC News	ABC News Prime: Biden's transition team; Possible COVID-19 vaccine update; Trump refuses to concede	Refute	*Neutral*	Neutral
Brian Tyler Cohen	Top CNN host finally LOSES IT, rips Trump adviser for lying ON AIR	Irrel	*Refute*	Refute
CBS News	Joe Biden and Kamala Harris deliver remarks on the economy in Delaware	Neutral	*Irrel*	Irrel
CNN	Trump threatens to deny New York a vaccine. See governor's response	*Irrel*	Refute	Irrel
Fox Business	Steve Bannon: Trump won't allow the election to be stolen	*Endorse*	Neutral	Endorse
Fox Business	Mick Mulvaney: If Trump can't win Arizona, he can't win the race	Neutral	Irrel	Endorse
Fox Business	GOP Rep Darrell Issa talks being re-elected in California	Neutral	Irrel	Endorse
JDMOONAN	Jenna Ellis tells Rep. Cynthia A. Johnson WHAT her JOB is! YOU WORK FOR THE PEOPLE!!!!!!!	*Endorse*	Neutral	Endorse
Newsmax TV	Jenna Ellis and Giuliani call out reporters, FBI	*Endorse*	Neutral	Endorse
NewsNOW from FOX	WE WILL WIN: Kevin McCarthy Reminds Reporters That 2022 Is COMING UP	Irrel	Endorse	Neutral
NewsNOW from FOX	"WE HAVE THE EVIDENCE" Rudy Giuliani Says Dead People Voted Big Time In Election 2020	Neutral	*Endorse*	Endorse
NewsNOW from FOX	"MASSIVE FRAUD" Rudy Giuliani Says Major LAWSUITS Will Be Happening	Neutral	*Endorse*	Endorse
NewsNOW from FOX	MEDIA BLACKOUT: Trump Supporters Attacked At MAGA March, Media Silent On Coverage	Neutral	Irrel	Endorse
NowThis News	Donald Trump Never Saw This Coming	*Irrel*	Neutral	Irrel
NowThis News	Rudy Giuliani's Star Voter Fraud Witness at Michigan Hearing	Neutral	*Refute*	Refute
NowThis News	How Republicans Have Been Rigging the Vote	Refute	*Irrel*	Irrel
PBS NewsHour	Shields and Brooks on Ginsburg's legacy, Trump's election rhetoric	*Refute*	Neutral	Refute
PBS NewsHour	Tamara Keith and Amy Walter on Biden's Cabinet picks and Trump's fraud claims	*Refute*	Neutral	Refute
PBS NewsHour	PBS NewsHour live episode, Nov. 20, 2020	Neutral	*Refute*	Refute
PBS NewsHour	PBS NewsHour full episode, Nov. 19, 2020	Neutral	Refute	NA
Sky News Australia	Nigel Farage: No doubt there was 'industrial scale' ballot harvesting in US Election	Neutral	*Endorse*	Endorse
The Choice	Zerlina, and The Mehdi Hasan Show | Live | The Choice on Peacock	Irrel	*Refute*	Refute
Yang Speaks	Biden Wins. Yang Gang moves to Georgia. Nina Turner joins. | Andrew Yang | Yang Speaks	Refute	Neutral	Irrel
Zooming In with Simone Gao	Narrow But Clear Path. Will PA Secure Victory for Trump? An Interview With Alan Dershowitz	Neutral	*Endorse*	Endorse

We re-analyze our main results with the human labeled data instead of the topic model data. As

---

30. At the time of our review, one of the 25 discrepancies had been set to private, precluding our ability to review its content.

illustrated in Figure 21, we continue to find persistent positive associations between self-reported skepticism and recommendations of content that either promotes or neutrally reports on Trump's narrative of election fraud, but only weak evidence for the labels of correcting the misinformation or the irrelevant categories.
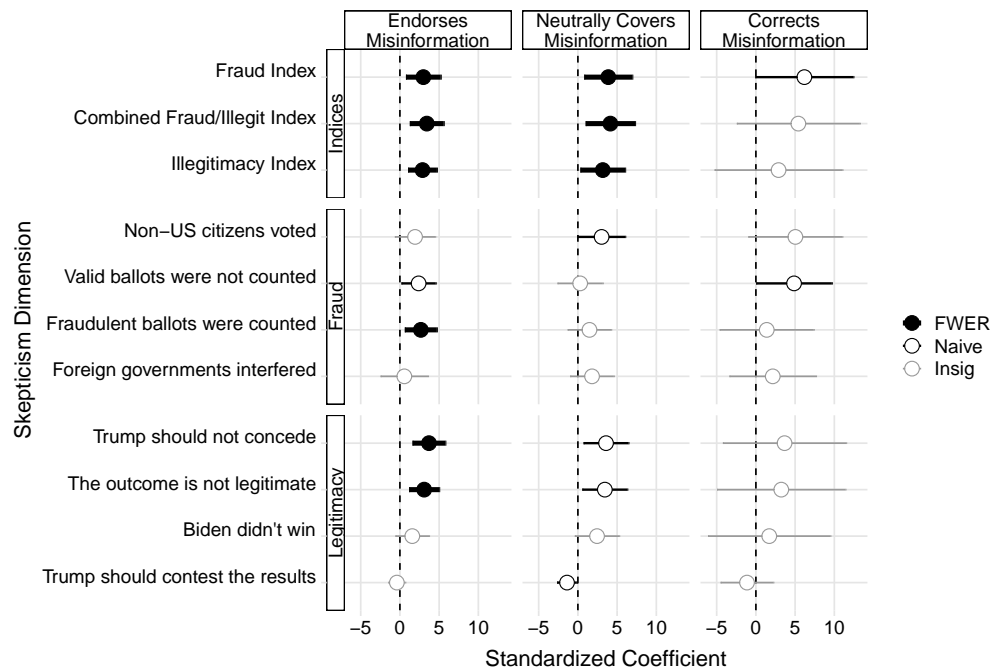
Figure 21: Re-estimating main results replacing the LDA-based outcome with the human labeled data for 6,006 videos. All other recommendations are included, setting the outcome label value to zero. Light gray confidence intervals and points indicate insignificant estimates, dark gray indicates significance at the 95% level of confidence, and black with white points indicates significant relationships after adjusting for the family-wise error rate (FWER).

Subsection B.1: Prevalence of Election-Related Content

Our main findings indicate that YouTube's algorithm recommended significantly more fraud-related content to users skeptical about the legitimacy of the election. But how prevalent was this content on the platform overall? We begin with a description of the prevalence of videos about the 2020 U.S. Presidential Election on YouTube. We then use text analysis to summarize the content of these videos by applying topic models to their transcripts, identifying the content that is about fraud or election-related conspiracies.

We begin with a random sample of YouTube videos generated using the method described in **ytrandomsample**. We subset to videos estimated to be in English. We then select videos published between August 1, 2020 and December 1, 2020 that mention "election" in the video description, video title, or video tags. In Figure 22, we show the proportion of election-related videos on YouTube, the proportion of views they receive, and the estimated proportion of comments they generate. We find that election-related videos peaked in the week of the election, as expected. Additionally, we find that in the lead up to the election, election-related videos received a proportionately higher number of views and comments, meaning that these videos were more engaging than the average set of videos on the platform. Volume, views, and comments peaked around the November 3rd election and the week after while votes were still being tallied.
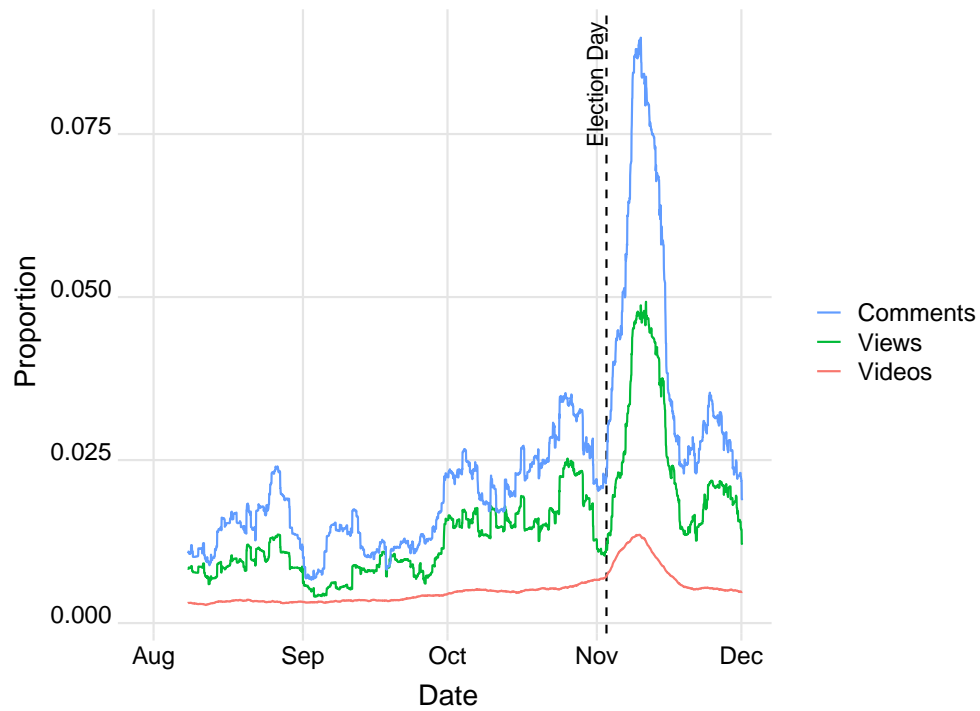
Figure 22: Prevalence of election-related videos. We plot the proportion of videos in a random sample that are election-related (red), the proportion of views that are of election-related videos (green), and the proportion of comments that are on election-related videos (blue).

Appendix C: Scorer Description

One of the challenges associated with unsupervised methods like LDA is how to manually identify topics of interest at scale. To overcome this challenge and estimate our correlation coefficients for a variety of choices of k, we developed a scorer that searched for fraud-related keywords among the bigrams (concatenated with underscores) associated with each topic. To be as encompassing as possible, we relied on regular expressions, a sequence of characters that are used by R to search for this pattern in longer strings of text. Special characters such as pipes, carrots, and dollar signs specify additional search logic, such as OR statements, beginning of lines, or end of lines respectively. For example, the regular expression `((voter|election|voting)_)*fraud` would search for the bigrams `voter_fraud`, `election_fraud`, `voting_fraud`, and `fraud`. The regular expressions we used include:

- `((voter|election|voting)_)*fraud`
- `(stopthesteal|stop_steal|stolen_election)`
- `(sharpiegate|sharpie_gate)`
- `(countthevotes|count_votes)`
- `(stopthecount|stop_count)`
- `maidengate`
- `(deadvoters|dead_voters)`
- `(rigged_vote|rigged_election|stole_election)`
- `poll_watchers`
- `ballot_audit`
- `^mail_in$`
- `absentee`
- `recount`
- `^wayne$|michigan`
- `irregular(ity|ities)`
- `philadelphia|pennsylvania`
- `sidney|powell`
- `rudy|giuliani`
- `discrepanc(y|ies)`
- `maricopa|arizona`
- `fraud`
- `recount`
- `dominion|smartMatic`
- `fake_news`
- `mainstream_(media|news)`

The counties and cities (Maricopa County, Arizona; Wayne County, Michigan; Philadelphia, Pennsylvania) are those which were the focus of targeted recount efforts by the Trump campaign. Rudy Giuliani and Sidney Powell were two of Trump's primary lawyers overseeing the legal

efforts for overturning the election results. Finally, Dominion and SmartMatic were two voting machine software producers at the center of conspiracy theories that the machines themselves were responsible for shifting votes toward Biden. We also look for co-occurrence of these words with "fraud", "fake_news", and "mainstream_media" to capture the dimension of the fraud narrative about how the mainstream media is lying to the nation.

The scorer extracts the values of $\phi_{w,k}$ for each of these keywords w for each topic k, which captures the topic-word probability. It then sums these probabilities to capture the net probability of fraud-related keywords occurring in a given topic k. To aggregate these keyword probabilities up to a recommended video, the scorer then multiplies the topic-document probability $\theta_{k,d}$ by the summed $\phi_{w,k}$ and sums over each topic to obtain an overall proportion of fraud-related content for a given video.

The scorer is a useful tool for quickly identifying which topics are most likely to be of interest. While we follow the advice of **roberts2014stm** and manually review every topic, the scorer nevertheless streamlines the process when dealing with robustness checks in which we vary the total number of topics from between 50 and 500, and re-estimate the LDA models on documents with and without lemmatization, run on corpora that include or omit transcripts. However, just because the scorer helps focus our attention based on the appearance of keywords, it doesn't mean that any video with a non-zero score is about the topic of substantive interest.

As illustrated in Figure 23, the top 5 highest scoring topics for $k = 150$, run on the meta data without lemmatization cover the spectrum of relevance. The highest scoring topic (#108 which we use in our main analysis) is clearly relevant, while several of the subsequent topics are clearly not about election fraud in the 2020 U.S. presidential election. These mistakes are driven by the scorer identifying only one or two key terms out of the total bag of terms we describe above.

Topic #108

Score: 3.76

news

fraud, giuliani, rudy_giuliani, voter_fraud, michigan, arizona, ballots, powell, election_fraud, sidney, mail_in, sidney_powell, recount, ballot, fake_news, in_ballots, in_michigan, irregularities, fraud_claims, fraud_in, fraud_election, giuliani_trump, dominion, absentee, georgia_recount, trump_rudy, fraud_trump, fraud_voter, fraudulent, hand_recount, arizona_election, voterfraud, giuliani_giuliani, fraud_voting, voting_fraud, michigan_state, fraud_vote

trump

follow

election

donald

donald_trump

president

live

politics

georgia

Topic #138

Score: 1.43

fifa

bet

team

wayne, rudy, rudymancuso, rudy_mancuso

wayne

awards

career

google

best

mode

player

Topic #89

Score: 0.82

china

ntd

us

chinese

toys

videos

family

fraud, stop_steal, dominion, vote_fraud, fraud_vote, fraud_ntd, voting_fraud, fraud_election, fraud_voting, michigan, dominion_voting, voterfraud, arizona_state, voter_fraud, sidney_powell, wayne, recounts, rudy, michigan_state, in_michigan

tom

toy

scp

Topic #29

Score: 0.42

biden

election

trump

news

joe

joe_biden

debate

presidential

president

us

rudy, michigan, in_arizona, giuliani_rudy, dominion_voting, voter_fraud, recounts, rudy_giuliani, mainstream_media, stop_steal, giuliani, recount, vote, fraud, irregularities, fraud, arizona, state

Topic #25

Score: 0.3

crime

images

police

getty

getty_images

true

case

true_crime

murder

unsolved

recounted, cases_recounted, recounted_reenactments, philadelphia, michigan, arizona, new_dominion, in_philadelphia

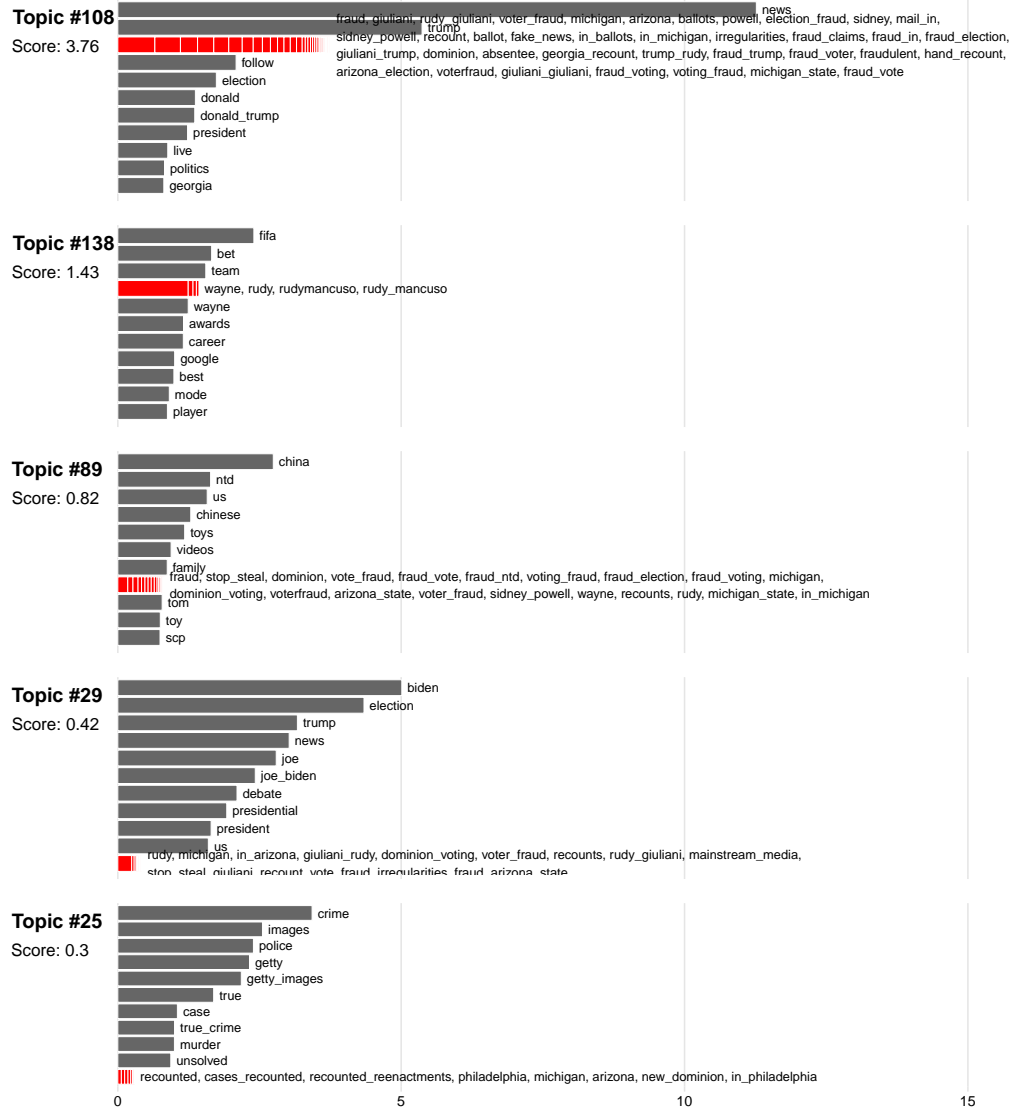0   5   10   15

Figure 23: Top 10 terms associated with the top 5 highest scoring topics. X-axes indicate the $\phi_{w,k}$ values for each term, with the fraud keywords given in red.

Appendix D: Robustness

Appendix D.1   Placebo Test

Our main results used an LDA topic model with 150 topics, combined with human labeled data, to identify content that promoted Trump's false claims of election fraud. Specifically, our main analysis focuses on a specific topic (#108) which we demonstrate is about election fraud, and in particular endorses conspiracies about various aspects of the 2020 U.S. presidential election. By documenting a systematic positive association between skepticism about the legitimacy of the election and the number of recommendations that were about this topic, we argue we uncover evidence of a pernicious side effect of YouTube's recommendation algorithm.

However, our results might be spurious if the participants who were more skeptical about the election were also being recommended disproportionately more content about other topics that spiked in popularity around the same time. As illustrated in Figure 24, there was another topic whose popularity rose and fell in tandem with the prevalence of videos about election fraud: the viral video game called "Among Us". We identify this topic in our data (#143, highest scoring terms and videos summarized in Figure 25) and use it as a placebo and test whether it is similarly positively associated with participant skepticism about the legitimacy of the presidential election.
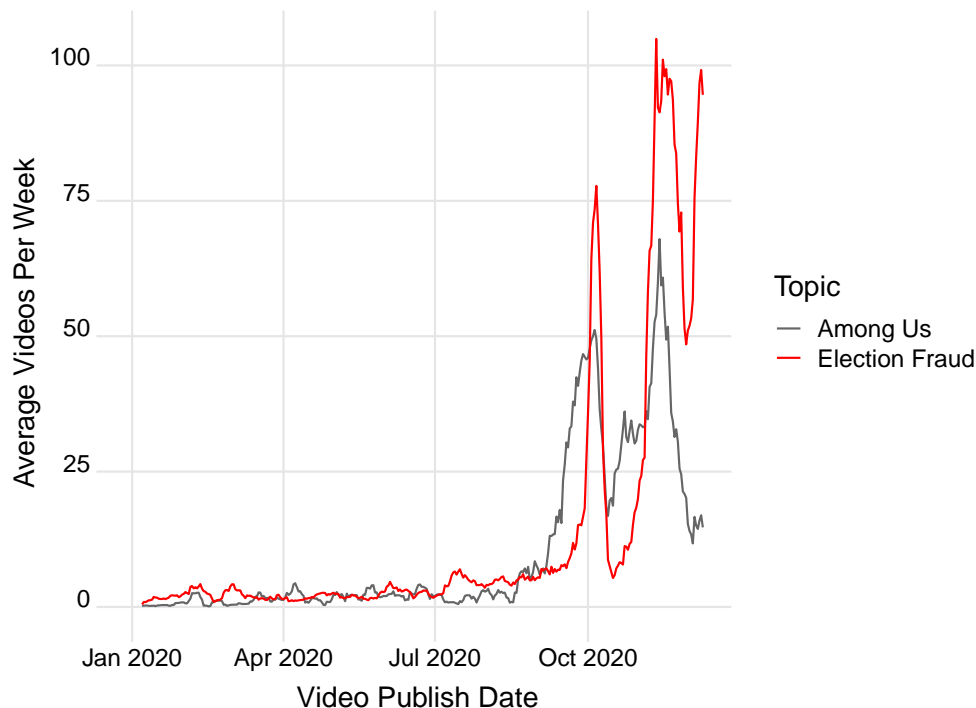
Figure 24: Total number of weekly videos published over the course of 2020 about Among Us (topic #143 in gray) and election fraud (topic #108 in red).

Unlike the strong positive associations we document for our fraud topic, we find no similar evidence when examining the Among Us topic (see Figure 26).

Appendix D.2   Topic Robustness

The main results cherry picked one topic for analysis out of 150 from the LDA model, based on the appearance of fraud-related keywords and validated by examining the highest scoring videos.

Topic #143

Topic Scoring Terms

us	
among	
among_us	
funny	
impostor	
twitch	
gameplay	
tv	
moments	
us_among	
toast	
imposter	
funny_moments	
resident	
twitch_tv	
resident_evil	
sykkuno	
us_funny	
disguised	
iq	

Phi: 0.0   2.5   5.0   7.5

Topic Scoring Videos

Among us be like
we pulled off an INSANE 500 IQ impostor strategy
cleanest impostor play i've ever done in among us
200 IQ Impostor Moments | Among us
Among Us: The MedBay Scan
Among Us But With 1000 Buff Impostors
they thought the game was over but...
I'VE BEEN GASLIT!! | Among Us w/ The Derp Crew & Friends
winning with 500 IQ vent plays as impostor...
My simple deduction trick saved us from the impostor
the most ambitious Among Us crossover event...
the 300 IQ solo impostor clutch...
I thought this impostor game was over but...
the 1500 IQ impostor backstab she never expected...
the 6300 IQ DOLPHIN DIVE sacrifice play...
using 1100 IQ to catch famous Youtubers in Among Us...
going off on a 3300 IQ calculated killing spree....
Among Us:Why Does No One Cares About The Reactor Meltdown
THE LUCKIEST IMPOSTOR GAME | Among Us w/ Friends
the 1400 IQ impostor DOUBLE KILL combo...

Theta: 0   25   50   75   100

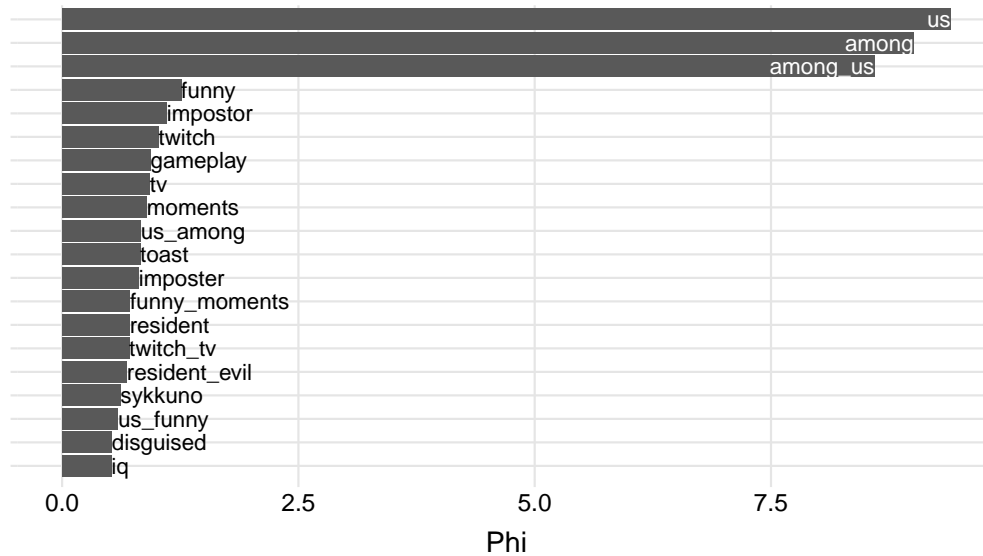Figure 25: Highest scoring terms and videos for Topic #143.

But there are 149 other topics we can analyze from this one LDA model; hundreds more that we calculated choosing different values of the total number of topics *k* (including 20, 50, 100, 150, 200, 250, 300, 400, and 500 topics); and thousands more generated by running the LDA on lemmatized versus raw text, and on the meta-data only, the transcripts only, or a combination of both. These decisions are typically relegated to a black box with the preferred topic chosen and justified with little in the way of systematic procedures.

The benefit of our scorer is that, not only does it help us quickly identify the topics that are most likely of interest, but in doing so it generates a standardized, continuous measure of the concept

Figure 26: Coefficients on election skepticism on the number of recommendations for content about the Among Us video game.
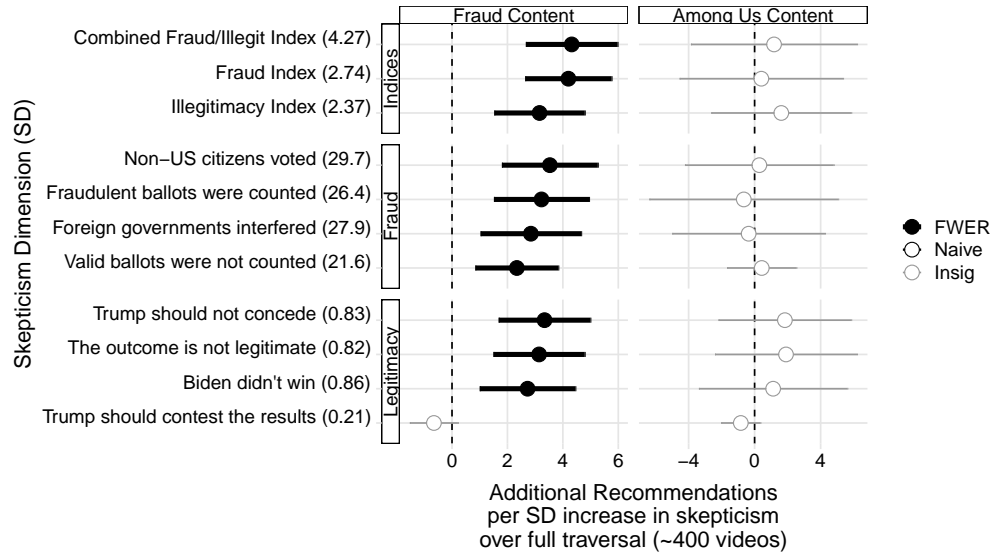
of interest. As such, a more general test of our conclusion applies our scorer to every topic, predicts the relationship between recommendations and skepticism, and models the strength of this relationship as a function of how highly the topic scored on our fraud metric. While the scorer by itself does not ensure all identified topics are necessarily correct, we should expect that the strength of the statistical association between skepticism and the fraud recommendations should be increasing in the fraud score.

Figure 27 plots the fraud score on the x-axes, and the t-statistic for the coefficient linking a dimension of skepticism with the proportion of fraud-related content recommended to a given user. Points are shaded to indicate whether the t-statistic indicates a statistically significant association at the 95% confidence level (un-adjusted). As illustrated, for all measures of skepticism except for those related to Trump's decision to contest or concede the election, the relationship between statistically significant positive correlations between skepticism and being recommended fraud-related content is positive. Substantively, this plot captures the intuition of our underlying placebo test – topics that are more likely to actually be about election fraud are those where we find the strongest positive relationships between skepticism and recommendations.
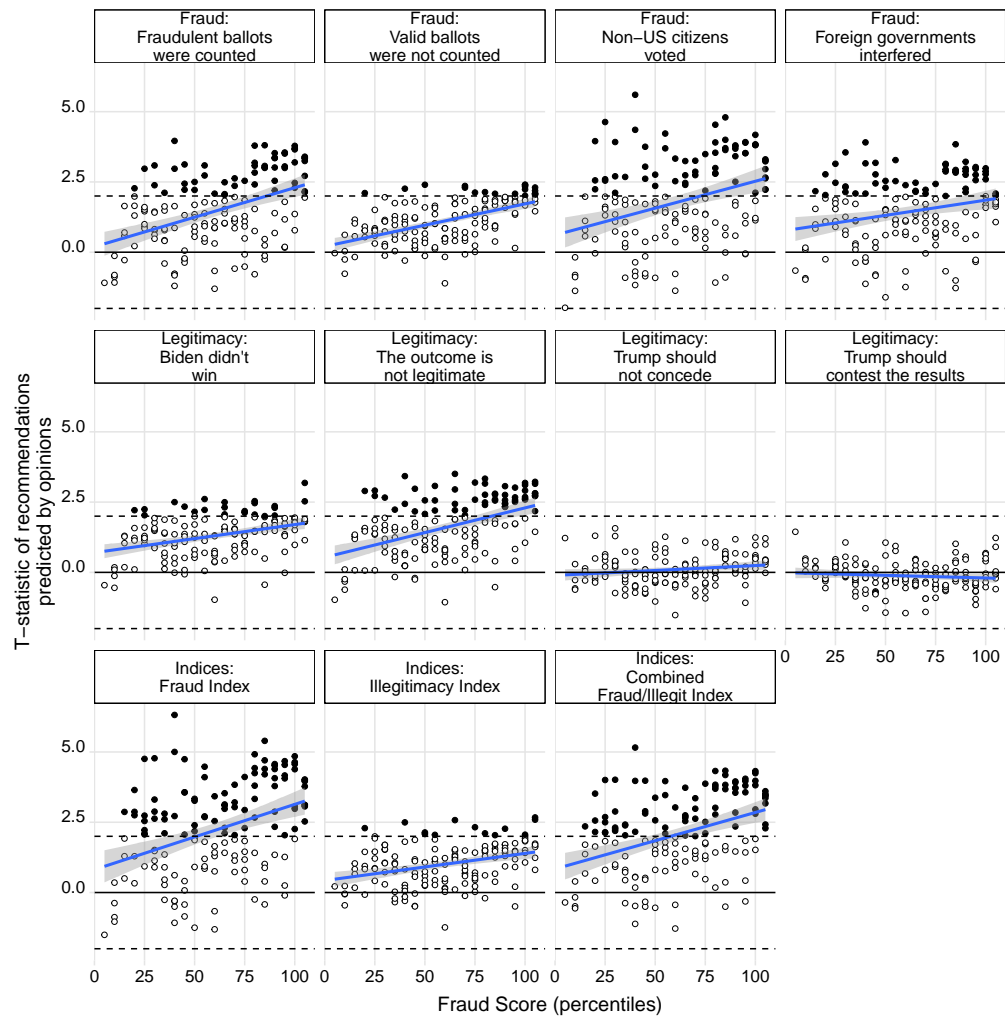
Figure 27: Placebo checks of whether the statistical association between skepticism and fraud-related recommendations (y-axes) is stronger for topics that score higher on the fraud scale (x-axes).

Appendix D.3   Specification Robustness

Our main results control only for the seed video, week, and traversal rule. As discussed in the paper, we do not include participant covariates as controls in order to accurately describe the extent to which YouTube differentially recommended videos about election fraud to participants most concerned about the legitimacy of the election. Below, we examine the robustness of our results to the inclusion of these participant covariates, as well as to the choice of whether to run the regression on the recommendation-level data (i.e., where each row indexes a participant-traversal step-recommended video), the traversal step-level data (i.e., where we average recommendations by participant-traversal step), or the participant-level data (i.e., where we average recommendations by the participant). As illustrated in Figure 28, adding participant covariates reduces the strength of the association between skepticism and recommendations, although overall doesn't remove the significant positive association entirely.
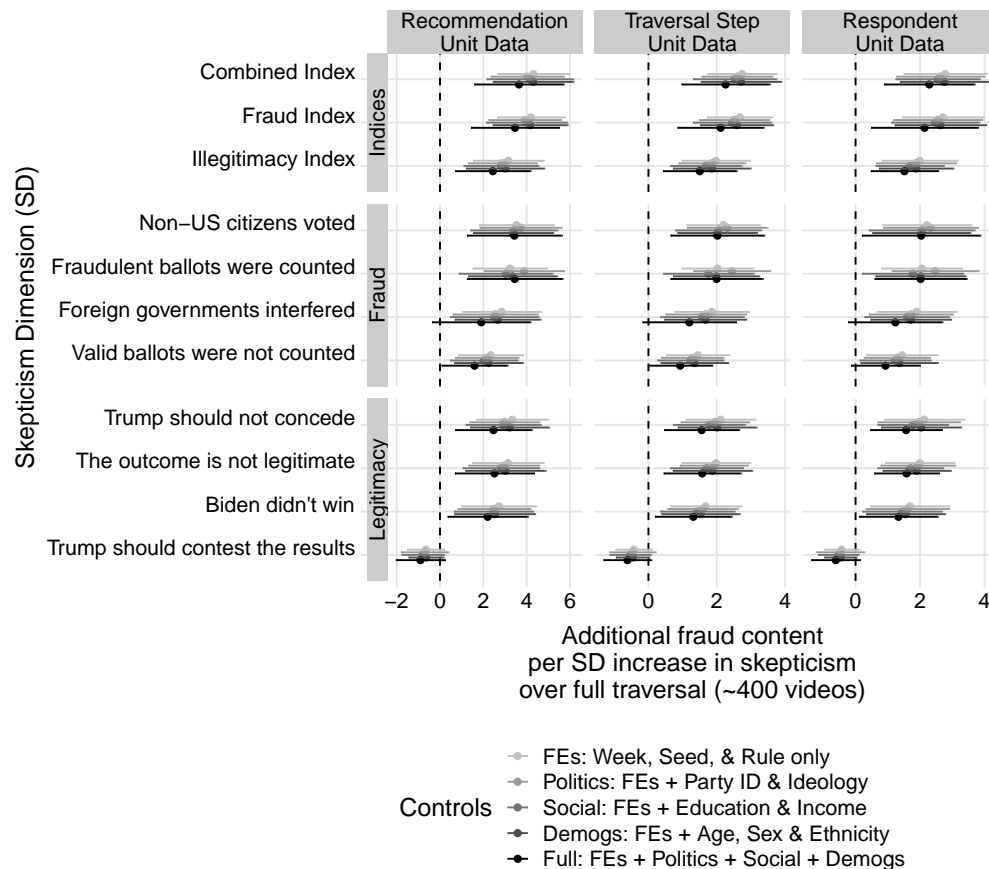
Figure 28: Robustness to the inclusion of participant covariates.

How to interpret the sensitivity of the correlations to the inclusion of these participant-level controls? Substantively, the results suggest that the algorithm is less able to differentially recommend videos about election fraud to two users sharing the same demographic profile who differ in terms of their concern about election fraud. Does this mean that the algorithm did not influence the information environment of our participants via its recommendations in the fall of 2020? Given that we randomly assigned users to a seed video and a traversal rule, that we observe a significant increase in fraud-related recommendations among our most skeptical participants indicates that the algorithm did influence the information environment. But by controlling for user covariates, we demonstrate that the algorithm is less able to infer different user beliefs among – for example – moderate white

male college educated middle income Millennial Democrats. In other words, these characteristics are proxies for the personalized data used by the algorithm to infer preferences.

Appendix D.4   User Watch Histories

We offered an additional $5 inducement for users to provide us with their watch histories, and included detailed instructions for how to download this information from their YouTube account (described above in SI Section 1). A total of 153 respondents provided us with watch histories. We then estimated the ideology of every video these participants had ever watched on YouTube which was still actively hosted on the site at the time of data collection, using the method described in [REDACTED]. Figure 29 visualizes the average ideology (x-axes) by participant's self-reported ideology (y-axes), focusing only on those videos that are categorized as either "News & Politics" or "People & Blogs".[31] As illustrated, there is some evidence that self-reported liberals spend more time watching liberal News & Politics than self-reported conservatives, although when it comes to People & Blogs, all groups appear to watch roughly the same distribution.
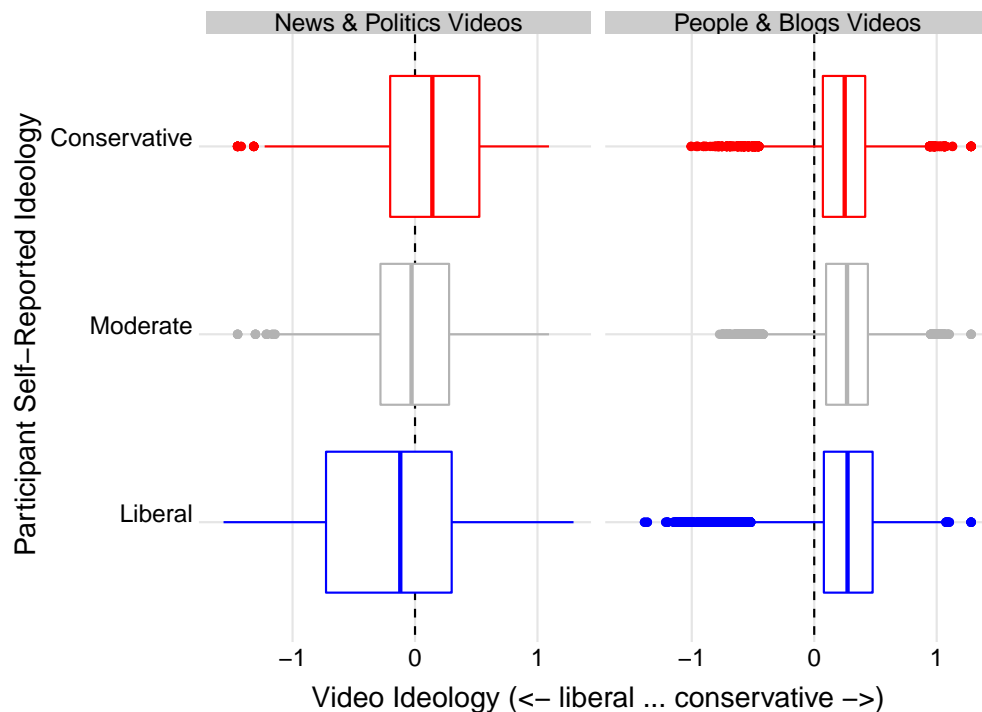
Figure 29: Distribution of ideology of videos watched by participants prior to taking our survey.

We test whether our results are robust to the inclusion of user watch histories by comparing our main results that predict recommendations for fraud-related content as a function of participant concern that fraudulent ballots were being counted. We summarize our results in Table 3, illustrating that the independent association between fraud recommendations and participant concern persists even after controlling for different measures of conservative content in watch histories. As such, we can conclude that there is an algorithmic bias which is based on more than the participants' expressed preferences from previously observed behaviors on the platform.

But does this mean that watch histories don't matter at all? To investigate, we run a final regression in which we predict fraud recommendations as a function of participant concern interacted with

---

31. These labels are applied by YouTube. More information can be found at https://techpostplus.com/youtube-video-categories-list-faqs-and-solutions/.

the logged count of videos in their watch histories. We would expect that the algorithm should be better able to suggest fraud-related videos to users concerned about election fraud who have richer information embedded in their watch histories, particularly among the categories of News & Politics and People & Blogs content. We plot the marginal effects in Figure 30, supporting this conclusion. Specifically, the more personalized information available to the algorithm, the better able it is to recommend fraud-related content to fraud-concerned users.
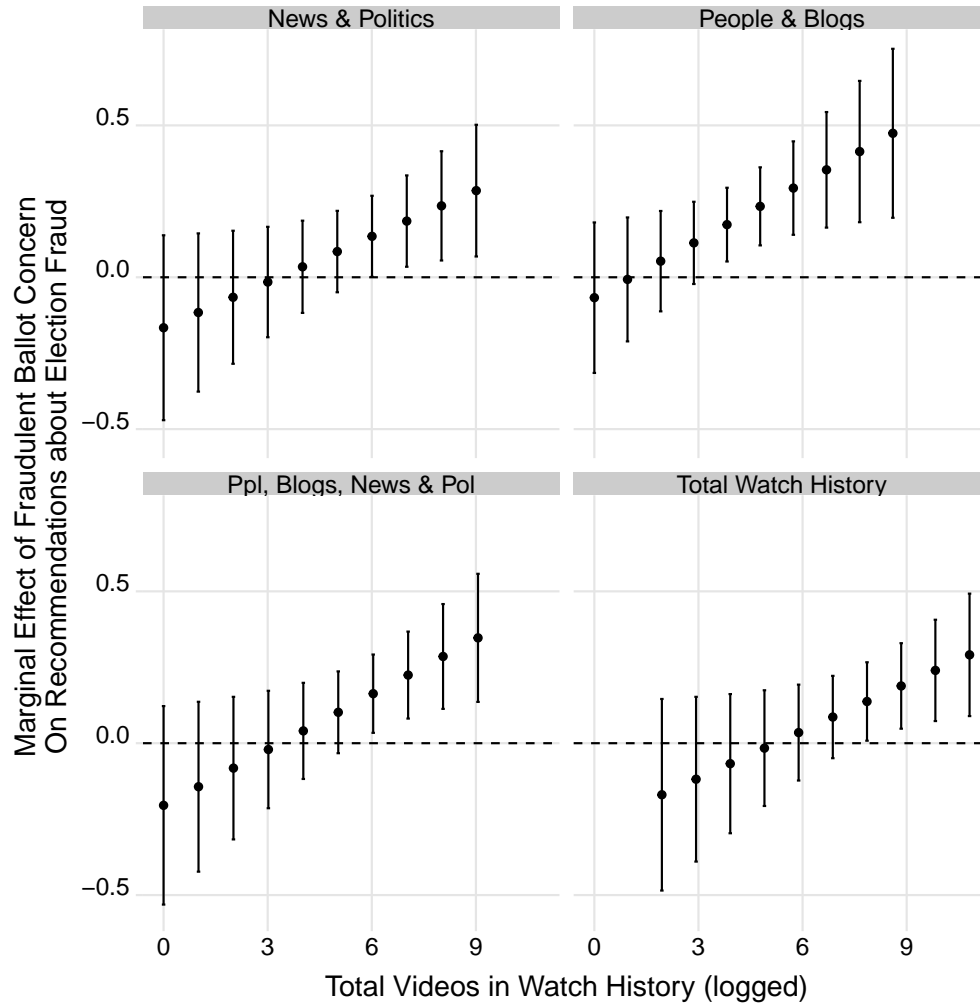
Figure 30: Marginal effects of concern about fraudulent ballots being counted on prevalence of recommendations of fraud-related content (y-axes) among participants with smaller or larger numbers of videos in their watch histories, separated out by News & Politics (first panel), People & Blogs (second panel), News, Politics, People & Blogs (third panel), and all videos in the watch histories (fourth panel). All interaction terms are statistically significant.

Table 3: Regression table testing robustness to controlling for watch histories.

Dependent Variable:	scale(TOPIC_VAL_108)			
Model:	(1)	(2)	(3)	(4)
Variables				
scale(elec2020_conc_fraud_ballot_count)	0.1333**	0.1256***	0.1499***	0.1034**
	(0.0509)	(0.0372)	(0.0501)	(0.0416)
YOB	-0.0042	-0.0070	0.0011	-0.0027
	(0.0047)	(0.0064)	(0.0056)	(0.0046)
educHighschoolgraduate	0.2685	0.1090	0.4815*	0.2714
	(0.4200)	(0.2527)	(0.2620)	(0.2841)
educSomecollegebutnodegree	-0.2765	-0.4431*	-0.1029	-0.3003
	(0.2909)	(0.2261)	(0.2520)	(0.2537)
educCollegeDegree	0.2312	0.1919	0.5730	0.2736
	(0.2789)	(0.3080)	(0.3364)	(0.3341)
educGraduateDegree	-0.0901	-0.0233	0.5518	0.2520
	(0.3738)	(0.3173)	(0.3296)	(0.2870)
male	-0.0213	-0.1206	-0.0305	-0.0358
	(0.1037)	(0.1877)	(0.1700)	(0.1516)
ethnicityBlack	-0.1464	-0.1716	-0.1406	-0.0393
	(0.1027)	(0.1250)	(0.1677)	(0.1968)
ethnicityAsian	-0.0740	-0.1529	-0.1375	-0.0578
	(0.2092)	(0.1866)	(0.1654)	(0.1337)
ethnicityNativeAmerican	0.0879	0.0051	0.0768	-0.0847
	(0.1682)	(0.1975)	(0.2288)	(0.1898)
ethnicitySomeotherrace	-0.2885	-0.5361	-0.6623*	-0.6297**
	(0.1691)	(0.3410)	(0.3394)	(0.3022)
inc40kto80k	-0.0943	-0.0757	0.0199	0.0262
	(0.1777)	(0.1637)	(0.2050)	(0.1682)
incMorethan$80k	-0.1456	-0.2446	-0.2671	-0.2795
	(0.1886)	(0.2205)	(0.2345)	(0.2166)
ideo3Moderate	-0.0747	-0.0217	-0.0325	0.1245
	(0.2189)	(0.1923)	(0.1982)	(0.1572)
ideo3Conservative	-0.1035	0.0276	-0.0093	0.1365
	(0.1053)	(0.1702)	(0.2110)	(0.2064)
overallHistIdeo		-0.0684		
		(0.4495)		
newPolHistIdeo			0.3989	
			(0.2457)	
peopleBlogsHistIdeo				0.3306
				(0.2413)
Fixed-effects				
week	Yes	Yes	Yes	Yes
travRule	Yes	Yes	Yes	Yes
seedVid	Yes	Yes	Yes	Yes
Fit statistics				
Observations	338	153	135	126
R^2	0.49433	0.59754	0.66022	0.66680
Within R^2	0.09190	0.18012	0.25422	0.24882

Clustered (seedVid) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Appendix E: Reverse Causality

Our main analysis predicts the content recommended to our participants as a function of their skepticism about the legitimacy of the 2020 US presidential election. However, the actual sequencing of our study asked the participants to complete the traversal task first, and then asked for their views on election legitimacy. As such, it is possible that the content that they were suggested while taking our survey may have caused them to update their views on the legitimacy of the election.

As discussed in our paper, we believe this is unlikely due to 1) the speed at which participants completed the traversal task and 2) the implausibility of a small thumbnail and video title influencing their beliefs. The first point is measurable, and we emphasize that the vast majority of respondents only spent a few seconds on each video. They were required to wait a minimum of 5 seconds in order for the browser plugin to work correctly, but were not asked to spend any more time on the video than that. Figure 31 plots the distribution of seconds users spent on each video in the traversal task, dropping extreme outliers where it was the clear the respondent started the task and then was distracted by something else, only returning after several hours (or in some cases, days). As illustrated, the duration of time participants spent on each traversal step is typically short, which we argue limits the plausibility of the reverse causality story. (I.e., spending less than 30 seconds on a video is unlikely to influence the participant's views on election integrity.)
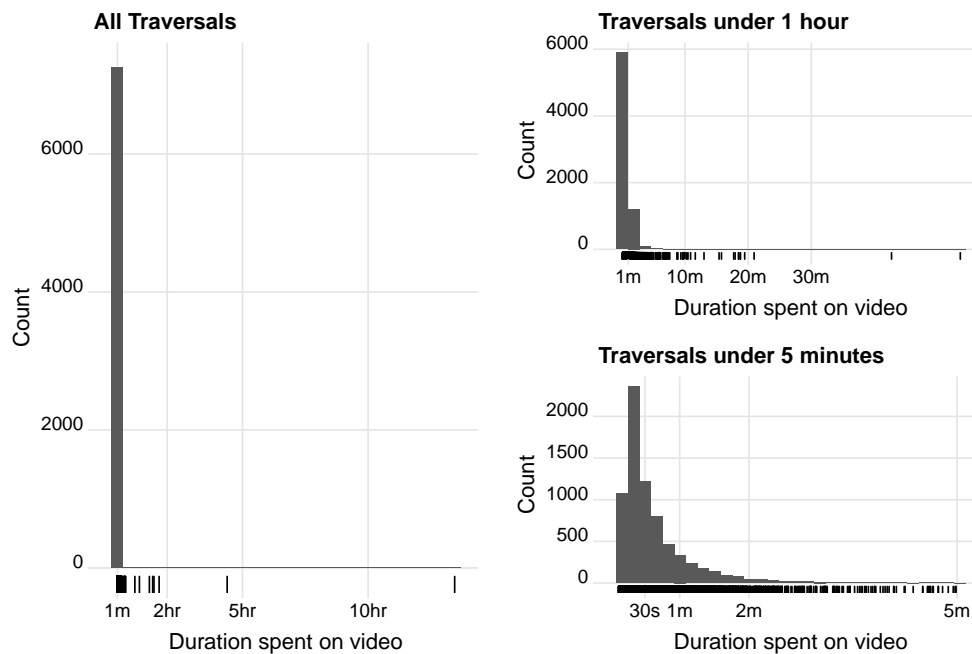
Figure 31: Histogram of amount of time spent on each video in the traversal task.

We can also examine whether videos more associated with fraud were more interesting to participants by predicting duration spent on a given video as a function of content. Figure 32 plots the relationship between the amount of time users spent on a given video and the probability that video was about Topic #108, the most likely topic containing content endorsing Trump's claims about election fraud. As illustrated, there is no systematic relationship between content of the videos and the amount of time participants spent on them across Democrats, Independents, and Republicans. A regression that predicts time spent as a function of the current video's election fraud content, controlling for participant demographics, confirms this conclusion.

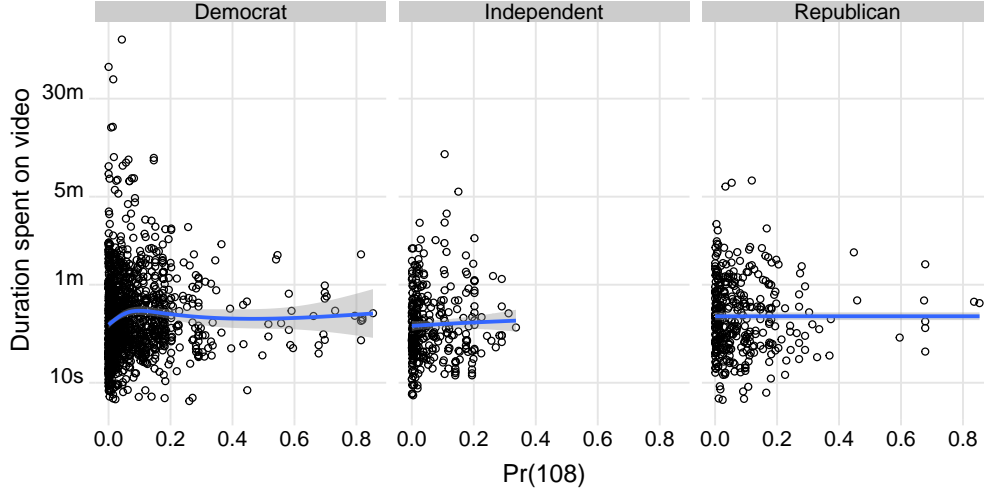Figure 32: Scatterplot of time spent on each video by probability the video was about topic 108, broken out by party affiliation.

As a final test, we can reverse the main regression and predict concern as a function of the fraud content of the videos that were actually watched by participants during the traversal task, interacted with how long they spent watching them. Formally, for participant i watching video j in week w at traversal step t:

$$Skep_i = \beta_1 \text{fraud}_{t,j} + \beta_2 duration_{i,t,j} + \beta_3 \text{fraud}_{t,j} \times duration_{i,t,j} + \delta_w + \alpha_s + \gamma_r + \lambda_t + \varepsilon_{i,t,j,w} \quad (4)$$

where $Skep_i$ is the level of concern about fraudulent ballots being counted (scale ranging from 0 to 100) expressed by participant i, $\text{fraud}_{t,j}$ is the topic loading for Topic #108 for video j watched at traversal step t, $duration_{i,t,j}$ is the amount of time participant i spent watching video j at step t, and δ_w, α_s, γ_r and λ_t are fixed effects for the week, seed video, traversal rule and traversal step, respectively. If consumption of the fraud content actually does influence beliefs, we would expect to see higher levels of user concern among those who spent more time watching videos about election fraud – i.e., $\beta_3 > 0$.

We plot the marginal effects from this regression in Figure 33. Overall, we find a positive but statistically insignificant relationship (top-left panel of Figure 33. However, as the figure makes clear, there are some extreme outliers when using the full data as a handful of participants spent extreme durations of time on a single traversal step, likely due to leaving the task and returning to complete it later. We re-estimate the same specification on increasingly small subsets of the data, limiting attention to where the support of the moderator is denser. Doing so continues to indicate that there is no systematic relationship between duration of time spent on a video about election fraud and the skepticism of the user. If anything, the association is negative among the densest part of our data where users spent less than a minute on a given video.

As a final test, we re-run our main results, dropping participants who spent longer than a certain threshold of time on the videos on average. The idea being that, if the reverse causality story is true, it would be less true among those who only spent a very short period of time on each video overall. Thus if our main results are driven by participants whose beliefs were influenced by the videos they were randomly assigned to click on, re-estimating our findings only among those who could not have possibly spent enough time on each video to have it influence their beliefs should produce null results. Yet as illustrated in Figure 34, the main results of a positive association between user concern about fraudulent ballots being counted and recommendations for content about election fraud persist even as we drop increasingly greater numbers of respondents who spent too long on average. While the point estimate in our most conservative sample is more noisily estimated, this is
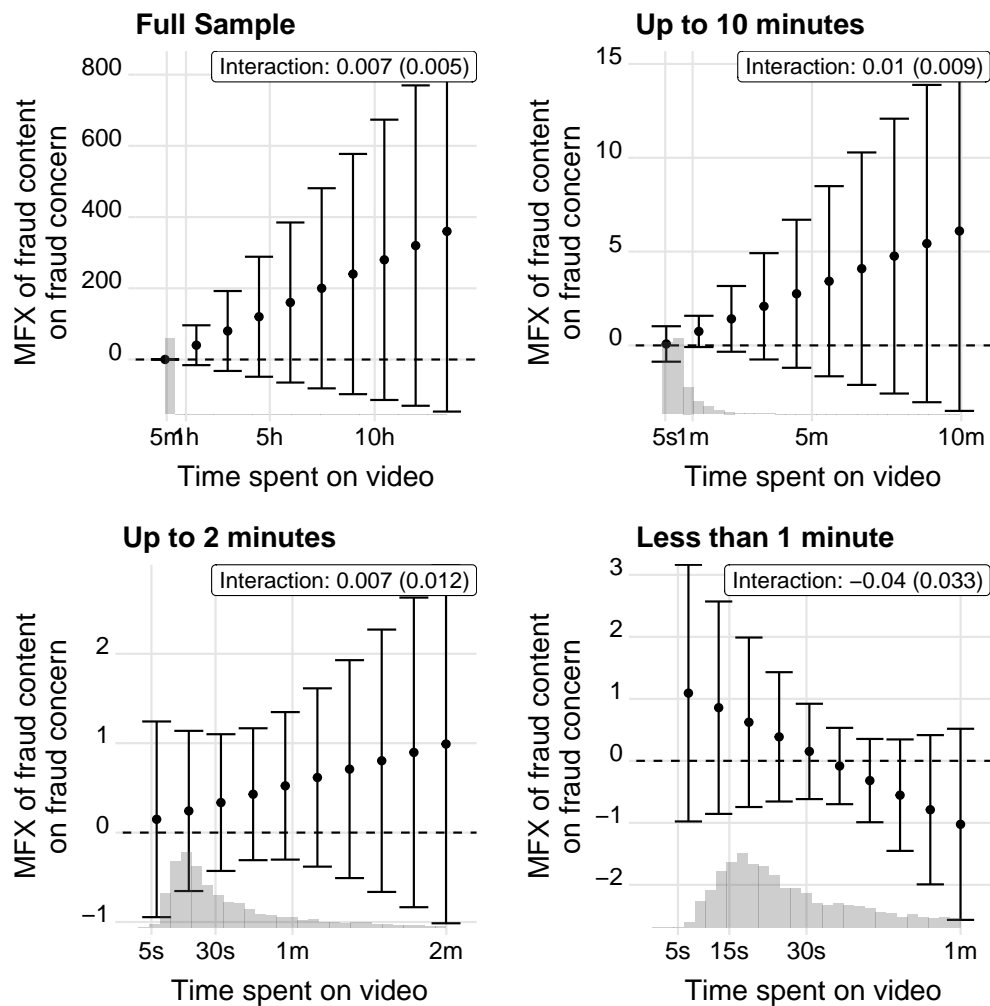
Figure 33: Marginal effects of watching fraud content on beliefs about fraudulent ballots being counted, among those who spend less or more time watching said videos.
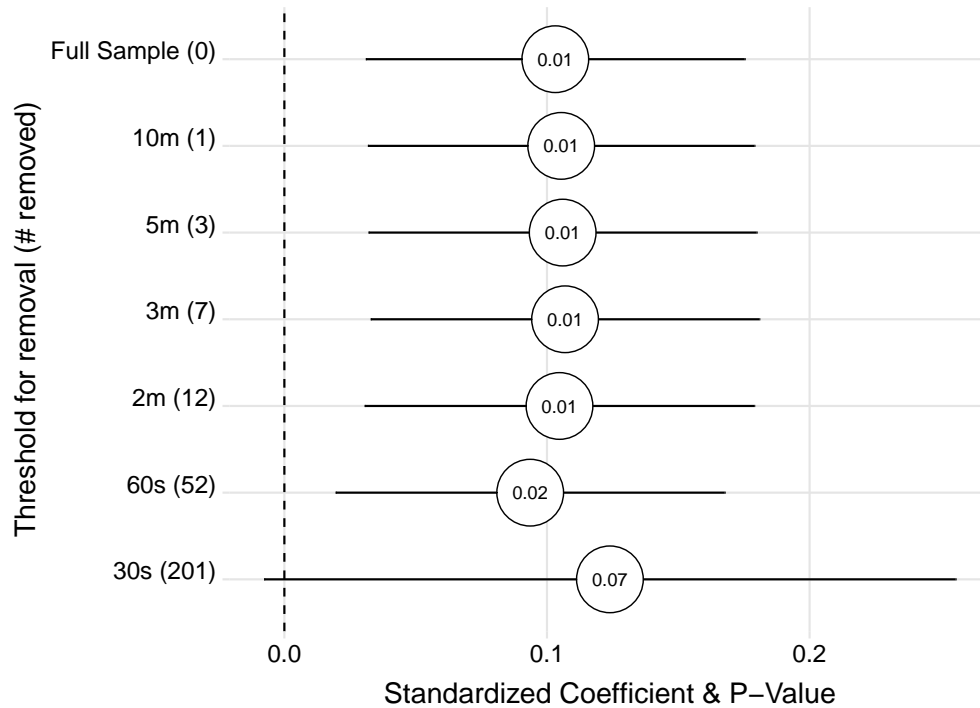
Figure 34: Main results re-estimated on increasingly conservative subsets of the sample where we drop any respondents who spent longer than a certain amount of time on clicked videos on average (y-axis indicates the threshold for removal, along with the total number of participants dropped at this threshold in parentheses). The x-axis displays the standardized coefficients describing the association between concern about fraudulent ballots being counted and the proportion of recommendations about election fraud. Substantively, these coefficients correspond to similar magnitudes found in the main results, i.e. approximately 3 additional videos suggested to users for each standard deviation increase in the concern measure (roughly 26 points on a concern scale ranging from 0 to 100). P-values are indicated in text.

due to dropping almost two-thirds of our respondents who spent longer than 30 seconds on average on each video. Substantively, the pattern holds, suggesting that reverse causality is unlikely to be driving our results.

The second point regarding the influence of videos about election fraud that were recommended but not viewed is less amenable to empirical investigation since we do not know whether the participants paid attention to the list of recommendations they were shown. We argue that, even if they did closely examine the list of videos displayed as small thumbnails on the right with truncated titles, these are unlikely to contain sufficient information to influence the respondents' beliefs.

Appendix F: Missing Data

We did not process the traversal data until January of 2021. This delay meant that we were unable to obtain information on certain recommendations due to them having been taken down by either the creator or YouTube, or set to private. We cannot assume that these missing recommendations are missing completely at random (MCAR) or even missing at random (MAR). Indeed, it is very likely that the very qualities of videos that we are interested in – their endorsement of election fraud-related misinformation – is prognostic of their missingness. We know that YouTube intervened to remove election misinformation starting December 8th, and that it had already removed over 8,000 channels between September and December 9th, 2020 (https://blog.youtube/news-and-events/supporting-the-2020-us-election/). If many of the videos that are missing in our data are missing because they contain fraud-related content, we would be undercounting the total size of the problem, and likely underestimating the strength of the relationship between user skepticism and algorithmic recommendations.

Upon closer inspection, however, there isn't overly alarming evidence of this being an issue in our data. First, there are only 3,065 total instances where we are missing a recommendation for a given user on a given traversal step, out of 146,735 total participant-step-recommendation in total, or roughly 2%. This corresponds to 1,528 unique recommendations that are missing out of a total of 49,415, or roughly 3.1%. Furthermore, this missingness does not appear to be systematically associated with the degree to which respondents were concerned that fraudulent ballots were being counted, as displayed in Figure 35.
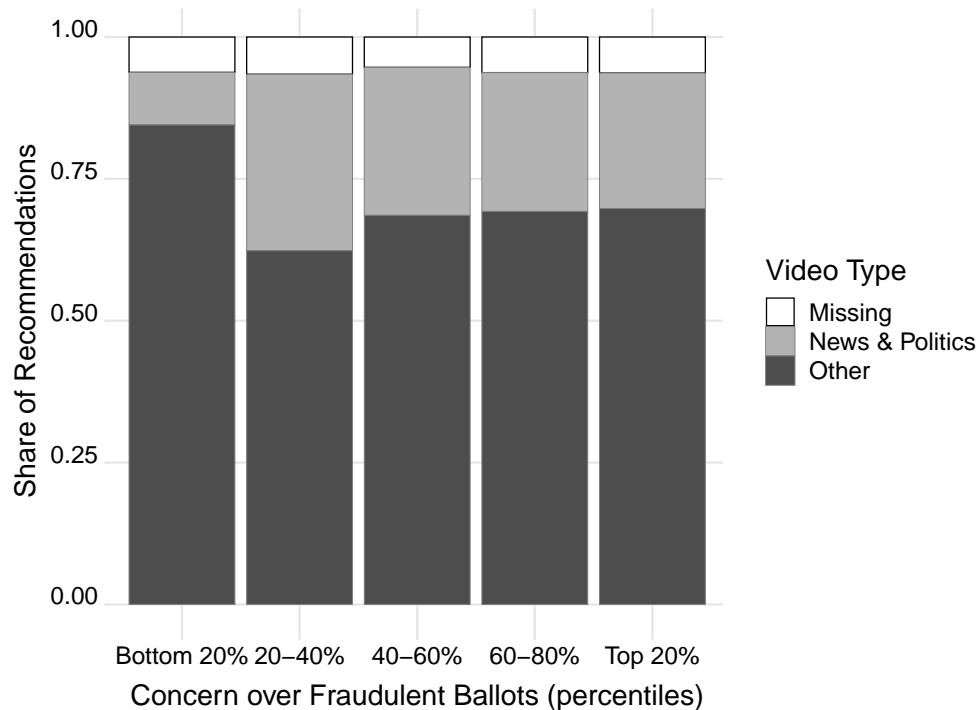
Figure 35: Proportion of recommendations that are classified as either missing at the time of data processing (white), categorized by YouTube as about News & Politics (light gray), or categorized as anything else (dark gray), broken out by the quantiles of concern that fraudulent ballots were being counted (x-axis).

Second, when we predict missingness as a function of user skepticism, we find no systematic relationship, regardless of the specification decisions. If anything, missingness is negatively associated with our measures of skepticism, although this relationship is sensitive to the choice of

controls and appears primarily driven by the concern over foreign government interference, which was not one of the main predictors in the main results. We plot the coefficient estimates and naive 95% confidence intervals in Figure 36. None of the statistically significant estimates persist when we adjust for the family-wise error correction rate.
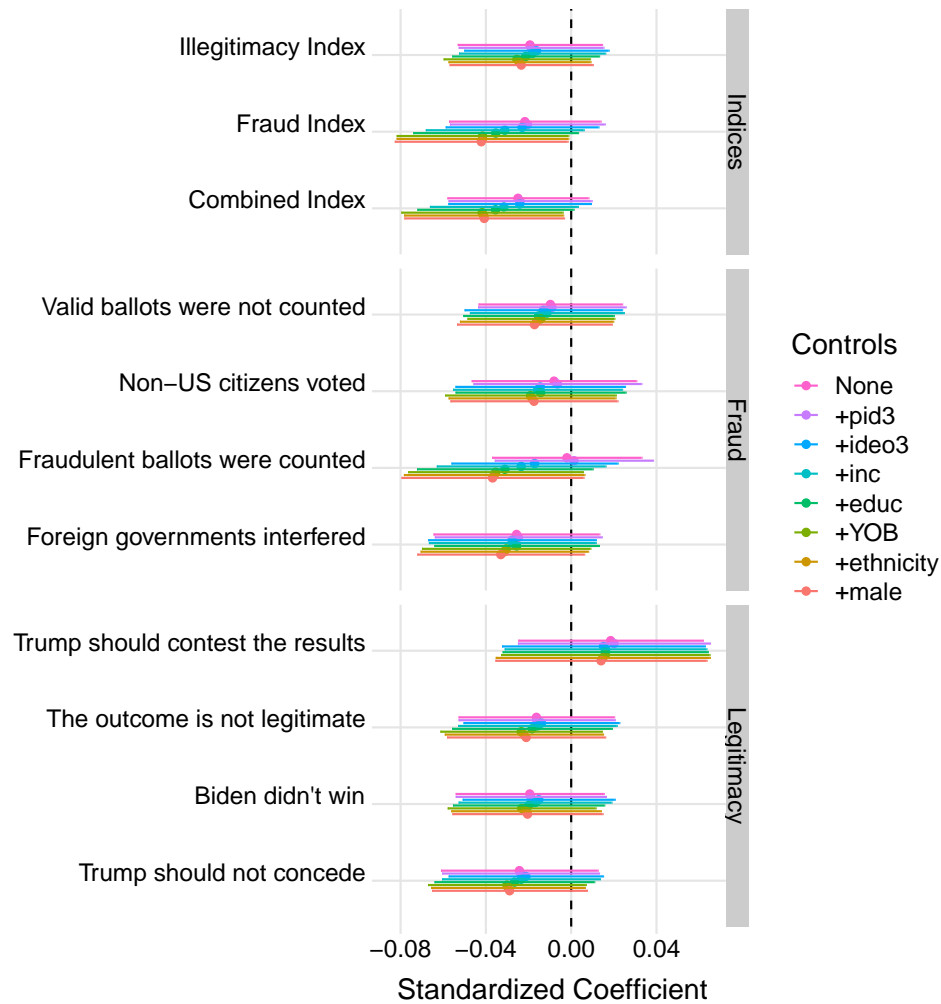
Figure 36: Coefficients (x-axis) estimating the relationship between skepticism about the election's legitimacy and whether the video has been removed from YouTube (y-axis) using different specifications (colors).

Appendix G: Regression Tables

We include the raw regression tables for our main results below, as well as a robustness check for the inclusion of participant controls.

Table 4: Regression table summarising concern questions from main results.

Model:	(1)	(2)	(3)	(4)
Variables				
fraud-ballot-count	0.181***			
	(0.055)			
valid-ballot-nocount		0.128**		
		(0.049)		
non-us-vote			0.195***	
			(0.061)	
foreign-int				0.167***
				(0.054)
Fixed-effects				
week	Yes	Yes	Yes	Yes
seedVid	Yes	Yes	Yes	Yes
travRule	Yes	Yes	Yes	Yes
traversal_step				
Fit statistics				
Observations	338	338	336	338
R^2	0.339	0.327	0.344	0.336
Within R^2	0.038	0.021	0.043	0.033

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 5: Regression table summarising legitimacy questions from main results.

Model:	(1)	(2)	(3)	(4)
Variables				
bidenLost	0.149***			
	(0.054)			
Trump-not-concede		0.188***		
		(0.055)		
Trump-contest			-0.038	
			(0.024)	
illegit				0.176***
				(0.047)
Fixed-effects				
week	Yes	Yes	Yes	Yes
seedVid	Yes	Yes	Yes	Yes
travRule	Yes	Yes	Yes	Yes
traversal_step				
Fit statistics				
Observations	338	338	338	338
R^2	0.331	0.343	0.314	0.339
Within R^2	0.027	0.044	0.002	0.038

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 6: Regression table summarising indices from main results.

Model:	(1)	(2)	(3)
Variables			
combinedFraud	0.239***		
	(0.056)		
combinedIllegit		0.175***	
		(0.052)	
combinedFull			0.244***
			(0.057)
Fixed-effects			
week	Yes	Yes	Yes
seedVid	Yes	Yes	Yes
travRule	Yes	Yes	Yes
traversal_step			
Fit statistics			
Observations	336	338	336
R^2	0.358	0.339	0.364
Within R^2	0.064	0.038	0.072

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 7: Regression table summarising controls for the fraud ballots concern question.

Dependent Variable:	scale(TOPIC$_1$08)				
	recc	recc_linweight	recc_expweight	travs	resp
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
fraud_ballot_count	0.069***	0.067***	0.069***	0.133***	0.178***
	(0.022)	(0.022)	(0.023)	(0.045)	(0.063)
male	0.007	-0.0001	-0.006	0.021	0.031
	(0.040)	(0.040)	(0.041)	(0.081)	(0.114)
YOB	-0.0003	-0.0006	-0.0007	-0.002	-0.003
	(0.002)	(0.002)	(0.002)	(0.004)	(0.005)
ethnicityBlack	-0.041	-0.043	-0.040	-0.067	-0.091
	(0.056)	(0.055)	(0.054)	(0.111)	(0.149)
ethnicityAsian	-0.082	-0.082	-0.072	-0.144	-0.195
	(0.055)	(0.055)	(0.058)	(0.111)	(0.163)
ethnicityNativeAmerican	0.019	0.026	0.020	0.024	0.021
	(0.064)	(0.064)	(0.064)	(0.123)	(0.175)
ethnicitySomeotherrace	-0.082	-0.093	-0.099	-0.165	-0.227
	(0.073)	(0.073)	(0.071)	(0.139)	(0.218)
inc40$k to$80k	-0.048	-0.057	-0.060	-0.099	-0.133
	(0.071)	(0.070)	(0.069)	(0.139)	(0.215)
incMorethan$80k	-0.068	-0.074	-0.073	-0.112	-0.146
	(0.058)	(0.059)	(0.058)	(0.116)	(0.173)
educHighschoolgraduate	-0.021	-0.0005	-0.002	-0.068	-0.094
	(0.102)	(0.104)	(0.111)	(0.213)	(0.305)
educSomecollegebutnodegree	-0.124	-0.115	-0.125	-0.221	-0.293
	(0.090)	(0.089)	(0.094)	(0.194)	(0.289)
educCollegeDegree	0.026	0.033	0.025	0.082	0.109
	(0.100)	(0.100)	(0.102)	(0.213)	(0.267)
educGraduateDegree	-0.102	-0.095	-0.102	-0.162	-0.220
	(0.108)	(0.107)	(0.110)	(0.229)	(0.328)
pid3_Ind	-0.080**	-0.073*	-0.073*	-0.173**	-0.228*
	(0.041)	(0.041)	(0.042)	(0.085)	(0.120)
pid3_Rep	0.097**	0.111**	0.133**	0.181*	0.238*
	(0.047)	(0.048)	(0.052)	(0.096)	(0.140)
ideo3Moderate	-0.062	-0.064	-0.063	-0.136	-0.179
	(0.057)	(0.059)	(0.065)	(0.119)	(0.172)
ideo3Conservative	-0.136***	-0.133***	-0.134***	-0.271***	-0.362**
	(0.044)	(0.043)	(0.044)	(0.089)	(0.144)
Fixed-effects					
week	Yes	Yes	Yes	Yes	Yes
seedVid	Yes	Yes	Yes	Yes	Yes
travRule	Yes	Yes	Yes	Yes	Yes
traversal_step	Yes	Yes	Yes		
Fit statistics					
Observations	143,355	143,103	143,355	6,694	338
R^2	0.072	0.073	0.077	0.225	0.399
Within R^2	0.013	0.013	0.014	0.059	0.126

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Appendix H: Recommendation Rank

Recommendations can appear in different positions to the right of a given video, and are ordered based on user interest (**covington2016deep**). Our main results ignore this placement. Below, we re-run our main results on the most granular version of our data where rows index participant (i)-traversal step (t)-recommendation (j), weighting the recommendations by the inverse of their rank (i.e., videos that appear at the top of the list are given the greatest weight while those that appear further down are given less weight). By using the inverse, we effectively are using hyperbolic weights such that the first recommendation's weight is twice that of the second, thrice that of the third, and so on. Alternative tests that calculate linear weights (i.e., 21 - the recommendation rank for the top 20 recommendations) are substantively similar. Formally, our main specification can be modified as follows:

$$y_{i,t,j,w} = \beta_1 Skep_i + \delta_w + \alpha_s + \gamma_r + \lambda_t + \varepsilon_{i,t,j,w} \tag{5}$$

where δ_w, α_s, γ_r and λ_t are fixed effects for the week, seed video, traversal rule and traversal step, respectively. In this setting, we cluster the standard errors at the level of the participant. To facilitate substantive interpretation, we convert the standardized coefficients to reflect the total number of additional fraud videos recommended for each standard deviation increase in skepticism over the full traversal for our participants.
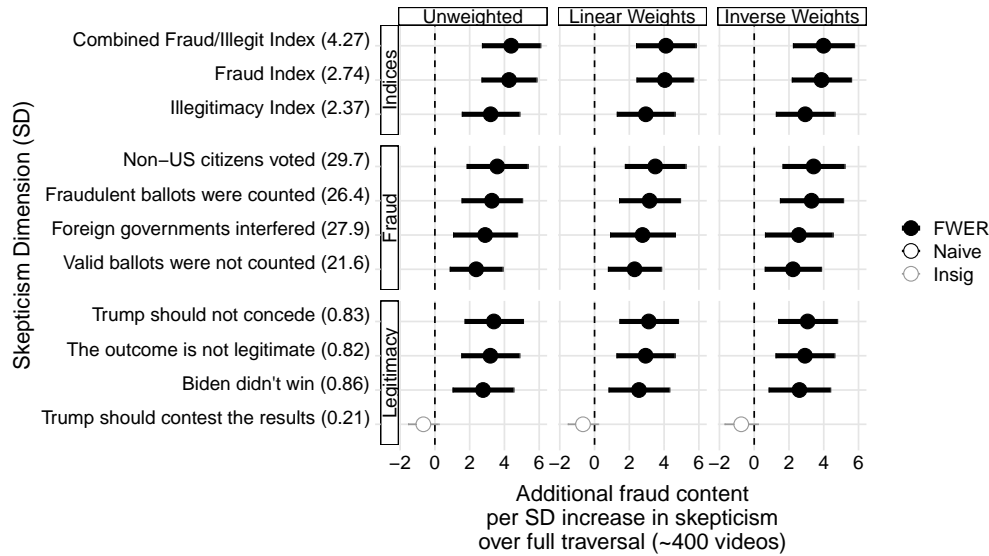

Figure 37: Robustness to choice to weighted by recommendation rank, either via linear weights (21 - the rank of the recommendation, center column), or via inverse weights (1/the rank, right column).

We can also interact user skepticism with the recommendation rank itself to investigate whether the relationship is stronger for videos that are suggested higher in the list. Specifically, we estimate the following specification on the most granular version of our data where rows index participant (i)-traversal step (t)-recommendation (j).

$$y_{i,t,j,w} = \beta_1 Skep_i + \beta_2 rank_j + \beta_3 Skep_i \times rank_j + \delta_w + \alpha_s + \gamma_r + \lambda_t + \varepsilon_{i,t,j,w} \tag{6}$$

As above, we cluster our standard errors at the participant. If the recommendation algorithm not only suggests more fraud content to our skeptical participants, but also positions these recommendations higher in the list, we would expect the β_3 coefficient to be negative (lower values indicate higher ranks in the list).
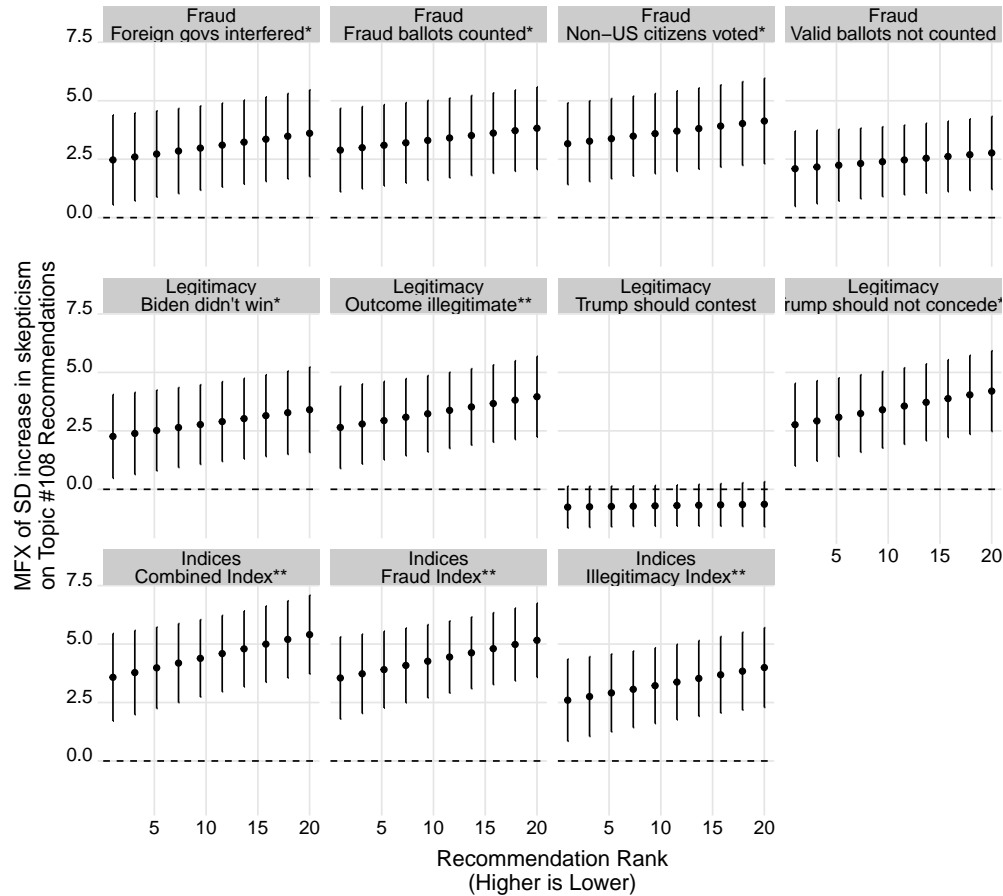
Figure 38: Marginal effect of participant skepticism on fraud recommendations across the position of the recommendation in the list.

We plot these as marginal effects in Figure 38, converting the standardized coefficients to be expressed in terms of total number of fraud videos recommended over the course of the traversal task. As illustrated, we find no evidence to support the conclusion that more skeptical participants were suggested fraud content that appeared higher up the list of recommendations. If anything, there is some evidence of the opposite, wherein the fraud content appears lower down on the recommendation sidebar. Importantly, the marginal effects do not indicate that the differential recommendations to election skeptics are null or negative at higher ranks. Instead, they suggest that the additional fraud-related recommendations for participants concerned that fraudulent ballots were being counted ranges from roughly 2.5 videos in the highest rank to roughly 3.5 in the lowest rank.

Appendix I: Content Curation

Our main analysis focuses exclusively on YouTube's recommendation algorithm and finds significant but modest evidence of participants who were skeptical about the legitimacy of the election being suggested content about election fraud. We concluded our paper by acknowledging that our results are, in a way, unsurprising. Specifically, YouTube's algorithm does what it is designed to do: suggested content that its users are most interested in. As such, a natural question is whether YouTube should modify its algorithm, or instead curate its content, if it wants to avoid contributing to potentially harmful public misinformation. To clearly differentiate these theoretical solutions, one might imagine an algorithmic tweak that incorporates a predicted measure of misinformation in the video metadata, and uses this misinformation score to penalize the appearance of certain videos in the list of a user's recommendations. Conversely, content curation refers to YouTube's explicit efforts to identify and remove harmful or offensive content from its platform, effectively curating the library of possible videos that a user might be recommended.

Properly adjudicating between the efficacy of algorithmic tweaks versus content curation is hampered by the opacity of YouTube policy, particularly when it comes to the trade secret operation of its algorithms. In general, researchers have found little evidence of recommendation algorithms independently contributing to other social outcomes such as echo chambers or extreme content. And in spite of our findings in this paper, it is not clear how or even whether a private company should adjust its algorithms to avoid promoting socially harmful content.

Conversely, what might be achieved with more strict content moderation? We offer suggestive evidence by exploiting the publicized decision of YouTube to crack down on election misinformation, removing content alleging that fraud or errors changed the election outcome, after the December 8th "Safe Harbor" deadline. To do so, we collected a 10% random sample of tweets over the fall of 2020 and searched for those that linked to YouTube videos. We identify 33 million tweets in this sample that linked to YouTube videos. Of those videos, 467,085 were identified to be election-related by selecting videos whose title, tags, or description contained the word "election." We identified 69,230 election-fraud related videos using the keywords referenced in our Methods section. We estimated the fraud content of these videos and plot the 7-day rolling average of the share of tweets linking to election fraud videos on YouTube out of all tweets that link to election-related YouTube videos between August 2020 and February 2021 in Figure 39. As illustrated, there is suggestive descriptive evidence that YouTube's decision to crack down on election misinformation worked, as the availability of fraud-related YouTube videos dropped off precipitously. This aligns with prior research showing that when YouTube implemented changes to its recommendation system in 2019, shares of alt-right and conspiracy content decreased on other platforms (**buntain_yt_recs**). Findings like those in (**chen_and_nyhanadl2021**) indicate that users often find harmful content on other platforms rather than by the recommendation system, so interventions on the platform like YouTube's content moderation intervention could decrease both exposure on the platform via its recommendation system and engagement with that content off-platform.
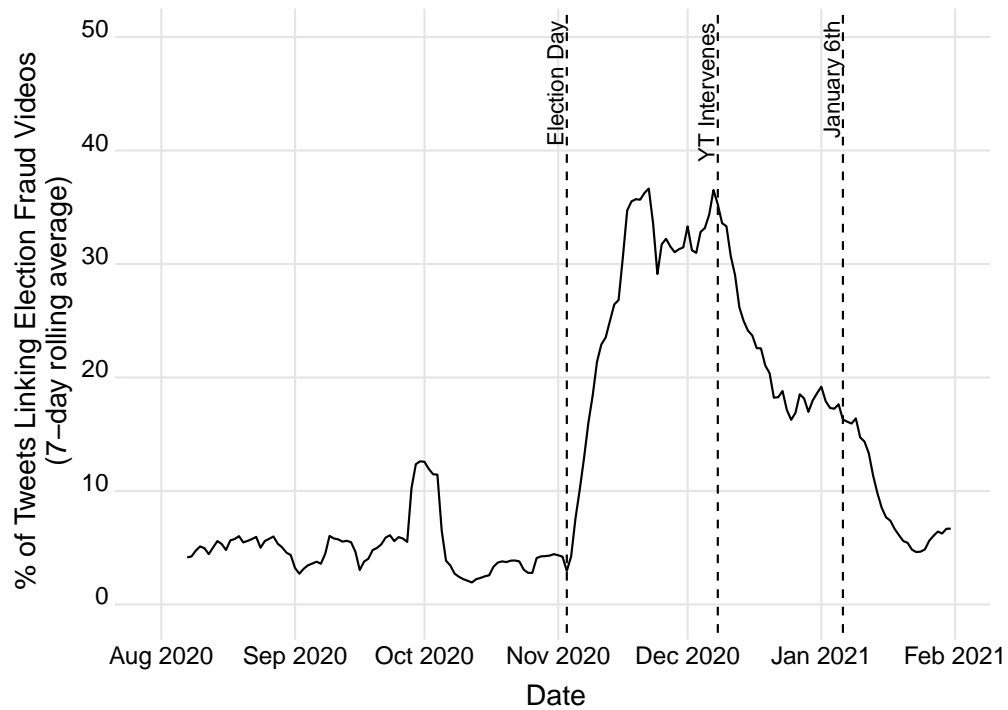
Figure 39: Proportion of tweets that link to a YouTube video about election fraud out of all tweets that link to YouTube videos (y-axis, 7-day rolling average) over time (x-axis). The impact of YouTube's decision to crack down on election fraud content after the Safe Harbor deadline (December 8th) is readily apparent.

Appendix J: Starting Videos

We outline the starting videos and the selection procedure by which we obtained them. We selected videos that were popular on Reddit during the time our survey was conducted, under the assumption that these were videos that online individuals would have been likely to encounter in their day-to-day browsing, improving the ecological validity of our data. Videos were selected such that five were liberal, five were moderate, and five were conservative, based on the ideology scoring method described in (**lai2022estimating**). In addition, we selected seven non-political videos from the categories sports, music, and video games. Throughout the period the survey was in the field, videos were taken down (either by YouTube or by the video owners), requiring us to replace them in the middle of a given week. These videos are enumerated in Table 8.

Table 8: Starting Videos

Category	Video ID	Video Title
Liberal	4b-dannQQ0Q	Jordan Klepper vs. Trump Supporters | The Daily Show
Liberal	q6YyCxK6EnM	Donald Trump: Divider-in-Chief
Liberal	q0lRgmlJOLw	Fed up Fox News host cuts into Trump's speech with blistering fact check
Liberal	z9toCXEjTog	Trump Admin's New Idea To Fight Covid: More Tax Cuts!
Liberal	V1Tsn3D2i0w	How Safe Are U.S. Election Results?
Moderate	KYS2KuBo3Zo	Krystal and Saagar: Feds place Ghislaine Maxwell on SUICIDE WATCH
Moderate	0NdT4qbDIe4	Don Lemon of CNN in 2013
Moderate	VDMkTHwKpbc	FBI Announces Iran, Russia Obtained Voter Registration Information, Interfering In Election
Moderate	ha-7SETmJD4	KGW: What it's like to be a Black officer policing Portland protests | Raw interview
Moderate	_A1cmqbI31M	A Response to Sam Harris on BLM, police violence, and the merits of conversation.
Conservative	lPntJ4k_sXI	WATCH Joe Biden's LAME Trump Rant
Conservative	b7b1NMMoqR4	EXCLUSIVE: Trump takes swipes at Biden in explosive 'Hannity' interview
Conservative	7wUNWjj8a0w	Don Lemon's Virus Lies DEBUNKED!
Conservative	CWdCFd_lgyA	Jessica Doty Whitaker murdered for saying all lives matter, gun sales soar, many people shot
Conservative	hzvWNs0vDAE	Bernie/Biden Task Force Ends In Extreme Failure
Music	Km4BayZykwE	J Balvin - Si Tu Novio Te Deja Sola ft. Bad Bunny (Official Video)
Sports	geFU0y6f-y0	Best post-match interview ever?!| Akinfenwa celebrates Wycombe's promotion
Sports	5JtFrD0mN20	Best Table tennis rally | Must watch !!!
Video Games	AJdQMv3uHtY	JOEL & ELLIE || VENGEANCE
Video Games	LKZ294vf7gI	Rush Racing 2 Trailer
Video Games	GjugTk9ovcI	UFC 4 Official Reveal Trailer
Video Games	bPPc-_BKV4k	WWE 2K18 but it's Quarantine