nature portfolio

https://doi.org/10.1038/s41586-026-10098-2

Supplementary information

The political effects of X's feed algorithm

In the format provided by the authors and unedited

Supplementary Information

Contents

Numbering of SI exhibits (Figures and Tables). The Supplementary Information (SI) is organized into four sections (1–4). Figures and Tables are numbered consecutively *within each section*. Exhibit labels use the format ⟨*section*⟩.⟨*item*⟩, where the first number indicates the SI section and the second number indicates the sequential figure or table number within that section. For example, **Fig. 2.9** denotes Figure 9 in Section 2, and **Table 1.18** denotes Table 18 in Section 1.

1 Experiment

1.1 Ethical Considerations

This research is the result of a collaboration among academics who are independent of the 𝕏 platform. We obtained ethical approval from the Ethics Committee of the University of St. Gallen, Switzerland.

To ensure ethical integrity, our research design incorporates the following key elements: First, participants provided informed consent and were compensated for their participation. Second, the feed settings assigned to participants were limited to options freely available on the platform. This means that no artificial manipulations of users' social media content were introduced. Third, the duration and nature of the intervention were clearly disclosed to participants in advance, although specific hypotheses that we aimed to test were withheld to minimize experimenter demand effects. Lastly, we arranged to share the findings of our research with participants through the YouGov platform, which served as our primary channel for communication with participants throughout the experiment.

Nevertheless, we recognize the ethical considerations associated with experimentally manipulating social media feeds and the potential influence this may have on participants' political attitudes. By randomly assigning participants to different feed settings for seven weeks, we inevitably affected their content exposure and potentially their political views—an effect that was observed in the subset of participants initially using the chronological feed. We were unable to fully mitigate these effects through debriefing after the experiment for two main reasons. First, at the time of the post-treatment survey, the experimental results were not yet available, making it impossible to provide immediate debriefing on the observed effects. Second, given that the experiment involved sustained exposure to varying feed settings over a seven-week period and led to changes in the accounts followed by a subset of participants, any subsequent debriefing might not fully offset the potential cumulative effects of prolonged exposure to distinct information environments.

We carefully weighed these concerns against the broader societal importance of understanding how social media algorithms influence political attitudes. Considering that these algorithms already shape the daily experiences of hundreds of millions of users with limited public oversight or scientific scrutiny, we concluded that the potential benefits for society gained by obtaining this knowledge through our experiment outweigh the risks to individual participants.

1.2 Details of Experimental Design

The experiment was pre-registered with the American Economic Association's registry for randomized controlled trials (AEARCTR-0011464).[1] We discuss the key elements of experimental design in the main text and the Methods Section. The flow chart, presented in the Extended Data Fig. 1, summarizes the structure of the experiment and the sample size at each stage. Here, we provide supplementary information on the compensation scheme and on how different elements of the experiment appeared on the 𝕏 platform.

Compensation scheme: Survey respondents were compensated in points, a currency used within the YouGov platform, with an exchange rate of 1,000 points equaling 1 USD. For the pre-treatment survey, they received 500 points (USD 0.50), with clear information that they would earn 2,500 points (USD 2.50) for using 𝕏 with the assigned feed setting between the two surveys and returning for the post-treatment survey. Additionally, participants could earn an extra 2,000 points (USD 2) for using the Chrome extension during the pre-treatment survey. After completing the post-treatment survey, they were paid the 2,500 points (USD 2.50) that had been announced already during the pre-treatment survey. Again, they could earn additional compensation for running the Chrome extension, this time 2,500 points (USD 2.50). Therefore, participants could earn up to 10,000 points (USD 10) if they completed both surveys and used the Chrome extension both pre- and post-treatment. For ethical considerations, sharing the 𝕏 handle (identifying information) was not incentivized.

𝕏 **platform:** Fig. 1.1 shows a screenshot of the 𝕏 feed page in its usual layout. 1.2 provides a screenshot while our Google Chrome extension collects posts.

[1] In addition to the experiment described in this paper, the pre-analysis plan also included an experiment involving undergraduate students at the University of St. Gallen. All students enrolled in a large introductory economics course were invited to complete the pre-treatment survey. We anticipated participation from approximately 100 students. However, it became clear that very few undergraduates actively used 𝕏, and only 38 students enrolled in the pre-treatment survey. As a result, we discontinued this experiment without conducting a post-treatment survey and chose to focus on the larger study involving US-based 𝕏 users.

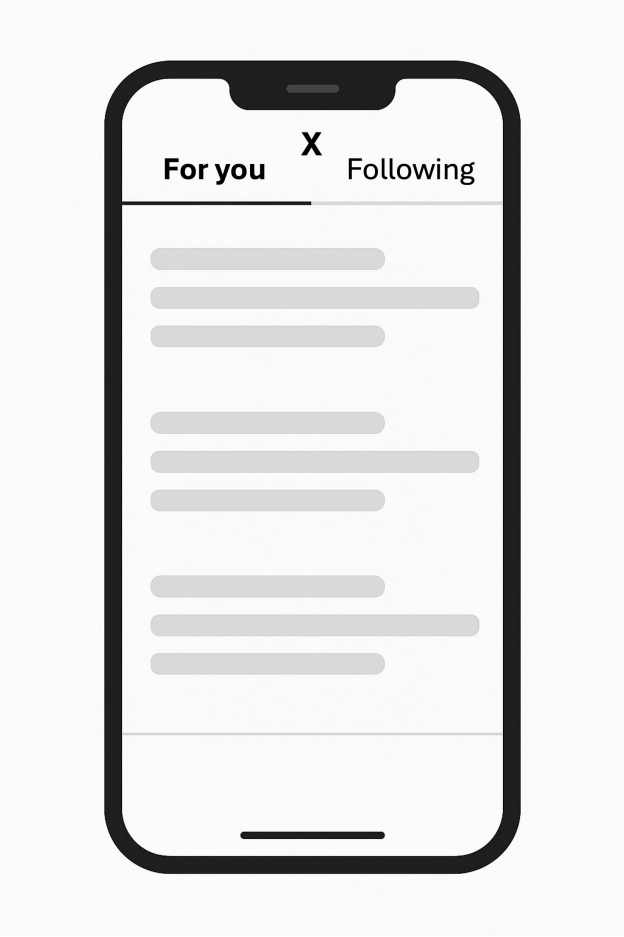


Fig. 1.1. Schematic Illustration of the 𝕏 Homepage on a Mobile Device.

At the top, the “For You” and “Following” tabs correspond to the algorithmic feed and the chronological feed of posts, respectively. By clicking on the respective tab, users choose the feed setting. As of July 2023, the authors’ tests showed that the setting remained unchanged when a user logged out and then reconnected on the same device.

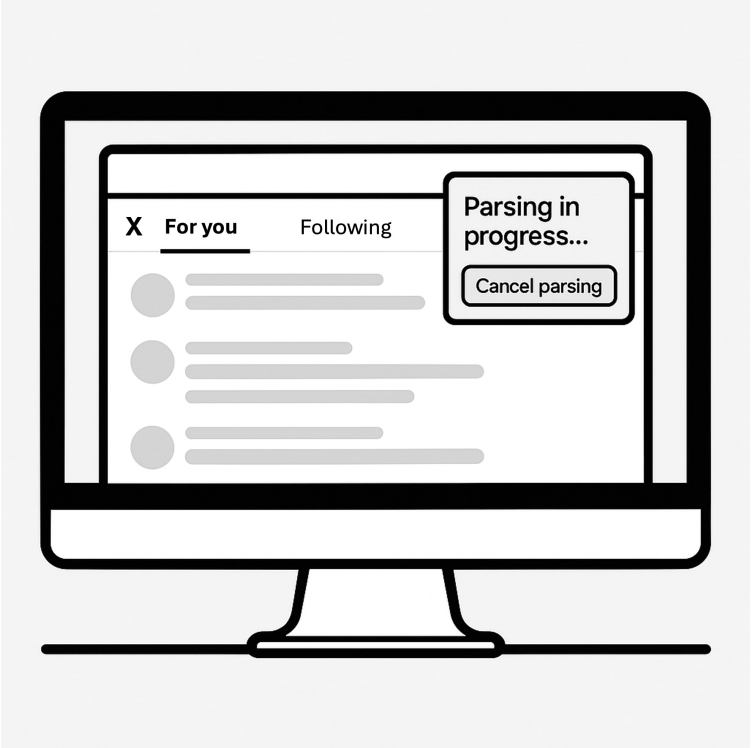


Fig. 1.2. Schematic Illustration of the Chrome Extension Running on a User's 𝕏 Feed on a Desktop Computer.

To launch the extension, the user clicked on an icon at the upper right of the Chrome browser (which appeared after the user had installed our extension) on their desktop computer. Then, a small box appeared with a button labelled "Start parsing." Once clicked, the extension collected 100 posts for each feed setting: algorithmic ("For You") and chronological ("Following"). Once the data were collected, a CSV file was downloaded to the user's device. The respondent then uploaded this file as a survey response.

1.3 Description of Survey-Based Outcome Variables

In this section, we describe the survey-based outcomes. Table 1.1 provides definitions for each individual outcome variable by topic. To address concerns related to multiple hypothesis testing, we use the individual outcome variables as inputs to construct several composite outcomes for each topic using Principal Component Analysis (PCA) or taking the intersection of all answers. Table 1.2 reports the factor loadings for each PCA composite outcome variable and the proportion of the variance explained by the first component.

The rebranding of Twitter to 𝕏 took place during our study period. Accordingly, the platform is referred to as Twitter in the pre-treatment survey and as Twitter/𝕏 in the post-treatment survey.

User Engagement	
Variable	**Question Asked**
Label: User engagement (PCA) **Coding:** First component of a principal component analysis performed on the two variables below. Positively correlates with higher user engagement compared to pre-treatment levels.	See Table 1.2 for information on factor loadings.
Label: Use 𝕏 no less than before **Coding:** Takes value 1 if frequency increased or stayed the same between pre- and post-treatment, 0 otherwise.	**Pre-treatment:** *How often do you use Twitter?* **Post-treatment:** *How often have you used Twitter/X over the past few weeks?* **Options:** *"Several times a day"*, *"Once a day"*, *"Several times a week"*, *"Once a week"*, *"Several times a month"*, *"Less than once a month"*
Label: Post on 𝕏 no less than before **Coding:** Takes value 1 if frequency increased or stayed the same between pre- and post-treatment, 0 otherwise.	**Pre-treatment:** *How often do you tweet?* **Post-treatment:** *How often have you tweeted over the past few weeks?* **Options:** *"Several times a day"*, *"Once a day"*, *"Several times a week"*, *"Once a week"*, *"Several times a month"*, *"Once a month or less"*, *"Never"*

Table 1.1. Survey Questions and Variable Definitions

Partisanship and Feelings toward Political Parties	
Variable	**Question Asked**
Label: Partisanship and affective polarisation (PCA) **Coding:** First component of a principal component analysis performed on affective polarisation and post-treatment Republican partisanship. This variable is only defined for Democrats and Republicans. Positively correlates with pre-treatment Republican partisanship.	See Table 1.2 for information on factor loadings.
Label: Positive feelings toward Republicans **Coding:** Takes values from 0 (very cold) to 100 (very warm).	**Pre-treatment:** *In the list that follows, rate that party using something we call the feeling thermometer. Ratings between 50 degrees and 100 degrees mean that you feel favourable and warm toward the party. Ratings between 0 and 50 degrees mean that you don't feel favourable toward the party and that you don't care too much for that party. You would rate the party at the 50 degree mark if you don't feel particularly warm or cold toward the party.* **Post-treatment:** Same question. **Options:** 0 to 100, or *"Don't know"*
Label: Positive feelings toward Democrats **Coding:** Takes values from 0 (very cold) to 100 (very warm).	
Label: Affective polarisation **Coding:** Feelings toward Republicans minus feelings toward Democrats, scaled to take values between 0 and 100; higher values indicate more favourable views toward Republicans. This variable is only defined for Democrats and Republicans.	
Label: Republican partisanship **Coding:** 1 if Republican, 0 otherwise.	**Pre-treatment:** Generally speaking, do you usually think of yourself as a Democrat, a Republican, an Independent, or what? **Post-treatment:** Same question. **Options:** *"Republican"*, *"Democrat"*, *"Independent"*, *"Some other party"*

Table 1.1 Continued: Survey Questions and Variable Definitions

Conservative Policy Priorities	
Variable	**Question Asked**
Label: Conservative policy priorities (PCA) **Coding:** First component of a principal component analysis, performed on indicator variables for each policy priority which take value 1 if the priority was in the respondent's top 3, and 0 otherwise. Positively correlates with pre-treatment Republican partisanship.	**Post-treatment:** *Of the following, what are the most important issues that you want your government to address? Please rank your top 3 issues below.* **Options:** *"Abortion", "Education", "Crime", "Gun control", "Crisis in Ukraine", "Healthcare", "Immigration", "Inflation", "Don't know"* See Table 1.2 for information on factor loadings.
Label: Conservative policy priorities share **Coding:** Average of the three indicator variables related to *"Inflation"*, *"Immigration"*, and *"Crime"*. These priorities were identified as Conservative by correlation with pre-treatment partisanship.	

Table 1.1 Continued: Survey Questions and Variable Definitions

Attitudes toward the Trump Criminal Investigations	
Variable	**Question Asked**
Label: Investigations are unacceptable (all) **Coding:** This variable takes value 1 if the respondent gave a pro-Trump to all four questions, 0 otherwise.	**Post-treatment:** *"Thinking about the current indictments and investigations into Donald Trump's actions while in office... Do you think these indictments and investigations..."* (1) *"...are or are not a deliberate attempt to stop the Trump campaign?"* (2) *"...do or do not uphold the rule of law?"* (3) *"... protect or undermine democracy?"* (4) *"... are or are not an attack on people like you?"*
Labels: • Investigations are an attempt to stop the campaign • Investigations are against the rule of law • Investigations undermine democracy • Investigations are an attack on people like me **Coding:** Takes value 1 for a pro-Trump answer, 0 otherwise.	

Table 1.1 Continued: Survey Questions and Variable Definitions

Attitudes to the War in Ukraine	
Variable	**Question Asked**
Label: Pro-Kremlin attitudes on Ukraine War (PCA) **Coding:** First component of a principal component analysis performed on the five variables below. Positively correlates with pre-treatment Republican partisanship.	See Table 1.2 for information on factor loadings.
Label: Pro-Putin sentiment **Coding:** Takes value 0 if *"Strongly unfavourable"*, 1 otherwise (i.e., split along the median).	**Post-treatment:** *"Now, we are providing you with a list of names and organisations. Please tell us if you have a favourable or unfavourable opinion of each person or organisation?"* **Options:** *"Strongly unfavourable"*, *"Somewhat unfavourable"*, *"Neutral"*, *"Somewhat favourable"*, *"Very favourable"*, *"Don't know"*
Label: Anti-Zelensky sentiment **Coding:** Takes value 1 if *"Strongly unfavourable"*, *"Somewhat unfavourable"*, or *"Neutral"*, 0 otherwise (i.e., split along the median).	
Label: Republicans handle the crisis better **Coding:** Takes the value 1 if the respondents chose *"Republican"*, and 0 otherwise.	**Post-treatment:** *"Between the Democratic Party and the Republican Party, who in your opinion is best able to handle the crisis in Ukraine?"* **Options:** *"Republican"*, *"Democrat"*, *"Both equally"*, *"Neither"*

Table 1.1 Continued: Survey Questions and Variable Definitions

Attitudes to the War in Ukraine (continued)	
Variable	**Question Asked**
Label: Disapproval of Biden policy in Ukraine **Coding:** Takes value 1 if *"Strongly disapprove"*, *"Somewhat disapprove"*, or *"Neither approve nor disapprove"*, 0 otherwise (i.e., split along the median).	**Post-treatment:** *"Do you approve or disapprove of the job the US government is doing on handling the crisis in Ukraine?"* **Options:** *"Strongly disapprove"*, *"Somewhat disapprove"*, *"Neither approve nor disapprove"*, *"Somewhat approve"*, *"Strongly approve"*, *"Don't know"*
Label: Republican measures for Ukraine **Coding:** Each option has its indicator variable, which takes the value 1 if the respondent selected this option, and 0 otherwise. We run a principal component analysis on all the indicator variables and keep the first component. Positively correlates with pre-treatment Republican partisanship.	**Post-treatment:** *Which actions should the US take or keep taking to help Ukraine? You can select multiple options."* **Options:** *"Sending more US troops to European countries other than Ukraine"*, *"Creating a no-fly zone over Ukraine"*, *"Placing economic sanctions on Russia"*, *"Taking Ukrainian refugees"*, *"Sending military aid to Ukraine, such as weapons and equipment"*, *"Sending US troops to Ukraine"*, *"None of these"*, *"Don't know"*

Table 1.1 Continued: Survey Questions and Variable Definitions

Well-being	
Variable	**Question Asked**
Label: Low well-being (PCA) **Coding:** First component of a principal component analysis performed on life dissatisfaction and unhappiness. Positively correlates with pre-treatment life dissatisfaction and unhappiness.	See Table 1.2 for information on factor loadings.
Label: Life dissatisfaction **Coding:** We inverted the values for our analyses for easier visual interpretation. Takes values between 1 (best possible life) and 11 (worst possible life).	**Post-treatment:** *"Please imagine a ladder, with steps numbered from 0 at the bottom to 10 at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time?"* **Options:** *0* to *10* , or *"Not sure"*
Label: Unhappiness **Coding:** We inverted the values for our analyses for easier visual interpretation. Takes values between 1 (*"Very happy"*) and 4 (*"Not at all happy"*).	**Post-treatment:** *"Taking all things together, would you say you are..."*. **Options:** *"Very happy"*, *"Rather happy"*, *"Not very happy"*, *"Not at all happy"*, *"Don't know"*

Table 1.1 Continued: Survey Questions and Variable Definitions

Other Outcomes	
Variable	**Question Asked**
Label: Negative **X** experience **Coding:** Takes values from 1 (*"Very positive"*) to 5 (*"Very negative"*).	**Post-treatment:** *Recently, have you found your time spent on Twitter/X a positive or a negative experience?* **Options:** *"Very positive"*, *"Somewhat positive"*, *"Neutral"*, *"Somewhat negative"*, *"Very negative"*
Label: Anti-NATO sentiment **Coding:** Takes value 1 if *"Strongly unfavourable"*, *"Somewhat unfavourable"*, or *"Neutral"*, 0 otherwise (i.e., split along the median).	**Post-treatment:** *"Now, we are providing you with a list of names and organisations. Please tell us if you have a favourable or unfavourable opinion of each person or organisation?"* **Options:** *"Strongly unfavourable"*, *"Somewhat unfavourable"*, *"Neutral"*, *"Somewhat favourable"*, *"Very favourable"*, *"Don't know"*
Label: Importance of the war in Ukraine **Coding:** Takes value 0 if *"Not important at all"*, *"One of the least important topics"*, *"A moderately important topic"*, 1 otherwise (i.e., split along the median).	**Post-treatment:** *"Regarding issues that you want the President and US Congress to address, how important do you consider the crisis in Ukraine?"* **Options:** *"Not important at all"*, *"One of the least important topics"*, *"A moderately important topic"*, *"One of the most important topics"*, *"The most important topic"*, *"Don't know"*
Label: Did not solve a puzzle for Ukraine **Coding:** The correct answer was 9. Many respondents answered 8, 9, or 10. If the answer given was between 8 and 10, we assume the respondents were indeed trying to correctly answer the puzzle. Takes value 1 if the respondent answered something other than 8,9 or 10; 0 otherwise.	**Post-treatment:** *"This question is optional. Now, you can solve a math puzzle to make a donation to the Red Cross of Ukraine. That is, if you correctly solve the puzzle, we will donate 1 USD to the Red Cross of Ukraine. Whether or not you solve the puzzle will not affect your own remuneration. If you want to solve it, please write the solution in the field. Otherwise, just leave the field blank. How many odd numbers are higher than 12 but smaller than 30?"*

Table 1.1 Continued: Survey Questions and Variable Definitions

Variable	Factor loading	Explained Variance (%)
User engagement (PCA)		60.72
Use 𝕏 no less than before	0.707	
Post on 𝕏 no less than before	0.707	
Partisanship and affective polarisation (PCA)		91.00
Affective polarisation	0.707	
Republican partisanship	0.707	
Conservative policy priorities (PCA)		28.53
Crime	0.436	
Immigration	0.442	
Inflation	0.423	
Crisis in Ukraine	-0.050	
Don't know	-0.004	
Education	-0.179	
Abortion	-0.324	
Healthcare	-0.354	
Gun control	-0.414	
Pro-Kremlin attitudes on Ukraine War (PCA)		54.10
Anti-Zelensky sentiment	0.515	
Disapproval of Biden policy in Ukraine	0.503	
Republican measures for Ukraine	0.500	
Pro-Putin sentiment	0.351	
Republicans handle the crisis better	0.329	
All attitudes on policies and current news (PCA)		73.59
Conservative policy priorities (PCA)	0.583	
Pro-Kremlin attitudes on Ukraine War (PCA)	0.574	
Trump investigations are unacceptable (all)	0.575	
Low well-being (PCA)		83.92
Life dissatisfaction	0.707	
Unhappiness	0.707	

Table 1.2. Factor Loadings of PCA Composite Outcomes

1.3.1 Handling of Missing Data for Covariates and Outcomes

Respondents who answered "Don't know" or "Not sure" to a question are coded as missing values for covariates. For any covariates with missing values, an additional variable indicating missing data is included in the regression model as a covariate, and missing values are recoded to an arbitrary constant that does not appear among the response values, such as -99. If the outcome variable is missing, the observation is not included in the analysis.

1.4 Summary Statistics

In this section, we present summary statistics.

Table 1.3 presents the means and standard deviations of pre-treatment characteristics for respondents in our main sample, alongside those of a representative sample of U.S. Twitter/𝕏 users, based on data from the American National Election Studies (ANES) Social Media Study. This study surveyed a nationally representative group of Americans about their political views and social media use between 2020 and 2022. At the time, 𝕏 was known as Twitter, and respondents were asked directly: "Do you currently use Twitter?" Of the 5,750 respondents, 935 answered "Yes." We treat these individuals as representative of the broader U.S. 𝕏 user population.[2]

Compared to the ANES sample of Twitter/𝕏 users, our sample includes a higher proportion of males (52% vs. 57%, $p < 0.001$), a larger share of well-educated individuals (58% vs. 49%, $p < 0.001$) and white respondents (78% vs. 66%, $p < 0.001$), and is older on average (50 vs. 43 years, $p < 0.001$). Additionally, our sample contains more Independents (32% vs. 21%, $p < 0.001$) and fewer Republicans (21% vs. 31%, $p < 0.001$). To address these differences, some specifications apply regression weights to align our sample with the ANES population. This adjustment does not alter our main findings (see SI Section 2.2.2 for details on the weighting procedure and Fig. 2.10 for the weighted results).

[2]We focus on five variables that are common to both the ANES survey and our pre-treatment survey: gender, race, age, education, and partisanship. Because the coding of these variables differs between surveys, we manually recode them to enable valid comparisons. For partisanship, we rely on the following ANES question: "When it comes to politics, would you describe yourself as. . . " with response options including "Very liberal," "Somewhat liberal," "Closer to liberals," "Neither liberal nor conservative," "Closer to conservatives," "Somewhat conservative," and "Very conservative." We classify respondents who chose any of the first three options as Democrats, those who selected "Neither liberal nor conservative" as Independents, and those who chose one of the last three options as Republicans. In our own survey, respondents who identify as "Republican" or "Democrat" are retained as such, while those who select "Independent" or "Some other party" are coded as Independents.

Fig. 1.3 displays patterns of 𝕏 usage and posting frequency among respondents in our sample. By design, the sample includes active users who browse 𝕏 at least several times a month. Within this group, platform engagement is high: 65.8% of respondents report using 𝕏 at least once per day, including 44.3% who use it several times a day and 21.5% who use it once daily. An additional 20.2% use the platform several times a week, while only 6.5% report using it just several times a month. Despite frequent browsing, posting behaviour is more heterogeneous. A notable share of respondents are relatively inactive in terms of content creation: 13.1% report never posting, and another 23.1% post only once a month or less. At the same time, a sizeable minority are frequent posters, namely, 16.7% post several times a day, and another 10.0% post once daily. Between these extremes, 18.5% post several times a week, 8.2% once a week, and 10.5% several times a month.

Table 1.4 presents the summary statistics for the outcome variables collected in the post-treatment survey, separately for the four groups of respondents based on the combination of their initial feed settings (algorithmic or chronological) and their randomly assigned treatment feed settings (also, algorithmic or chronological). The table reveals patterns consistent with our main findings.[3]

We now describe outcomes related to the accounts users follow. These data are available for 2,387 of the 2,600 participants who had provided their 𝕏 handle. For the remaining 212 participants, the handle was either invalid or the account inactive; therefore, we were unable to collect data on the accounts they followed (see Extended Data Fig. 1). Table 1.5 presents means and standard deviations for pre-treatment characteristics, comparing participants for whom we were able to collect data on followed accounts to those for whom we were not. The two groups are similar across most observed dimensions. The share of male participants is the same (52%). There are modest differences in the shares with a college degree (57% vs. 59%, $p = 0.351$) and identifying as white (79% vs. 77%, $p = 0.022$). Participants with data on followed accounts are also older on average (52 vs. 49 years, $p < 0.001$). Patterns of 𝕏 use and self-reported well-being are also comparable across groups. There are differences in hard news consumption (71% vs. 67%, $p = 0.003$) and affective polarisation scores (39 vs. 42, $p < 0.001$). To ensure our results on the feeds are not specific to those for whom we could collect data on followed accounts, we show robustness to reweighting this selected sample to resemble the overall sample (see SI Section 2.2.2 for technical

[3] See also Extended Data Table 1, which presents pre-treatment variables by initial feed setting and is discussed in the main text.

details on how we construct the weights). This adjustment does not alter our main findings (see Fig. 2.14). Table 1.6 presents summary statistics of the accounts users follow post-treatment, by initial feed setting and treatment assignment. Again, the table reveals patterns consistent with our main findings about the asymmetric effects of algorithmic exposure.

We now turn to the content shown to users on their feed. The feed data come from participants who completed the survey on a laptop or desktop computer and chose to run the Google Chrome extension during the pre-treatment survey, the post-treatment survey, or both (784 participants in the first wave and 599 in the second). These data yield 928 unique participants for whom at least one chronological feed sample was collected, and 922 for whom at least one algorithmic feed sample was available. The number of observations for chronological and algorithmic feeds is not identical because, for a small number of users, the Chrome extension did not run to completion due to network interruptions. Table 1.7 presents means and standard deviations for pre-treatment variables by Chrome extension use. The demographic characteristics of those who ran the extension differ from those who opted out. Participants who ran the extension were older (61 vs. 49 years, $p = 0.001$) and more likely to be male (55% vs. 51%, $p = 0.151$). Racial composition and education levels were similar across groups, with 78% identifying as white and 58% holding a four-year college degree or higher. In terms of $\mathbb{X}$ usage patterns, extension users reported lower soft news consumption (55% vs. 62%, $p = 0.163$) and higher hard news consumption (73% vs. 69%, $p < 0.001$). They also reported lower levels of life dissatisfaction (4.28 vs. 4.63, $p = 0.263$), while average unhappiness did not differ across groups. Politically, extension users were somewhat less likely to identify as Democrats (43% vs. 47%, $p = 0.513$). To ensure that our results on feeds are not specific to those who opted to run the Chrome extension, we reweight the sample with available feed data to resemble the overall sample (see SI Section 2.2.2 for technical details on how we construct the weights). This adjustment does not alter our main findings (see Table 2.13).

Finally, Table 1.8 presents summary statistics for the characteristics of the content shown to users in the chronological and algorithmic feeds. We discuss this comparison in the main text and present it graphically in main text Fig. 3.

		Sample	
		Main Sample	**ANES Twitter Users**
Variable		**Mean (Standard Deviation)**	
General Information			
Gender	Share of Males	0.52 (0.50)	0.57 (0.50)
Education	Share of 4-year college or above	0.58 (0.49)	0.49 (0.50)
Race	Share of Whites	0.78 (0.42)	0.66 (0.48)
Age		50.42 (14.90)	43.47 (14.94)
Political Attitudes			
Partisanship	Share of Democrats	0.46 (0.50)	0.48 (0.50)
	Share of Independents	0.32 (0.47)	0.21 (0.41)
	Share of Republicans	0.21 (0.41)	0.31 (0.46)
Number of Observations (Users)		4965	933

Table 1.3. Summary Statistics Comparing Main Sample and ANES Twitter Users.

The means and standard deviations (in parentheses) or proportions for selected demographic and political variables, comparing respondents in the Main Sample and the ANES Twitter Users.

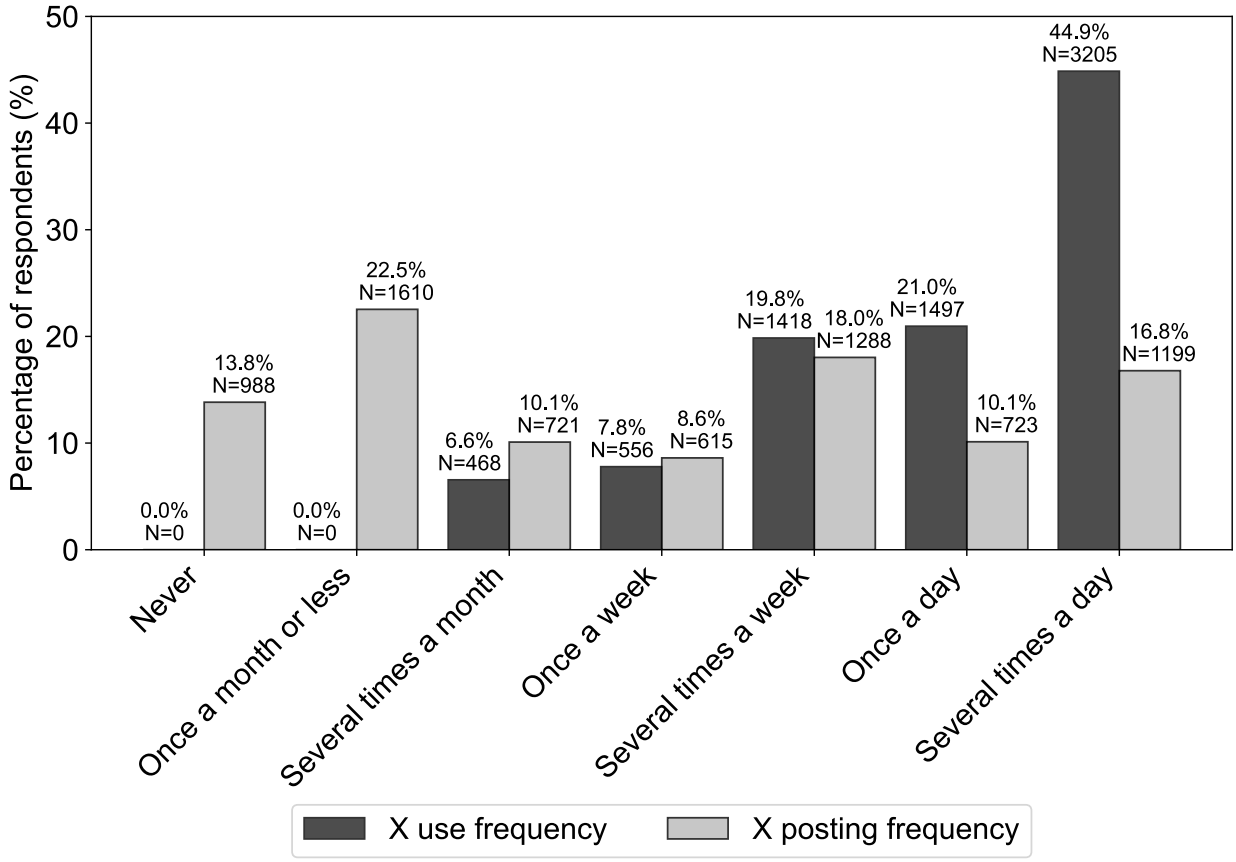


Fig. 1.3. 𝕏 use and posting frequency distributions among our respondents

	Initial Feed to Treatment-Assignment Feed			
	Chronological to Chronological	Chronological to Algorithmic	Algorithmic to Chronological	Algorithmic to Algorithmic
Variable	**Mean (Standard Deviation)**			
User Engagement				
User engagement (PCA)	69.98 (35.42)	74.82 (33.14)	71.88 (34.25)	73.69 (33.72)
Use 𝕏 no less than before	0.77 (0.42)	0.82 (0.38)	0.79 (0.41)	0.81 (0.40)
Post on 𝕏 no less than before	0.62 (0.49)	0.66 (0.47)	0.63 (0.48)	0.66 (0.48)
Affective polarisation and partisanship				
Partisanship and affective polarisation (PCA)	30.46 (34.12)	32.89 (35.39)	35.00 (35.08)	36.51 (36.06)
Affective polarisation	33.13 (29.90)	36.08 (31.88)	37.74 (30.55)	38.75 (31.48)
Self-declared Republican	0.19 (0.39)	0.19 (0.39)	0.22 (0.41)	0.23 (0.42)
Conservative Policy Priorities				
Conservative policies (PCA)	41.92 (33.27)	46.61 (34.34)	48.14 (31.61)	47.27 (32.11)
Shares Republican priorities	0.38 (0.35)	0.42 (0.36)	0.44 (0.33)	0.43 (0.34)
Attitudes toward the Trump Criminal Investigations				
Investigations are unacceptable (all)	0.20 (0.40)	0.25 (0.43)	0.22 (0.42)	0.23 (0.42)
Investigations are against the rule of law	0.27 (0.45)	0.33 (0.47)	0.32 (0.47)	0.32 (0.47)
Investigations undermine democracy	0.30 (0.46)	0.35 (0.48)	0.33 (0.47)	0.35 (0.48)
Investigations are an attack on people like me	0.28 (0.45)	0.33 (0.47)	0.34 (0.47)	0.35 (0.48)
Investigations are an attempt to stop the campaign	0.36 (0.48)	0.40 (0.49)	0.42 (0.49)	0.44 (0.50)
Attitudes to the War in Ukraine				
Pro-Kremlin attitudes on Ukraine War (PCA)	34.31 (29.49)	39.32 (31.87)	40.20 (30.72)	40.79 (30.68)
Anti-Zelensky sentiment	0.33 (0.47)	0.41 (0.49)	0.40 (0.49)	0.41 (0.49)
Pro-Putin sentiment	0.20 (0.40)	0.25 (0.43)	0.29 (0.45)	0.28 (0.45)
Republicans handle the Ukraine crisis better	0.14 (0.35)	0.19 (0.39)	0.21 (0.41)	0.22 (0.41)
Disapproval of Biden policy in Ukraine	0.41 (0.49)	0.47 (0.50)	0.48 (0.50)	0.48 (0.50)
Republican measures for Ukraine	49.83 (28.87)	52.78 (31.24)	53.53 (30.43)	54.10 (29.99)
Well-being				
Low well-being (PCA)	34.49 (19.79)	36.47 (21.15)	35.49 (20.78)	34.92 (20.49)
Life dissatisfaction	4.44 (2.05)	4.68 (2.15)	4.66 (2.15)	4.59 (2.12)
Unhappiness	2.05 (0.68)	2.09 (0.73)	2.03 (0.72)	2.02 (0.72)
Other outcomes				
Negative 𝕏 experience	2.77 (1.07)	2.84 (1.10)	2.62 (1.04)	2.57 (1.04)
Anti-NATO sentiment	0.38 (0.49)	0.40 (0.49)	0.41 (0.49)	0.42 (0.49)
Importance of the war in Ukraine	0.40 (0.49)	0.41 (0.49)	0.40 (0.49)	0.41 (0.49)
Did not solve a puzzle for Ukraine	0.14 (0.35)	0.14 (0.34)	0.13 (0.34)	0.15 (0.36)
Number of Observations (Users)	626	580	1841	1918

Table 1.4. Summary Statistics for Post-Treatment Survey-Based Outcomes.

The means and standard deviations (in parentheses) of survey-based outcomes collected in the post-treatment survey by the combination of the initial feed settings and the treatment assignment.

		List of Accounts the User Follows	
Variable	**Category**	**Collected**	**Could Not Collect**
General Information			
Gender	Share of Males	0.52 (0.50)	0.52 (0.50)
Race	Share of Whites	0.79 (0.41)	0.77 (0.42)
Education	Share 4-year college or more	0.57 (0.49)	0.59 (0.49)
Age		51.51 (14.23)	49.44 (15.50)
𝕏 **Use**			
Time on 𝕏		4.93 (1.24)	4.87 (1.22)
𝕏 posting frequency		2.98 (2.03)	2.88 (2.11)
News consumption	Soft news consumption	0.62 (0.48)	0.61 (0.49)
	Hard news consumption	0.71 (0.45)	0.67 (0.47)
Well-being			
Life dissatisfaction		4.67 (2.14)	4.50 (2.08)
Unhappiness		2.06 (0.72)	2.02 (0.72)
Political Attitudes			
Affective polarisation		35.51 (31.53)	39.57 (31.34)
Partisanship	Share of Democrats	0.46 (0.50)	0.47 (0.50)
	Share of Independents	0.31 (0.46)	0.27 (0.45)
	Share of Republicans	0.20 (0.40)	0.23 (0.42)
	Share of Some other party	0.03 (0.18)	0.03 (0.17)

Table 1.5. Summary Statistics by Whether We Were Able to Collect Data on Followed Accounts.

The means and standard deviations (in parentheses) for pre-treatment variables, split by whether data for the accounts that participants follow could be successfully collected.

	Initial Feed to Treatment-Assignment Feed			
	Chronological to Chronological	Chronological to Algorithmic	Algorithmic to Chronological	Algorithmic to Algorithmic
Variable	**Mean (Standard Deviation)**			
Followed Accounts by Account Type				
Political activist accounts	0.12 (0.13)	0.14 (0.16)	0.12 (0.16)	0.13 (0.16)
News outlet accounts	0.11 (0.09)	0.11 (0.11)	0.09 (0.10)	0.09 (0.10)
Entertainment accounts	0.45 (0.23)	0.45 (0.25)	0.48 (0.24)	0.47 (0.24)
Other accounts	0.22 (0.15)	0.21 (0.14)	0.22 (0.16)	0.22 (0.16)
Followed Accounts by Political Content				
Conservative accounts	0.08 (0.17)	0.12 (0.22)	0.10 (0.19)	0.10 (0.19)
Liberal accounts	0.24 (0.22)	0.22 (0.22)	0.18 (0.21)	0.20 (0.22)
Conservative political activist accounts	0.03 (0.08)	0.06 (0.12)	0.05 (0.10)	0.05 (0.11)
Liberal political activist accounts	0.08 (0.11)	0.08 (0.13)	0.07 (0.13)	0.07 (0.13)
Conservative news outlet accounts	0.01 (0.03)	0.01 (0.04)	0.01 (0.03)	0.01 (0.03)
Liberal news outlet accounts	0.03 (0.04)	0.03 (0.05)	0.02 (0.04)	0.02 (0.04)
Number of Observations (Users)	286	270	907	924

Table 1.6. Summary Statistics for the Accounts the User Follows.

The means and standard deviations (in parentheses) of the accounts the user follows by the initial feed setting and the treatment assignment.

		Chrome Extension Use	
Variable	**Category**	**Ran the Extension**	**Did Not Run the Extension**
General Information			
Gender	Share of Males	0.60 (0.49)	0.50 (0.50)
Race	Share of Whites	0.79 (0.41)	0.78 (0.42)
Education	Share 4-year college or more	0.62 (0.49)	0.57 (0.49)
Age		49.88 (14.73)	50.55 (14.98)
𝕏 **Use**			
Time on 𝕏		4.93 (1.21)	4.89 (1.24)
𝕏 posting frequency		2.67 (2.04)	2.98 (2.07)
News consumption	Soft news consumption	0.64 (0.48)	0.61 (0.49)
	Hard news consumption	0.72 (0.45)	0.69 (0.46)
Well-being			
Life dissatisfaction		4.83 (2.11)	4.54 (2.11)
Unhappiness		2.12 (0.69)	2.03 (0.72)
Political Attitudes			
Affective polarisation		33.95 (30.65)	38.32 (31.60)
Partisanship	Share of Democrats	0.44 (0.50)	0.47 (0.50)
	Share of Independents	0.34 (0.47)	0.28 (0.45)
	Share of Republicans	0.18 (0.38)	0.22 (0.42)
	Share of Some other party	0.04 (0.20)	0.03 (0.17)
Number of Observations (Users)		803	4162

Table 1.7. Summary Statistics by Chrome Extension Use

The means and standard deviations (in parentheses) for pre-treatment variables, broken down by whether the respondent ran our Chrome extension or not.

	Content across Feed Settings	
	Chronological	Algorithm
Variable	**Mean (Standard Deviation)**	
Engagement for Posts Shown		
Number of likes (in thousands)	2.78 (22.23)	13.21 (41.02)
Number of reposts (in thousands)	0.45 (3.76)	1.84 (5.44)
Number of comments (in thousands)	0.15 (1.81)	0.72 (3.40)
Total engagement (in thousands)	3.37 (26.27)	15.72 (47.27)
Post Shown by Account Type		
Political activist accounts	0.22 (0.41)	0.27 (0.45)
News accounts	0.27 (0.44)	0.11 (0.32)
Entertainment accounts	0.42 (0.49)	0.52 (0.50)
Other accounts	0.05 (0.21)	0.05 (0.21)
Posts Shown by Political Content		
Conservative content	0.15 (0.35)	0.17 (0.38)
Liberal content	0.32 (0.47)	0.33 (0.47)
Conservative political activist accounts	0.08 (0.28)	0.11 (0.31)
Conservative news accounts	0.03 (0.18)	0.01 (0.11)
Liberal political activist accounts	0.13 (0.34)	0.16 (0.37)
Liberal news accounts	0.09 (0.28)	0.04 (0.19)
Number of Unique Users	928	922
Number of Observations	134706	133826

Table 1.8. Summary Statistics for Content Shown in User Feeds.

The means and standard deviations (in parentheses) of content characteristics shown in the same users' chronological and algorithmic feeds collected by the users with our Google Chrome extension.

1.5 Annotations of User Feed Data

In this section, we explain how we annotate the posts that users see in their feeds. We begin in SI Section 1.5.1 by describing our baseline annotations, which are generated using the open-source large language model Llama 3.3 70B. To ensure robustness, we also develop simpler content proxies based on word frequencies, presented in SI Sections 1.5.2 and 1.5.3. We then validate these annotations by assessing model agreement, their correlation with users' pre-treatment partisanship, and comparisons with human annotations, as detailed in SI Section 1.5.4.

1.5.1 Baseline: Llama 3 Annotations

In our baseline results, we use the open-source large language model Llama 3.3 70B Instruct (henceforth Llama 3) to classify the feed data collected through a Google Chrome extension by political leaning (conservative or liberal) and by type (posts from political activists, entertainment accounts, or news media outlets). For consistency and replicability, the model's temperature is set to zero. We create content measures at the account level rather than the post level, as post-level annotations are often noisier. Supplying the large language model with multiple posts from the same account provides added context, improving annotation accuracy.

In particular, we classify the accounts that appear in the participants' feeds along these two dimensions:

- **Political Leaning**

  The prompt that we used is as follows:

  I will show you the name, description, and a sample of tweets from a Twitter account.

  Your task is to classify the account's political leaning based on this information.

  Choose only one of the following labels:

  - Conservative

  - Liberal

  - Cannot say

  Do not explain your choice or add anything else.

  Account name: [...]

  Description: [...]

Sample tweets: [...]

Label:

We present Llama 3 with one hundred randomly selected posts from the account (or all posts if fewer than one hundred were available).

- **Type of Content Shared or Produced**

  The prompt is as follows:

  You will be shown the name, description, and a sample of posts from a Twitter account.

  Classify the account into one of the following five categories:

  - News: Official news outlets or accounts primarily focused on reporting current events.

  - Political activist: Accounts primarily focused on partisan political commentary or advocacy.

  - Entertainment: Accounts centered on content like memes, jokes, pop culture, art, books, music, movies, video games, sports, cooking, science, business promotions, animals, lifestyle, motivation, and personal development.

  - Official: Accounts belonging to elected officials or government representatives.

  - Other: Accounts that don't clearly fit any of the above.

  Respond with only one of these five labels (News, Politics, Entertainment, Official, Other). Do not include any explanation.

  Account name: [...]

  Description: [...]

  Sample tweets: [...]

  Label:EV

  We present Llama 3 with up to one hundred randomly selected posts from the account (or all posts if fewer than one hundred were available).

SI Fig. 1.4 displays word clouds of the terms most strongly associated with each annotation class, based on coefficients from a penalized multinomial logistic regression model.

Posts from conservative-leaning accounts commonly reference terms like "wokeness" and "libs," and frequently share content from sources such as "foxnews.com," "conservativebrief.com,"

and “breitbart.com.” Prominent conservative figures also appear frequently, including Jack Posobiec (“poso”), Phillip Buchanan (“catturd”), and Dan Bongino (“bongino”). In contrast, liberal-leaning accounts often emphasize words like “democracy” and “resistance,” and use hashtags such as “#tuckfrump” and “#demvoice1.” These accounts typically link to outlets like *The New York Times*, *The Washington Post*, and *The Guardian*, and feature liberal commentators such as the Krassenstein brothers (“krassenstein”), Ron Filipkowski (“filipkowski”), and George Takei (“takei”).

Accounts categorized as “News” tend to reference mainstream media outlets, including *CNN*, *Fox News*, the *BBC*, *Reuters*, and *The Economist*. Those labeled “Political Activism” frequently employ strongly political terms like “maga” and “libertarian.” Meanwhile, “Entertainment” accounts are characterized by mentions of gaming and streaming (e.g., “gamer,” “geek,” “stream”), soft news platforms (e.g., *The Onion*), online shopping (e.g., “giveaway,” “amazon.com”), and sports-related terms (e.g., “football,” “league”). Lastly, “Official” accounts frequently refer to elected officials such as senators and governors, as well as corporate entities like SpaceX.

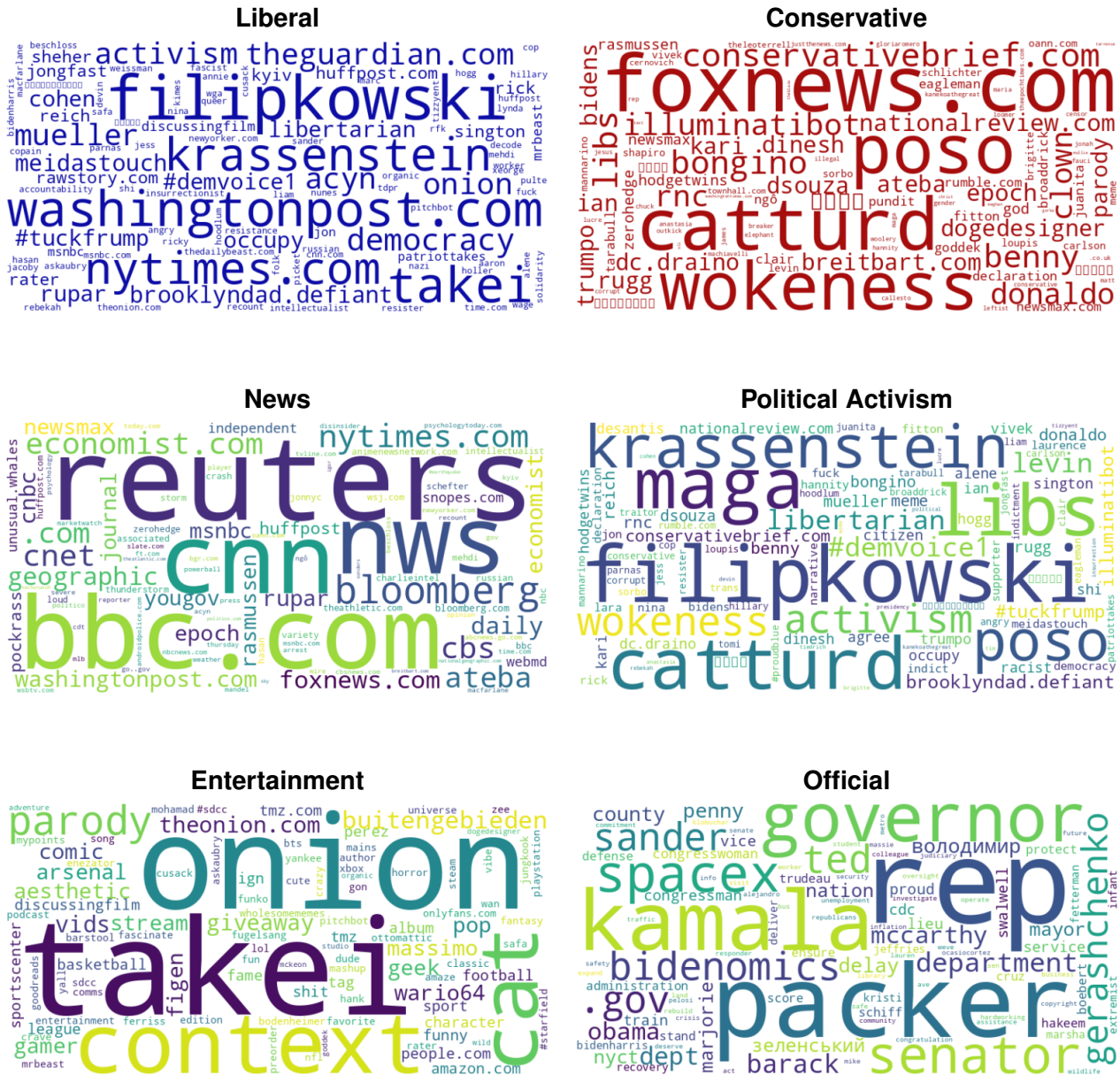


Fig. 1.4. Clouds of the Words Most Associated with Annotations.

The highest-scoring terms for each class, scaled by their corresponding model coefficients. A penalized multinomial logistic regression model with an L2 penalty is trained to predict Llama 3 annotations at the post level using word frequencies.

1.5.2 Robustness of Content Annotations: ML Classifier for Political Leaning

We built a machine-learning (ML) classifier to predict the author of a post's party affiliation based on word frequencies. We train our model on the universe of posts by congressmen on 𝕏 between January 2023 and July 2023.[4] The dataset contains 353,742 messages from 968 unique accounts officially affiliated with the Republican or the Democratic Party (we remove Independents and accounts for which party affiliation is missing). We lowercase the text and retain adjectives, nouns, proper nouns, and verbs. We remove the pre-defined list of English stopwords provided by the scikit-learn Python library, as well as the words "thank," "follow," "hour," "minute," "debate," and "begin" which are ubiquitous in our data and uninformative. To prevent geographical identifiers from driving our predictions, we also remove mentions of US states, counties, and cities. These may appear in the forms of hashtags (e.g., #ny21), nouns (e.g., New York), and adjectives (e.g., new yorker). We then split the text into unigrams and bigrams and keep those that appear in more than 0.01% and less than 1% of all posts (the latter threshold drops 287 very frequent unigrams and bigrams). We also keep hashtags, (cleaned) URLs, and account mentions—as they are a large part of the way users communicate on 𝕏. The final vocabulary contains 24,170 distinct phrases.

We then specify a penalized multinomial logistic regression with an L2-penalty. We use 80% of the data to train the model and keep 20% as a test set. We train the model via five-fold cross-validation. Table 1.9 shows the model's performance metrics on the test set. It has an F1-Score of 0.87, and precision and recall are balanced across party labels. Furthermore, since the model relies on word frequencies for prediction, its inner workings are easily interpretable. Fig. 1.5 shows the top predictive hashtags and phrases for Republican and Democrat accounts. Typical Democrat hashtags include *#inflationreductionact* and *#medicareforall*. *"climate crisis"* and *"gun violence"* are among the most predictive bigrams for Democrats. Typical Republican hashtags include *#bidenbordercrisis* and *#bidenflation*. *"illegal immigrant"* and *"reckless spending"* are prominent Republican bigrams.

The model performs well, so we then use it to predict the partisanship of the posts in our feed data, following the same text processing steps. For each post, we obtain an estimated probability of being written by a "Republican." If this probability is above 50%, we label the post as "Conservative." If it is below 50%, we label it as "Liberal." In some rare cases, the text processing may remove all words from the post. We are then unable to predict the post's political leaning and label

[4] For further details, see the Congress Tweets database: `https://alexlitel.github.io/congresstweets/`, accessed June 17, 2025.

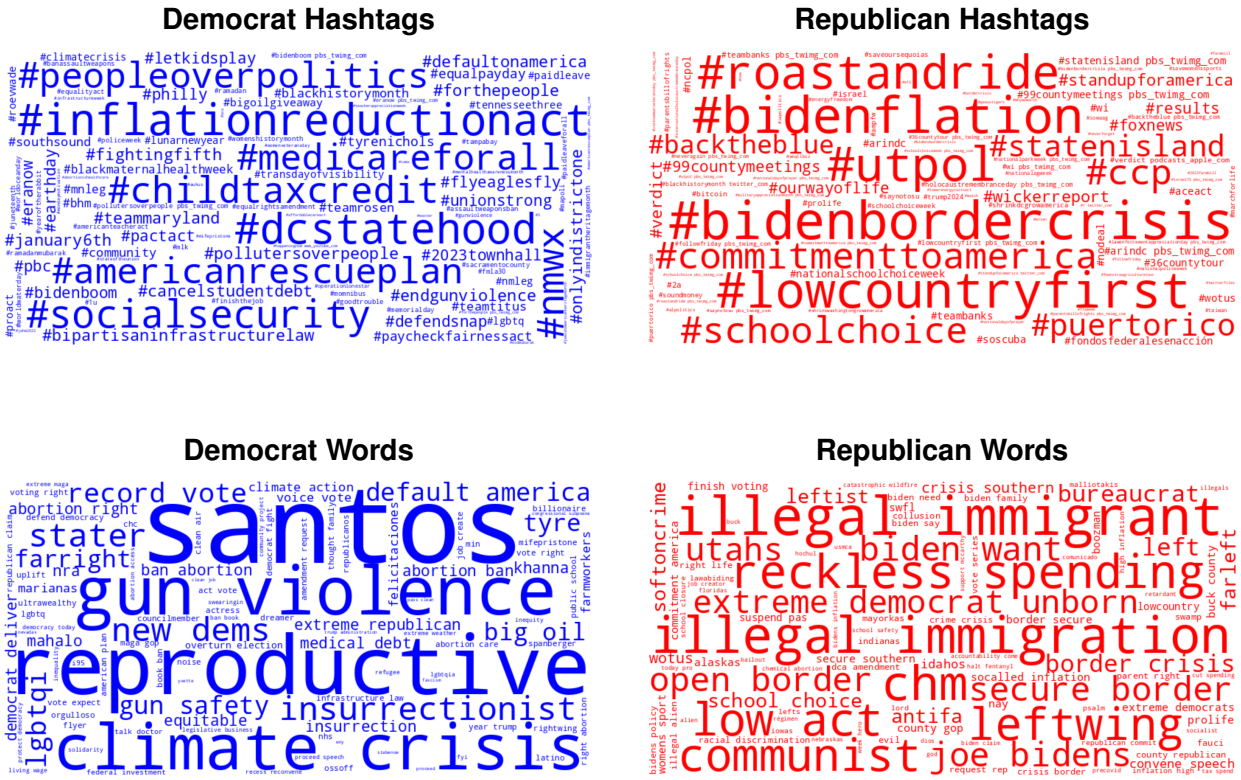


Fig. 1.5. Clouds of the Hashtags and Phrases Most Predictive of Party Affiliation.

The highest-scoring hashtags and phrases for each class, scaled by their associated coefficients. A penalized multinomial logistic regression model (L2 penalty) is trained to predict the party affiliation of the author of a given post.

it as “Cannot say.”

	Precision	Recall	F1-Score	Support
Democrat	0.88	0.86	0.87	35,364
Republican	0.87	0.88	0.87	35,385
Accuracy			0.87	70,749

Table 1.9. The ML-classifier Performance by Political Affiliation

1.5.3 Robustness of Content Annotations: Topic Model for the Type of Content Shared or Produced

We train a topic model to infer the topics in the posts based on word frequencies. We first process the text similarly to the ML Classifier (as described in SI Section 1.5.2). That is, we lowercase the text and retain adjectives, nouns, proper nouns, and verbs. We remove the pre-defined list of English stopwords provided by the scikit-learn Python library, as well as the words “thank”, “follow”, “hour”, “minute”, “debate”, and “begin” which are ubiquitous in our data and uninformative. We also remove all mentions of US states, counties, and cities to prevent geographical identifiers from driving our predictions. These may appear in the forms of hashtags (e.g., #ny21), nouns (e.g., new york), and adjectives (e.g., new yorker). We then split the text into unigrams and keep those that appear in more than 0.01% and less than 1% of all posts (the latter threshold drops 148 very frequent unigrams). We also keep hashtags, (cleaned) URLs, and account mentions—as they are a large part of the way users communicate on 𝕏. The final vocabulary contains 12,853 distinct unigrams.

We then rely on the Latent Dirichlet Allocation model (47). LDA models treat documents as distributions over topics and topics as distributions over words. As for most topic models, the main hyperparameter is the number of topics that the researcher needs to specify. We specify two topics, as we are primarily interested in distinguishing between political content and entertainment, and expect these two dimensions of the text to be the most prominent. We train the model for 500 iterations of the Gibbs sampling algorithm. Fig. 1.6 shows that the negative likelihood flattens from iteration 50 onward and shows no sign of improvement by iteration 400. Table 1.10 presents the top twenty words associated with each topic. As expected, the first topic refers to “Politics”, while the second topic corresponds to “Entertainment”.

We classify the content of a post in the following way. To label posts as “News”, we rely on a comprehensive database of US-based media outlets, their associated Twitter handles, and URL domains (48). We label a post as “News” if it satisfies at least one of these three conditions: (i) it was produced by a news media account, (ii) it mentions a news media account, or (iii) it links to a newspaper URL domain. We label posts as “Official” if they were posted by a politician appearing in the Congress Tweets database. For the remainder of the posts, we label them as “Politics” or “Entertainment” if their topic share in that category exceeds 50%. In some rare cases, the text processing may remove all words from the post. We are then unable to compute the post’s topic

share and label it as “Other”.

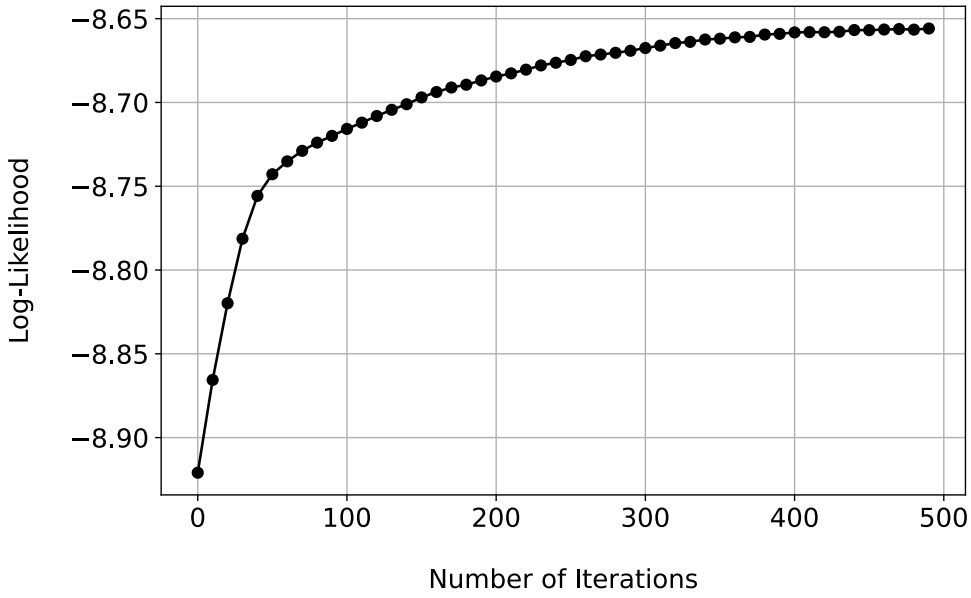


Fig. 1.6. Log-Likelihood of the LDA Model.

Log-likelihood per word of the LDA model over iterations of the Gibbs sampling algorithm.

Topic	Top Words
Politics	’desantis’, ’charge’, ’federal’, ’attack’, ’russian’, ’rep’, ’kill’, ’judge’, ’leader’, ’health’, ’climate’, ’county’, ’military’, ’student’, ’political’, ’bidens’, ’lie’, ’care’, ’arrest’, ’cnn’
Entertainment	’sport’, ’cat’, ’football’, ’film’, ’summer’, ’player’, ’favorite’, ’update’, ’weekend’, ’dog’, ’buy’, ’drop’, ’tonight’, ’barbie’, ’sale’, ’music’, ’hot’, ’food’, ’fun’, ’song’

Table 1.10. LDA Topics for the Posts in the Feed Data.

The two estimated LDA topics based on the text of the posts users saw in their feeds. The first column corresponds to the label. The second column presents the most representative words for each topic.

1.5.4 Validation of Content Annotations

We evaluate the quality of our annotations using different approaches. First, we demonstrate that Llama 3's annotations align closely with alternative indicators of ideological slant and content type. In Fig. 1.7, we present the average share of posts labelled as "Conservative" by two alternative measures: (i) the ML classifier introduced in SI Section 1.5.2 and (ii) a URL-based measure of ideological slant based on (49) separately for accounts annotated by Llama 3 as "Conservative", "Cannot say" and "Liberal".[5] We find that accounts Llama 3 labelled as "Conservative" tend to post a high proportion of conservative content according to both proxies, while those labelled "Liberal" post substantially less conservative content. Turning to content types, Fig. 1.8 shows that accounts labelled "Entertainment" by Llama 3 post very little political content according to our topic model, whereas accounts labelled as "Political Activist" by Llama 3 post frequently about politics.

Second, Fig. 1.9 illustrates that all three conservative content measures are strongly correlated with users' pre-treatment partisanship: self-identified Republicans are typically exposed to far more conservative content than Independents or Democrats in their feeds. Note that Llama 3 annotations more clearly distinguish between partisan groups than the alternative proxies. This is in part due to the model categorizing accounts unrelated to politics as "Cannot say", while other methods necessarily assign a political label to all posts or accounts. For this reason, Llama 3 annotations are our preferred approach.

Third, we assess the quality of our Llama 3 annotations with four US-based human annotators. The human annotators were presented with the exact same information, and each annotator labelled 500 accounts that appeared in the feed content. Annotations were blind (the LLM labels were not shown). Each of the 500 observations were drawn at random from our database.[6] All human annotators saw the same 500 samples so that we could ensure the annotation quality through the inter-annotator agreement. Interannotator agreement is high for political leanings with a Krippendorff's Alpha of 0.69. For content types, agreement is lower at 0.49, suggesting this task is more challenging and ambiguous. In practice, as detailed below, most disagreements are confined to the "Entertainment" and "Other" categories, which are not central to this study and are

[5] While URL-based measures offer a transparent and widely used approach to inferring ideology, they are limited to accounts that share at least one URL.

[6] When sampling the 500 observations, we imposed balance across the partisan leaning (Liberal, Conservative, Cannot Say) and account types as annotated by Llama 3 8B Instruct, a smaller and faster large language model than used for the baseline annotations.

often confused with each other. The exact instructions given to the human annotators are listed in SI Section 4.1.

To determine the political leaning of an account, we first aggregate four individual human annotations by converting their categorical labels into numerical values: “Conservative” is assigned +1, “Liberal” is assigned -1, and “Cannot say” is assigned 0. We then calculate the average score for each account across the four annotations. This average is converted back into a final “gold standard” label using the following rule: if the average is greater than 0, the label is “Conservative”; if it is less than 0, the label is “Liberal”; and if it is exactly 0, the label is “Cannot say.” For content type annotation, we assign the “gold standard” label based on the majority vote among human annotators. In cases of a tie, the annotation provided by the highest-performing annotator–measured by agreement with the political leaning gold standard–is selected as the final label. Due to the relatively low inter-annotator agreement regarding content type, the reliability of this “gold standard” label should be interpreted with greater caution.

Figures 1.10 and 1.11 display the confusion matrices comparing Llama 3’s annotations with the human “gold standard” for political bias and content categories, respectively. Table 1.11 presents precision, recall, and F1-scores for Llama 3 and contrasts these metrics with the average performance of human annotators when evaluated against the same reference labels.

Regarding the classification of political leanings, Llama 3 achieved an overall accuracy of 80% (compared to 33% by random chance). It was most effective at identifying Liberal content, with an F1-score of 0.84 and a precision of 0.87. The “Cannot Say” category demonstrated high recall (0.87) but lower precision (0.67), indicating that the model frequently labelled partisan content as ambiguous. For Conservative content, the model showed the highest precision (0.94) but lower recall (0.68), suggesting a more cautious labelling strategy that may have overlooked some true Conservative instances. Across all three labels, Llama 3’s F1-scores are very similar to the average human F1-scores.

Regarding the classification of content types, Llama 3 achieved an overall accuracy of 75% (compared to 20% by random chance). The model performs well across all categories, with F1-scores above 70%, with the exception of the category “Other,” which was often confused with “Entertainment.” The average performance across human annotators also shows a low F1-score for this category, suggesting that the boundary between what constitutes “Entertainment” and “Other” content is ambiguous. Note that our results throughout the paper focus on “News” and “Political Activist” accounts, and the other categories are of lesser importance in the context of our

study. Overall, Llama 3's F1-scores are very similar to the average human F1-scores across all five labels.

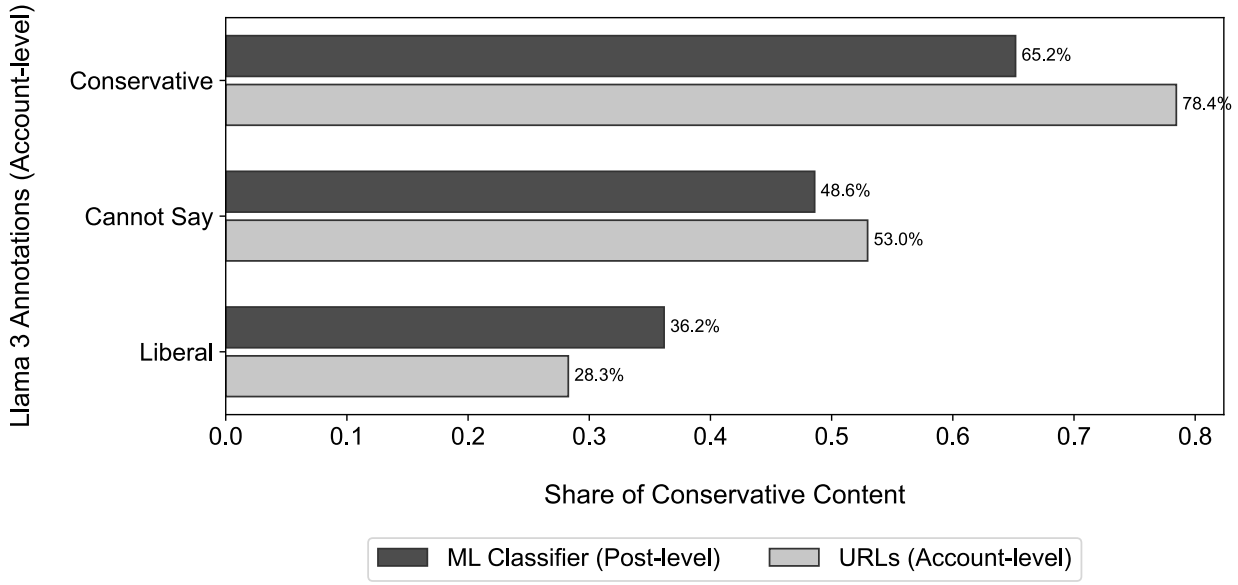


Fig. 1.7. Conservative Content with Alternative Proxies Against Llama 3 Annotations.

The average share of posts for a given Llama 3 annotation labelled as conservative by two other proxies. The ML Classifier makes predictions at the post-level and is plotted in dark gray. The political leaning of an account can also be inferred by the URLs shared by this account and is plotted in light gray. As expected, the agreement between Llama 3 and these alternative proxies is high.

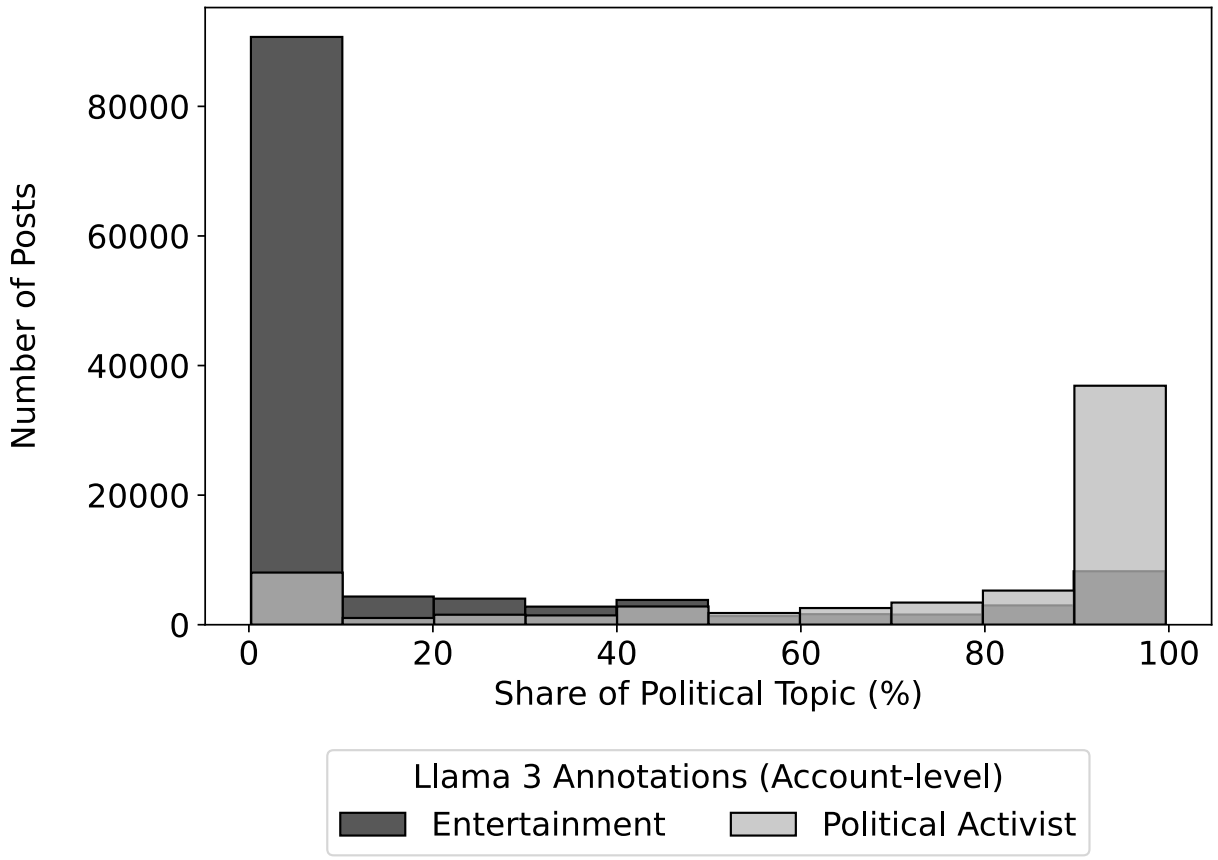


Fig. 1.8. Share of Political Topic Against Llama 3 Annotations.

The distribution of the topic share for the topic “Politics” for posts shared by accounts annotated by Llama 3 as “Entertainment” (in dark grey) and “Political Activist” (in light grey). As expected, accounts labelled as “Entertainment” by Llama 3 post very little political content (as measured by our topic model), while accounts labelled as “Political Activist” by Llama 3 mainly post political content.

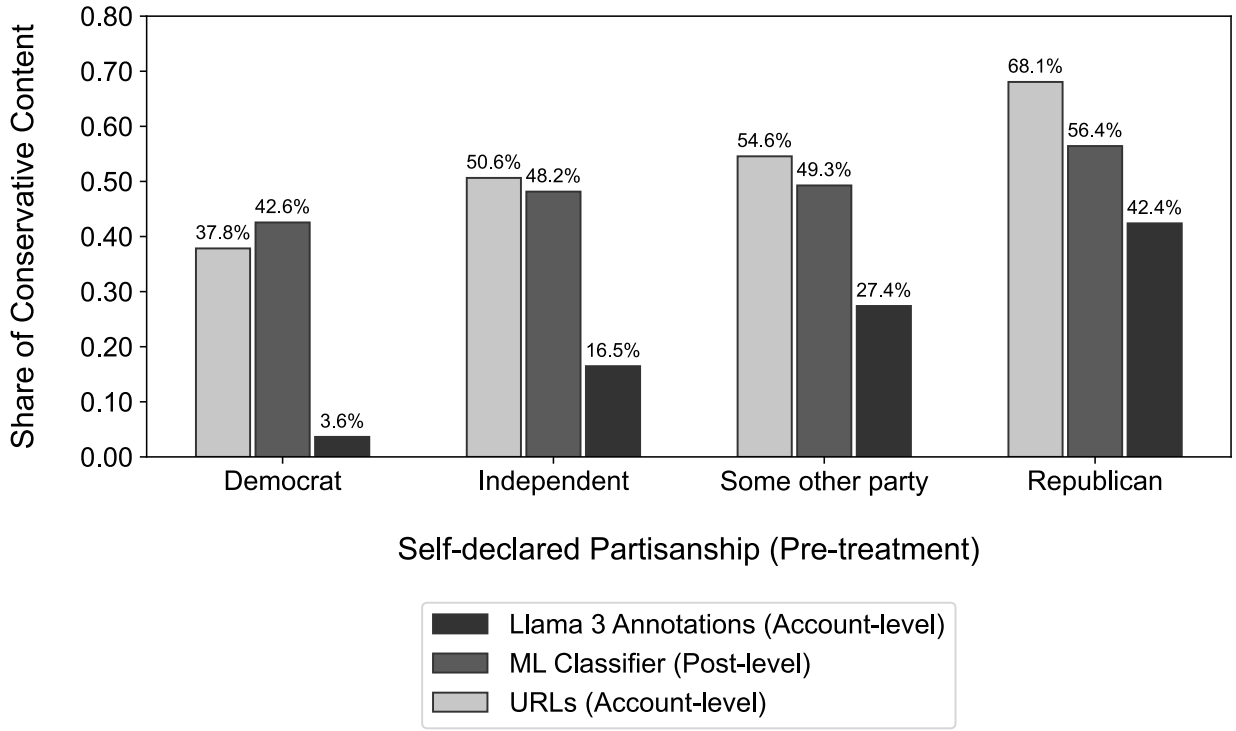


Fig. 1.9. Conservative Content by Partisanship (Pre-treatment).

The average share of conservative content shown in the feed by respondents' self-declared partisanship in the pre-treatment survey. We measure conservative content in three distinct ways: using Llama 3 (at the account-level), using an ML Classifier (at the post-level), and using the URLs shared in the posts (at the account-level). As expected, all three measures show that conservative respondents are exposed to higher shares of conservative content in their feeds.

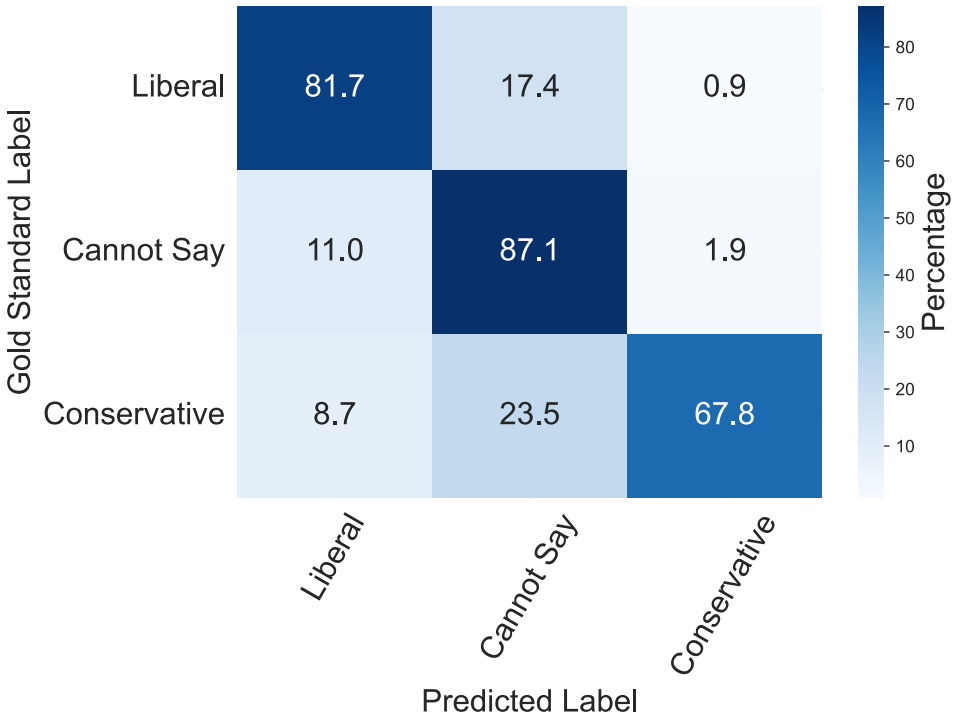


Fig. 1.10. Confusion Matrix: Political Leaning of the Accounts in the Feed Data. Llama 3 vs. human annotations.

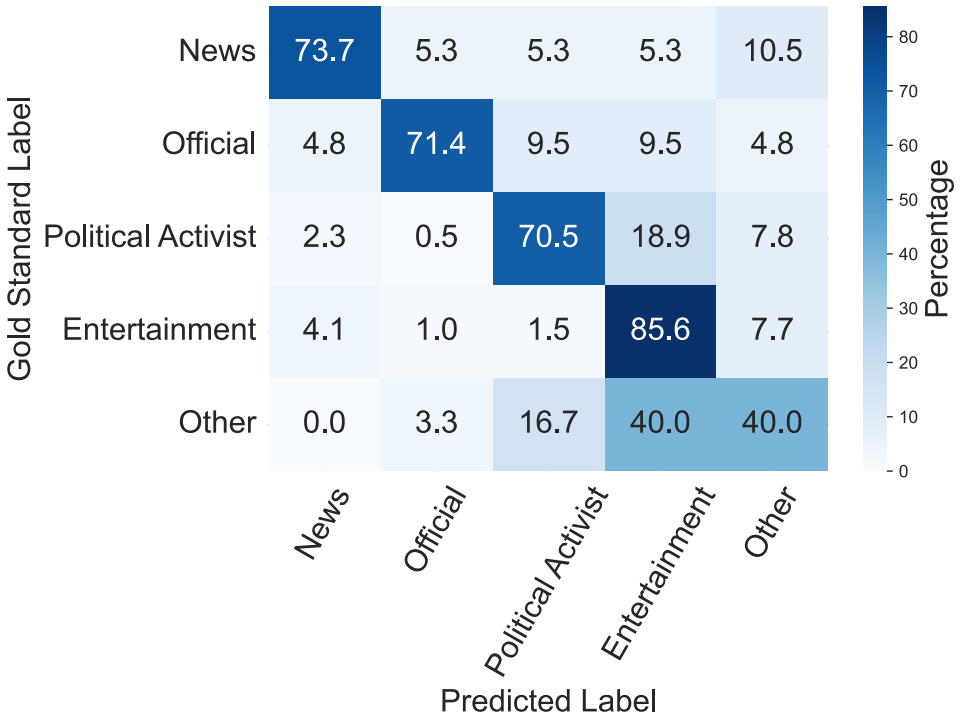


Fig. 1.11. Confusion Matrix: Content Type of the Accounts in the Feed Data. Llama 3 vs. human annotations.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Llama 3 Performance			**Average Human Performance**		
Label	**Support**	**Precision**	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
Panel A: Political Leaning (Account-level)							
Liberal	230	0.87	0.82	0.84	0.97	0.74	0.84
Cannot Say	155	0.67	0.87	0.76	0.64	0.96	0.76
Conservative	115	0.94	0.68	0.79	0.94	0.67	0.78
Panel B: Content Type (Account-level)							
News	38	0.67	0.74	0.70	0.78	0.65	0.69
Official	21	0.71	0.71	0.71	0.83	0.76	0.79
Political Activist	217	0.93	0.71	0.80	0.95	0.79	0.86
Entertainment	194	0.74	0.86	0.80	0.95	0.60	0.70
Other	30	0.24	0.40	0.30	0.25	1.00	0.39

Table 1.11. Llama 3 vs. Average Human (Accounts in the Feed Data)

Llama 3 performance and average performance of human annotators against “gold standard” labels (as defined in SI Section 1.5.4). Columns 3 to 5 report Llama 3’s performance. Columns 5 to 8 report the average performance of human annotators.

1.6 Annotations of Followed Accounts

In this section, we explain how we annotate the accounts that users follow. We begin by describing in SI Section 1.6.1 our baseline annotations, which are generated using the open-source large language model Llama 3.3 70B. We then validate these annotations by examining their correlation with users' pre-treatment partisanship and comparisons with human annotations in SI Section 1.6.2.

1.6.1 Llama 3 Annotations of Followed Accounts

We annotate the accounts followed by participants who shared their 𝕏 handle with us using the open-source large language model Llama 3.3 70B Instruct, using the same methodology as for annotating the feeds from Google Chrome extension (see SI Section 1.5).[7]

As with the feed annotation, we classify the accounts that participants follow along these two dimensions:

- **Political Leaning**

  The prompt that we used is as follows:

  I will show you the name and description from a Twitter account.

  Your task is to classify the account's political leaning based on this information.

  Choose only one of the following labels:

  - Conservative

  - Liberal

  - Cannot say

  Do not explain your choice or add anything else.

  Account name: [...]

  Description: [...]

  Label:

[7] Note that, contrary to the feed data, we cannot build simpler proxies of partisanhip and content type with word frequencies for the accounts users follow because each account is associated with very little data, namely an account name and a brief description.

We present Llama 3 with ten randomly selected posts from the account (or all posts if less than ten were available).

- **Type of Content Shared or Produced**

  The prompt is as follows:

  “You will be shown the name and description of a Twitter account.

  Classify the account into one of the following five categories:

  - News: Official news outlets or accounts primarily focused on reporting current events.

  - Political activist: Accounts primarily focused on partisan political commentary or advocacy.

  - Entertainment: Accounts centered on content like memes, jokes, pop culture, art, books, music, movies, video games, sports, cooking, science, business promotions, animals, lifestyle, motivation, and personal development.

  - Official: Accounts belonging to elected officials or government representatives.

  - Other: Accounts that don’t clearly fit any of the above.

  Respond with only one of these five labels (News, Politics, Entertainment, Official, Other). Do not include any explanation.

  Account name: [...]

  Description: [...]

  Label:”

1.6.2 Validation of Annotation of Followed Accounts

We assess the quality of our annotations through several methods. First, Fig. 1.12 shows that the slant of accounts followed by users strongly correlates with their pre-treatment partisanship. Republican users follow "Conservative" accounts 24.6% of the time, compared to 10.9% for Independents and just 2.0% for Democrats. A mirrored trend appears for accounts annotated as "Liberal": they constitute 4.7% of the followed accounts for Republicans, 14.9% for Independents, and 28.7% for Democrats. In contrast, the proportion of accounts labelled as "Cannot say" remains relatively stable across all partisan groups. These patterns indicate that our annotations effectively capture the well-known tendency of individuals to follow ideologically aligned accounts.

Second, we rely on human annotators. As for the feed data (see SI Section 1.5.4), the human annotators were presented with the exact same information, and each annotator labelled 500 accounts that some users followed. Annotations were blind (the LLM labels were not shown). Each of the 500 observations were drawn at random from our database.[8] All human annotators saw the same 500 samples so that we could ensure the annotation quality through the inter-annotator agreement. As for the feed data, human annotators show strong agreement on political leanings, with a Krippendorff's Alpha of 0.74. In contrast, agreement on content type is lower, with an Alpha of 0.41, indicating once again that this task is more challenging and ambiguous. As detailed below, most disagreements are between the "Entertainment" and "Other" categories (which are not the main focus of this study). The exact instructions given to the human annotators are in SI Sections 4.2. We construct "gold standard" human labels using the same procedure as for the feed data (see SI Section 1.5.4).

Figures 1.13 and 1.14 display the confusion matrices comparing Llama 3's annotations with the human "gold standard" for political bias and content categories, respectively. Table 1.12 presents precision, recall, and F1-scores for Llama 3 and contrasts these metrics with the average performance of human annotators when evaluated against the same reference labels.

Regarding the classification of political leanings, Llama 3 achieved an overall accuracy of 85% (compared to 33% by random chance). It displays F1-scores of 0.89, 0.78, and 0.86 for the labels "Liberal," "Cannot say," and "Conservative," respectively. Across all political labels, Llama 3's F1-scores are higher than the average human annotator's performance.

For content type classification, Llama 3 achieved an overall accuracy of 71% (compared to 20%

[8] When sampling the 500 observations, we imposed balance across the partisan leaning (Liberal, Conservative, Cannot Say) and account types as annotated by Llama 3 8B Instruct.

by random chance). The model performed reliably across the major categories, with F1-scores of 0.76 for “News” and 0.77 for “Political Activist,” which are the two categories of central interest in our study. The “Official” category had an F1-score of 0.67, and “Entertainment” reached 0.71. The weakest performance was on the “Other” category, with an F1-score of 0.46, reflecting a common confusion with “Entertainment”. Notably, the average human performance on “Other” is similarly low (F1 = 0.47), highlighting the inherent ambiguity in this classification. As with political leaning, Llama 3’s F1-scores for content type are closely aligned with human averages.

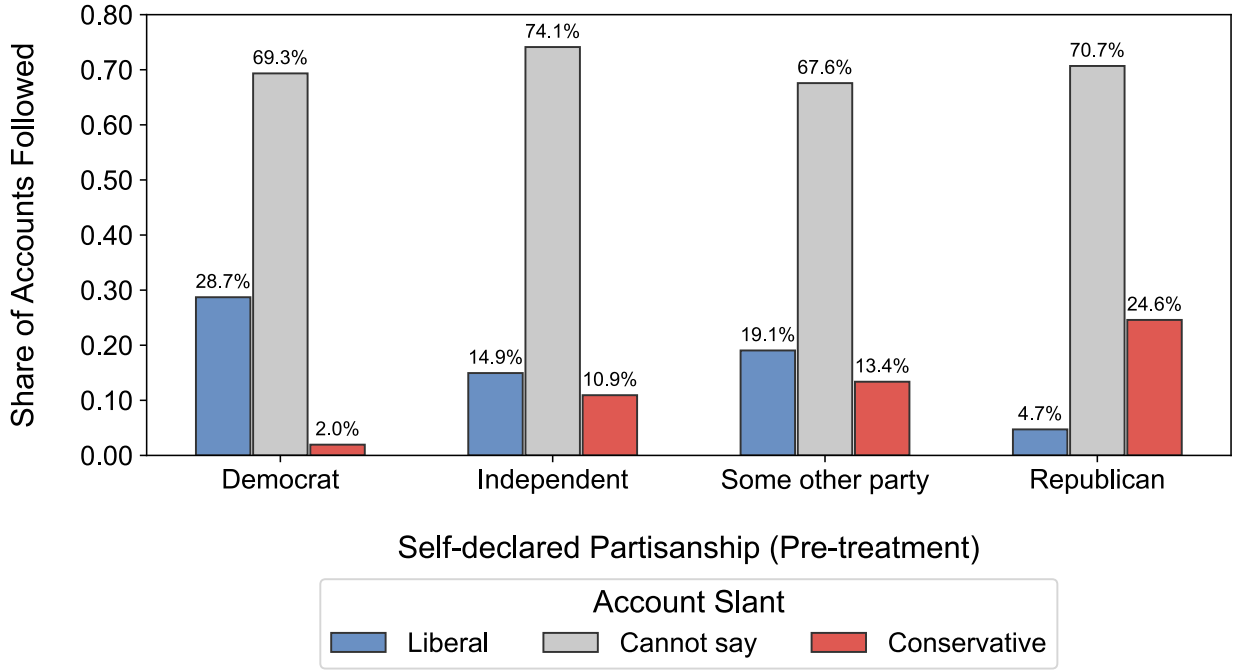


Fig. 1.12. Slant of Accounts Followed by Pre-treatment Partisanship.

The mean share of conservative accounts (according to Llama 3 annotations) that the respondent follows. We distinguish respondents by their self-declared partisanship in the pre-treatment survey. As expected, the more right-wing the respondent, the more they follow accounts labelled as conservative by Llama 3.

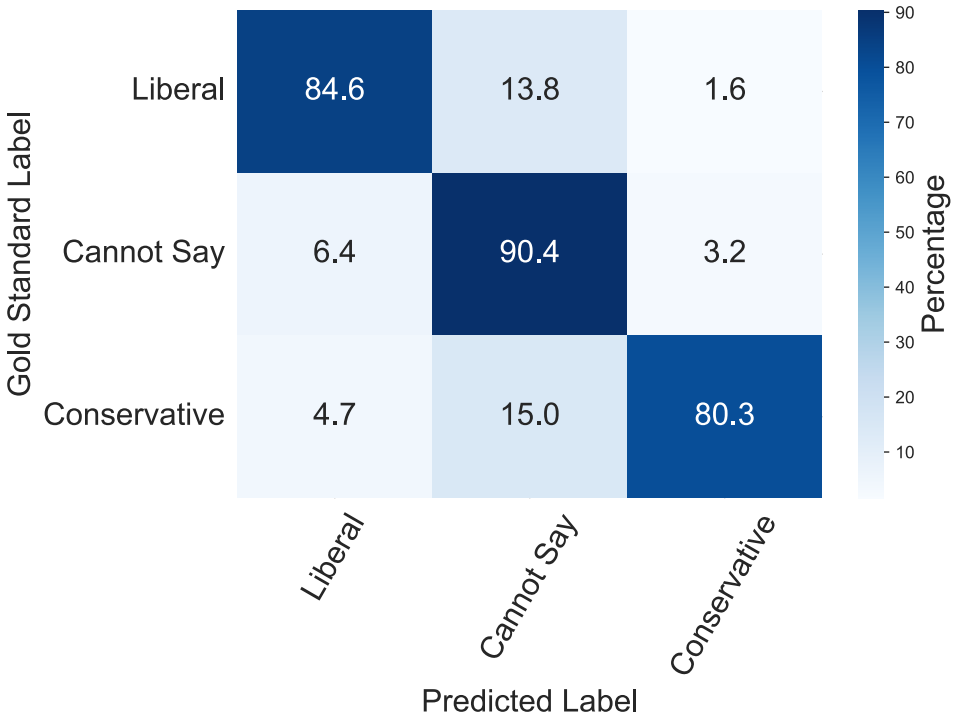


Fig. 1.13. Confusion Matrix: Political Leaning of Followed Accounts. Llama 3 vs. human annotations.

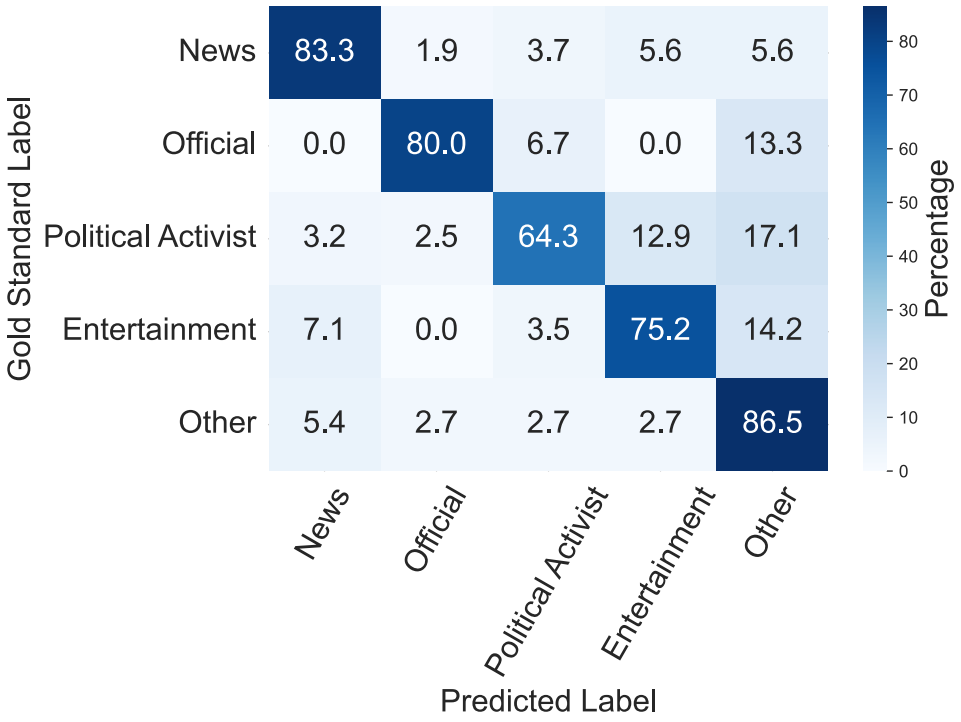


Fig. 1.14. Confusion Matrix: Content Types of Followed Accounts. Llama 3 vs. human annotations.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Llama 3 Performance			**Average Human Performance**		
Label	**Support**	**Precision**	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
Panel A: Political Leaning (Account-level)							
Liberal	247	0.94	0.85	0.89	0.99	0.77	0.86
Cannot Say	125	0.68	0.90	0.78	0.59	0.79	0.61
Conservative	127	0.93	0.80	0.86	0.96	0.66	0.77
Panel B: Content Type (Account-level)							
News	54	0.70	0.83	0.76	0.88	0.74	0.79
Official	15	0.57	0.80	0.67	0.84	0.57	0.65
Political Activist	280	0.96	0.64	0.77	0.94	0.71	0.80
Entertainment	113	0.68	0.75	0.71	0.85	0.54	0.63
Other	37	0.32	0.86	0.46	0.33	0.99	0.47

Table 1.12. Llama 3 vs. Average Human (Followed Accounts)

Llama 3 performance and average performance of human annotators against "gold standard" labels (as defined in SI Section 1.5.4). Columns 3 to 5 report Llama 3's performance. Columns 5 to 8 report the average performance of human annotators.

1.7 Covariate Balance

In this section, we document balance across treatment groups. Table 1.13 presents the balance of pre-treatment variables between the two experimental groups: participants assigned to the algorithmic feed and those assigned to the chronological feed. Table 1.14 displays covariate balance across experimental groups separately for participants initially on the chronological feed and those initially on the algorithmic feed. For continuous and ordinal variables, both tables report the p-value from a t-test comparing the means between respondents assigned to the two feed settings. For categorical variables, we report the p-value from the corresponding Chi-squared test. P-values are not adjusted for multiple hypothesis testing.

Table 1.13 shows that the randomisation was largely successful, with no statistically significant differences at the 10% level between users assigned to chronological versus algorithmic feeds across demographics, $\mathbb{X}$ usage patterns, well-being measures, and partisanship. There is one marginally significant difference ($p = 0.08$) in the share of users who were initially on the algorithmic feed (75% vs 77%), which we account for in our analysis by controlling for initial feed settings. The overall balance in observable characteristics, particularly in partisanship, provides a strong basis for attributing any differences in outcomes to the treatment rather than to pre-existing differences between groups.

Turning to the covariate balance by previous feed setting in Table 1.14, we find that for users initially on the chronological feed (1,206 respondents), there are no statistically significant differences between treatment groups, even at the 10% level. Among users initially on the algorithmic feed (3,759 respondents), we observe three marginally significant differences: for those assigned to the chronological feed, $\mathbb{X}$ use frequency is higher (3.88 vs 8.82, $p = 0.09$), life dissatisfaction is greater (4.67 vs 4.55, $p = 0.09$), and the share of users who consume hard news on $\mathbb{X}$ is slightly higher (0.69 vs 0.66, $p = 0.05$). Our results are robust to including these variables as controls (see SI Sections 2.3 and 2.4).

The overall balance within each initial feed setting group, combined with the robustness of our results to the inclusion of controls, suggests that we identify the causal effects of switching feed settings in both directions.

		Treatment Assignment:			
		Chronological	Algorithm		
Variable		**Mean**		**Difference**	**p-value**
General Information					
Gender	Share of Males	0.52	0.52	0.00	0.82
Age		50.39	50.48	-0.09	0.83
Race	Share of Whites	0.78	0.78	0.01	0.54
Education	Share of 4-year college +	0.59	0.57	0.01	0.32
𝕏 **Use**					
Initial feed setting	Share on the Algorithm	0.75	0.77	-0.02	0.08
𝕏 use frequency		4.92	4.87	0.05	0.13
𝕏 posting frequency		2.91	2.94	-0.03	0.63
News consumption	Share of soft news	0.61	0.62	0.00	0.77
	Share of hard news	0.70	0.68	0.02	0.13
Well-being					
Life Dissatisfaction		4.61	4.56	0.04	0.46
Unhappiness		2.03	2.05	-0.01	0.58
Political Attitudes					
Affective polarisation		36.88	38.45	-1.57	0.15
Partisanship	Share of Democrats	0.48	0.45	0.03	0.25
	Share of Independents	0.29	0.29	-0.01	
	Share of Republicans	0.21	0.22	-0.01	
	Share of Some other party	0.03	0.03	-0.01	
Number of Observations		2467	2498		

Table 1.13. Covariate Balance For the Baseline Sample of Respondents.

Differences in means of pre-treatment covariates by treatment group. For continuous and ordinal variables, the p-value of a t-test of the difference between the mean for respondents assigned to the chronological feed and the mean for those assigned to the algorithmic feed is reported. For categorical variables, the p-value of a Chi-squared test is reported. p-values are not adjusted for multiple hypothesis testing.

		Initial Feed Setting: Chronological				Initial Feed Setting: Algorithm			
		Treatment Assignment: Chrono	Treatment Assignment: Algorithm			Treatment Assignment: Chrono	Treatment Assignment: Algorithm		
		Mean		Diff.	p-value	Mean		Diff.	p-value
General Information									
Gender		0.48	0.52	-0.04	0.15	0.53	0.52	0.01	0.55
Age		51.91	51.31	0.60	0.47	49.88	50.23	-0.35	0.47
Race	Share of Whites	0.81	0.82	-0.01	0.57	0.77	0.76	0.01	0.38
Education	Share of 4-year college +	0.62	0.64	-0.02	0.52	0.58	0.55	0.02	0.16
𝕏 **Use**									
𝕏 use frequency		5.04	5.05	-0.01	0.89	4.88	4.82	0.07	0.09
𝕏 posting frequency		2.87	2.91	-0.03	0.78	2.92	2.95	-0.02	0.71
News consumption	Share of soft news	0.62	0.65	-0.03	0.31	0.61	0.61	0.00	0.84
	Share of hard news	0.72	0.74	-0.02	0.49	0.69	0.66	0.03	0.05
Well-being									
Unhappiness		2.02	2.08	-0.06	0.13	2.04	2.03	0.00	0.85
Life Dissatisfaction		4.44	4.62	-0.18	0.15	4.67	4.55	0.12	0.09
Political Attitudes									
Affective polarisation		33.12	35.97	-2.86	0.19	38.15	39.17	-1.02	0.42
Partisanship	Democrat	0.49	0.46	0.03	0.62	0.47	0.45	0.02	0.26
	Independent	0.29	0.32	-0.03		0.28	0.28	0.00	
	Republican	0.19	0.20	-0.01		0.22	0.23	-0.01	
	Some other party	0.04	0.03	0.00		0.03	0.04	-0.01	
Number of Observations		626	580			1841	1918		

Table 1.14. Covariate Balance for the Baseline Sample of Respondents by Initial Feed Setting.

Differences in means of pre-treatment covariates by treatment group, separately for each initial feed setting. For continuous and ordinal variables, the p-value of a t-test of the difference between the mean for respondents assigned to the chronological feed and the mean for those assigned to the algorithmic feed is reported. For categorical variables, the p-value of a Chi-squared test is reported. p-values are not adjusted for multiple hypothesis testing.

1.8 Compliance and Attrition

In this section, we first describe the measures of compliance, namely whether participants used the experimentally assigned setting during the experiment, and then describe attrition, namely how many participants did not return for the post-treatment survey, across treatment groups.

1.8.1 Compliance

Extended Data Fig. 2 shows that the experiment achieved high compliance rates across different measures. The Chrome extension data, which directly observed users' feed settings, indicate that 93.49% of participants were using their assigned feed setting during the pre-treatment survey, with compliance remaining very high at 89.15% during the post-treatment survey. Self-reported compliance data from the post-treatment survey also reflect generally high compliance rates, with 85.38% of participants indicating that they used their assigned feed setting either "Always" (58.09%) or "Most of the time" (27.29%). Only a small fraction reported that they complied "Rarely" (3.85%) or "Never" (1.29%). Note that respondents were explicitly informed that their self-declared compliance would not affect their compensation. These rates, across both objective and self-reported measures, suggest that our intention-to-treat estimates closely approximate the average treatment effects.

We further investigate how compliance varied across treatment groups. Tables 1.15 and 1.16 document the numbers and percentages of (observed and self-reported) compliers per treatment assignment and by previous feed setting. When focusing on respondents who declare having complied "Always" or "Most of the time", there is no evidence of differential compliance between treatment and control groups: the p-value is 0.11 for respondents who were initially on the chronological feed, and 0.63 for those who were initially on the algorithmic feed. However, respondents initially on the algorithmic feed were 15 percentage points more likely to declare having complied "Most of the time" rather than "Always" if they were assigned to the chronological feed ($p < 0.001$). This suggests that compliance may have been slightly lower (with a difference between "Always" and "Most of the time") among respondents for whom we switched the feed algorithm off.

As discussed in the Methods Section, to rule out that our reported null effects for respondents for whom the feed algorithm was switched off are driven by lower compliance rates, we estimate the LATE using a conservative definition of compliance. Specifically, in this alternative definition, we assume that respondents who self-report almost always complying actually adhere to the as-

signed setting at any moment with probability 0.5. The results remain unchanged (see SI Fig. 2.13, compared to Extended Data Fig. 3).

		Treatment Assignment:			
		Chronological	Algorithm		
Variable		**Count**		**Difference**	**p-value**
Complied (indicator)		2111	2128	-17	0.70
Self-declared Compliance	Always	1303	1581	-278	0.00
	Most of the time	808	547	261	
	Sometimes	235	236	-1	
	Rarely	96	95	1	
	Never	25	39	-14	
Observed Compliance	Pre-treatment Survey	276	335	-59	0.00
	Post-treatment Survey	236	298	-62	0.00
Variable		**Percentage (%)**		**Difference**	**p-value**
Complied (indicator)		86	85	1	0.70
Self-declared Compliance	Always	53	63	-10	0.00
	Most of the time	33	22	11	
	Sometimes	10	9	0	
	Rarely	4	4	0	
	Never	1	2	-1	
Observed Compliance	Pre-treatment Survey	0.90	0.97	-0.07	0.00
	Post-treatment Survey	0.83	0.95	-0.12	0.00

Table 1.15. Compliance Across Treatment Groups.

Counts (upper panel) and percentages (lower panel) of compliers, according to observed and self-reported measures of compliers. For categorical variables, the p-value of a Chi-squared test is reported. For continuous or binary outcomes, the p-value of a t-test is reported. Complied (indicator) takes value one if the respondent declared complying with their feed assignment “Always” or “Most of the time.”

		Initial Feed Setting:							
		Chronological				Algorithm			
		Treatment Assignment:				Treatment Assignment:			
		Chrono	Algorithm			Chrono	Algorithm		
		Count		Diff.	p-value	Count		Diff.	p-value
Complied (indicator)		551	492	59	0.11	1560	1636	-76	0.63
Self-declared Compliance	Always	379	333	46	0.39	924	1248	-324	0.00
	Most of the time	172	159	13		636	388	248	
	Sometimes	56	59	-3		179	177	2	
	Rarely	15	21	-6		81	74	7	
	Never	4	8	-4		21	31	-10	
Observed Compliance	Pre-treatment Survey	67	61	6	0.12	209	274	-65	0.00
	Post-treatment Survey	59	62	-3	0.56	177	236	-59	0.00
		Percentage (%)		Diff.	p-value	Percentage (%)		Diff.	p-value
Complied (indicator)		88.00	85.00	3.00	0.11	85.00	85.00	-1.00	0.63
Self-declared Compliance	Always	61.00	57.00	4.00	0.39	50.00	65.00	-15.00	0.00
	Most of the time	27.00	27.00	0.00		35.00	20.00	15.00	
	Sometimes	9.00	10.00	-1.00		10.00	9.00	0.00	
	Rarely	2.00	4.00	-1.00		4.00	4.00	1.00	
	Never	1.00	1.00	-1.00		1.00	2.00	-0.00	
Observed Compliance	Pre-treatment Survey	86.00	94.00	-8.00	0.12	91.00	97.00	-6.00	0.00
	Post-treatment Survey	88.00	91.00	-3.00	0.56	81.00	96.00	-15.00	0.00

Table 1.16. Compliance Across Treatment Groups by Initial Feed Setting.

Counts (upper panel) and percentages (lower panel) of compliers, according to observed and self-reported measures of compliers. For categorical variables, the p-value of a Chi-squared test is reported. For continuous or binary outcomes, the p-value of a t-test is reported. Complied (indicator) takes value one if the respondent declared complying with their feed assignment “Always” or “Most of the time.”

1.8.2 Attrition

Table 1.17 analyses differential attrition between treatment groups. Table 1.18 presents a similar analysis, separately for participants initially on the chronological feed and those initially on the algorithmic feed. In both tables, For continuous and ordinal variables, we report the p-value from a t-test comparing the means of respondents assigned to the chronological and algorithmic feeds. For categorical variables, we report the p-value from a Chi-squared test. P-values are not adjusted for multiple hypothesis testing.

Table 1.17 shows that, in total, 21.8% of pre-treatment survey respondents did not participate in the post-treatment survey. This attrition did not differ between treatment groups. The attrition rates for the chronological feed (21.88%) and the algorithmic feed (21.72%) differed by only 0.16 percentage points, which was not statistically significant ($p = 0.87$). Attrition was also non-differential across demographic characteristics, $\mathbb{X}$ usage patterns, well-being measures, and partisanship, with no statistically significant differences across socio-demographic groups.

In Table 1.18, we examine attrition separately for participants initially on chronological and algorithmic feeds. Among users who were initially on the chronological feed, those assigned to switch to the algorithmic feed had a somewhat higher attrition rate (23.18%) compared to those who remained on the chronological feed (19.74%), although this difference is only marginally significant ($p = 0.10$). For users initially on the algorithmic feed, the pattern reverses but is smaller and not statistically significant ($p = 0.27$): those assigned to switch to the chronological feed experienced slightly higher attrition (22.58%) than those who remained on the algorithmic feed (21.26%). Within each initial feed group, attrition patterns are generally balanced across demographics, $\mathbb{X}$ usage patterns, well-being measures, and partisanship, with only posting frequency showing marginal significance ($p = 0.08$) in the sample of participants initially on the algorithmic feed. These patterns suggest that, while switching feed settings may slightly increase attrition, this effect is modest and unlikely to drive our main findings on the asymmetric effects of algorithmic exposure.

We confirm that selective attrition does not bias our results by applying a formal Lee bounds test. Lee bounds (50) provide a range within which the true treatment effect lies in the presence of non-random selective attrition. The estimator relies on two reasonable assumptions: random treatment assignment (which holds in our study) and monotonicity. Monotonicity implies that assignment to the treatment group can influence attrition in only one direction, which is a plausible

assumption. We conduct this analysis using the Stata package `leebounds` (51), not including additional controls. We conduct the estimation separately by initial feed setting.

Extended Data Fig. 6 presents the results for the baseline survey outcomes and Fig. 1.15 for the accounts followed by the participants. As in all other coefficient plots, we standardise each outcome variable. The results confirm that selective attrition is not a concern in our study.

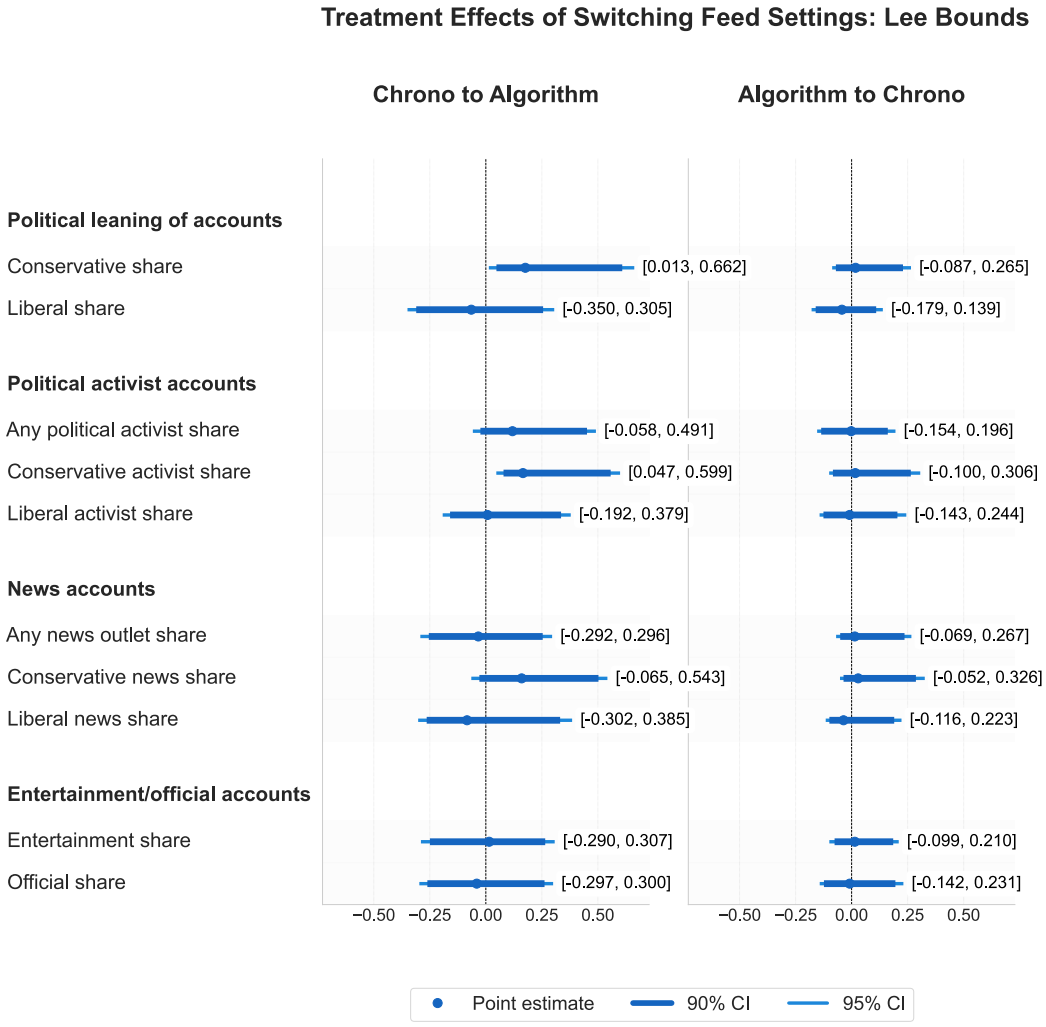


Fig. 1.15. Followed Accounts. ITT Estimates by Initial Feed Setting: Lee Bounds Estimates.

Lee bounds for the unconditional point estimates of switching the algorithm on (left panel) and switching it off (right panel). 90% and 95% CIs are reported. All outcomes are standardized. “Chrono” and “Algorithm” stand for the chronological and algorithmic feeds, respectively. The left panel: estimates for respondents with a chronological feed (Chrono) as an initial feed setting and who were assigned to the algorithm. The right panel: estimates for respondents with the algorithm as an initial feed setting and who were assigned to the chronological feed. A similar figure for the survey outcomes is presented in Extended Data Fig. 6.

		Treatment Assignment:			
		Chronological	Algorithm		
Variable		**Attritors (%)**		**Difference (%)**	**p-value**
Full sample		21.88	21.72	0.16	0.87
General Information					
Gender	Female	21.54	21.18	0.36	0.79
	Male	22.20	22.20	-0.01	
Race	Other	24.05	23.64	0.40	0.88
	White	21.26	21.14	0.12	
Education	4y college +	21.75	21.74	0.01	0.78
	Community college –	22.06	21.68	0.38	
𝕏 **Use**					
Initial feed setting	Algorithm	22.58	21.26	1.32	0.22
	Chronological	19.74	23.18	-3.44	
𝕏 use frequency	Several times a day	23.49	22.72	0.77	0.45
	Once a day	20.00	20.35	-0.35	
	Several times a week	20.46	19.54	0.92	
	Once a week	19.64	22.98	-3.34	
	Several times a month	23.04	24.66	-1.62	
𝕏 posting frequency	Several times a day	24.68	21.83	2.85	0.21
	Once a day	21.32	23.27	-1.95	
	Several times a week	21.45	18.21	3.24	
	Once a week	24.05	25.62	-1.58	
	Several times a month	20.92	18.38	2.54	
	Once a month or less	20.46	21.19	-0.74	
	Never	21.14	25.78	-4.63	
Well-being					
Happiness	Not at all happy	22.45	16.39	6.06	0.67
	Not very happy	23.50	25.23	-1.73	
	Rather happy	20.51	21.17	-0.66	
	Very happy	23.95	21.64	2.31	
Political Attitudes					
Partisanship	Democrat	21.79	22.19	-0.40	0.98
	Independent	20.25	20.13	0.12	
	Republican	23.89	22.69	1.20	
	Some other party	24.21	22.32	1.89	
Number of Observations		3158	3191		

Table 1.17. Differential Attrition Between Treatment Groups.

Attrition rates by treatment group. For continuous and ordinal variables, the p-value of a t-test for the difference between the share of attritors among respondents assigned to the chronological feed and the share of attritors among respondents assigned to the algorithm is reported. For categorical variables, the p-value of a Chi-squared test is reported. p-values are not adjusted for multiple hypothesis testing.

		Initial Feed Setting: Chronological				Initial Feed Setting: Algorithm			
		Treatment Assignment: Chrono	Treatment Assignment: Algorithm			Treatment Assignment: Chrono	Treatment Assignment: Algorithm		
		Attritors (%)		Diff.	p-value	Attritors (%)		Diff.	p-value
Full Sample		19.74	23.18	-3.44	0.10	22.58	21.26	1.32	0.27
General Information									
Gender	Female	18.75	23.06	-4.31	0.90	22.54	20.60	1.94	0.85
	Male	20.79	23.29	-2.50		22.62	21.87	0.75	
Race	Other	20.27	23.31	-3.04	0.79	25.04	23.71	1.33	0.62
	White	19.62	23.15	-3.53		21.83	20.46	1.37	
Education	4y college +	20.04	22.64	-2.60	0.90	22.36	21.42	0.94	0.79
	Community college –	19.26	24.10	-4.84		22.88	21.07	1.81	
𝕏 **Use**									
𝕏 use frequency	Several times a day	21.34	25.78	-4.44	0.98	24.30	21.55	2.75	0.27
	Once a day	20.37	21.38	-1.01		19.88	20.04	-0.16	
	Once a week	12.50	18.37	-5.87		21.59	24.12	-2.53	
	Several times a week	17.74	19.08	-1.34		21.16	19.65	1.51	
	Several times a month	16.28	25.00	-8.72		24.84	24.60	0.25	
𝕏 posting frequency	Never	20.79	26.73	-5.94	0.63	21.26	25.50	-4.24	0.08
	Once a day	19.75	27.69	-7.94		21.85	22.13	-0.29	
	Once a week	17.65	27.69	-10.05		26.29	25.00	1.29	
	Several times a day	21.97	21.97	0.00		25.55	21.78	3.77	
	Several times a month	14.29	19.75	-5.47		23.24	17.92	5.32	
Well-being									
Happiness	Not at all happy	17.39	7.14	10.25	0.35	24.00	19.15	4.85	0.56
	Not very happy	20.83	29.71	-8.88		24.24	23.69	0.55	
	Rather happy	17.97	20.59	-2.62		21.40	21.35	0.05	
	Very happy	22.58	27.14	-4.56		24.39	20.08	4.30	
Political Attitudes									
Partisanship	Democrat	20.31	22.35	-2.04	0.52	22.30	22.14	0.16	0.99
	Independent	15.74	21.03	-5.29		21.71	19.82	1.88	
	Republican	24.68	27.85	-3.17		23.65	21.22	2.43	
	Some other party	15.38	25.00	-9.62		27.54	21.59	5.95	
Number of Observations		780	755			2378	2436		

Table 1.18. Differential Attrition Between Treatment Groups by Initial Feed Setting.

Attrition rates by treatment group and initial feed setting. For continuous and ordinal variables, the p-value of a t-test for the difference between the mean for respondents assigned to the chronological feed and the mean for those assigned to the algorithmic feed is reported. For categorical variables, the p-value of a Chi-squared test is reported. p-values are not adjusted for multiple hypothesis testing.

2 Additional Results

- Section 2.1 presents the bar chart summaries of raw data for all individual outcomes.
- Section 2.2 provides additional methodological details about the empirical approach.
- Section 2.3 presents coefficient plots for the results of various robustness exercises for unconditional and GRFs specifications.
- Section 2.4 presents tables with linear regression results for all outcomes across various alternative specifications.
- Section 2.5 presents the tests for whether the selection into the initial feed setting could explain asymmetry in results for turning the algorithm on and off.
- Section 2.6 presents regression results for the analysis of the content promoted by the algorithm.
- Section 2.7 presents the analysis of how experimental treatments affected on content of the users' chronological feed.

2.1 Descriptive Results: Raw-Data Summaries for All Individual Outcomes

Fig. 1 illustrates our main results using raw-data summaries, plotting the mean unconditional values of all outcomes by treatment assignment separately for samples with different initial feed settings. In this section, we present such raw-data summaries for all (rather than some) individual outcome variables. Each figure in this section summarizes results for a group of related outcomes. In each figure, the upper panel shows results for participants initially on the chronological feed, while the lower panel focuses on respondents initially on the algorithmic feed. For each sample, the outcomes for the chronological feed treatment assignment are shown in grey, and the corresponding outcomes for the algorithmic feed treatment assignment are in red.

We organize the graphic presentation of these raw-data summaries according to the following groups of outcomes:

- Fig. 2.1 presents all user engagement outcomes;
- Fig. 2.2 presents all outcomes related to affective polarisation and partisanship;
- Fig. 2.3 presents all policy priorities outcomes;
- Fig. 2.4 presents all opinions about investigations against Trump;
- Fig. 2.5 presents all outcomes related to the war in Ukraine;
- Fig. 2.6 presents life dissatisfaction outcomes;

- Fig. 2.7 presents the composition of the accounts that users follow;
- Fig. 2.8 presents other survey-based outcomes.

2.1.1 User Engagement

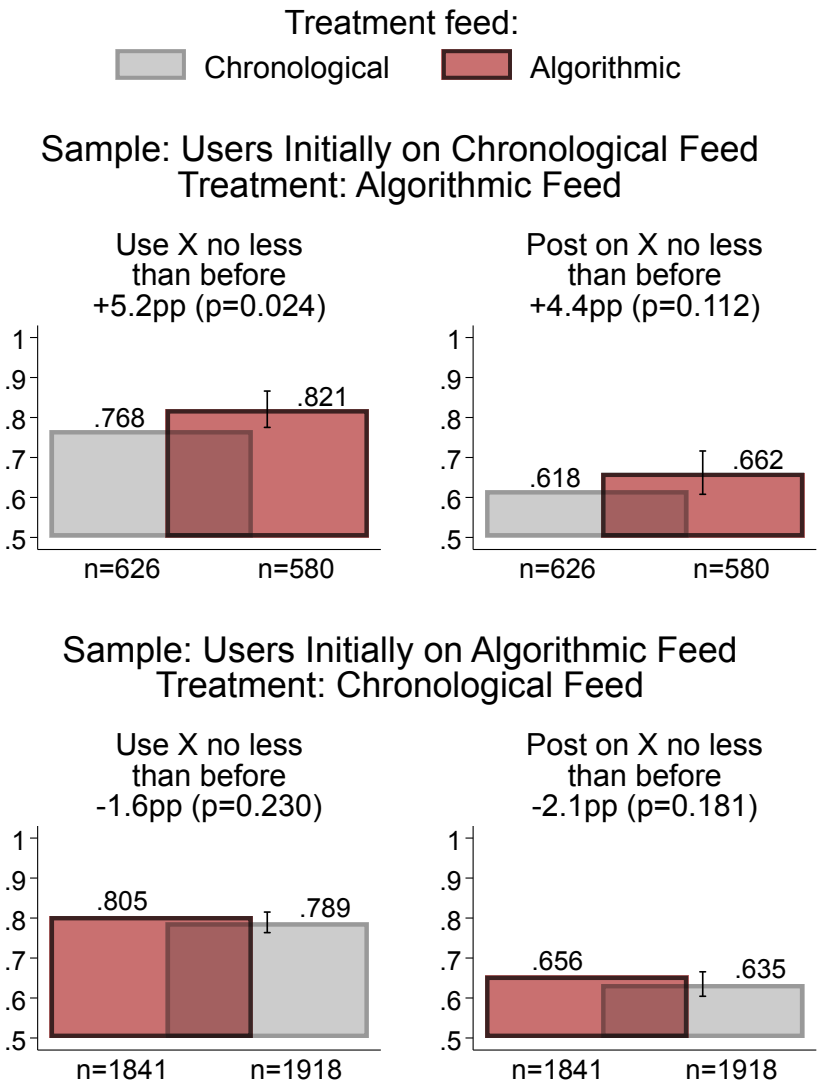


Fig. 2.1. Outcomes: User Engagement and Experience on the Platform.

Means of individual variables measuring user engagement on 𝕏 by assigned treatment feed in the two samples: (1) the top panel summarizes outcomes for the sample of users initially on the chronological feed; (2) the bottom panel summarizes outcomes for the sample of participants initially on the algorithmic feed. The difference in the height of the two bars in each chart represents the intention-to-treat (ITT) effect. Error bars show 95% CIs for the treatment effect relative to the control mean (Control mean + ATE $\pm$ 1.96 $\times$ SE). The title of each chart presents the treatment-effect magnitude and its *p*-value. See the regression results for these outcomes in Table 2.1.

2.1.2 Affective Polarisation and Self-Declared Partisanship

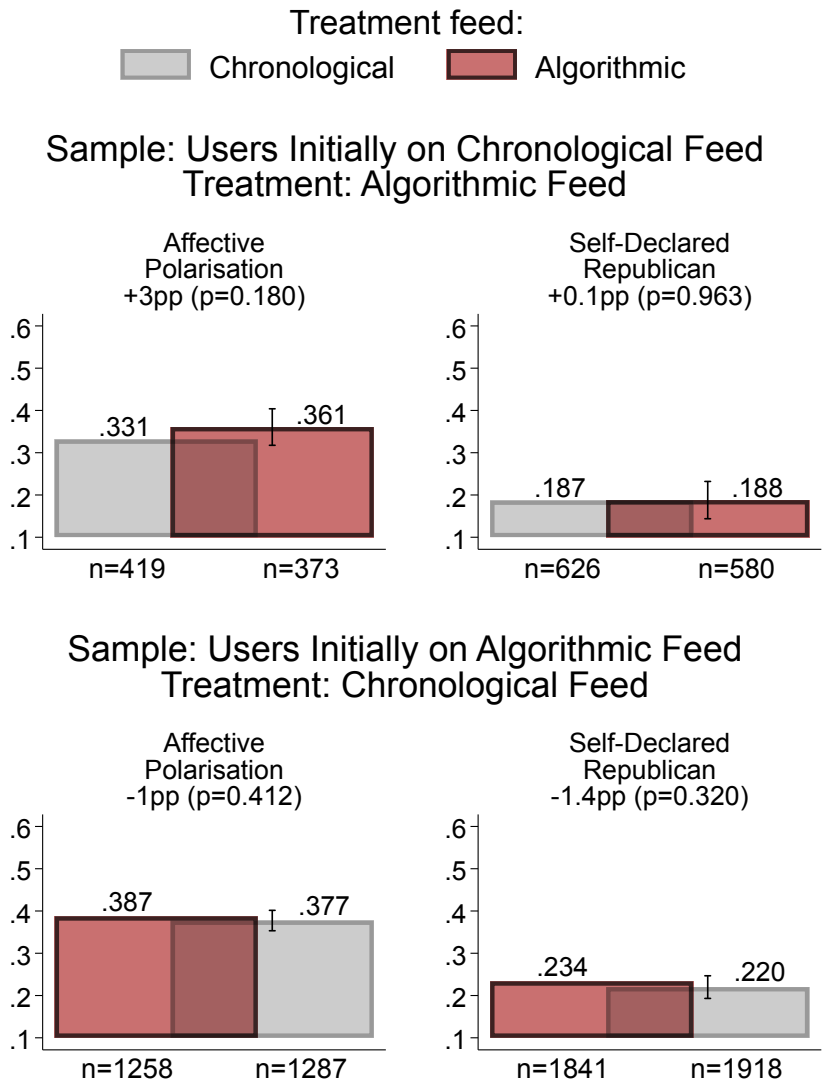


Fig. 2.2. Outcomes: Affective Polarisation and Self-Declared Partisanship.

Means of affective polarisation and partisanship by assigned treatment feed in the two samples: (1) the top panel summarizes outcomes for the sample of users initially on the chronological feed; (2) the bottom panel summarizes outcomes for the sample of participants initially on the algorithmic feed. Affective polarisation was rescaled to range from 0 to 1. The difference in the height of the two bars in each chart represents the intention-to-treat (ITT) effect. Error bars show 95% CIs for the treatment effect relative to the control mean (Control mean + ATE $\pm$ 1.96 $\times$ SE). The title of each chart presents the treatment-effect magnitude and its *p*-value. See the regression results for these outcomes in Table 2.2.

2.1.3 Conservative Policy Priorities

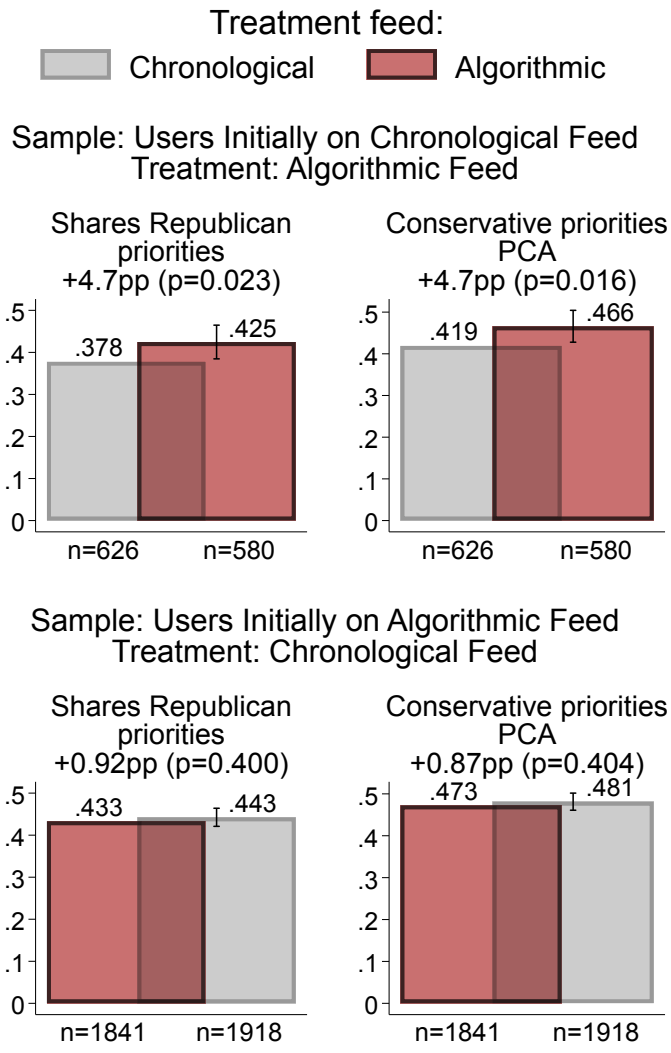


Fig. 2.3. Outcomes: Conservative Policy Priorities.

Means of individual variables measuring conservative policy priorities by assigned treatment feed in the two samples: (1) the top panel summarizes outcomes for the sample of users initially on the chronological feed; (2) the bottom panel summarizes outcomes for the sample of participants initially on the algorithmic feed. The difference in the height of the two bars in each chart represents the intention-to-treat (ITT) effect. Error bars show 95% CIs for the treatment effect relative to the control mean (Control mean + ATE $\pm$ 1.96 $\times$ SE). The title of each chart presents the treatment-effect magnitude and its *p*-value. See the regression results for these outcomes in Table 2.3.

2.1.4 Attitudes toward the Trump Criminal Investigations

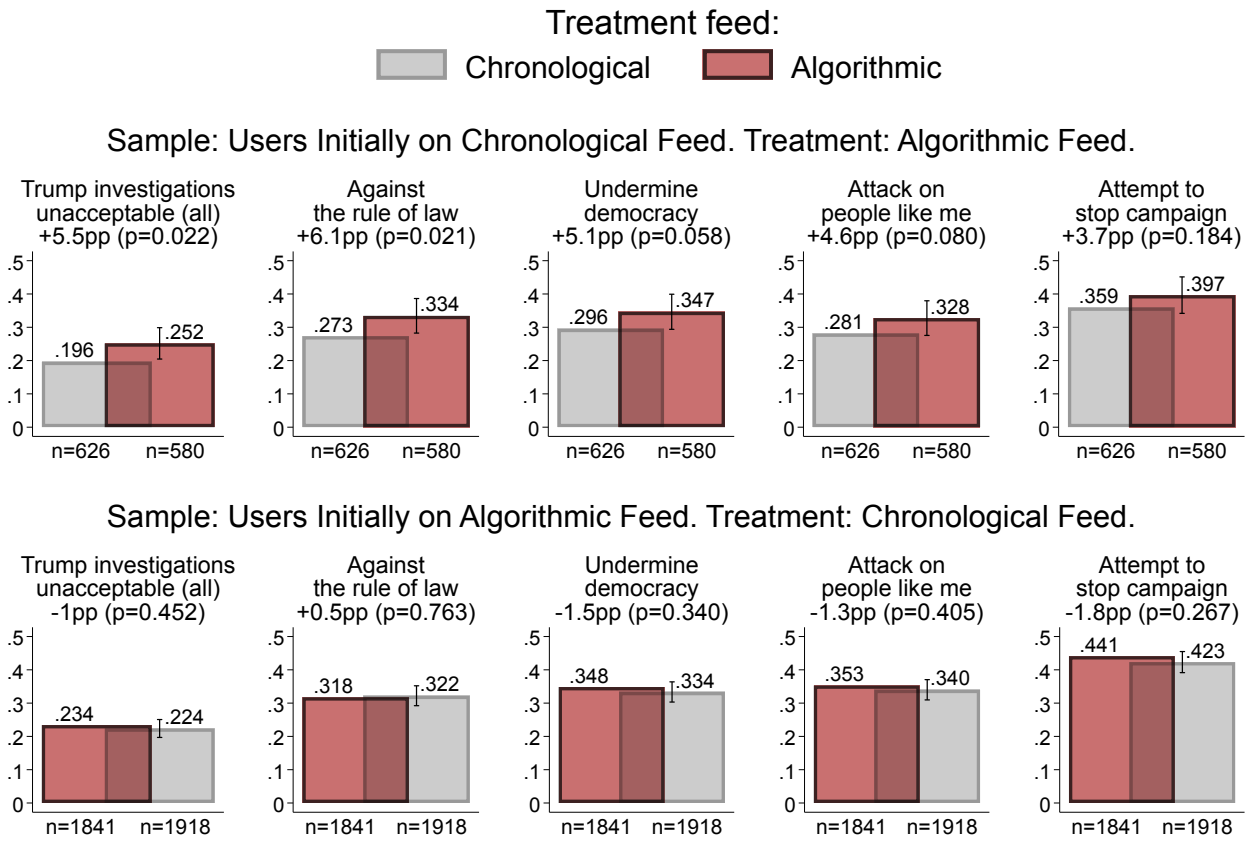


Fig. 2.4. Outcomes: Attitude to the Trump Criminal Investigations.

Means of individual variables measuring attitudes toward the Trump Criminal Investigations by assigned treatment feed in the two samples: (1) the top panel summarizes outcomes for the sample of users initially on the chronological feed; (2) the bottom panel summarizes outcomes for the sample of participants initially on the algorithmic feed. The difference in the height of the two bars in each chart represents the intention-to-treat (ITT) effect. Error bars show 95% CIs for the treatment effect relative to the control mean (Control mean + ATE $\pm$ 1.96 $\times$ SE). The title of each chart presents the treatment-effect magnitude and its *p*-value. See the regression results for these outcomes in Table 2.4.

2.1.5 Attitudes toward the War in Ukraine

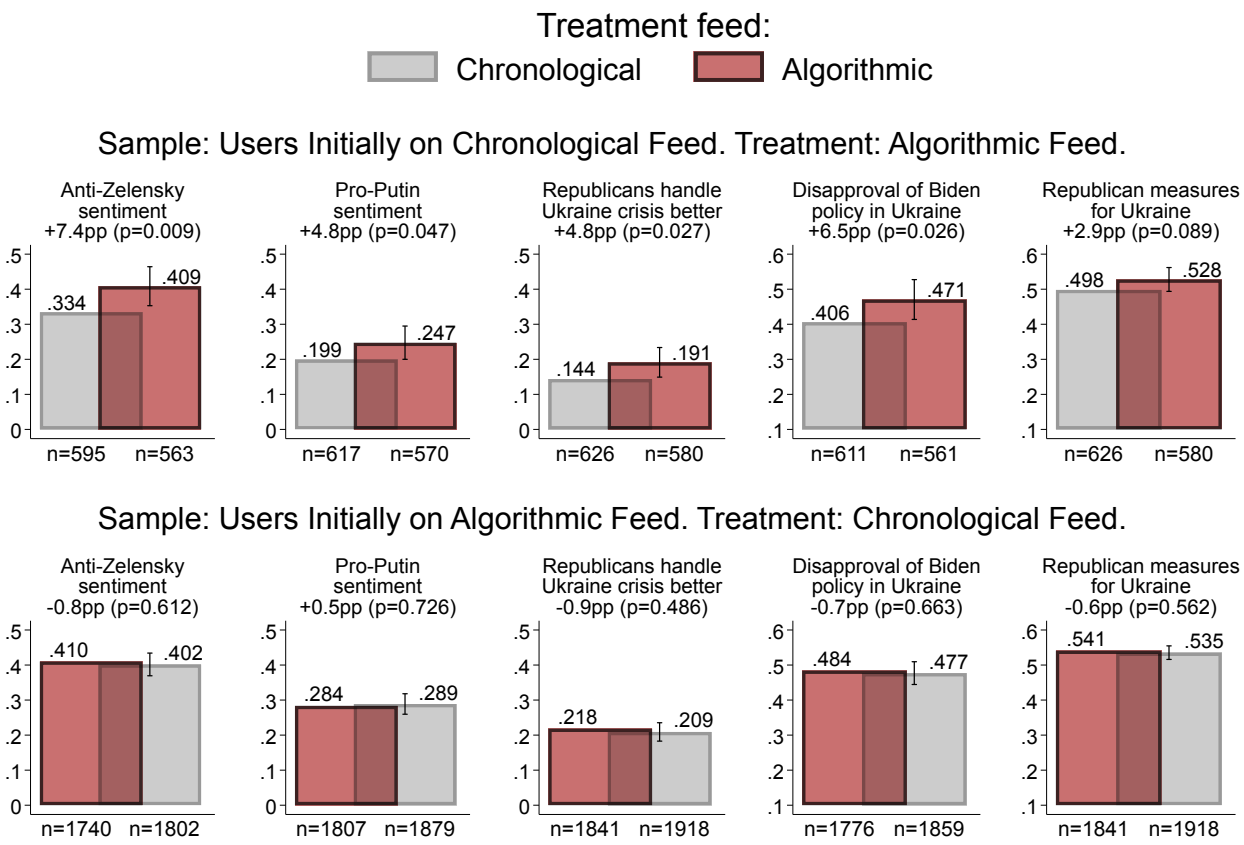


Fig. 2.5. Outcomes: Attitudes toward the US Involvement in Ukraine War.

Means of individual variables measuring attitudes toward the war in Ukraine by assigned treatment feed in the two samples: (1) the top panel summarizes outcomes for the sample of users initially on the chronological feed; (2) the bottom panel summarizes outcomes for the sample of participants initially on the algorithmic feed. The difference in the height of the two bars in each chart represents the intention-to-treat (ITT) effect. Error bars show 95% CIs for the treatment effect relative to the control mean (Control mean + ATE $\pm$ 1.96 $\times$ SE). The title of each chart presents the treatment-effect magnitude and its *p*-value. See the regression results for these outcomes in Table 2.5.

2.1.6 Life Dissatisfaction and Unhappiness

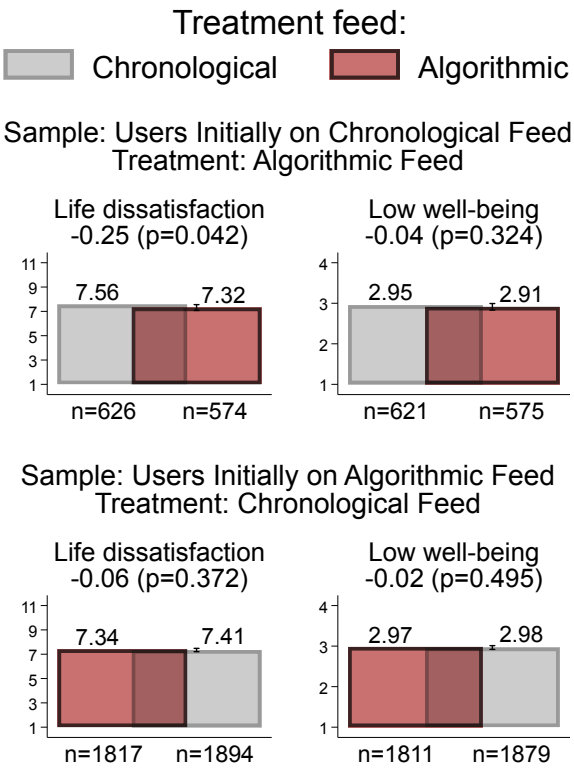


Fig. 2.6. Outcomes: Life Dissatisfaction and Unhappiness.

Means of individual variables measuring dissatisfaction and unhappiness by assigned treatment feed in the two samples: (1) the top panel summarizes outcomes for the sample of users initially on the chronological feed; (2) the bottom panel summarizes outcomes for the sample of participants initially on the algorithmic feed. The difference in the height of the two bars in each chart represents the intention-to-treat (ITT) effect. Error bars show 95% CIs for the treatment effect relative to the control mean (Control mean+ATE$\pm$1.96$\times$SE). The title of each chart presents the treatment-effect magnitude and its *p*-value. See the regression results for these outcomes in Table 2.6.

2.1.7 Accounts the User Follows

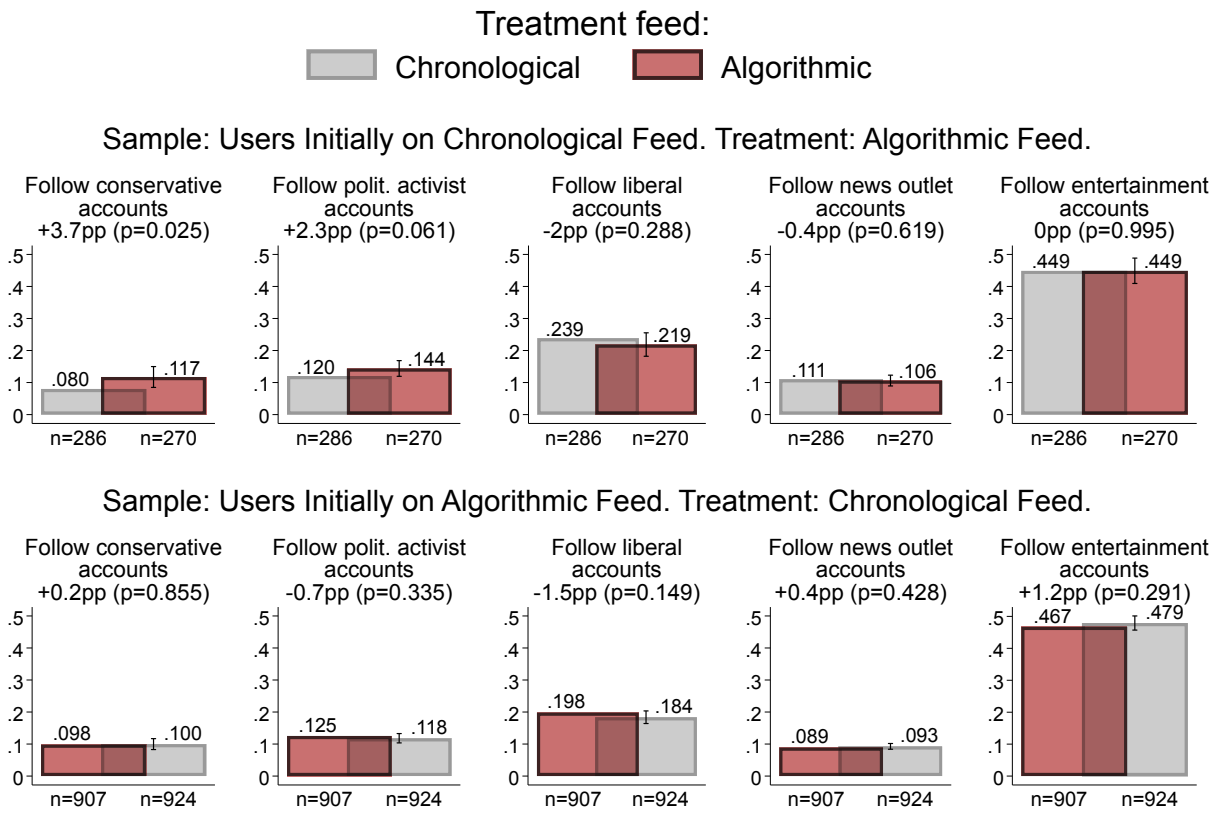


Fig. 2.7. Outcomes: Accounts the User Follows.

Means of characteristics of the accounts the users follow by assigned treatment feed in the two samples: (1) the top panel summarizes outcomes for the sample of users initially on the chronological feed; (2) the bottom panel summarizes outcomes for the sample of participants initially on the algorithmic feed. The difference in the height of the two bars in each chart represents the intention-to-treat (ITT) effect. Error bars show 95% CIs for the treatment effect relative to the control mean (Control mean + ATE $\pm$ 1.96 $\times$ SE). The title of each chart presents the treatment-effect magnitude and its *p*-value. See the regression results for these outcomes in Table 2.8.

2.1.8 Additional outcomes

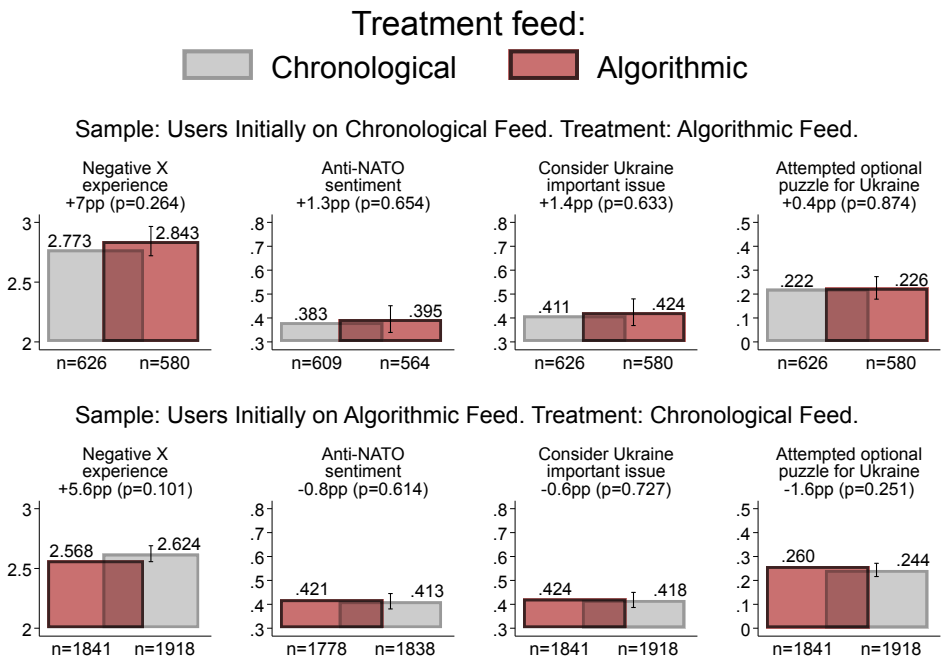


Fig. 2.8. Other Outcomes: Negative 𝕏 experience, Sentiment against NATO, Importance of Ukraine war, Puzzle for Ukraine.

Means for additional outcomes by assigned treatment feed in the two samples: (1) the top panel summarizes outcomes for the sample of users initially on the chronological feed; (2) the bottom panel summarizes outcomes for the sample of participants initially on the algorithmic feed. The difference in the height of the two bars in each chart represents the intention-to-treat (ITT) effect. Error bars show 95% CIs for the treatment effect relative to the control mean (Control mean + ATE $\pm$ 1.96 $\times$ SE). The title of each chart presents the treatment-effect magnitude and its *p*-value.

2.2 Methodological details

The main results reported in the article are intention-to-treat (ITT) and LATE estimates. We explain our ITT and LATE specifications in the Methods Section. In this subsection, we provide more details on the Generalized Random Forests, the methodology used for the choice of covariates in our ITT and LATE estimates (Section 2.2.1) and the reweighing of observations using inverse probability weights based on logistic regressions, which we use in order to make out sample resemble the representative sample of ANES Twitter users and to address section into initial feed setting (Section 2.2.2).

2.2.1 Estimates via Generalized Random Forests

OLS regressions with controls are standard practice but come with two limitations. First, the researcher chooses the relevant controls. Second, the specifications impose restrictive functional form assumptions. In particular, they impose that covariates have a linear effect on outcomes and that treatment effects are constant. To flexibly model treatment effect heterogeneity, we rely on Generalized Random Forests (52). In all our coefficient plots (whether in the main text, the Extended Data, or the Supplementary Information), we report ITTs (or LATEs) conditional on controls estimated using Generalized Random Forests (GRFs).

GRFs combine machine-learning models (in this case, random forests) with semi-parametric inference. In doing so, they provide a flexible framework for estimating heterogeneous treatment effects while retaining desirable statistical properties such as consistency and asymptotic normality. Importantly, the framework leverages orthogonality in its estimating equations to ensure that target parameters like the Average Treatment Effect can be estimated efficiently, achieving $\sqrt{n}$-rate convergence under appropriate regularity conditions, despite the nuisance components (i.e., the propensity score and outcome regression) being estimated non-parametrically via random forests.

Denote W the treatment assignment indicator. To estimate the CATE $\Delta(x) = \mathbb{E}[\,Y(1) - Y(0) \mid X = x\,]$, GRFs adopt a partially linear model:

$$Y_i = \Delta(X_i)\,W_i + f(X_i) + \varepsilon_i, \quad \mathbb{E}[\varepsilon_i \mid X_i, W_i] = 0,$$

where $f(X_i)$ captures the baseline response. This model can be transformed via residualisation:

$$Y_i - m(X_i) = \Delta(X_i)(W_i - e(X_i)) + \varepsilon_i,$$

where $m(x) = \mathbb{E}[Y \mid X = x]$ is the regression model and $e(x) = \mathbb{E}[W \mid X = x]$ is the propensity score. The GRF algorithm proceeds by estimating $m(x)$ and $e(x)$ and then regressing residualized outcomes on residualized treatments within a localized neighborhood, effectively solving:

$$\Delta(x) := \arg\min_{\Delta} \sum_{i=1}^{n} \alpha_i(x) \left((Y_i - \hat{m}^{(-i)}(X_i)) - \Delta(W_i - \hat{e}^{(-i)}(X_i)) \right)^2.$$

The weights $\alpha_i(x)$ play a critical role in localizing estimation. GRFs adapt Breiman's random forest algorithm to treatment effect estimation by constructing trees that split on covariates to maximise treatment effect heterogeneity (rather than variance in outcomes). For a target point x, the forest defines a set of adaptive neighborhoods: a training point i receives a weight $\alpha_i(x)$ proportional to how often X_i appears in the same leaf as x across the ensemble. These weights guide the localized regression used to estimate $\Delta(x)$ and can be interpreted as an adaptive kernel derived from the data.

While $\hat{\Delta}(x)$ provides local treatment effect estimates, GRF supports the estimation of global summaries (e.g., ATE, best linear projection) using doubly robust scores. The Augmented Inverse Probability Weighted (AIPW) estimator:

$$\hat{\Delta}_{\text{AIPW}} = \frac{1}{n} \sum_{i=1}^{n} \left(\Delta(X_i) + \frac{W_i - e(X_i)}{e(X_i)(1 - e(X_i))} \left(Y_i - \mu(X_i, W_i) \right) \right),$$

combines a model-based CATE estimate $\Delta(X_i)$ with a correction term involving the residual and the debiasing weight. This construction ensures consistency if either the propensity score $e(X_i)$ or the outcome model $\mu(X_i, W_i)$ is correct, granting the estimator its doubly robust property. These scores, denoted Γ_i, can be used to compute precise summaries with root-n convergence rates and valid confidence intervals, making them essential for inference with GRFs.

This framework was also extended to population weights and instrumental variable strategies, and was implemented in the R package `grf`. In our context, the treatment assignment is being assigned to the other feed setting relative to the initial setting (i.e., switching from Algorithm to Chrono, or from Chrono to Algorithm). We denote this variable *Switch*. It takes value 1 if the user switched, and zero otherwise. Given our experimental setup, we know that the propensity score is 50% and we do not need to estimate it. We consider the following pre-treatment covariates to model the regression model: previous feed setting, gender, age, $\mathbb{X}$ use indicators, educational attainment, race, $\mathbb{X}$ use frequency, $\mathbb{X}$ posting frequency, self-declared partisanship, the device used to answer the pre-treatment survey (e.g., laptop, smartphone, etc.), life dissatisfaction, unhappiness, and affective polarisation. In total, these represent 32 covariates. We grow causal

forests with the seed 1234, default parameters, and automatically tune hyperparameters for better performance (tune.parameters = "all"). We are mainly interested in estimating $\mathbb{E}[Y(\textit{Switch} = 1) - Y(\textit{Switch} = 0) \mid \textit{InitialAlgo}_i = 0]$ and $\mathbb{E}[Y(\textit{Switch} = 1) - Y(\textit{Switch} = 0) \mid \textit{InitialAlgo}_i = 1]$, which have the same interpretation as β_1 and β_2 in Equation (1). We also report additional heterogeneity by pre-treatment partisanship in Figures 4 and 2.11.

2.2.2 Population ITT Estimates with Inverse Probability Weights

To demonstrate that our results are not driven by sample selection, we apply weights to our observations in some specifications when estimating the ITT (see Equation (1)). The type of weights depends on the sample. For the main sample of respondents, we construct (i) weights that align the sample with the U.S.-based population of $\mathbb{X}$ users (ANES Twitter users), and (ii) feed-setting-specific weights that align the sample with the population of users on the alternative feed setting. For the subsample of users for whom we observe followed accounts, we similarly construct (i) weights that align the subsample with the U.S.-based $\mathbb{X}$ user population, and (ii) weights that align it with the main sample of respondents. For the subsample of users with feed data, we construct weights that align it with the main sample of respondents.

Let Y_i be an indicator variable equal to zero for observations of the target population and to one for observations of the selected sample and $\mathbf{X_i}$ be a vector of predictors. We estimate the following logistic regression equation:

$$\Pr(Y_i = 1 \mid \mathbf{X}_i) = \frac{1}{1 + \exp(-\mathbf{X_i}'\Gamma)}. \tag{3}$$

We construct weights w_i as the odds ratio of the resulting estimated probabilities:

$$w_i = \frac{1 - \widehat{\Pr}(Y_i = 1 \mid \mathbf{X}_i)}{\widehat{\Pr}(Y_i = 1 \mid \mathbf{X}_i)}. \tag{4}$$

To build weights for the US-based population of $\mathbb{X}$ users, we rely on the ANES Social Media Study conducted between 2020 and 2022 in the United States. The study surveys a representative sample of the US population on their political views and social media activity. In particular, the survey explicitly asked respondents: "Do you currently use Twitter?". 935 out of 5750 respondents answered "yes" and represent our main target population. We pool these observations with our YouGov respondents and predict the probability of being a YouGov respondent. We use age, a binary variable for educational attainment, a binary variable for race, a binary variable for gender, and a categorical variable for partisanship as predictors. The choice of predictors is determined

by the overlap in variables between the ANES survey and our pre-treatment survey (see Section 1.4).

For all other weights, we use all pre-treatment survey responses as predictors: gender, an indicator variable for high education, an indicator variable for race, age, affective polarisation, 𝕏 use and posting frequency, all 𝕏 use purposes, life satisfaction, unhappiness, as well as the category indicators for Republican, Independent, Democrat, or other partisan affiliation.

2.3 Robustness to Specification Choices: Coefficient Plots

This section demonstrates the robustness of the results presented in the main text Figures 2 and 4. In particular, we present the following coefficient plots:

- Fig. 2.9: Intention-to-treat (ITT) estimates for all individual survey-based outcomes, i.e., in addition to the aggregate outcomes shown in the main text;
- Fig. 2.10: Intention-to-treat (ITT) estimates for baseline survey-based outcomes reweighted to resemble the US 𝕏 user population according to the American National Election Studies (ANES) survey;
- Fig. 2.11: Intention-to-treat (ITT) estimates for baseline survey-based outcomes separately for Republicans and Independents (i.e., in addition to the combined results for Republicans and Independents presented in Extended Data Fig. 4);
- Fig. 2.12: Intention-to-treat (ITT) estimates for baseline survey-based outcomes for respondents for whom we can confirm compliance through observation (i.e., those who ran our custom Google Chrome extension);
- Fig. 2.13: Local Average Treatment Effect (LATE) estimates for baseline survey-based outcomes assuming respondents who declared complying “Most of the time” complied 50% of the time;
- Fig. 2.14: Intention-to-treat (ITT) estimates for the accounts users follow reweighted to resemble the full sample of respondents (we observe the followed accounts only for respondents who provided their 𝕏 handle);
- Fig. 2.15: Intention-to-treat (ITT) estimates for the accounts users follow reweighted to resemble the US 𝕏 user population according to the American National Election Studies (ANES) survey;
- Fig. 2.16: Local Average Treatment Effect (LATE) estimates for the accounts users follow;
- Fig. 2.17: Local Average Treatment Effect (LATE) estimates for the accounts users follow, assuming that respondents who reported complying “Always” did so consistently, while those who reported complying “Most of the time” complied 50% of the time;

- Fig. 2.18: Intention-to-treat (ITT) estimates for the accounts users follow separately for Republicans and Independents vs. Democrats;
- Fig. 2.19: Intention-to-treat (ITT) estimates for the accounts users follow for respondents for whom we can confirm compliance through observation (i.e., those who ran our custom Google Chrome extension).

Table 2.16 reports the detailed numerical results for each coefficient plot presented in the paper, including the point estimates, standard errors, 90% and 95% confidence intervals, and the number of observations for each specification..

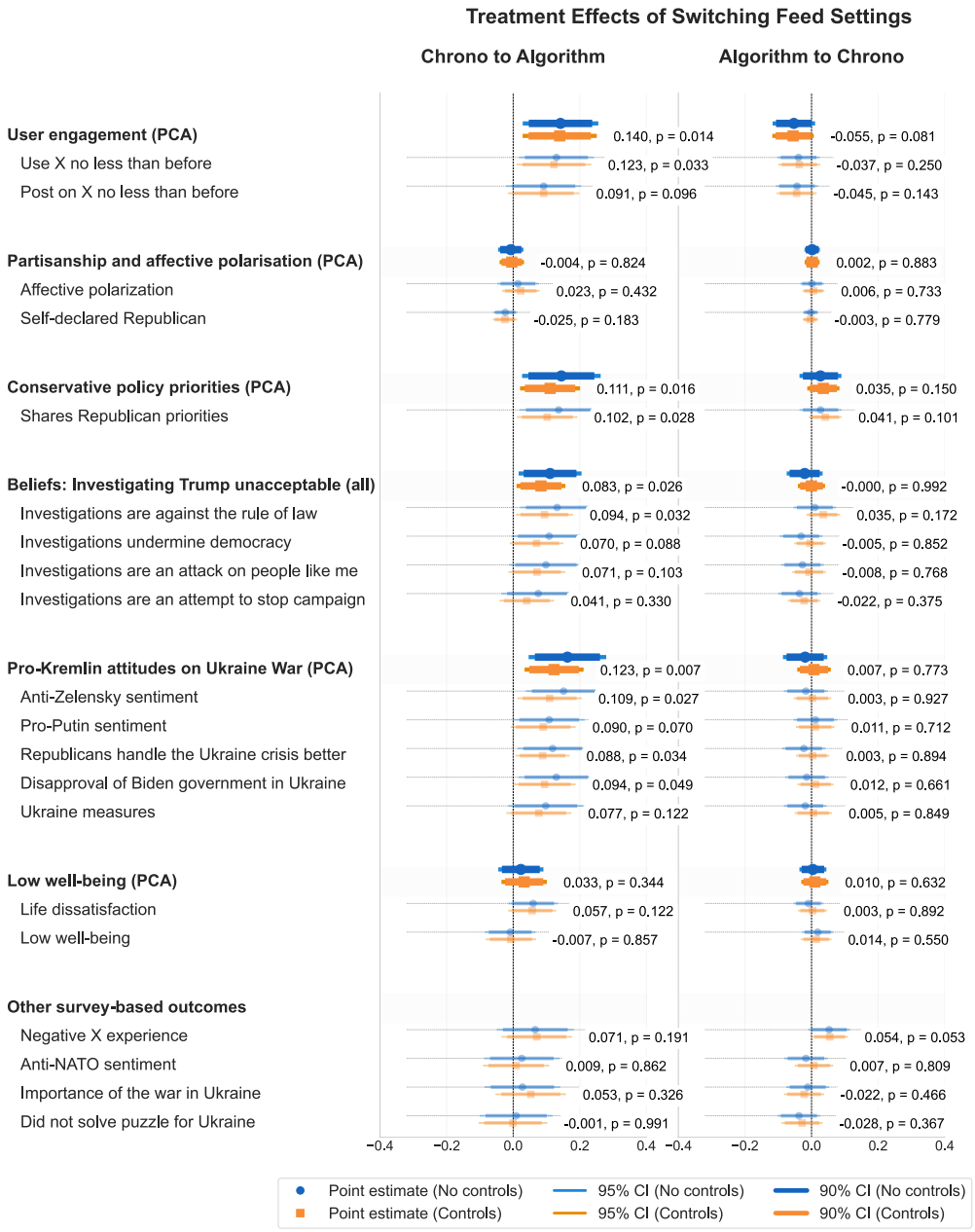


Fig. 2.9. All Survey-Based Outcomes. ITT Estimates by Initial Feed Setting.

Intention-to-treat (ITT) estimates of switching the algorithm on (left panel) and switching it off (right panel). Left panel: the effect of moving from the chronological to the algorithmic feed for users initially on the chronological feed. Right panel: the effect of moving in the opposite direction for users initially on the algorithmic feed. “Chrono” and “Algorithm” refer to the chronological and algorithmic feeds, respectively. 90% and 95% CIs are reported. For each outcome, the results of two specifications are reported. In blue: the unconditional estimates with robust standard errors, controlling only for the initial feed setting and, where applicable, pre-treatment outcome levels. In orange: the conditional estimates, controlling for pre-treatment covariates using Generalized Random Forests (GRFs). All outcomes are standardized.

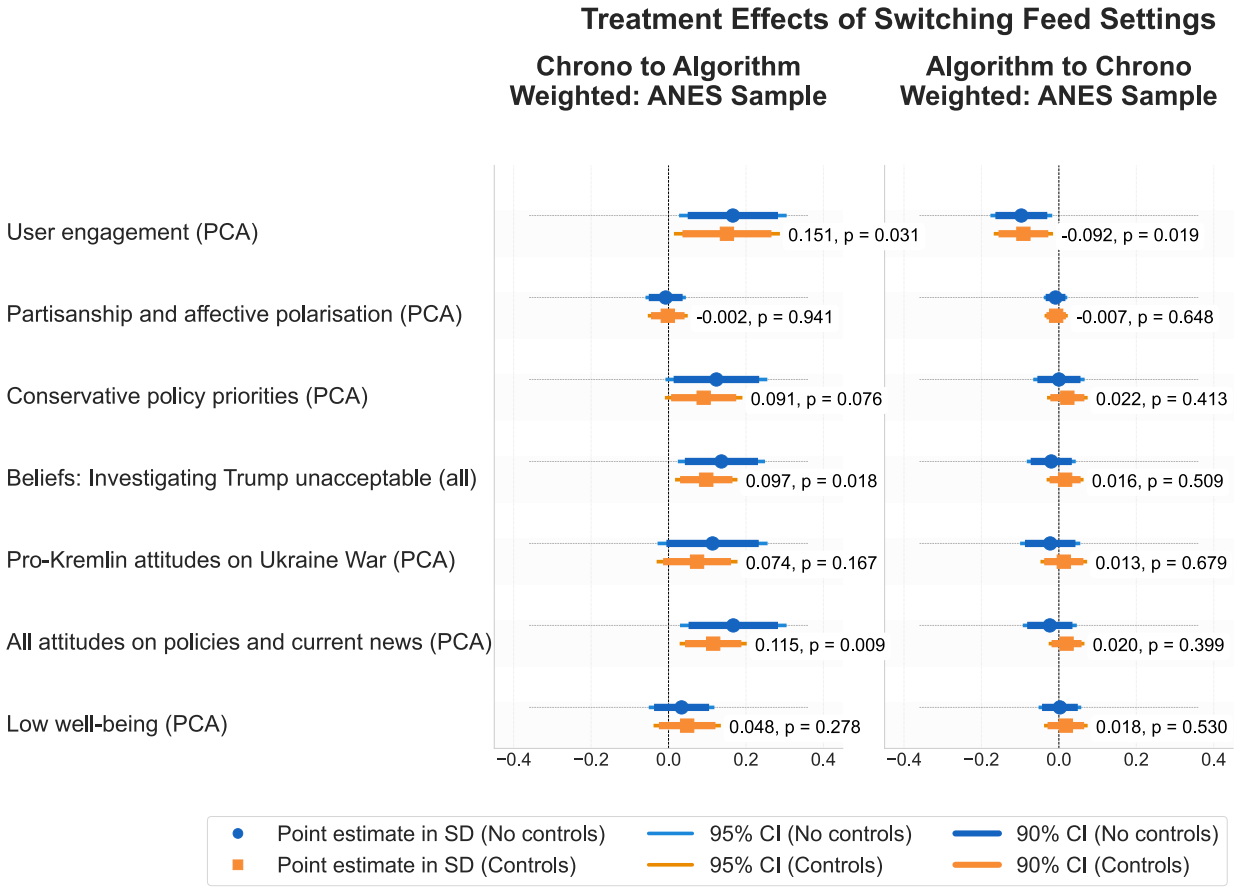


Fig. 2.10. Engagement and Political Attitudes. ITT Estimates by Initial Feed Setting: Reweighted for Twitter/𝕏 population (ANES).

Replication of main-text Fig. 2 with respondents re-weighted to mimic the population of Twitter/𝕏 between 2020 and 2022 (see SI Section 2.2.2 for how we compute the weights). Intention-to-treat (ITT) estimates. Left panel: ITT estimates of switching the algorithm on. Right panel: ITT estimates of switching the algorithm off. 90% and 95% CIs. For each outcome, the results of two specifications are reported. In blue: the unconditional estimates with robust standard errors, controlling only for the initial feed setting and, where applicable, pre-treatment outcome levels. In orange: the conditional estimates, controlling for pre-treatment covariates using Generalized Random Forests (GRFs). All outcomes are standardized. “Chrono” and “Algorithm” stand for the chronological and algorithmic feeds, respectively.

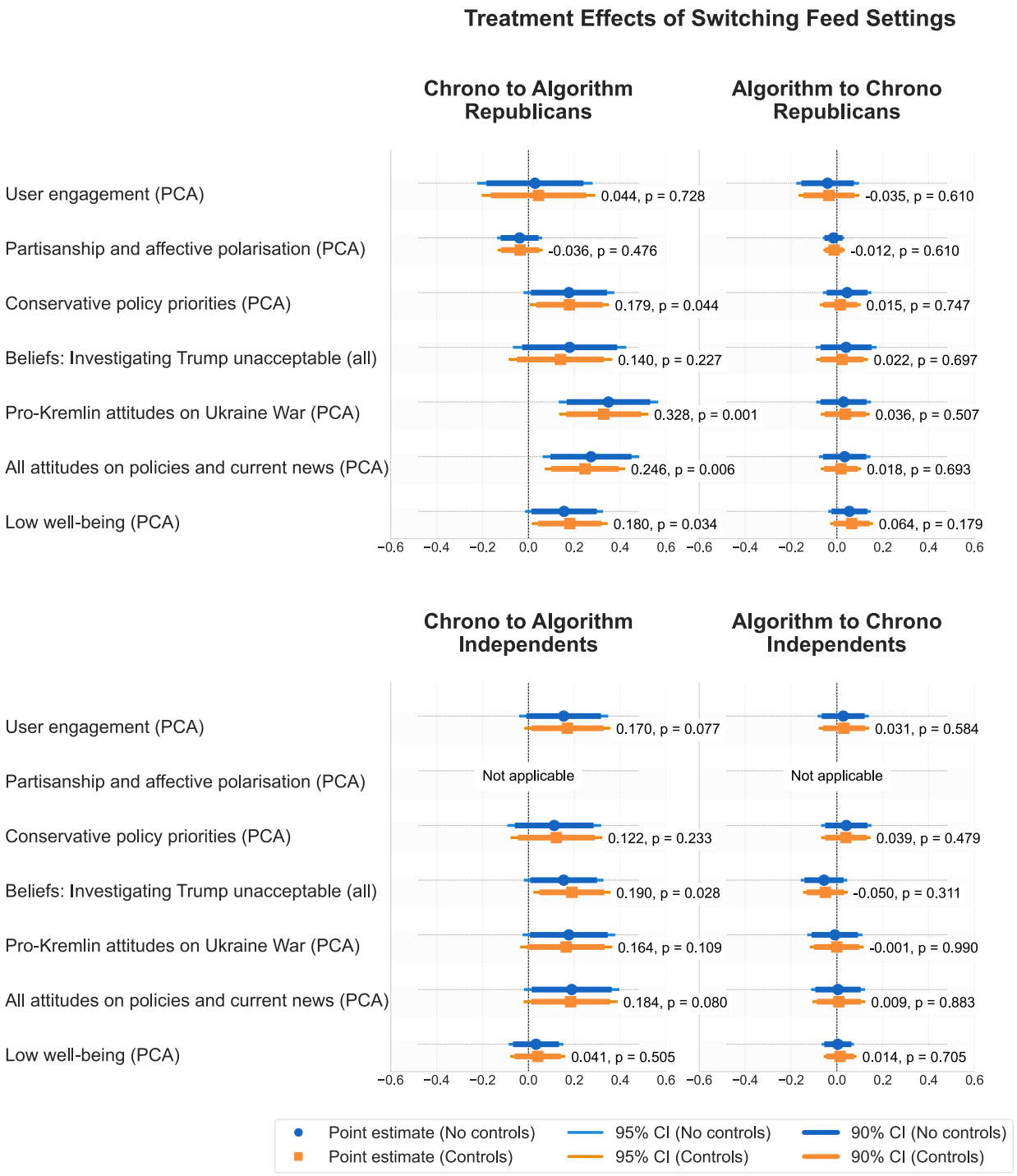


Fig. 2.11. Engagement and Political Attitudes. ITT Estimates by Initial Feed Setting: Republicans vs. Independents.

Intention-to-treat (ITT) estimates separately for (1) self-declared Republicans and (2) self-declared Independents. ITT estimates of switching the algorithm on (left panel) and switching it off (right panel). “Chrono” and “Algorithm” refer to the chronological and algorithmic feeds, respectively. 90% and 95% CIs are reported. In blue: the unconditional estimates with robust standard errors, controlling only for the initial feed setting and, where applicable, pre-treatment outcome levels. In orange: the conditional estimates, controlling for pre-treatment covariates using Generalized Random Forests (GRFs). Affective polarisation is undefined for Independents. All outcomes are standardized. ITT estimates jointly for Republicans and Independents are shown in Extended Fig. 4 and for the full sample in main-text Fig. 2.

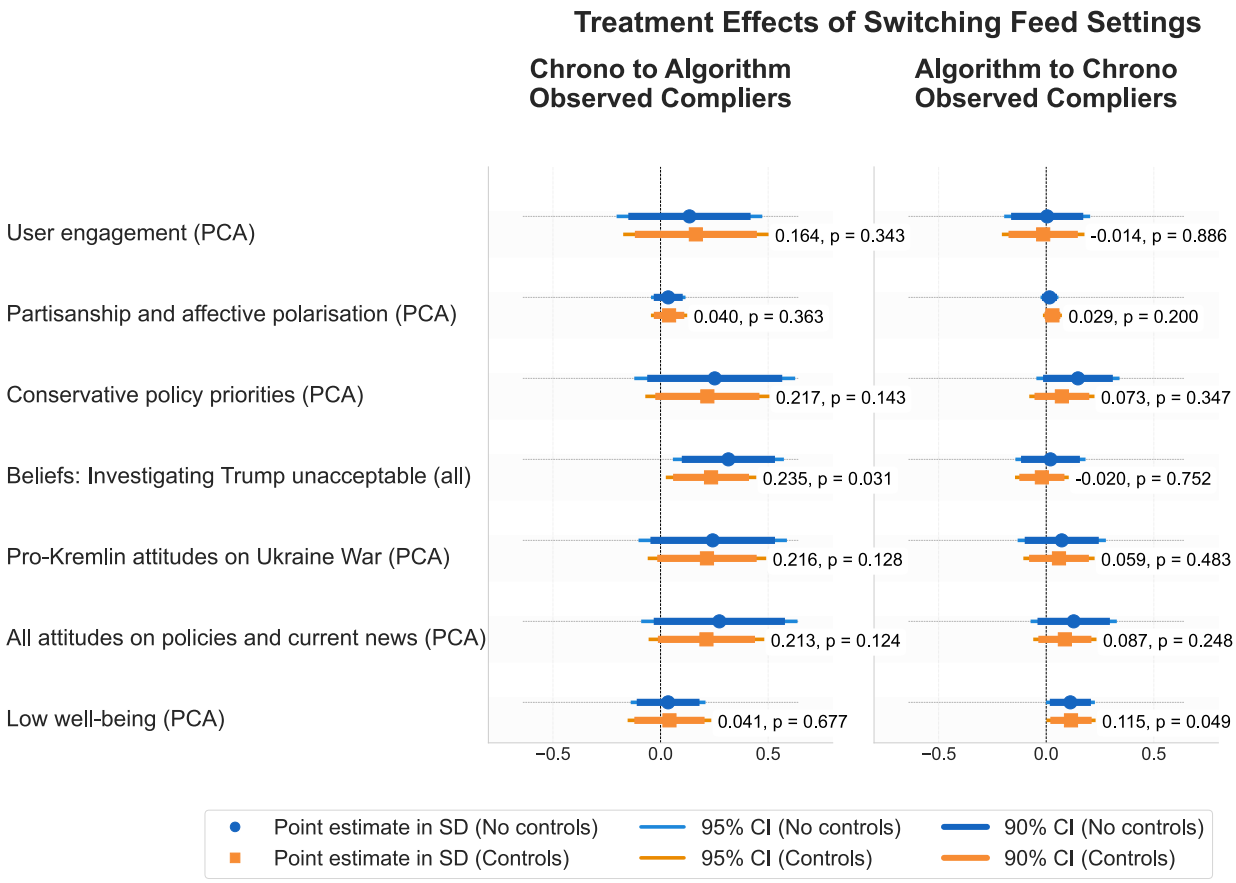


Fig. 2.12. Engagement and Political Attitudes. ITT Estimates by Initial Feed Setting: Observed Compliers.

Replication of main-text Fig. 2 restricted to the sample of users for whom we directly observe compliance (via the Google Chrome extension, n=599). Intention-to-treat (ITT) estimates. Left panel: ITT estimates of switching the algorithm on. Right panel: ITT estimates of switching the algorithm off. 90% and 95% CIs. For each outcome, the results of two specifications are reported. In blue: the unconditional estimates with robust standard errors, controlling only for the initial feed setting and, where applicable, pre-treatment outcome levels. In orange: the conditional estimates, controlling for pre-treatment covariates using Generalized Random Forests (GRFs). All outcomes are standardized. “Chrono” and “Algorithm” stand for the chronological and algorithmic feeds, respectively.

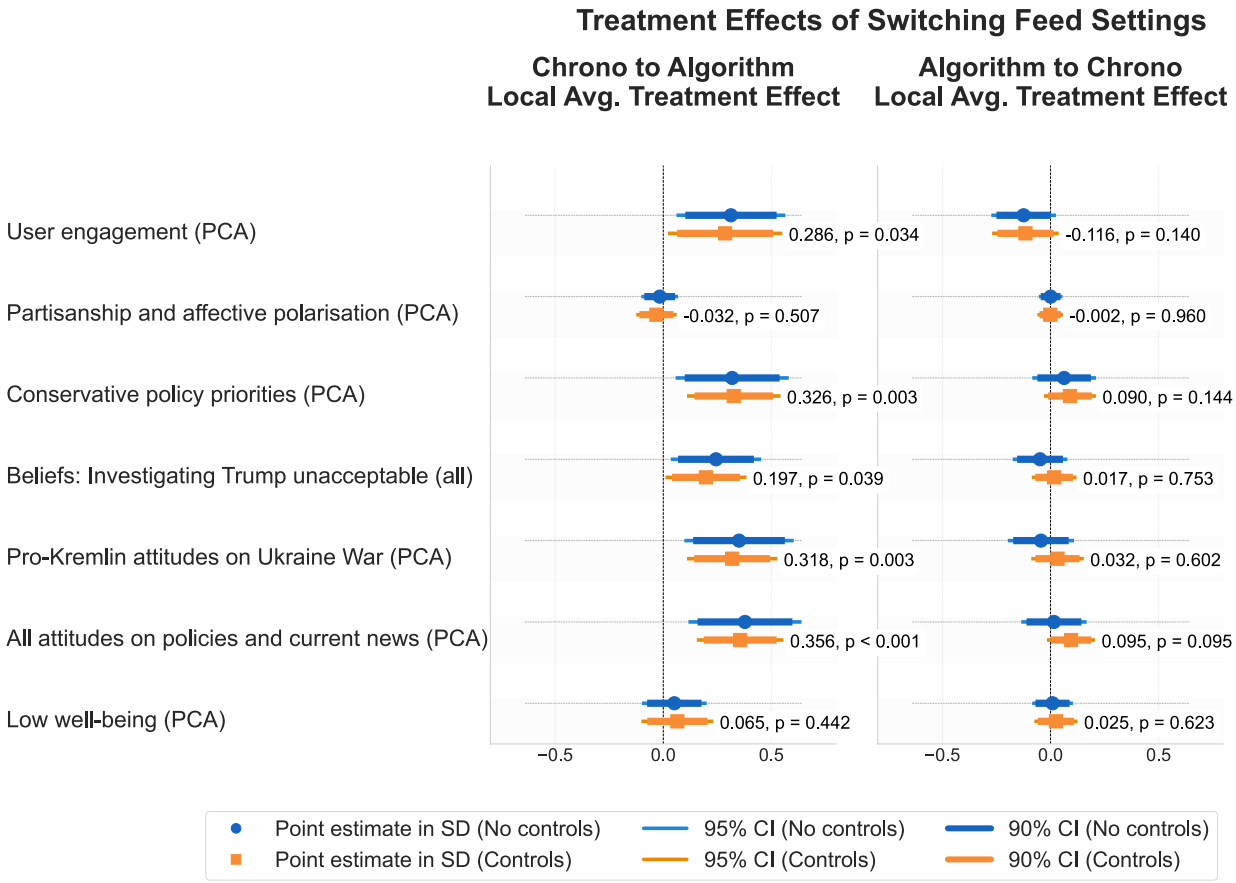


Fig. 2.13. Engagement and Political Attitudes. LATE Estimates by Initial Feed Setting: Robustness.

Replication of Extended Data Fig. 3. Local Average Treatment Effect (LATE) estimates of switching the algorithm on (left panel) and switching it off (right panel). We assume that respondents who reported complying "Always" did so consistently, while those who reported complying "Most of the time" complied 50% of the time. Left panel: LATE estimates of switching the algorithm on. Right panel: LATE estimates of switching the algorithm off. 90% and 95% CIs. For each outcome, the results of two specifications are reported. In blue: the unconditional estimates with robust standard errors, controlling only for the initial feed setting and, where applicable, pre-treatment outcome levels. In orange: the conditional estimates, controlling for pre-treatment covariates using Generalized Random Forests (GRFs). All outcomes are standardized. "Chrono" and "Algorithm" stand for the chronological and algorithmic feeds, respectively.

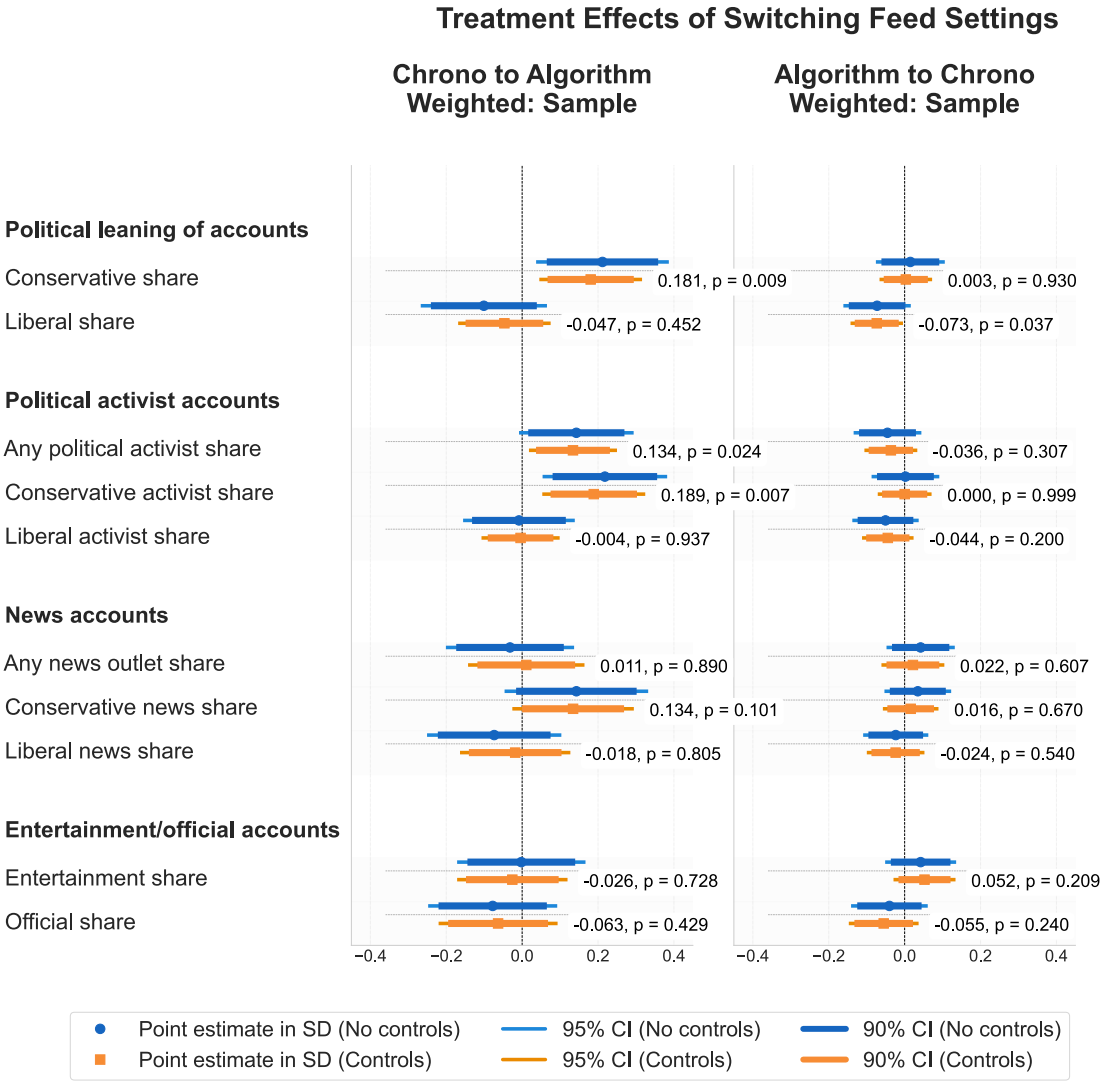


Fig. 2.14. Followed Accounts. ITT Estimates by Initial Feed Setting: Reweighted to Resemble the Full Sample.

Replication of main-text Fig. 4 with respondents re-weighted to mimic the full sample of respondents (see SI Section 2.2.2 for how we compute the weights). Intention-to-treat (ITT) estimates. Left panel: ITT estimates of switching the algorithm on. Right panel: ITT estimates of switching the algorithm off. 90% and 95% CIs. For each outcome, the results of two specifications are reported. In blue: the unconditional estimates with robust standard errors, controlling only for the initial feed setting and, where applicable, pre-treatment outcome levels. In orange: the conditional estimates, controlling for pre-treatment covariates using Generalized Random Forests (GRFs). All outcomes are standardized. “Chrono” and “Algorithm” stand for the chronological and algorithmic feeds, respectively.

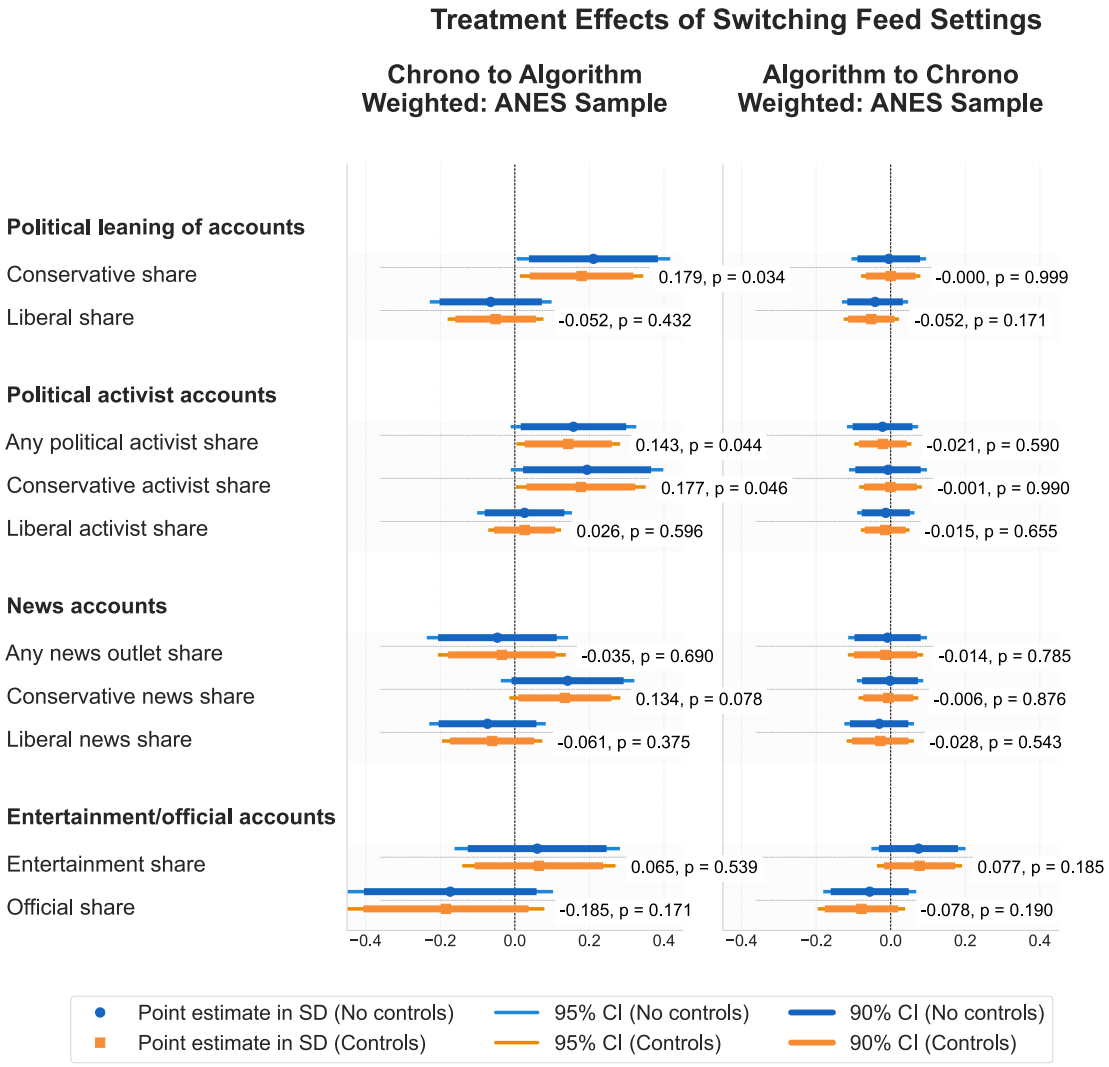


Fig. 2.15. Followed Accounts. ITT Estimates by Initial Feed Setting: Reweighted for Twitter/𝕏 population (ANES).

Replication of main-text Fig. 4 with respondents re-weighted to mimic the population of Twitter/𝕏 between 2020 and 2022 (see SI Section 2.2.2 for how we compute the weights). Intention-to-treat (ITT) estimates. Left panel: ITT estimates of switching the algorithm on. Right panel: ITT estimates of switching the algorithm off. 90% and 95% CIs. For each outcome, the results of two specifications are reported. In blue: the unconditional estimates with robust standard errors, controlling only for the initial feed setting and, where applicable, pre-treatment outcome levels. In orange: the conditional estimates, controlling for pre-treatment covariates using Generalized Random Forests (GRFs). All outcomes are standardized. “Chrono” and “Algorithm” stand for the chronological and algorithmic feeds, respectively.

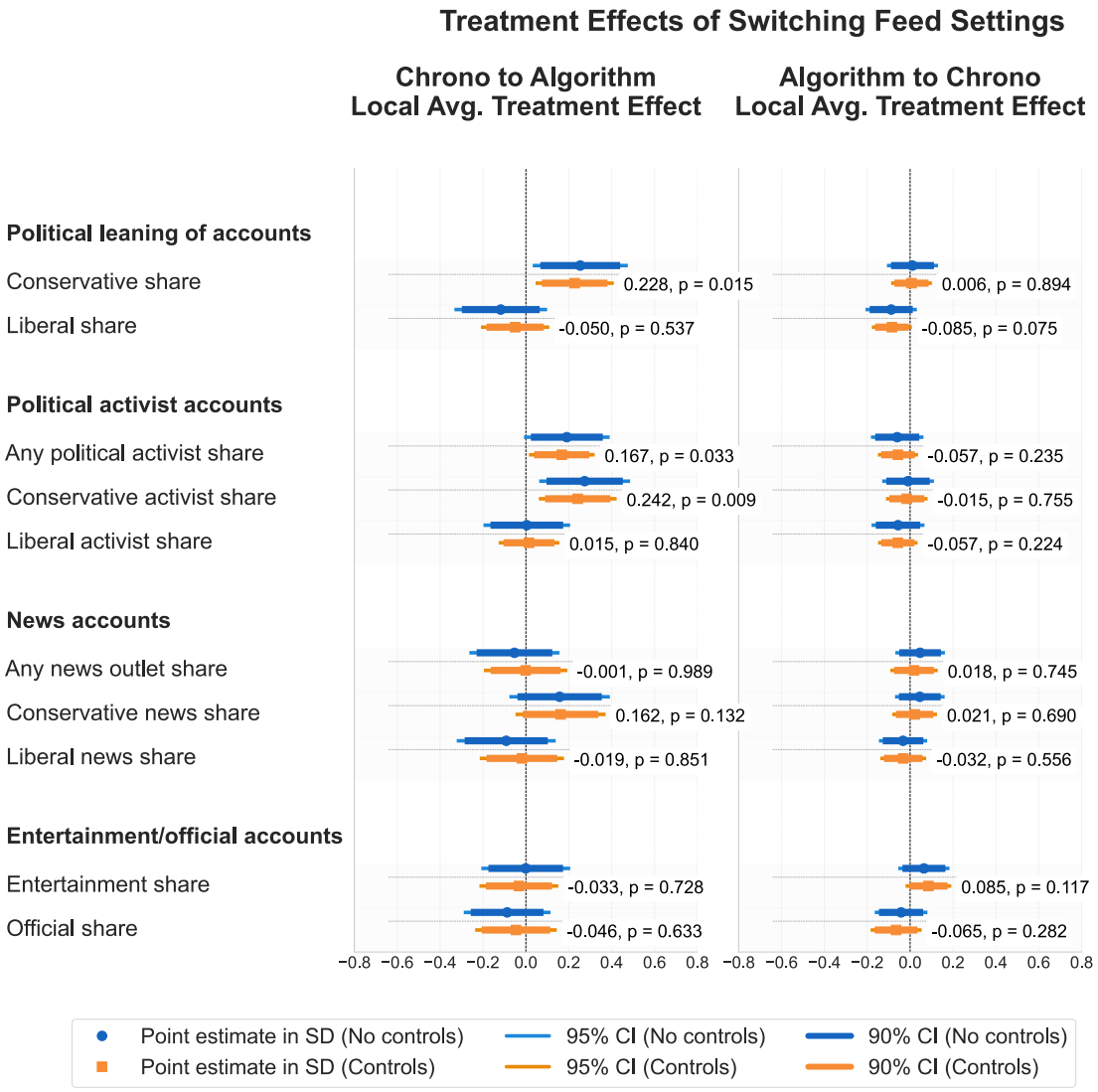


Fig. 2.16. Followed Accounts. LATE Estimates by Initial Feed Setting.

Local Average Treatment Effect (LATE) estimates. We assume respondents who declared complying "Most of the time" and "Always" did so consistently. Left panel: LATE estimates of switching the algorithm on. Right panel: LATE estimates of switching the algorithm off. 90% and 95% CIs. For each outcome, the results of two specifications are reported. In blue: the unconditional estimates with robust standard errors, controlling only for the initial feed setting and, where applicable, pre-treatment outcome levels. In orange: the conditional estimates, controlling for pre-treatment covariates using Generalized Random Forests (GRFs). All outcomes are standardized. "Chrono" and "Algorithm" stand for the chronological and algorithmic feeds, respectively.

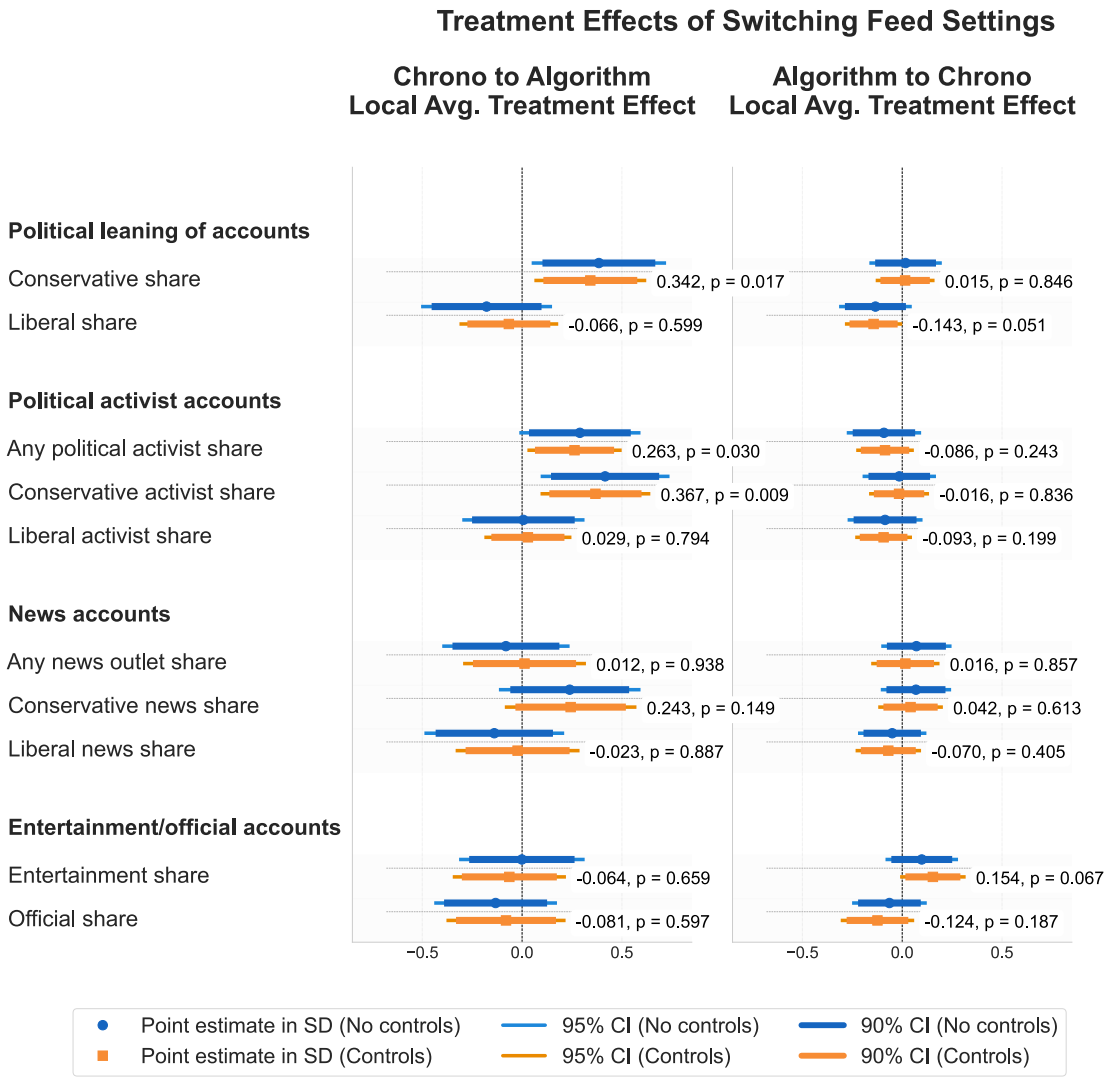


Fig. 2.17. Followed Accounts. LATE Estimates by Initial Feed Setting: Robustness.

Local Average Treatment Effect (LATE) estimates. We assume that respondents who reported complying "Always" did so consistently, while those who reported complying "Most of the time" complied 50% of the time. Left panel: LATE estimates of switching the algorithm on. Right panel: LATE estimates of switching the algorithm off. 90% and 95% CIs. For each outcome, the results of two specifications are reported. In blue: the unconditional estimates with robust standard errors errors, controlling only for the initial feed setting and, where applicable, pre-treatment outcome levels. In orange: the conditional estimates, controlling for pre-treatment covariates using Generalized Random Forests (GRFs). All outcomes are standardized. "Chrono" and "Algorithm" stand for the chronological and algorithmic feeds, respectively.

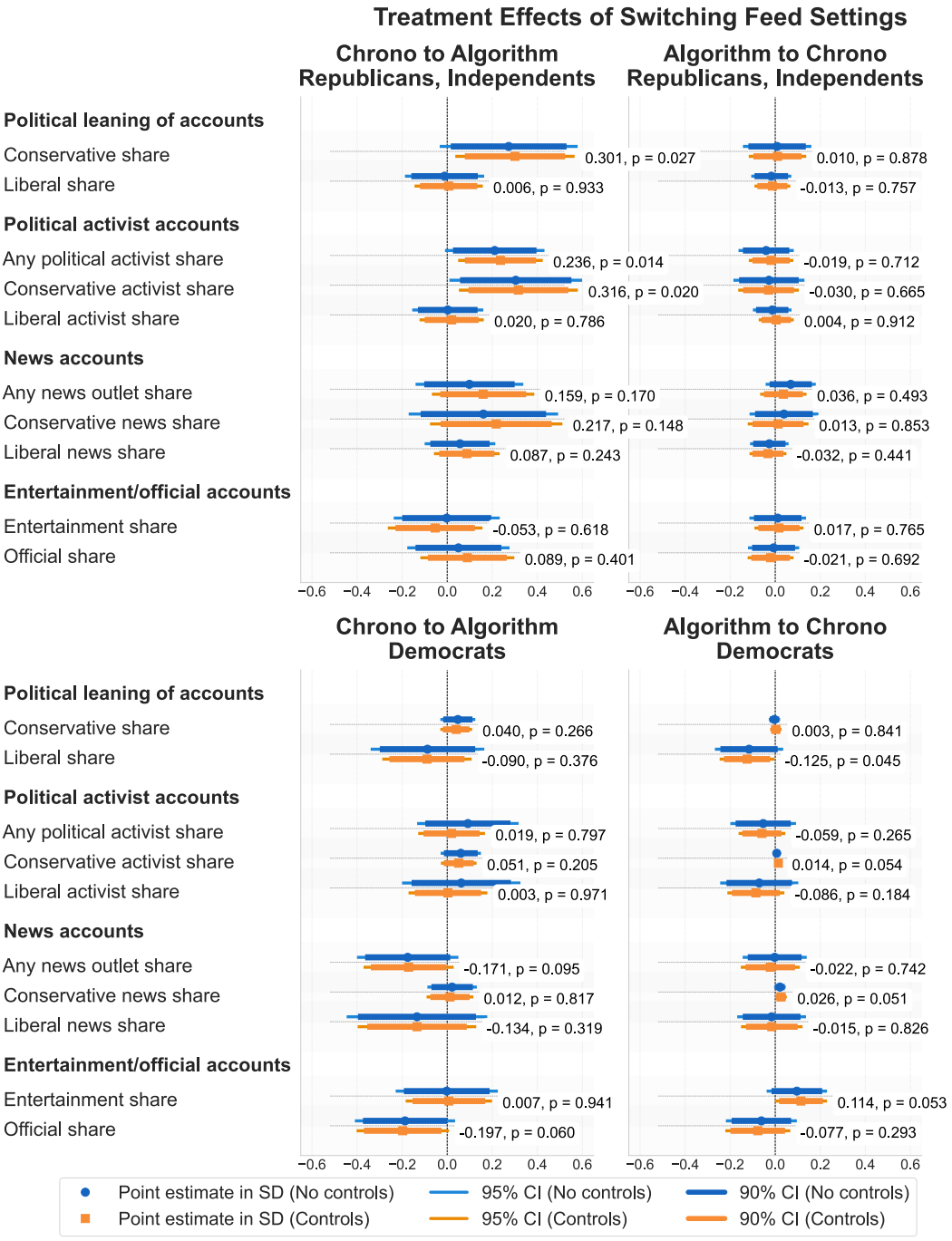


Fig. 2.18. Followed Accounts. ITT Estimates by Initial Feed Setting: Republicans and Independents vs. Democrats.

Replication of main-text Fig. 4 separately in the subsamples of (1) self-declared Republicans and Independents and (2) self-declared Democrats. Left panel: Intention-to-treat (ITT) estimates of switching the algorithm on. Right panel: ITT estimates of switching the algorithm off. “Chrono” and “Algorithm” refer to the chronological and algorithmic feeds, respectively. 90% and 95% CIs are reported. In blue: the unconditional estimates with robust standard errors, controlling only for the initial feed setting and, where applicable, pre-treatment outcome levels. In orange: the conditional estimates, controlling for pre-treatment covariates using Generalized Random Forests (GRFs).

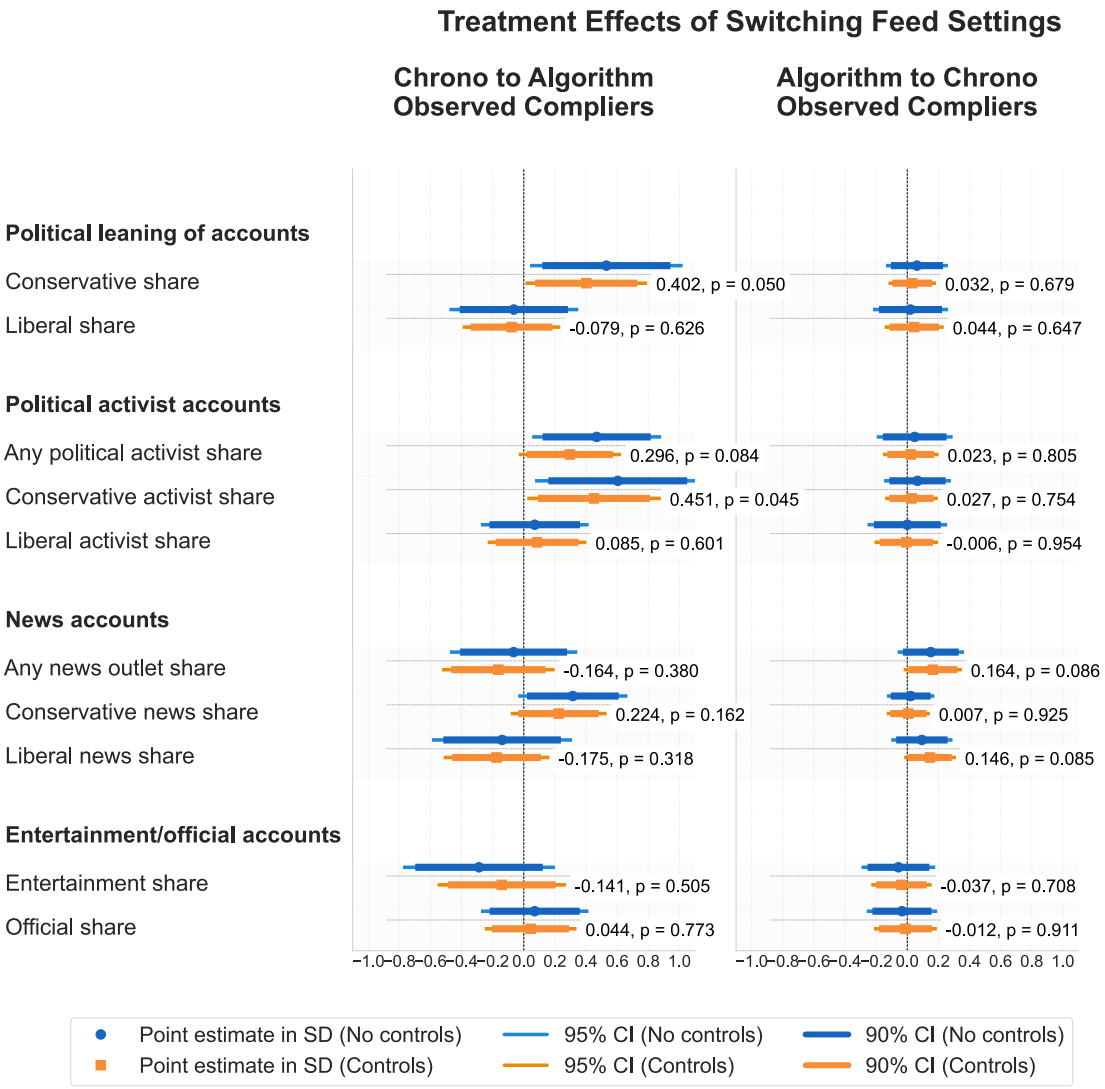


Fig. 2.19. Followed Accounts. ITT Estimates by Initial Feed Setting: Observed Compliers.

Replication of main-text Fig. 4 restricted to the sample of users for whom we directly observe compliance (via the Google Chrome extension, n=599). Intention-to-treat (ITT) estimates. Left panel: ITT estimates of switching the algorithm on. Right panel: ITT estimates of switching the algorithm off. 90% and 95% CIs. For each outcome, the results of two specifications are reported. In blue: the unconditional estimates with robust standard errors, controlling only for the initial feed setting and, where applicable, pre-treatment outcome levels. In orange: the conditional estimates, controlling for pre-treatment covariates using Generalized Random Forests (GRFs). All outcomes are standardized. “Chrono” and “Algorithm” stand for the chronological and algorithmic feeds, respectively.

2.4 Robustness to Linear Specification Choices: Tables for All Outcomes

This section presents detailed results of linear regression analysis for all outcomes in the post-treatment survey, including composite and individual outcomes, as well as for the outcomes related to the characteristics of the accounts the users follow. All tables are structured in the same way as follows:

- For completeness, Column 1 shows the intention-to-treat (ITT) estimate for the algorithmic feed in the full sample, without differentiating the treatment effects by the initial feed setting: $Y_i = \alpha + \beta\ \textit{Treatment Algo}_i + \mathbf{X_i'}\Gamma + \epsilon_i$. This specification forces the ITT estimate β to be the weighted average of β_1 and $-\beta_2$ from Equation (1).
- Column 2 presents the treatment effect estimates separately based on respondents' pre-treatment feed setting, β_1 and β_2 in Equation (1).
- Column 3 replicates column 2, controlling for demographics (gender, age, race, educational attainment), pre-treatment $\mathbb{X}$ use (frequency of use, frequency of posts, purposes for which the respondent uses $\mathbb{X}$), and political party affiliation.
- Column 4 replicates column 3 and additionally controls for the *predicted* initial feed setting; see SI Section 2.5 for details on how we predicted the initial feed setting.
- Column 5 replicates column 3 additionally controlling for post-treatment $\mathbb{X}$ use frequency. While post-treatment $\mathbb{X}$ use is endogenous, robustness to controlling for it indicates that the effects are not driven exclusively by changes in $\mathbb{X}$ use. (We present this specification in all tables with the exception of Table 2.1, where the $\mathbb{X}$ use variables are the outcomes. In this table, column 5 shows the local average treatment effect (LATE) estimates, γ_1 and γ_2 in Equation (2).)
- Column 6 shows the local average treatment effect (LATE) estimates, γ_1 and γ_2 in Equation (2). (Again, this is for all tables except Table 2.1, where the $\mathbb{X}$ use variables are the outcomes. In this table, column 6 shows the LATE estimate with the same controls as in column 3).
- Column 7 shows the LATE estimate with the same controls as in column 3 (in Table 2.1, there is no column 7).

All confidence intervals are computed using robust standard errors. Each panel in each table corresponds to a specific outcome variable. At the bottom of each panel, we report *p*-values for two

tests: (i) a test of the equality of the effects of being on the algorithmic feed between switching it off and turning it on, $\beta_1 = -\beta_2$; and (ii) a test of the equality of outcome levels between respondents who were on the algorithmic feed only during the experiment (those initially on the chronological feed and randomised into the algorithmic feed) and those who were initially on the algorithmic feed and remained on it, $\beta_1 = \beta_3$.

This section presents tables for the following groups of outcomes:

- Table 2.1 for all user engagement outcomes;
- Table 2.2 for affective polarisation and partisanship;
- Table 2.3 for policy priorities;
- Table 2.4 for opinions on investigations against Trump;
- Table 2.5 for attitudes toward the war in Ukraine;
- Table 2.6 for life satisfaction;
- Table 2.7 for other survey-based outcomes;
- Table 2.8 for the composition of the accounts that users follow.

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Dependent variable:	**User engagement compared to pre-treatment 1st component of PCA (positively correlated to user engagement)**					
Treatment Algo	2.640***					
	(0.958)					
β_1: Initial Chrono × Treatment Algo		4.920**	4.735**	4.737**	6.754**	6.500**
		(1.955)	(1.953)	(1.953)	(2.682)	(2.674)
β_2: Initial Algo × Treatment Chrono		-1.893*	-1.940*	-1.942*	-2.704*	-2.769*
		(1.099)	(1.099)	(1.099)	(1.568)	(1.564)
β_3: Initial Algo		3.963**	3.822**	3.839**	5.173***	5.012***
		(1.592)	(1.595)	(1.596)	(1.905)	(1.903)
Predicted Initial Algo				-9.386		
				(38.292)		
Mean dependent variable	72.68	72.68	72.68	72.68	72.68	72.68
P-value $\beta_1 = -\beta_2$		0.18	0.21	0.21	0.19	0.23
P-value $\beta_1 = \beta_3$		0.54	0.56	0.57	0.41	0.43
Panel B. Dependent variable:	**Use frequency compared to pre-treatment Increases or stays the same (indicator)**					
Treatment Algo	0.023**					
	(0.011)					
β_1: Initial Chrono × Treatment Algo		0.053**	0.047**	0.048**	0.072**	0.065**
		(0.023)	(0.023)	(0.023)	(0.032)	(0.031)
β_2: Initial Algo × Treatment Chrono		-0.014	-0.015	-0.015	-0.019	-0.021
		(0.013)	(0.013)	(0.013)	(0.019)	(0.018)
β_3: Initial Algo		0.029	0.029	0.030	0.041*	0.040*
		(0.019)	(0.019)	(0.019)	(0.023)	(0.023)
Predicted Initial Algo				-0.622		
				(0.453)		
Mean dependent variable	0.80	0.80	0.80	0.80	0.80	0.80
P-value ($\beta_1 = -\beta_2$)		0.14	0.22	0.22	0.15	0.23
P-value ($\beta_1 = \beta_3$)		0.20	0.31	0.33	0.16	0.25
Panel C. Dependent variable:	**Posting frequency compared to pre-treatment Increases or stays the same (indicator)**					
Treatment Algo	0.028**					
	(0.013)					
β_1: Initial Chrono × Treatment Algo		0.046*	0.047*	0.047*	0.062*	0.065*
		(0.027)	(0.027)	(0.027)	(0.037)	(0.037)
β_2: Initial Algo × Treatment Chrono		-0.022	-0.025	-0.025	-0.032	-0.035
		(0.015)	(0.015)	(0.015)	(0.022)	(0.021)
β_3: Initial Algo		0.041*	0.050**	0.049**	0.054**	0.063**
		(0.022)	(0.022)	(0.022)	(0.026)	(0.026)
Predicted Initial Algo				0.534		
				(0.528)		
Mean dependent variable	0.64	0.64	0.64	0.64	0.64	0.64
P-value ($\beta_1 = -\beta_2$)		0.45	0.46	0.46	0.47	0.49
P-value ($\beta_1 = \beta_3$)		0.85	0.91	0.94	0.74	0.94
Observations	4965	4965	4965	4965	4965	4965
Outcome Pre-treatment Level Control	✓	✓	✓	✓	✓	✓
Demographic, 𝕏 Use, and Partisan Controls			✓	✓		✓
IV (LATE) Estimate Instead of ITT					✓	✓

Table 2.1. 𝕏 User Engagement

Regression estimates: Col. 1 reports the intent-to-treat effect for the full sample. Col. 2 shows treatment effects by pre-treatment feed setting (β_1, β_2 from eq. 1). Col. 3 adds demographic, platform usage, and partisan controls. Col. 4 includes the predicted feed setting (see SM 2.5). Col. 5 shows LATE estimates (γ_1, γ_2 from eq. 2). Col. 6 presents LATE with controls. All standard errors are robust.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Dependent variable:	**Affective polarization and partisanship**						
	1st component of PCA (positively correlated with Republican partisanship)						
Treatment Algo	-0.098						
	(0.354)						
β_1: Initial Chrono × Treatment Algo		-0.259	-0.197	-0.198	-0.197	-0.362	-0.275
		(0.685)	(0.683)	(0.685)	(0.683)	(0.958)	(0.953)
β_2: Initial Algo × Treatment Chrono		0.066	0.112	0.117	0.112	0.092	0.157
		(0.415)	(0.414)	(0.414)	(0.414)	(0.581)	(0.577)
β_3: Initial Algo		0.384	0.286	0.285	0.287	0.323	0.228
		(0.503)	(0.507)	(0.509)	(0.508)	(0.610)	(0.612)
Predicted Initial Algo					-0.554		
					(18.150)		
Mean dependent variable	34.65	34.65	34.65	34.65	34.65	34.65	34.65
P-value $\beta_1 = -\beta_2$		0.81	0.92	0.92	0.92	0.81	0.92
P-value $\beta_1 = \beta_3$		0.30	0.44	0.44	0.44	0.37	0.51
Observations	3308	3308	3308	3308	3308	3308	3308
Panel B. Dependent variable:	**Affective political polarization**						
	0 = Very warm toward Dems / 100 = Very warm toward Reps						
Treatment Algo	0.067						
	(0.489)						
β_1: Initial Chrono × Treatment Algo		0.445	0.686	0.685	0.680	0.623	0.959
		(0.988)	(0.931)	(0.932)	(0.931)	(1.382)	(1.298)
β_2: Initial Algo × Treatment Chrono		0.047	0.314	0.314	0.314	0.066	0.439
		(0.564)	(0.532)	(0.533)	(0.533)	(0.790)	(0.743)
β_3: Initial Algo		0.063	0.162	0.152	0.170	0.137	0.230
		(0.754)	(0.713)	(0.715)	(0.714)	(0.910)	(0.860)
Predicted Initial Algo					-11.963		
					(26.243)		
Mean dependent variable	37.27	37.27	37.27	37.27	37.27	37.27	37.27
P-value $\beta_1 = -\beta_2$		0.67	0.35	0.35	0.35	0.67	0.35
P-value $\beta_1 = \beta_3$		0.66	0.52	0.52	0.53	0.65	0.46
Observations	3308	3308	3308	3308	3308	3308	3308
Panel C. Dependent variable:	**Partisanship**						
	Self-declared Republican post-treatment (indicator)						
Treatment Algo	-0.002						
	(0.005)						
β_1: Initial Chrono × Treatment Algo		-0.013	-0.013	-0.013	-0.014	-0.018	-0.019
		(0.009)	(0.009)	(0.009)	(0.009)	(0.013)	(0.012)
β_2: Initial Algo × Treatment Chrono		-0.001	-0.002	-0.002	-0.002	-0.002	-0.002
		(0.006)	(0.006)	(0.006)	(0.006)	(0.008)	(0.008)
β_3: Initial Algo		0.007	0.007	0.008	0.005	0.005	0.005
		(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)
Predicted Initial Algo					1.082***		
					(0.221)		
Mean dependent variable	0.31	0.31	0.31	0.31	0.31	0.31	0.31
P-value ($\beta_1 = -\beta_2$)		0.18	0.16	0.16	0.14	0.18	0.16
P-value ($\beta_1 = \beta_3$)		0.02	0.01	0.01	0.02	0.03	0.02
Observations	3371	3371	3371	3371	3371	3371	3371
Outcome Pre-treatment Level Control	✓	✓	✓	✓	✓	✓	✓
Demographic, 𝕏 Use, & Partisan Controls			✓	✓	✓		✓
Post-treatment 𝕏 Use Controls				✓			
IV (LATE) Estimate Instead of ITT						✓	✓

Table 2.2. Affective polarization and partisanship

Regression estimates: Col. 1 reports the intent-to-treat effect for the full sample, col. 2 treatment effects by pre-treatment feed setting (β_1, β_2 from eq. 1). Col. 3 adds demographic, platform usage, and partisan controls. Col. 4 includes the predicted feed setting (see SM 2.5). Col. 5 controls for post-treatment 𝕏 use frequency. Col. 6 and 7 show LATE estimates (γ_1, γ_2 from eq. 2) without and with controls.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Dependent variable:	**Conservative priorities (PCA)**						
	1st component of PCA (positively correlated with Republican partisanship)						
Treatment Algo	0.556						
	(0.919)						
β_1: Initial Chrono × Treatment Algo		4.689**	3.470**	3.394**	3.394**	6.437**	4.762**
		(1.949)	(1.551)	(1.546)	(1.537)	(2.682)	(2.130)
β_2: Initial Algo × Treatment Chrono		0.868	1.705**	1.731**	1.772**	1.240	2.427**
		(1.040)	(0.855)	(0.854)	(0.838)	(1.484)	(1.217)
β_3: Initial Algo		5.350***	2.903**	2.839**	2.212*	5.938***	3.123**
		(1.518)	(1.234)	(1.231)	(1.226)	(1.806)	(1.468)
Predicted Initial Algo					382.404***		
					(30.512)		
Mean dependent variable	46.84	46.84	46.84	46.84	46.84	46.84	46.84
P-value $\beta_1 = -\beta_2$		0.01	0.00	0.00	0.00	0.01	0.00
P-value $\beta_1 = \beta_3$		0.68	0.66	0.66	0.35	0.80	0.29
Panel B. Dependent variable:	**Conservative priorities share**						
	Share of priorities indicated by the respondent that are Republican						
Treatment Algo	0.005						
	(0.010)						
β_1: Initial Chrono × Treatment Algo		0.047**	0.034**	0.033**	0.034**	0.064**	0.047**
		(0.020)	(0.017)	(0.016)	(0.016)	(0.028)	(0.023)
β_2: Initial Algo × Treatment Chrono		0.009	0.018*	0.018**	0.018**	0.013	0.025*
		(0.011)	(0.009)	(0.009)	(0.009)	(0.016)	(0.013)
β_3: Initial Algo		0.055***	0.031**	0.030**	0.024*	0.061***	0.033**
		(0.016)	(0.013)	(0.013)	(0.013)	(0.019)	(0.016)
Predicted Initial Algo					4.028***		
					(0.323)		
Mean dependent variable	0.43	0.43	0.43	0.43	0.43	0.43	0.43
P-value ($\beta_1 = -\beta_2$)		0.02	0.01	0.01	0.01	0.02	0.01
P-value ($\beta_1 = \beta_3$)		0.61	0.81	0.83	0.47	0.88	0.40
Observations	4965	4965	4965	4965	4965	4965	4965
Demographic, 𝕏 Use, & Partisan Controls			✓	✓	✓		✓
Post-treatment 𝕏 Use Controls				✓			
IV (LATE) Estimate Instead of ITT						✓	✓

Table 2.3. Conservative Policy Priorities

Regression estimates: Col. 1 reports the intent-to-treat effect for the full sample, col. 2 treatment effects by pre-treatment feed setting (β_1, β_2 from eq. 1). Col. 3 adds demographic, platform usage, and partisan controls. Col. 4 includes the predicted feed setting (see SM 2.5). Col. 5 controls for post-treatment 𝕏 use frequency. Col. 6 and 7 show LATE estimates (γ_1, γ_2 from eq. 2) without and with controls.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Dependent variable:	**Investigations are unacceptable (all) Always provided a pro-Trump answer (indicator)**						
Treatment Algo	0.021*						
	(0.012)						
β_1: Initial Chrono × Treatment Algo		0.055**	0.043**	0.042**	0.042**	0.076**	0.059**
		(0.024)	(0.019)	(0.019)	(0.019)	(0.033)	(0.026)
β_2: Initial Algo × Treatment Chrono		-0.010	-0.002	-0.001	-0.001	-0.015	-0.002
		(0.014)	(0.011)	(0.011)	(0.011)	(0.020)	(0.016)
β_3: Initial Algo		0.038**	0.020	0.019	0.012	0.049**	0.027
		(0.019)	(0.015)	(0.015)	(0.015)	(0.022)	(0.018)
Predicted Initial Algo					4.255***		
					(0.382)		
Mean dependent variable	0.23	0.23	0.23	0.23	0.23	0.23	0.23
P-value ($\beta_1 = -\beta_2$)		0.10	0.06	0.07	0.06	0.11	0.07
P-value ($\beta_1 = \beta_3$)		0.39	0.16	0.16	0.06	0.28	0.11
Panel B. Dependent variable:	**Investigations against Trump are Against the rule of law (indicator)**						
Treatment Algo	0.012						
	(0.013)						
β_1: Initial Chrono × Treatment Algo		0.061**	0.046**	0.045**	0.046**	0.084**	0.064**
		(0.026)	(0.021)	(0.021)	(0.021)	(0.036)	(0.029)
β_2: Initial Algo × Treatment Chrono		0.005	0.016	0.017	0.017	0.007	0.023
		(0.015)	(0.013)	(0.013)	(0.012)	(0.022)	(0.018)
β_3: Initial Algo		0.044**	0.014	0.014	0.006	0.053**	0.019
		(0.021)	(0.017)	(0.017)	(0.017)	(0.025)	(0.020)
Predicted Initial Algo					4.561***		
					(0.444)		
Mean dependent variable	0.32	0.32	0.32	0.32	0.32	0.32	0.32
P-value ($\beta_1 = -\beta_2$)		0.03	0.01	0.01	0.01	0.03	0.01
P-value ($\beta_1 = \beta_3$)		0.45	0.07	0.08	0.03	0.26	0.04
Panel C. Dependent variable:	**Investigations against Trump Undermine democracy (indicator)**						
Treatment Algo	0.024*						
	(0.013)						
β_1: Initial Chrono × Treatment Algo		0.051*	0.035*	0.033*	0.034*	0.070*	0.047*
		(0.027)	(0.020)	(0.020)	(0.020)	(0.037)	(0.028)
β_2: Initial Algo × Treatment Chrono		-0.015	-0.001	-0.001	-0.001	-0.021	-0.002
		(0.015)	(0.012)	(0.012)	(0.012)	(0.022)	(0.017)
β_3: Initial Algo		0.053**	0.022	0.022	0.013	0.064**	0.028
		(0.021)	(0.016)	(0.016)	(0.016)	(0.025)	(0.019)
Predicted Initial Algo					4.937***		
					(0.438)		
Mean dependent variable	0.34	0.34	0.34	0.34	0.34	0.34	0.34
P-value ($\beta_1 = -\beta_2$)		0.24	0.16	0.17	0.15	0.25	0.16
P-value ($\beta_1 = \beta_3$)		0.94	0.47	0.49	0.23	0.83	0.36
Observations	4965	4965	4965	4965	4965	4965	4965
Demographic, $\mathbb{X}$ Use, & Partisan Controls			✓	✓	✓		✓
Post-treatment $\mathbb{X}$ Use Controls				✓			
IV (LATE) Estimate Instead of ITT						✓	✓

Table 2.4. Attitudes to the Trump Criminal Investigations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel D. Dependent variable:	**Investigations against Trump are Attack on people like me (indicator)**						
Treatment Algo	0.022						
	(0.013)						
β_1: Initial Chrono × Treatment Algo		0.046*	0.033	0.032	0.032	0.064*	0.045
		(0.027)	(0.022)	(0.022)	(0.021)	(0.036)	(0.030)
β_2: Initial Algo × Treatment Chrono		-0.013	-0.003	-0.002	-0.001	-0.018	-0.004
		(0.016)	(0.013)	(0.013)	(0.013)	(0.022)	(0.019)
β_3: Initial Algo		0.072***	0.036**	0.035**	0.025	0.082***	0.042**
		(0.021)	(0.017)	(0.017)	(0.017)	(0.025)	(0.021)
Predicted Initial Algo					5.980***		
					(0.478)		
Mean dependent variable	0.34	0.34	0.34	0.34	0.34	0.34	0.34
P-value ($\beta_1 = -\beta_2$)		0.28	0.23	0.24	0.22	0.29	0.23
P-value ($\beta_1 = \beta_3$)		0.26	0.88	0.88	0.70	0.50	0.87
Panel E. Dependent variable:	**Investigations against Trump are Attempt to stop the campaign (indicator)**						
Treatment Algo	0.024*						
	(0.014)						
β_1: Initial Chrono × Treatment Algo		0.037	0.020	0.019	0.018	0.051	0.027
		(0.028)	(0.022)	(0.022)	(0.022)	(0.038)	(0.030)
β_2: Initial Algo × Treatment Chrono		-0.018	-0.005	-0.005	-0.004	-0.026	-0.007
		(0.016)	(0.013)	(0.013)	(0.013)	(0.023)	(0.019)
β_3: Initial Algo		0.082***	0.041**	0.040**	0.028	0.092***	0.045**
		(0.022)	(0.018)	(0.018)	(0.017)	(0.027)	(0.021)
Predicted Initial Algo					7.210***		
					(0.482)		
Mean dependent variable	0.42	0.42	0.42	0.42	0.42	0.42	0.42
P-value ($\beta_1 = -\beta_2$)		0.55	0.56	0.58	0.56	0.57	0.57
P-value ($\beta_1 = \beta_3$)		0.06	0.26	0.26	0.61	0.16	0.43
Observations	4965	4965	4965	4965	4965	4965	4965
Demographic, $\mathbb{X}$ Use, & Partisan Controls			✓	✓	✓		✓
Post-treatment $\mathbb{X}$ Use Controls				✓			
IV (LATE) Estimate Instead of ITT						✓	✓

Table 2.4 End: Attitudes to the Trump Criminal Investigations

Regression estimates: Col. 1 reports the intent-to-treat effect for the full sample, col. 2 treatment effects by pre-treatment feed setting (β_1, β_2 from eq. 1). Col. 3 adds demographic, platform usage, and partisan controls. Col. 4 includes the predicted feed setting (see SM 2.5). Col. 5 controls for post-treatment $\mathbb{X}$ use frequency. Col. 6 and 7 show LATE estimates (γ_1, γ_2 from eq. 2) without and with controls.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel B. Dependent variable:	**Pro-Kremlin attitudes on Ukraine War**						
	1st component of PCA (positively correlated with Republican partisanship)						
Treatment Algo	1.766*						
	(0.907)						
β_1: Initial Chrono × Treatment Algo		5.009***	4.063***	4.147***	4.002***	6.823***	5.536***
		(1.827)	(1.439)	(1.442)	(1.413)	(2.498)	(1.964)
β_2: Initial Algo × Treatment Chrono		-0.594	0.491	0.499	0.498	-0.839	0.691
		(1.044)	(0.841)	(0.840)	(0.825)	(1.474)	(1.184)
β_3: Initial Algo		6.476***	3.477***	3.483***	2.839**	7.408***	4.044***
		(1.417)	(1.158)	(1.156)	(1.145)	(1.684)	(1.374)
Predicted Initial Algo					385.811***		
					(29.340)		
Mean dependent variable	39.57	39.57	39.57	39.57	39.57	39.57	39.57
P-value $\beta_1 = -\beta_2$		0.04	0.01	0.01	0.01	0.04	0.01
P-value $\beta_1 = \beta_3$		0.34	0.63	0.58	0.32	0.76	0.31
Observations	4596	4596	4596	4596	4596	4596	4596
Panel B. Dependent variable:	**Anti-Zelensky sentiment**						
	Unfavorable sentiment (indicator)						
Treatment Algo	0.025*						
	(0.014)						
β_1: Initial Chrono × Treatment Algo		0.074***	0.059**	0.060**	0.058**	0.102***	0.082**
		(0.028)	(0.025)	(0.025)	(0.025)	(0.039)	(0.034)
β_2: Initial Algo × Treatment Chrono		-0.008	0.005	0.005	0.005	-0.012	0.007
		(0.017)	(0.015)	(0.015)	(0.015)	(0.023)	(0.021)
β_3: Initial Algo		0.076***	0.035*	0.034*	0.027	0.089***	0.043*
		(0.023)	(0.020)	(0.020)	(0.020)	(0.027)	(0.024)
Predicted Initial Algo					4.551***		
					(0.516)		
Mean dependent variable	0.40	0.40	0.40	0.40	0.40	0.40	0.40
P-value ($\beta_1 = -\beta_2$)		0.05	0.03	0.02	0.03	0.05	0.03
P-value ($\beta_1 = \beta_3$)		0.95	0.23	0.21	0.13	0.67	0.13
Observations	4700	4700	4700	4700	4700	4700	4700
Panel A. Dependent variable:	**Pro-Putin sentiment**						
	Favorable sentiment (indicator)						
Treatment Algo	0.009						
	(0.013)						
β_1: Initial Chrono × Treatment Algo		0.048**	0.039*	0.040*	0.038*	0.065**	0.053*
		(0.024)	(0.023)	(0.023)	(0.022)	(0.033)	(0.031)
β_2: Initial Algo × Treatment Chrono		0.005	0.010	0.010	0.010	0.007	0.014
		(0.015)	(0.014)	(0.014)	(0.014)	(0.021)	(0.019)
β_3: Initial Algo		0.084***	0.045**	0.046**	0.039**	0.091***	0.050**
		(0.019)	(0.018)	(0.018)	(0.018)	(0.023)	(0.021)
Predicted Initial Algo					4.005***		
					(0.489)		
Mean dependent variable	0.27	0.27	0.27	0.27	0.27	0.27	0.27
P-value ($\beta_1 = -\beta_2$)		0.06	0.07	0.06	0.07	0.06	0.07
P-value ($\beta_1 = \beta_3$)		0.08	0.73	0.79	0.95	0.31	0.90
Observations	4873	4873	4873	4873	4873	4873	4873
Demographic, $\mathbb{X}$ Use, & Partisan Controls			✓	✓	✓		✓
Post-treatment $\mathbb{X}$ Use Controls				✓			
IV (LATE) Estimate Instead of ITT						✓	✓

Table 2.5. Attitudes to the War in Ukraine

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel C. Dependent variable:	**Who handles Ukraine crisis better**						
	Republicans handle the crisis better (indicator)						
Treatment Algo	0.020*						
	(0.011)						
β_1: Initial Chrono × Treatment Algo		0.048**	0.037**	0.037**	0.036**	0.065**	0.050**
		(0.022)	(0.017)	(0.017)	(0.017)	(0.030)	(0.024)
β_2: Initial Algo × Treatment Chrono		-0.009	0.001	0.001	0.001	-0.013	0.001
		(0.013)	(0.011)	(0.011)	(0.011)	(0.019)	(0.015)
β_3: Initial Algo		0.075***	0.046***	0.046***	0.040***	0.084***	0.052***
		(0.017)	(0.014)	(0.014)	(0.014)	(0.020)	(0.016)
Predicted Initial Algo					2.956***		
					(0.366)		
Mean dependent variable	0.20	0.20	0.20	0.20	0.20	0.20	0.20
P-value ($\beta_1 = -\beta_2$)		0.13	0.07	0.06	0.07	0.14	0.07
P-value ($\beta_1 = \beta_3$)		0.15	0.55	0.57	0.77	0.41	0.94
Observations	4965	4965	4965	4965	4965	4965	4965
Panel D. Dependent variable:	**Disapproval of US government in Ukraine**						
	Disapproval of Biden policy in Ukraine (indicator)						
Treatment Algo	0.022						
	(0.014)						
β_1: Initial Chrono × Treatment Algo		0.065**	0.057**	0.059**	0.056**	0.089**	0.078**
		(0.029)	(0.025)	(0.025)	(0.024)	(0.040)	(0.034)
β_2: Initial Algo × Treatment Chrono		-0.007	0.009	0.009	0.010	-0.010	0.013
		(0.017)	(0.014)	(0.014)	(0.014)	(0.024)	(0.020)
β_3: Initial Algo		0.078***	0.043**	0.043**	0.035*	0.090***	0.050**
		(0.023)	(0.020)	(0.020)	(0.020)	(0.027)	(0.024)
Predicted Initial Algo					4.613***		
					(0.507)		
Mean dependent variable	0.47	0.47	0.47	0.47	0.47	0.47	0.47
P-value ($\beta_1 = -\beta_2$)		0.09	0.02	0.02	0.02	0.09	0.02
P-value ($\beta_1 = \beta_3$)		0.57	0.49	0.44	0.29	0.96	0.26
Observations	4807	4807	4807	4807	4807	4807	4807
Panel F. Dependent variable:	**Republican measures for Ukraine**						
	1st component of PCA (positively correlated with Republican partisanship)						
Treatment Algo	1.204						
	(0.857)						
β_1: Initial Chrono × Treatment Algo		2.948*	2.615*	2.621*	2.554*	4.047*	3.589*
		(1.735)	(1.526)	(1.531)	(1.514)	(2.384)	(2.092)
β_2: Initial Algo × Treatment Chrono		-0.572	0.249	0.249	0.302	-0.817	0.352
		(0.986)	(0.892)	(0.893)	(0.884)	(1.407)	(1.270)
β_3: Initial Algo		4.273***	2.172*	2.170*	1.620	4.878***	2.554*
		(1.341)	(1.211)	(1.212)	(1.205)	(1.598)	(1.440)
Predicted Initial Algo					305.811***		
					(30.475)		
Mean dependent variable	53.20	53.20	53.20	53.20	53.20	53.20	53.20
P-value ($\beta_1 = -\beta_2$)		0.23	0.10	0.11	0.10	0.24	0.11
P-value ($\beta_1 = \beta_3$)		0.37	0.73	0.73	0.47	0.64	0.51
Observations	4965	4965	4965	4965	4965	4965	4965
Demographic, $\mathbb{X}$ Use, & Partisan Controls			✓	✓	✓		✓
Post-treatment $\mathbb{X}$ Use Controls				✓			
IV (LATE) Estimate Instead of ITT						✓	✓

Table 2.5 End: Attitudes to the War in Ukraine

Regression estimates: Col. 1 reports the intent-to-treat effect for the full sample, col. 2 treatment effects by pre-treatment feed setting (β_1, β_2 from eq. 1). Col. 3 adds demographic, platform usage, and partisan controls. Col. 4 includes the predicted feed setting (see SM 2.5). Col. 5 controls for post-treatment $\mathbb{X}$ use frequency. Col. 6 and 7 show LATE estimates (γ_1, γ_2 from eq. 2) without and with controls.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Dependent variable:	**Low well-being**						
	1st component of PCA (positively correlated with low well-being)						
Treatment Algo	0.042						
	(0.364)						
β_1: Initial Chrono × Treatment Algo		0.474	0.519	0.586	0.514	0.652	0.714
		(0.712)	(0.710)	(0.712)	(0.710)	(0.979)	(0.974)
β_2: Initial Algo × Treatment Chrono		0.083	0.191	0.195	0.191	0.118	0.269
		(0.425)	(0.422)	(0.422)	(0.422)	(0.602)	(0.596)
β_3: Initial Algo		-0.339	-0.627	-0.607	-0.643	-0.277	-0.580
		(0.563)	(0.566)	(0.566)	(0.566)	(0.671)	(0.673)
Predicted Initial Algo					11.794		
					(19.431)		
Mean dependent variable	35.26	35.26	35.26	35.26	35.26	35.26	35.26
P-value $\beta_1 = -\beta_2$		0.50	0.39	0.35	0.39	0.50	0.39
P-value $\beta_1 = \beta_3$		0.17	0.06	0.05	0.05	0.20	0.08
Observations	4850	4850	4850	4850	4850	4850	4850
Panel A. Dependent variable:	**Life dissatisfaction**						
	1 = Best possible life / 10 = Worst possible life						
Treatment Algo	0.046						
	(0.041)						
β_1: Initial Chrono × Treatment Algo		0.127	0.134*	0.138*	0.136*	0.175	0.185*
		(0.081)	(0.081)	(0.081)	(0.081)	(0.111)	(0.111)
β_2: Initial Algo × Treatment Chrono		-0.020	-0.005	-0.005	-0.005	-0.028	-0.007
		(0.048)	(0.047)	(0.047)	(0.047)	(0.068)	(0.067)
β_3: Initial Algo		0.078	0.044	0.044	0.050	0.103	0.067
		(0.063)	(0.063)	(0.063)	(0.063)	(0.075)	(0.075)
Predicted Initial Algo					-3.504*		
					(1.858)		
Mean dependent variable	4.61	4.61	4.61	4.61	4.61	4.61	4.61
P-value ($\beta_1 = -\beta_2$)		0.25	0.17	0.15	0.16	0.26	0.17
P-value ($\beta_1 = \beta_3$)		0.47	0.18	0.17	0.21	0.39	0.16
Observations	4911	4911	4911	4911	4911	4911	4911
Panel B. Dependent variable:	**Unhappiness**						
	1 = Very happy / 4 = Not happy at all						
Treatment Algo	-0.012						
	(0.015)						
β_1: Initial Chrono × Treatment Algo		-0.006	-0.005	-0.002	-0.005	-0.009	-0.007
		(0.028)	(0.028)	(0.028)	(0.028)	(0.038)	(0.038)
β_2: Initial Algo × Treatment Chrono		0.013	0.016	0.016	0.016	0.019	0.023
		(0.017)	(0.017)	(0.017)	(0.017)	(0.024)	(0.024)
β_3: Initial Algo		-0.041*	-0.051**	-0.050**	-0.051**	-0.045*	-0.055**
		(0.022)	(0.023)	(0.023)	(0.022)	(0.027)	(0.027)
Predicted Initial Algo					-0.256		
					(0.779)		
Mean dependent variable	2.04	2.04	2.04	2.04	2.04	2.04	2.04
P-value ($\beta_1 = -\beta_2$)		0.83	0.74	0.66	0.74	0.82	0.73
P-value ($\beta_1 = \beta_3$)		0.14	0.05	0.04	0.06	0.21	0.10
Observations	4886	4886	4886	4886	4886	4886	4886
Outcome Pre-treatment Level Control	✓	✓	✓	✓	✓	✓	✓
Demographic, $\mathbb{X}$ Use, & Partisan Controls			✓	✓	✓		✓
Post-treatment $\mathbb{X}$ Use Controls				✓			
IV (LATE) Estimate Instead of ITT						✓	✓

Table 2.6. Life Dissatisfaction

Regression estimates: Col. 1 reports the intent-to-treat effect for the full sample, col. 2 treatment effects by pre-treatment feed setting (β_1, β_2 from eq. 1). Col. 3 adds demographic, platform usage, and partisan controls. Col. 4 includes the predicted feed setting (see SM 2.5). Col. 5 controls for post-treatment $\mathbb{X}$ use frequency. Col. 6 and 7 show LATE estimates (γ_1, γ_2 from eq. 2) without and with controls.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel E. Dependent variable	**Negative 𝕏 experience**						
	1 = Very positive user experience / 5 = Very negative user experience						
Treatment Algo	-0.030						
	(0.030)						
β_1: Initial Chrono × Treatment Algo		0.070	0.086	0.089	0.096	0.118	
		(0.063)	(0.059)	(0.059)	(0.086)	(0.081)	
β_2: Initial Algo × Treatment Chrono		0.056	0.050	0.048	0.080	0.071	
		(0.034)	(0.031)	(0.030)	(0.049)	(0.044)	
β_3: Initial Algo		-0.205***	-0.157***	-0.136***	-0.206***	-0.153***	
		(0.049)	(0.046)	(0.046)	(0.058)	(0.054)	
Predicted Initial Algo				-11.560***			
				(1.169)			
Mean dependent variable	2.65	2.65	2.65	2.65	2.65	2.65	
P-value ($\beta_1 = -\beta_2$)		0.08	0.04	0.04	0.07	0.04	
P-value ($\beta_1 = \beta_3$)		0.00	0.00	0.00	0.00	0.00	
Observations	4965	4965	4965	4965	4965	4965	
Panel E. Dependent variable	**Anti-NATO sentiment**						
	Unfavorable sentiment (indicator)						
Treatment Algo	0.009						
	(0.014)						
β_1: Initial Chrono × Treatment Algo		0.009	0.003	0.003	0.002	0.013	0.004
		(0.028)	(0.025)	(0.025)	(0.025)	(0.039)	(0.035)
β_2: Initial Algo × Treatment Chrono		-0.008	0.005	0.005	0.005	-0.011	0.007
		(0.016)	(0.015)	(0.015)	(0.015)	(0.023)	(0.021)
β_3: Initial Algo		0.036	0.004	0.003	-0.004	0.039	0.003
		(0.023)	(0.020)	(0.020)	(0.020)	(0.027)	(0.024)
Predicted Initial Algo					6.129***		
					(0.597)		
Mean dependent variable	0.41	0.41	0.41	0.41	0.41	0.41	0.41
P-value ($\beta_1 = -\beta_2$)		0.97	0.79	0.80	0.81	0.97	0.78
P-value ($\beta_1 = \beta_3$)		0.26	0.97	0.97	0.78	0.36	0.97
Observations	4789	4789	4789	4789	4789	4789	4789
Demographic, 𝕏 Use, & Partisan Controls			✓	✓	✓		✓
Post-treatment 𝕏 Use Controls				✓			
IV (LATE) Estimate Instead of ITT						✓	✓

Table 2.7. Other survey-based outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel H. Dependent variable:	**Importance of the War in Ukraine War is important (indicator)**						
Treatment Algo	0.008						
	(0.014)						
β_1: Initial Chrono × Treatment Algo		0.016	0.021	0.021	0.022	0.022	0.028
		(0.028)	(0.027)	(0.027)	(0.026)	(0.039)	(0.036)
β_2: Initial Algo × Treatment Chrono		-0.006	-0.015	-0.016	-0.015	-0.009	-0.022
		(0.016)	(0.015)	(0.015)	(0.015)	(0.023)	(0.022)
β_3: Initial Algo		0.008	0.030	0.031	0.037*	0.012	0.037
		(0.023)	(0.021)	(0.021)	(0.021)	(0.027)	(0.025)
Predicted Initial Algo					-4.892***		
					(0.620)		
Mean dependent variable	0.41	0.41	0.41	0.41	0.41	0.41	0.41
P-value ($\beta_1 = -\beta_2$)		0.76	0.86	0.88	0.80	0.77	0.88
P-value ($\beta_1 = \beta_3$)		0.73	0.67	0.64	0.52	0.72	0.76
Observations	4850	4850	4850	4850	4850	4850	4850
Panel H. Dependent variable:	**Puzzle for donations to Ukraine Did not solve the puzzle (indicator)**						
Treatment Algo	0.014						
	(0.012)						
β_1: Initial Chrono × Treatment Algo		0.005	0.009	0.007	0.008	0.006	0.012
		(0.024)	(0.023)	(0.023)	(0.023)	(0.033)	(0.031)
β_2: Initial Algo × Treatment Chrono		-0.016	-0.009	-0.009	-0.009	-0.023	-0.013
		(0.014)	(0.014)	(0.014)	(0.014)	(0.020)	(0.020)
β_3: Initial Algo		0.038*	0.019	0.019	0.017	0.042*	0.023
		(0.019)	(0.019)	(0.019)	(0.019)	(0.023)	(0.022)
Predicted Initial Algo					1.910***		
					(0.594)		
Mean dependent variable	0.25	0.25	0.25	0.25	0.25	0.25	0.25
P-value ($\beta_1 = -\beta_2$)		0.68	0.99	0.93	0.97	0.67	0.98
P-value ($\beta_1 = \beta_3$)		0.10	0.59	0.53	0.65	0.15	0.65
Observations	4965	4965	4965	4965	4965	4965	4965
Demographic, 𝕏 Use, & Partisan Controls			✓	✓	✓		✓
Post-treatment 𝕏 Use Controls				✓			
IV (LATE) Estimate Instead of ITT						✓	✓

Table 2.7 End: Other survey-based outcomes

Regression estimates: Col. 1 reports the intent-to-treat effect for the full sample, col. 2 treatment effects by pre-treatment feed setting (β_1, β_2 from eq. 1). Col. 3 adds demographic, platform usage, and partisan controls. Col. 4 includes the predicted feed setting (see SM 2.5). Col. 5 controls for post-treatment 𝕏 use frequency. Col. 6 and 7 show LATE estimates (γ_1, γ_2 from eq. 2) without and with controls.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Dependent variable:	**Conservative Followings Share of Conservative Followings**						
Treatment Algo	0.008						
	(0.008)						
β_1: Initial Chrono × Treatment Algo		0.038**	0.030**	0.031**	0.030**	0.048**	0.038**
		(0.017)	(0.014)	(0.014)	(0.014)	(0.021)	(0.017)
β_2: Initial Algo × Treatment Chrono		0.001	-0.000	0.000	0.001	0.002	-0.000
		(0.009)	(0.007)	(0.007)	(0.007)	(0.011)	(0.009)
β_3: Initial Algo		0.018	0.011	0.011	0.008	0.022*	0.015
		(0.012)	(0.010)	(0.010)	(0.010)	(0.013)	(0.011)
Predicted Initial Algo					2.004***		
					(0.252)		
Mean dependent variable	0.10	0.10	0.10	0.10	0.10	0.10	0.10
P-value $\beta_1 = -\beta_2$		0.04	0.05	0.04	0.04	0.04	0.05
P-value $\beta_1 = \beta_3$		0.19	0.12	0.11	0.06	0.14	0.09
Panel B. Dependent variable:	**Liberal Followings Share of Liberal Followings**						
Treatment Algo	0.004						
	(0.009)						
β_1: Initial Chrono × Treatment Algo		-0.023	-0.007	-0.008	-0.006	-0.029	-0.008
		(0.018)	(0.014)	(0.014)	(0.014)	(0.023)	(0.018)
β_2: Initial Algo × Treatment Chrono		-0.013	-0.014*	-0.014*	-0.015*	-0.017	-0.018*
		(0.010)	(0.008)	(0.008)	(0.008)	(0.013)	(0.010)
β_3: Initial Algo		-0.042***	-0.009	-0.010	-0.005	-0.042**	-0.008
		(0.015)	(0.012)	(0.012)	(0.012)	(0.016)	(0.013)
Predicted Initial Algo					-2.324***		
					(0.316)		
Mean dependent variable	0.20	0.20	0.20	0.20	0.20	0.20	0.20
P-value $\beta_1 = -\beta_2$		0.08	0.20	0.17	0.18	0.08	0.20
P-value $\beta_1 = \beta_3$		0.19	0.82	0.88	0.92	0.42	0.99
Panel C. Dependent variable:	**Political Activist Followings – All Share of Political Activist Followings**						
Treatment Algo	0.009						
	(0.006)						
β_1: Initial Chrono × Treatment Algo		0.020*	0.021**	0.020**	0.021**	0.026*	0.026**
		(0.011)	(0.010)	(0.010)	(0.010)	(0.014)	(0.013)
β_2: Initial Algo × Treatment Chrono		-0.006	-0.006	-0.006	-0.006	-0.007	-0.008
		(0.007)	(0.006)	(0.006)	(0.006)	(0.009)	(0.008)
β_3: Initial Algo		0.003	0.012	0.011	0.013*	0.006	0.015*
		(0.009)	(0.008)	(0.008)	(0.008)	(0.010)	(0.009)
Predicted Initial Algo					-0.344		
					(0.213)		
Mean dependent variable	0.12	0.12	0.12	0.12	0.12	0.12	0.12
P-value $\beta_1 = -\beta_2$		0.27	0.21	0.24	0.22	0.28	0.22
P-value $\beta_1 = \beta_3$		0.10	0.33	0.35	0.36	0.10	0.28
Observations	2387	2387	2387	2387	2387	2387	2387
Controls for # of Followings	✓	✓	✓	✓	✓	✓	✓
Demographic, 𝕏 Use, & Partisan Controls			✓	✓	✓		✓
Post-treatment 𝕏 Use Controls				✓			
IV (LATE) Estimate Instead of ITT						✓	✓

Table 2.8. Accounts Followed by the User

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel D. Dependent variable:	**Polit. Activist Followings – Conservative Share of Conservative Political Activist Followings**						
Treatment Algo	0.006						
	(0.004)						
β_1: Initial Chrono × Treatment Algo		0.022**	0.018**	0.018**	0.018**	0.028**	0.022**
		(0.009)	(0.008)	(0.008)	(0.008)	(0.011)	(0.010)
β_2: Initial Algo × Treatment Chrono		-0.001	-0.001	-0.001	-0.001	-0.001	-0.002
		(0.005)	(0.004)	(0.004)	(0.004)	(0.006)	(0.005)
β_3: Initial Algo		0.013**	0.008	0.008	0.007	0.015**	0.010*
		(0.006)	(0.005)	(0.005)	(0.005)	(0.007)	(0.006)
Predicted Initial Algo					0.834***		
					(0.143)		
Mean dependent variable	0.05	0.05	0.05	0.05	0.05	0.05	0.05
P-value $\beta_1 = -\beta_2$		0.03	0.06	0.06	0.06	0.03	0.06
P-value $\beta_1 = \beta_3$		0.23	0.18	0.17	0.13	0.16	0.14
Panel E. Dependent variable:	**Polit. Activist Followings – Liberal Share of Liberal Political Activist Followings**						
Treatment Algo	0.002						
	(0.005)						
β_1: Initial Chrono × Treatment Algo		-0.003	0.003	0.002	0.003	-0.003	0.003
		(0.009)	(0.008)	(0.008)	(0.008)	(0.011)	(0.010)
β_2: Initial Algo × Treatment Chrono		-0.004	-0.004	-0.004	-0.004	-0.005	-0.005
		(0.006)	(0.005)	(0.005)	(0.005)	(0.007)	(0.006)
β_3: Initial Algo		-0.009	0.004	0.003	0.006	-0.009	0.005
		(0.007)	(0.006)	(0.006)	(0.006)	(0.008)	(0.007)
Predicted Initial Algo					-1.171***		
					(0.186)		
Mean dependent variable	0.07	0.07	0.07	0.07	0.07	0.07	0.07
P-value $\beta_1 = -\beta_2$		0.53	0.90	0.83	0.86	0.53	0.89
P-value $\beta_1 = \beta_3$		0.42	0.87	0.82	0.65	0.55	0.88
Panel F. Dependent variable:	**News Accounts Followings Share of News Accounts Followings**						
Treatment Algo	-0.000						
	(0.006)						
β_1: Initial Chrono × Treatment Algo		-0.010	-0.015	-0.014	-0.015	-0.012	-0.019
		(0.011)	(0.011)	(0.011)	(0.011)	(0.014)	(0.014)
β_2: Initial Algo × Treatment Chrono		-0.002	-0.001	-0.001	-0.001	-0.003	-0.001
		(0.007)	(0.007)	(0.007)	(0.007)	(0.010)	(0.009)
β_3: Initial Algo		0.007	0.003	0.004	0.002	0.006	0.002
		(0.010)	(0.009)	(0.009)	(0.009)	(0.011)	(0.010)
Predicted Initial Algo					0.549**		
					(0.272)		
Mean dependent variable	0.23	0.23	0.23	0.23	0.23	0.23	0.23
P-value ($\beta_1 = -\beta_2$)		0.38	0.22	0.26	0.23	0.38	0.23
P-value ($\beta_1 = \beta_3$)		0.08	0.06	0.06	0.07	0.09	0.06
Observations	2387	2387	2387	2387	2387	2387	2387
Controls for # of Followings	✓	✓	✓	✓	✓	✓	✓
Demographic, 𝕏 Use, & Partisan Controls			✓	✓	✓		✓
Post-treatment 𝕏 Use Controls				✓			
IV (LATE) Estimate Instead of ITT						✓	✓

Table 2.8 Continued: Accounts Followed by the User

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel G. Dependent variable:	**News Accounts Followings – Conservative Share of Conservative News Accounts Followings**						
Treatment Algo	0.000						
	(0.001)						
β_1: Initial Chrono × Treatment Algo		0.004	0.003	0.004	0.003	0.005	0.004
		(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)
β_2: Initial Algo × Treatment Chrono		0.001	0.001	0.001	0.001	0.001	0.001
		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
β_3: Initial Algo		-0.000	-0.001	-0.001	-0.001	0.000	-0.001
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Predicted Initial Algo					0.255***		
					(0.044)		
Mean dependent variable	0.01	0.01	0.01	0.01	0.01	0.01	0.01
P-value $\beta_1 = -\beta_2$		0.12	0.16	0.13	0.15	0.12	0.16
P-value $\beta_1 = \beta_3$		0.07	0.05	0.04	0.03	0.07	0.05
Panel H. Dependent variable:	**News Accounts Followings – Liberal Share of Liberal News Accounts Followings**						
Treatment Algo	0.000						
	(0.002)						
β_1: Initial Chrono × Treatment Algo		-0.003	-0.000	-0.000	-0.000	-0.004	-0.000
		(0.004)	(0.003)	(0.003)	(0.003)	(0.005)	(0.004)
β_2: Initial Algo × Treatment Chrono		-0.001	-0.002	-0.002	-0.002	-0.001	-0.002
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
β_3: Initial Algo		-0.009***	-0.004	-0.004*	-0.003	-0.010***	-0.004
		(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Predicted Initial Algo					-0.438***		
					(0.073)		
Mean dependent variable	0.02	0.02	0.02	0.02	0.02	0.02	0.02
P-value $\beta_1 = -\beta_2$		0.35	0.62	0.60	0.58	0.35	0.61
P-value $\beta_1 = \beta_3$		0.04	0.15	0.16	0.24	0.10	0.24
Observations	2387	2387	2387	2387	2387	2387	2387
Panel I. Dependent variable:	**Entertainment Accounts Followings Share of Entertain. Accounts Followings**						
Treatment Algo	-0.008						
	(0.010)						
β_1: Initial Chrono × Treatment Algo		0.002	0.002	0.002	0.002	0.002	0.002
		(0.020)	(0.017)	(0.017)	(0.017)	(0.025)	(0.022)
β_2: Initial Algo × Treatment Chrono		0.011	0.015	0.015	0.015	0.014	0.020
		(0.011)	(0.010)	(0.010)	(0.010)	(0.015)	(0.013)
β_3: Initial Algo		0.019	-0.001	-0.001	-0.001	0.018	-0.003
		(0.015)	(0.013)	(0.013)	(0.013)	(0.017)	(0.015)
Predicted Initial Algo					-0.520		
					(0.370)		
Mean dependent variable	0.47	0.47	0.47	0.47	0.47	0.47	0.47
P-value $\beta_1 = -\beta_2$		0.58	0.39	0.39	0.39	0.57	0.38
P-value $\beta_1 = \beta_3$		0.31	0.83	0.82	0.88	0.43	0.75
Observations	2387	2387	2387	2387	2387	2387	2387
Controls for # of Followings	✓	✓	✓	✓	✓	✓	✓
Demographic, 𝕏 Use, & Partisan Controls			✓	✓	✓		✓
Post-treatment 𝕏 Use Controls				✓			
IV (LATE) Estimate Instead of ITT						✓	✓

Table 2.8 Continued: Accounts Followed by the User

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel J. Dependent variable:	**Official Accounts Followings Share of Official Accounts Followings**						
Treatment Algo	0.002						
	(0.005)						
β_1: Initial Chrono × Treatment Algo		-0.007	-0.004	-0.003	-0.004	-0.008	-0.004
		(0.009)	(0.009)	(0.009)	(0.009)	(0.011)	(0.011)
β_2: Initial Algo × Treatment Chrono		-0.004	-0.006	-0.006	-0.005	-0.005	-0.007
		(0.005)	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)
β_3: Initial Algo		-0.003	0.002	0.002	0.001	-0.003	0.003
		(0.008)	(0.007)	(0.008)	(0.008)	(0.009)	(0.008)
Predicted Initial Algo					0.555**		
					(0.219)		
Mean dependent variable	0.10	0.10	0.10	0.10	0.10	0.10	0.10
P-value $\beta_1 = -\beta_2$		0.30	0.36	0.36	0.37	0.30	0.35
P-value $\beta_1 = \beta_3$		0.61	0.39	0.40	0.47	0.53	0.36
Observations	2387	2387	2387	2387	2387	2387	2387
Controls for # of Followings	✓	✓	✓	✓	✓	✓	✓
Demographic, 𝕏 Use, & Partisan Controls			✓	✓	✓		✓
Post-treatment 𝕏 Use Controls				✓			
IV (LATE) Estimate Instead of ITT						✓	✓

Table 2.8 End: Accounts Followed by the User

Regression estimates: Col. 1 reports the intent-to-treat effect for the full sample, col. 2 treatment effects by pre-treatment feed setting (β_1, β_2 from eq. 1). Col. 3 adds demographic, platform usage, and partisan controls. Col. 4 includes the predicted feed setting (see SM 2.5). Col. 5 controls for post-treatment 𝕏 use frequency. Col. 6 and 7 show LATE estimates (γ_1, γ_2 from eq. 2) without and with controls.

2.5 Selection into Initial Feed Setting

The differences in estimated effects of switching the algorithm on and off could potentially be driven by treatment effect heterogeneity, as users who self-selected into the initial chronological feed setting may have different characteristics from those who self-selected into the algorithmic feed setting. To test this, we predict the probability of an initial algorithmic feed setting based on all pre-treatment characteristics using LASSO.[9] In the regression of the initial feed setting on all covariates chosen by the LASSO, the following covariates are statistically significant with $p < 0.05$: female gender, high education, white race, age, frequent $\mathbb{X}$ use, and not using $\mathbb{X}$ for entertainment purposes, as well as a dummy for the Democrat thermometer not having a missing value. All these variables positively predict the initial chronological feed setting.

First, we augment Equation (1) to directly estimate the difference in treatment effects between the two initial feed settings:

$$\begin{aligned} Y_i = \; & \alpha_1 + \xi_1 \left(\textit{Initial Chrono}_i \times \textit{Treatment Algo}_i \right) \\ & + \alpha_2 \textit{Treatment Algo}_i + \alpha_3 \textit{ Initial Algo}_i + \mathbf{X_i'}\Gamma + \epsilon_i. \end{aligned} \tag{5}$$

A comparison of Equations (1) and (5) shows that $\xi_1 = \beta_1 + \beta_2$. Specifically, ξ_1 estimates the difference in the effects of being on the algorithm between those initially on the chronological feed setting and those on the algorithmic feed setting.

To test whether the difference between β_1 and $-\beta_2$ is driven by the characteristics of users who self-select into different initial settings, we estimate two additional equations. We extend the list of covariates in Equation (5) to allow for heterogeneous treatment effects with respect to the predicted initial setting (*Predicted Initial Chrono*):

$$\begin{aligned} Y_i = \; & \alpha_1 + \xi_1 \left(\textit{Initial Chrono}_i \times \textit{Treatment Algo}_i \right) \\ & + \xi_2 \left(\textit{Predicted Initial Chrono}_i \times \textit{Treatment Algo}_i \right) \\ & + \alpha_2 \textit{Treatment Algo}_i + \alpha_3 \textit{Initial Algo}_i \\ & + \alpha_4 \textit{Predicted Initial Chrono}_i + \mathbf{X_i'}\Gamma + \epsilon_i. \end{aligned} \tag{6}$$

Furthermore, we estimate the heterogeneity with respect to the predicted setting without control-

[9] The variables chosen by the LASSO are gender, education, race, age, affective polarisation, $\mathbb{X}$ use and posting frequency, all $\mathbb{X}$ use purposes, life satisfaction, unhappiness, as well as the category indicators for Republican, Independent, or other partisan affiliation.

ling for the actual initial setting interacted with the treatment:

$$\begin{aligned} Y_i = \ & \alpha_1 + \xi_2 \left(\textit{Predicted Initial Chrono}_i \times \textit{Treatment Algo}_i \right) \\ & + \alpha_2 \textit{Treatment Algo}_i + \alpha_4 \textit{Predicted Initial Chrono}_i + \mathbf{X_i'}\Gamma + \epsilon_i. \end{aligned} \tag{7}$$

If the estimates of ξ_2 in Equation (7) have the same sign as ξ_1 in Equation (5), this would suggest that people who self-select into the chronological feed setting rather than the algorithmic feed setting are more likely to respond to being on the algorithm in line with our baseline estimates. This would suggest that the asymmetry in the effects of the algorithm when it is turned on and off is driven by the selection of different type of individuals into one of the two initial settings. Conversely, if ξ_2 in Equation (7) is centred at zero in contrast to ξ_1 from Equation (5), or ξ_2 has the opposite sign to ξ_1, this would imply that selection into the initial feed setting is not what is driving the asymmetry in the results.

In particular, if the ξ_2 estimates are zero, this indicates an absence of treatment heterogeneity with respect to pre-treatment characteristics that predict the initial feed setting. This is precisely what we find.

Fig. 2.20 presents coefficient plots for the estimates of ξ_1 from Equation (5) on the left, ξ_1 from Equation (6) in the middle, and ξ_2 from Equation (7) on the right (for detailed results, see Table 2.9). The estimates of ξ_1 do not depend on the inclusion of control for pre-treatment characteristics that predict selection into initial setting interacted with treatment (the estimates are similar in the left and middle panels), while the estimates of ξ_2 are close to zero (right panel). This confirms that our baseline estimates are not driven by treatment heterogeneity related to selection into the initial feed setting based on pre-treatment characteristics. Overall, we conclude that our main analysis captures the true differences in the effects of switching the algorithm on and off.

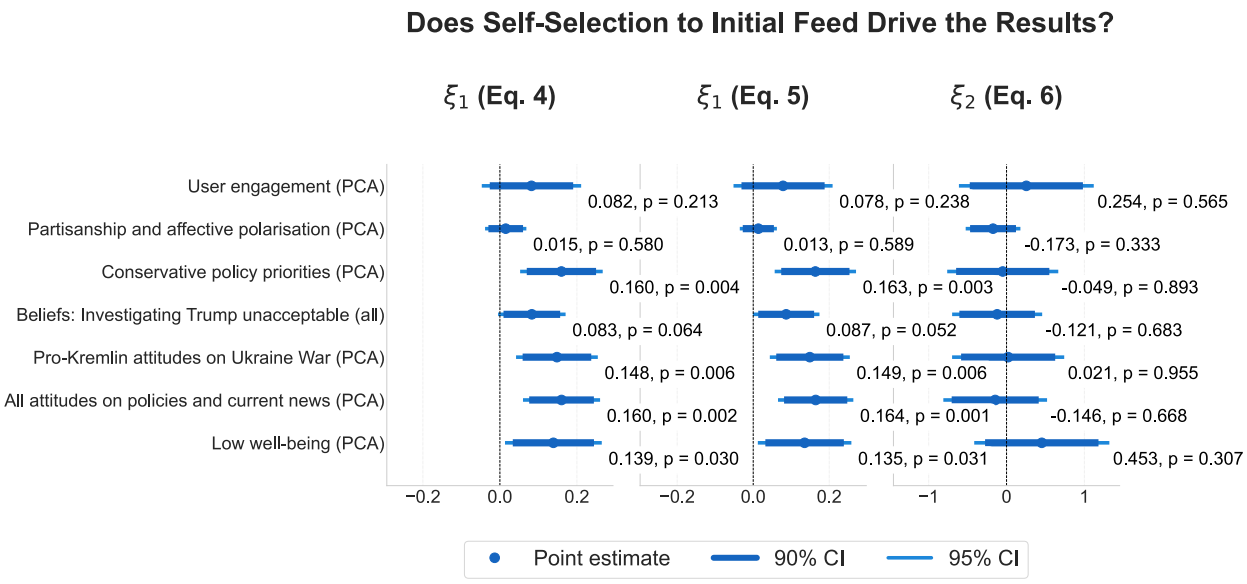


Fig. 2.20. Engagement and Political Attitudes. Checks for Heterogeneity by Initial Feed Setting.

Estimates of ξ_1 (left panel and middle panel) and ξ_2 (right panel) as described in Section 2.5. The left panel estimates the sum of the effects of switching the algorithm on and off for the baseline specification, illustrating the asymmetry of these effects. In the middle panel, we estimate the same sum, but with added controls for the predicted initial setting interacted with the treatment. The right panel presents treatment heterogeneity with respect to the predicted initial setting. Table 2.9 reports the numerical results underlying this coefficient plot.

Outcome	Coef.	95% CI	90% CI	SE	P-value	N
Panel A: ξ_1 (Eq. 5)						
User engagement (PCA)	0.082	[-0.047, 0.211]	[-0.026, 0.190]	0.066	0.213	4965
Partisanship and affective polarisation (PCA)	0.015	[-0.038, 0.069]	[-0.030, 0.060]	0.027	0.580	3337
Conservative policy priorities (PCA)	0.160	[0.053, 0.267]	[0.070, 0.250]	0.055	0.004	4965
Beliefs: Investigating Trump unacceptable (all)	0.083	[-0.005, 0.171]	[0.009, 0.156]	0.045	0.064	4965
Pro-Kremlin attitudes on Ukraine War (PCA)	0.148	[0.042, 0.254]	[0.059, 0.237]	0.054	0.006	4596
All attitudes on policies and current news (PCA)	0.160	[0.060, 0.260]	[0.076, 0.244]	0.051	0.002	4596
Low well-being (PCA)	0.139	[0.013, 0.265]	[0.033, 0.244]	0.064	0.030	4850
Panel B: ξ_1 (Eq. 6)						
User engagement (PCA)	0.078	[-0.052, 0.209]	[-0.031, 0.188]	0.066	0.238	4965
Partisanship and affective polarisation (PCA)	0.013	[-0.035, 0.062]	[-0.027, 0.054]	0.025	0.589	3337
Conservative policy priorities (PCA)	0.163	[0.056, 0.270]	[0.074, 0.253]	0.055	0.003	4965
Beliefs: Investigating Trump unacceptable (all)	0.087	[-0.001, 0.174]	[0.013, 0.160]	0.045	0.052	4965
Pro-Kremlin attitudes on Ukraine War (PCA)	0.149	[0.044, 0.254]	[0.061, 0.237]	0.054	0.006	4596
All attitudes on policies and current news (PCA)	0.164	[0.065, 0.263]	[0.081, 0.247]	0.051	0.001	4596
Low well-being (PCA)	0.135	[0.012, 0.258]	[0.032, 0.238]	0.063	0.031	4850
Panel C: ξ_2 (Eq. 7)						
User engagement (PCA)	0.254	[-0.612, 1.120]	[-0.472, 0.981]	0.442	0.565	4965
Partisanship and affective polarisation (PCA)	-0.173	[-0.525, 0.178]	[-0.468, 0.121]	0.179	0.333	3337
Conservative policy priorities (PCA)	-0.049	[-0.762, 0.664]	[-0.647, 0.550]	0.364	0.893	4965
Beliefs: Investigating Trump unacceptable (all)	-0.121	[-0.699, 0.457]	[-0.606, 0.365]	0.295	0.683	4965
Pro-Kremlin attitudes on Ukraine War (PCA)	0.021	[-0.699, 0.741]	[-0.584, 0.625]	0.367	0.955	4596
All attitudes on policies and current news (PCA)	-0.146	[-0.812, 0.520]	[-0.705, 0.413]	0.340	0.668	4596
Low well-being (PCA)	0.453	[-0.415, 1.320]	[-0.276, 1.181]	0.443	0.307	4850

Table 2.9. Does Self-Selection to Initial Feed Drive the Results? Detailed Results

Detailed estimates of ξ_1 (left panel and middle panel) and ξ_2 (right panel) as described in Section 2.5.

2.6 Content Promoted by the Algorithm

In this section, we utilize data collected via a Google Chrome extension to present regression results on the types of content the algorithm promotes to an average user and verify the robustness of these findings. Fig. 3 in the main text summarizes the content displayed to users on both algorithmic and chronological feeds by presenting simple means for the posts collected under the two feed settings. Here, we provide a formal analysis in support of these results and show which of them are statistically significant.

We estimate the following specification for all measures of content typology using a linear probability model (LPM):

$$Content_{ki} = \nu_i + \omega_t + \Delta\, Algorithm_{ki} + \epsilon_{ki}, \tag{8}$$

where i indexes users and k indexes posts collected via the Google Chrome extension for each user i. $Content_{ki}$ denotes the outcome variable, describing the type of content in the user's feed, for example, an indicator for whether a post originates from an account annotated as conservative (see SI Section 1.5 for details on content annotation). $Algorithm_{ki}$ is an indicator for the feed setting under which respondent i would see post k, equal to 1 if the setting is algorithmic and 0 if chronological. ν_i is a user fixed effect that captures variation related to the user's prior behaviour on the platform, including treatment status. ω_t is a fixed effect indicating whether the post was collected at the end of the pre-treatment or post-treatment survey. ϵ_{ki} is a heteroskedastic error term, clustered by user i.

The coefficient of interest Δ represents the average difference in the type of content respondent i would see under the algorithmic feed relative to the chronological feed.

Engagement variables—likes, reposts, comments—are count data. Accordingly, we estimate a Poisson pseudo-maximum likelihood (PPML) model instead of a linear regression:

$$Y_{ki} = \exp\left(\nu_i + \omega_t + \Delta\, Algorithm_{ki}\right) + \epsilon_{ki}. \tag{9}$$

Δ remains the coefficient of interest, now capturing the percentage change in expected engagement (likes, reposts, or comments) when switching from the chronological to the algorithmic feed.

The tables below present the results of the estimation of Δ in different samples:

- Table 2.10 uses the full sample of all respondents who submitted the feeds collected by the Google Chrome extension.

- Table 2.11 uses the subsample of respondents who were assigned to the algorithmically curated feed.
- Table 2.12 uses the subsample of respondents who were assigned to the chronological feed.
- Table 2.13 weights each data point to resemble the overall sample of respondents (see SI Section 2.2.2 for technical details on how we build the weights).
- Table 2.14 replicates the results in Tables 2.10, 2.11, 2.12, and 2.13 using simpler proxies of content and ideological slant (see SI Sections 1.5.2 and 1.5.3).

In each table, panels indicate groups of outcomes, and columns within each panel have different outcomes. The specification is always the same.

The results are robust to using different samples. Overall, the tables show that the algorithm promotes significantly more engaging content (see panel A in each table). Note that the distribution of engagement metrics is highly skewed: a median post on the algorithmic feed receives 933 likes, 137 reposts, and 56 comments, whereas the median post on the chronological feed receives only 40 likes, 8 reposts, and 3 comments, while, on average, as we discuss in the main text, posts shown on the algorithmic feed receive 13,212 likes, 1,836 reposts, and 722 comments, compared to 2,781 likes, 453 reposts, and 145 comments for posts on the chronological feed. The algorithm also promotes significantly more conservative content but not more liberal content (panel B). Furthermore, the algorithm significantly promotes the content produced by political activists and entertainment accounts but demotes content produced by media outlets (panel C). Among accounts of political activists, those who are conservative-leaning are promoted more ($p = 0.002$), whereas accounts of news media outlets are demoted less when they are conservative ($p = 0.005$; panel D).

In Fig. 2.21, we examine differences between partisan users in the type of content promoted by the algorithm. As noted in the main text, both (1) Democrats and (2) Republicans and Independents are exposed to a higher share of conservative posts among political content under the algorithmic feed, even though co-partisan political content is more prevalent for both groups. Additionally, posts from news organisations are demoted by the algorithm regardless of users' partisan affiliation. However, in the case of political activists, we observe signs of informational polarisation as a result of exposire to the algorithm: Democrats see 4 percentage points more posts from liberal activists (23.2% vs. 19.2%, $p < 0.001$) and only 1.3 percentage points more from conservative

activists (2.2% vs. 0.9%, $p < 0.001$) on the algorithmic feed compared to the chronological feed; whereas Republicans and Independents see 3.9 percentage points more posts from conservative activists (18.1% vs. 14.2%, $p < 0.001$) and 2.4 percentage points more from liberal activists (10.5% vs. 8.1%, $p < 0.001$) on the algorithmic feed compared to the chronological feed. Thus, the algorithm disproportionately promotes co-partisan activists.

Panel A. Dependent variable:	**Engagement (number of ...)**			
	Likes	**Reposts**	**Comments**	**All (sum)**
Δ: Algorithm	1.569***	1.406***	1.626***	1.553***
	(0.099)	(0.071)	(0.121)	(0.095)
Mean dependent variable	8005	1146	434	9548
Observations	264957	267142	265649	264208
Panel B. Dependent variable:	**Political content (indicator)**			
	Conservative		**Liberal**	
Δ: Algorithm	0.029***		0.010**	
	(0.004)		(0.005)	
Mean dependent variable	0.16		0.32	
Observations	268532		268532	
Panel C. Dependent variable:	**Account type (indicator)**			
	Political Activist	**News**	**Entertainment**	**Official**
Δ: Algorithm	0.059***	-0.155***	0.091***	0.004*
	(0.005)	(0.006)	(0.006)	(0.002)
Mean dependent variable	0.24	0.19	0.47	0.05
Observations	268532	268532	268532	268532
Panel D. Dependent variable:	**Partisan account type (indicator)**			
	Conservative		**Liberal**	
	Political Activist	**News**	**Political Activist**	**News**
Δ: Algorithm	0.028***	-0.018***	0.031***	-0.047***
	(0.003)	(0.002)	(0.004)	(0.003)
Mean dependent variable	0.10	0.02	0.15	0.06
Observations	268532	268532	268532	268532
Survey Wave Fixed Effect	✓	✓	✓	✓
User Fixed Effect	✓	✓	✓	✓

Table 2.10. 𝕏 Feed Content, Algorithmic vs. Chronogical Setting.

Descriptive results. Δ captures how the algorithmic feed content compares to the chronological feed content for a given user. See SI Section 2.6 for details on τ. For the engagement variables (top panel), a Poisson pseudo-maximum likelihood model is used. For all others, a linear probability model is estimated.

Panel A. Dependent variable:	**Engagement (number of ...)**			
	Likes	**Reposts**	**Comments**	**All (sum)**
Δ: Algorithm	1.644***	1.445***	1.805***	1.629***
	(0.095)	(0.075)	(0.096)	(0.092)
Mean dependent variable	8150	1171	438	9722
Observations	139910	141112	140279	139478
Panel B. Dependent variable:	**Political content (indicator)**			
	Conservative		**Liberal**	
Δ: Algorithm	0.023***		0.009	
	(0.004)		(0.007)	
Mean dependent variable	0.16		0.33	
Observations	141849		141849	
Panel C. Dependent variable:	**Account type (indicator)**			
	Political Activist	**News**	**Entertainment**	**Official**
Δ: Algorithm	0.059***	-0.161***	0.094***	0.006**
	(0.006)	(0.008)	(0.008)	(0.003)
Mean dependent variable	0.25	0.19	0.47	0.05
Observations	141849	141849	141849	141849
Panel D. Dependent variable:	**Partisan account type (indicator)**			
	Conservative		**Liberal**	
	Political Activist	**News**	**Political Activist**	**News**
Δ: Algorithm	0.027***	-0.022***	0.032***	-0.049***
	(0.004)	(0.003)	(0.005)	(0.005)
Mean dependent variable	0.09	0.02	0.15	0.06
Observations	141849	141849	141849	141849
Survey Wave Fixed Effect	✓	✓	✓	✓
User Fixed Effect	✓	✓	✓	✓

Table 2.11. 𝕏 Feed Content, Algorithmic vs. Chronogical Setting. (Subsample: Algorithmic Treatment).

Descriptive results. Δ captures how the algorithmic feed content compares to the chronological feed content for a given user. See SI Section 2.6 for details on τ. For the engagement variables (top panel), a Poisson pseudo-maximum likelihood model is used. For all others, a linear probability model is estimated. Regressions for the subsample of respondents assigned to the Algo treatment.

Panel A. Dependent variable:	**Engagement (number of ...)**			
	Likes	**Reposts**	**Comments**	**All (sum)**
Δ: Algorithm	1.492***	1.364***	1.454***	1.476***
	(0.169)	(0.122)	(0.203)	(0.162)
Mean dependent variable	7843	1117	430	9353
Observations	125047	126030	125370	124730
Panel B. Dependent variable:	**Political content (indicator)**			
	Conservative		**Liberal**	
Δ: Algorithm	0.035***		0.011	
	(0.006)		(0.007)	
Mean dependent variable	0.16		0.32	
Observations	126683		126683	
Panel C. Dependent variable:	**Account type (indicator)**			
	Political Activist	**News**	**Entertainment**	**Official**
Δ: Algorithm	0.059***	-0.148***	0.087***	0.001
	(0.007)	(0.009)	(0.008)	(0.003)
Mean dependent variable	0.24	0.19	0.47	0.05
Observations	126683	126683	126683	126683
Panel D. Dependent variable:	**Partisan account type (indicator)**			
	Conservative		**Liberal**	
	Political Activist	**News**	**Political Activist**	**News**
Δ: Algorithm	0.030***	-0.014***	0.029***	-0.045***
	(0.005)	(0.003)	(0.005)	(0.005)
Mean dependent variable	0.10	0.02	0.14	0.06
Observations	126683	126683	126683	126683
Survey Wave Fixed Effect	✓	✓	✓	✓
User Fixed Effect	✓	✓	✓	✓

Table 2.12. 𝕏 Feed Content, Algorithmic vs. Chronogical Setting. (Subsample: Chronological Treatment).

Descriptive results. Δ captures how the algorithmic feed content compares to the chronological feed content for a given user. See SI Section 2.6 for details on τ. For the engagement variables (top panel), a Poisson pseudo-maximum likelihood model is used. For all others, a linear probability model is estimated. Regressions for the subsample of respondents assigned to the Chrono treatment.

Panel A. Dependent variable:	**Engagement (number of ...)**			
	Likes	**Reposts**	**Comments**	**All (sum)**
Δ: Algorithm	1.566***	1.380***	1.610***	1.545***
	(0.110)	(0.095)	(0.148)	(0.109)
Mean dependent variable	8005	1146	434	9548
Observations	264957	267142	265649	264208
Panel B. Dependent variable:	**Political content (indicator)**			
	Conservative		**Liberal**	
Δ: Algorithm	0.032***		0.008	
	(0.004)		(0.005)	
Mean dependent variable	0.16		0.32	
Observations	268532		268532	
Panel C. Dependent variable:	**Account type (indicator)**			
	Political Activist	**News**	**Entertainment**	**Official**
Δ: Algorithm	0.058***	-0.149***	0.085***	0.005**
	(0.005)	(0.006)	(0.006)	(0.002)
Mean dependent variable	0.24	0.19	0.47	0.05
Observations	268532	268532	268532	268532
Panel D. Dependent variable:	**Partisan account type (indicator)**			
	Conservative		**Liberal**	
	Political Activist	**News**	**Political Activist**	**News**
Δ: Algorithm	0.030***	-0.018***	0.029***	-0.045***
	(0.003)	(0.002)	(0.004)	(0.003)
Mean dependent variable	0.10	0.02	0.15	0.06
Observations	268532	268532	268532	268532
Survey Wave Fixed Effect	✓	✓	✓	✓
User Fixed Effect	✓	✓	✓	✓

Table 2.13. 𝕏 Feed Content, Algorithmic vs. Chronogical Setting. (Inverse Probability Weights).

Descriptive results. Δ captures how the algorithmic feed content compares to the chronological feed content for a given user. See SI Section 2.6 for details on τ. For the engagement variables (top panel), a Poisson pseudo-maximum likelihood model is used. For all others, a linear probability model is estimated. All tweets are weighted by the inverse of the participant's probability of running the Google Chrome extension needed to collect the newsfeed data (the probability is predicted using all pre-treatment characteristics).

Dependent variable:	Political content (indicator)		Content type (indicator)				Partisan content type (indicator)			
	Conservative	Liberal	Pol. Activist	News	Entertainment	Official	Conservative		Liberal	
							Pol. Activist	News	Pol. Activist	News
Panel A. Full Sample.										
Δ: Algorithm	0.022***	-0.039***	0.033***	-0.108***	0.061***	0.005***	0.025***	-0.046***	0.004	-0.062***
	(0.003)	(0.003)	(0.005)	(0.005)	(0.006)	(0.001)	(0.003)	(0.002)	(0.003)	(0.003)
Mean dependent variable	0.47	0.47	0.32	0.12	0.53	0.01	0.15	0.05	0.16	0.07
Observations	268532	268532	268532	268532	268532	268532	268532	268532	268532	268532
Panel B. Subsample: Algorithmic Treatment.										
Δ: Algorithm	0.019***	-0.034***	0.030***	-0.116***	0.074***	0.005***	0.022***	-0.051***	0.005	-0.064***
	(0.005)	(0.005)	(0.006)	(0.007)	(0.008)	(0.001)	(0.003)	(0.003)	(0.004)	(0.004)
Mean dependent variable	0.47	0.47	0.32	0.12	0.53	0.01	0.15	0.05	0.16	0.07
Observations	141849	141849	141849	141849	141849	141849	141849	141849	141849	141849
Panel C. Subsample: Chronological Treatment.										
Δ: Algorithm	0.025***	-0.045***	0.036***	-0.099***	0.047***	0.005***	0.029***	-0.040***	0.003	-0.059***
	(0.004)	(0.005)	(0.007)	(0.006)	(0.008)	(0.001)	(0.004)	(0.003)	(0.004)	(0.004)
Mean dependent variable	0.47	0.47	0.31	0.11	0.53	0.01	0.15	0.05	0.15	0.06
Observations	126683	126683	126683	126683	126683	126683	126683	126683	126683	126683
Panel D. Inverse Probability Weighting.										
Δ: Algorithm	0.023***	-0.040***	0.034***	-0.105***	0.057***	0.006***	0.026***	-0.046***	0.004	-0.059***
	(0.003)	(0.004)	(0.005)	(0.005)	(0.006)	(0.001)	(0.003)	(0.002)	(0.003)	(0.003)
Mean dependent variable	0.47	0.47	0.32	0.12	0.53	0.01	0.15	0.05	0.16	0.07
Observations	268532	268532	268532	268532	268532	268532	268532	268532	268532	268532
Survey Wave Fixed Effect	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
User Fixed Effect	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 2.14. Replication of Tables S2.9 to S2.12 with Simpler Proxies of Partisanship and Content at the Post-level.

Replication of Tables S2.9 to S2.12. We replace Llama 3 annotations at the account level by measures of partisanship and content based on word frequencies at the post level. The measure of partisanship is based on an ML-classifier trained to predict the party affiliation of posts from US Congress members. The measure of content is based on an LDA topic model, as well as a comprehensive list of the 𝕏 handles of US media outlets and politicians. For further details, please refer to SI Sections 1.5.2 and 1.5.3.

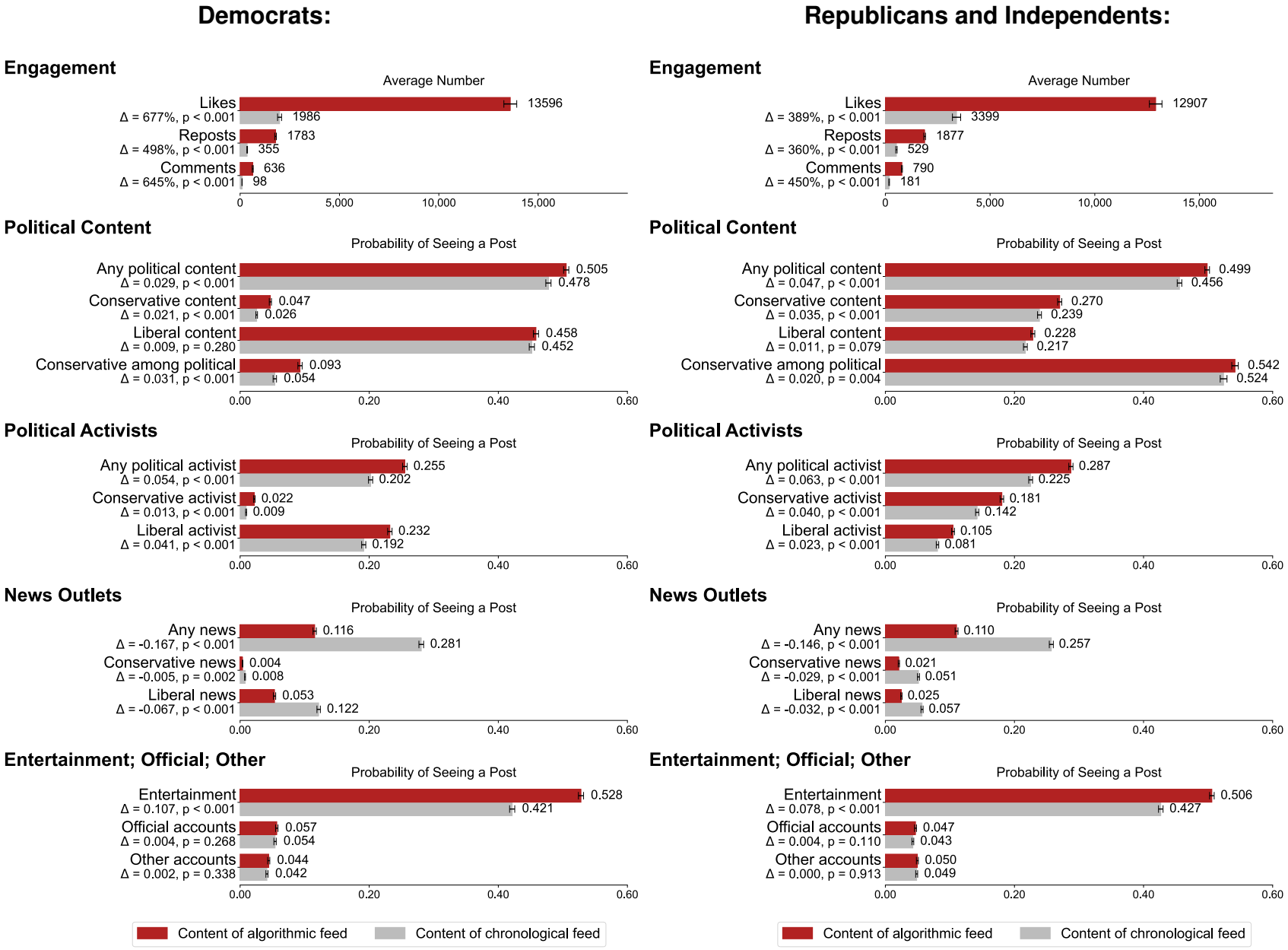


Fig. 2.21. Content of Feeds by Self-Declared Partisanship.

Replication of main-text Fig. 3 separately for self-declared Democrats (left) vs. self-declared Republicans and Independents (right). Average content shown to users in each feed setting: the chronological feed (grey) and the algorithmic feed (red). 95% CIs around subsample means (Mean ±1.96 × SE) are reported. Reported on the left, for each outcome, Δ and p-values are from regressions with user and time fixed effects, rather than raw comparison of means.

2.7 Effect of Experimental Treatments on Content of Chronological Feed

This section presents the empirical specification for analysing the effect of exposure to the algorithmic feed on the content of the chronological feed, operating through changes in the accounts followed by the users. We capture the feed content using the Google Chrome extension both at the end of the pre-treatment and post-treatment surveys. This enables us to measure treatment effects on the chronological feed, which reflects accounts the user follows, and also to estimate a placebo. Specifically, we expect no treatment effect on the chronological feed at the end of the pre-treatment survey, as users were only just assigned the new feed setting and had not yet had time to adjust accounts they follow. In contrast, after the treatment phase, users had time to modify followed accounts, so we would expect treatment effects on the accounts they follow to also appear in their chronological feed post-treatment.

Thus, we estimate the following equation:

$$\begin{aligned} \textit{Content}_{ki} = \; & \alpha + \beta_1 \left(\textit{Initial Chrono}_{ki} \times \textit{Treatment Algo}_{ki}\right) \times (\textit{Post Treat}_{ki}) \\ & + \beta_2 \left(\textit{Initial Algo}_{ki} \times \textit{Treatment Chrono}_{ki}\right) \times (\textit{Post Treat}_{ki}) \\ & + \delta_1 \left(\textit{Initial Chrono}_{ki} \times \textit{Treatment Algo}_{ki}\right) \times (1 - \textit{Post Treat}_{ki}) \\ & + \delta_2 \left(\textit{Initial Algo}_{ki} \times \textit{Treatment Chrono}_{ki}\right) \times (1 - \textit{Post Treat}_{ki}) \\ & + \beta_3 \, \textit{Initial Algo}_{ki} \times (\textit{Post Treat}_{ki}) \; + \; \delta_3 \, \textit{Initial Algo}_{ki} \times (1 - \textit{Post Treat}_{ki}) \\ & + \beta_4 \, \textit{Post Treat}_{ki} + \mathbf{X}_i' \Gamma + \epsilon_{ki}, \end{aligned} \tag{10}$$

i indexes users, and k indexes posts collected via the Google Chrome extension for each user i. We restrict the sample to posts shown to user i in the chronological feed setting. $\textit{Post Treat}_{ki}$ indicates whether the post was collected during the post-treatment survey, with this variable equal to 1 if the post was collected during the post-treatment survey and 0 if it was collected during the pre-treatment survey. β_1 and β_2 mirror β_1 and β_2 from Equation (1), capturing how the content in the chronological feed shown to user i changes as a result of the treatment. δ_1 and δ_2 are placebo estimates; they estimate how the chronological feed differs depending on treatment assignment but before the treatment phase. With successful randomisation, the content shown to different treatment groups should not differ at the end of the pre-treatment survey. As above, we control for the initial feed setting and allow this effect to differ between the post- and pre-treatment surveys.

In SI Table 2.8 and the main-text Fig. 4, we show that switching from an initial chronological feed to an algorithmic feed increases the likelihood that users follow accounts featuring conservative

content, political activist accounts, and, specifically, conservative political activist accounts.

As discussed in the main text, Extended Data Table 2 verifies that these changes translate into changes of the content that users would see on their chronological feeds by estimating Equation (10). For each outcome, the first column estimates Equation (10) without additional covariates (see Columns 1, 4, and 7). The second column for each outcome adds controls for $\mathbb{X}$ use (frequency of use and posting, and usage purposes; Columns 2, 5, and 8). The last column uses the full set of controls, including $\mathbb{X}$ use, demographic, and political controls (Columns 3, 6, and 9). This baseline analysis uses Llama 3 annotations.

Overall, Extended Data Table 2 displays changes consistent with those observed in the accounts users follow (Fig. 4): switching from a chronological to an algorithmic feed led to a change in the content shown in users' chronological feeds toward more conservative material and more content from political activists, particularly conservative political activists. In contrast, switching the algorithm off did not affect the content of the chronological feed.

Specifically, the magnitudes are as follows. β_1 estimated without controls indicates that a given post in the chronological feed is 6 percentage points more likely to feature conservative content and 4 percentage points more likely to be written by a conservative political activist. In our preferred specification, the point estimates are larger and more precisely estimated: controlling for pre-treatment $\mathbb{X}$ use, switching the algorithm on leads to a 7 percentage point increase in conservative content and a 4.7 percentage point increase in content from conservative political activists in the user's chronological feed.

At the same time, we find no effects of the treatment assignment for the feed before the treatment phase: as one would expect, estimates of δ_1 and δ_2 are insignificant.

SI Table 2.15 below, instead, uses simpler proxies of partisanship and content based on word frequencies (see SI Section 1.5). The results are noisier, but they point in the same direction, with content generated by conservative political activists significantly higher by approximately 3 percentage points in users' chronological feed, when the algorithm was turned on due to the experimental assignment, compared to when users were randomized to stay on the chronological feed.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Probability of seeing account in chronological feed that is:								
	Conservative			**Political Activist**			**Conservat. Polit. Activist**		
β_1: Initial Chrono	0.040*	0.041*	0.032	0.018	0.025	0.020	0.030*	0.034**	0.028*
× Treatment Algo	(0.021)	(0.021)	(0.020)	(0.030)	(0.026)	(0.025)	(0.017)	(0.016)	(0.016)
β_2: Initial Algo	-0.002	-0.002	-0.002	-0.022	-0.012	-0.016	-0.009	-0.003	-0.005
× Treatment Chrono	(0.012)	(0.012)	(0.011)	(0.018)	(0.016)	(0.016)	(0.010)	(0.010)	(0.009)
Placebo δ_1: Init. Chrono	-0.024	-0.027	-0.022	0.045	0.024	0.024	0.014	0.002	0.005
× Treat. Algo Pre-Treat	(0.018)	(0.018)	(0.016)	(0.029)	(0.024)	(0.023)	(0.016)	(0.014)	(0.013)
Placebo δ_2: Init. Algo	-0.004	-0.006	-0.007	0.012	0.004	0.004	0.007	0.003	0.003
× Treat. Chrono Pre-Treat	(0.011)	(0.011)	(0.010)	(0.017)	(0.016)	(0.015)	(0.010)	(0.009)	(0.009)
Initial Algo	0.009	0.012	0.003	0.000	-0.007	-0.004	0.003	0.001	-0.004
	(0.014)	(0.014)	(0.013)	(0.022)	(0.020)	(0.019)	(0.013)	(0.012)	(0.011)
Mean dependent variable	0.46	0.46	0.46	0.30	0.30	0.30	0.14	0.14	0.14
Observations	134706	134706	134706	134706	134706	134706	134706	134706	134706
𝕏 Use Controls		✓	✓		✓	✓		✓	✓
Demogr. & Polit. Controls			✓			✓			✓

Table 2.15. Accounts Shown in Chronological Feed: Simpler Proxies of Partisanship and Content

Regression estimates using simpler proxies of political content (see SI Sections 1.5.2 and 1.5.3) for the definition of simpler proxies). The baseline results with Llama 3 annotations are presented in Extended Data Table 2. Col. 1, 4, and 7 report estimates of equation 10 without controls, examining effects on Conservative content (1), political activist content (4), and Conservative political activist content (7). Col. 2, 5, and 8 add controls for pre-treatment 𝕏 use frequency and purpose. Col. 3, 6, and 9 include 𝕏 use, demographic, and political controls. For each outcome, coefficients β_1 and β_2 capture content changes in the chronological feed between pre- and post-treatment periods, while δ_1 and δ_2 serve as placebo estimates validating the randomization. In all specifications, we control for users' previous feed settings. The sample is restricted to posts shown in the chronological feed setting.

2.8 Detailed Results for Coefficient Plots

		Chrono to Algorithm					Algorithm to Chrono					
Outcome	**Controls**	**Coef.**	**95% CI**	**90% CI**	**SE**	**p-value**	**Coef.**	**95% CI**	**90% CI**	**SE**	**p-value**	**N**
Panel A: Engagement and Political Attitudes: ITT Estimates by Initial Feed Setting (Main Text Figure 2)												
User engagement (PCA)	No	0.142	[0.029, 0.256]	[0.047, 0.237]	0.058	0.014	-0.053	[-0.117, 0.011]	[-0.107, 0.000]	0.033	0.103	4965
	Yes	0.140	[0.028, 0.251]	[0.046, 0.233]	0.057	0.014	-0.055	[-0.118, 0.007]	[-0.108, -0.003]	0.032	0.081	
Partisanship and affective polarisation (PCA)	No	-0.007	[-0.046, 0.031]	[-0.039, 0.025]	0.019	0.703	0.000	[-0.023, 0.023]	[-0.019, 0.019]	0.012	0.987	3337
	Yes	-0.004	[-0.041, 0.032]	[-0.035, 0.027]	0.019	0.824	0.002	[-0.021, 0.024]	[-0.017, 0.021]	0.011	0.883	
Conservative policy priorities (PCA)	No	0.145	[0.027, 0.263]	[0.046, 0.244]	0.060	0.016	0.027	[-0.036, 0.090]	[-0.026, 0.080]	0.032	0.404	4965
	Yes	0.111	[0.021, 0.201]	[0.035, 0.186]	0.046	0.016	0.035	[-0.013, 0.084]	[-0.005, 0.076]	0.025	0.150	
Beliefs: Investigating Trump unacceptable (all)	No	0.111	[0.016, 0.205]	[0.032, 0.190]	0.048	0.022	-0.021	[-0.075, 0.033]	[-0.066, 0.025]	0.027	0.452	4965
	Yes	0.083	[0.010, 0.157]	[0.022, 0.145]	0.037	0.026	-0.000	[-0.041, 0.041]	[-0.034, 0.034]	0.021	0.992	
Pro-Kremlin attitudes on Ukraine War (PCA)	No	0.163	[0.046, 0.279]	[0.065, 0.261]	0.059	0.006	-0.019	[-0.086, 0.047]	[-0.075, 0.037]	0.034	0.569	4596
	Yes	0.123	[0.034, 0.211]	[0.048, 0.197]	0.045	0.007	0.007	[-0.043, 0.058]	[-0.035, 0.050]	0.026	0.773	
All attitudes on policies and current news (PCA)	No	0.176	[0.056, 0.296]	[0.075, 0.276]	0.061	0.004	-0.006	[-0.072, 0.060]	[-0.062, 0.049]	0.034	0.850	4596
	Yes	0.124	[0.040, 0.207]	[0.054, 0.194]	0.043	0.004	0.017	[-0.027, 0.061]	[-0.020, 0.054]	0.023	0.453	
Low well-being (PCA)	No	0.023	[-0.045, 0.091]	[-0.034, 0.080]	0.035	0.506	0.004	[-0.036, 0.044]	[-0.030, 0.038]	0.021	0.845	4850
	Yes	0.033	[-0.035, 0.101]	[-0.024, 0.090]	0.035	0.344	0.010	[-0.031, 0.051]	[-0.024, 0.044]	0.021	0.632	
Panel B: Accounts that Users Follow: ITT Estimates by Initial Feed Setting (Main Text Figure 4)												
Conservative share	No	0.200	[0.025, 0.375]	[0.054, 0.347]	0.089	0.025	0.009	[-0.083, 0.100]	[-0.068, 0.085]	0.047	0.855	2387
	Yes	0.172	[0.033, 0.310]	[0.055, 0.288]	0.071	0.015	0.007	[-0.064, 0.078]	[-0.053, 0.066]	0.036	0.854	
Liberal share	No	-0.093	[-0.264, 0.078]	[-0.236, 0.051]	0.087	0.288	-0.068	[-0.160, 0.024]	[-0.145, 0.009]	0.047	0.149	2387
	Yes	-0.041	[-0.166, 0.083]	[-0.146, 0.063]	0.063	0.517	-0.063	[-0.133, 0.007]	[-0.122, -0.004]	0.036	0.078	
Any political activist share	No	0.151	[-0.007, 0.308]	[0.019, 0.283]	0.080	0.061	-0.046	[-0.140, 0.048]	[-0.125, 0.033]	0.048	0.335	2387
	Yes	0.129	[0.009, 0.249]	[0.028, 0.230]	0.061	0.036	-0.037	[-0.108, 0.035]	[-0.097, 0.023]	0.037	0.314	
Conservative activist share	No	0.217	[0.049, 0.384]	[0.076, 0.357]	0.085	0.011	-0.007	[-0.100, 0.085]	[-0.085, 0.070]	0.047	0.875	2387
	Yes	0.185	[0.046, 0.324]	[0.068, 0.302]	0.071	0.010	-0.010	[-0.084, 0.064]	[-0.072, 0.052]	0.038	0.790	
Liberal activist share	No	0.003	[-0.156, 0.163]	[-0.131, 0.137]	0.081	0.967	-0.043	[-0.138, 0.051]	[-0.123, 0.036]	0.048	0.370	2387
	Yes	0.012	[-0.101, 0.124]	[-0.083, 0.106]	0.057	0.840	-0.036	[-0.107, 0.035]	[-0.096, 0.024]	0.036	0.319	
Any news outlet share	No	-0.042	[-0.209, 0.124]	[-0.182, 0.097]	0.085	0.619	0.036	[-0.053, 0.125]	[-0.039, 0.110]	0.045	0.428	2387

	Yes	-0.004	[-0.156, 0.148]	[-0.132, 0.123]	0.078	0.958	0.010	[-0.072, 0.092]	[-0.059, 0.079]	0.042	0.812	
Conservative news share	No	0.124	[-0.061, 0.309]	[-0.031, 0.280]	0.094	0.188	0.035	[-0.054, 0.124]	[-0.040, 0.109]	0.045	0.443	2387
	Yes	0.116	[-0.041, 0.273]	[-0.016, 0.248]	0.080	0.149	0.019	[-0.056, 0.094]	[-0.044, 0.082]	0.038	0.628	
Liberal news share	No	-0.073	[-0.255, 0.110]	[-0.226, 0.081]	0.093	0.437	-0.025	[-0.112, 0.061]	[-0.098, 0.047]	0.044	0.566	2387
	Yes	-0.022	[-0.172, 0.127]	[-0.148, 0.103]	0.076	0.770	-0.025	[-0.101, 0.052]	[-0.089, 0.039]	0.039	0.526	
Entertainment share	No	-0.001	[-0.164, 0.163]	[-0.138, 0.137]	0.084	0.995	0.049	[-0.042, 0.141]	[-0.028, 0.126]	0.047	0.291	2387
	Yes	-0.023	[-0.166, 0.119]	[-0.143, 0.096]	0.073	0.747	0.060	[-0.019, 0.140]	[-0.006, 0.127]	0.040	0.135	
Official share	No	-0.069	[-0.229, 0.091]	[-0.203, 0.065]	0.082	0.396	-0.032	[-0.127, 0.062]	[-0.112, 0.047]	0.048	0.503	2387
	Yes	-0.052	[-0.199, 0.094]	[-0.175, 0.071]	0.075	0.484	-0.046	[-0.132, 0.040]	[-0.118, 0.026]	0.044	0.291	
Panel C: Engagement and Political Attitudes. LATE Estimates by Initial Feed Setting (Extended Data Fig. 3)												
User engagement (PCA)	No	0.195	[0.039, 0.351]	[0.064, 0.326]	0.079	0.014	-0.076	[-0.167, 0.015]	[-0.152, 0.001]	0.046	0.103	4965
	Yes	0.178	[0.020, 0.336]	[0.046, 0.311]	0.080	0.027	-0.076	[-0.167, 0.015]	[-0.152, 0.001]	0.046	0.102	
Partisanship and affective polarisation (PCA)	No	-0.010	[-0.064, 0.043]	[-0.055, 0.034]	0.027	0.703	0.000	[-0.032, 0.033]	[-0.027, 0.027]	0.017	0.987	3337
	Yes	-0.014	[-0.069, 0.041]	[-0.060, 0.032]	0.028	0.618	-0.000	[-0.034, 0.033]	[-0.029, 0.028]	0.017	0.977	
Conservative policy priorities (PCA)	No	0.199	[0.036, 0.361]	[0.062, 0.335]	0.083	0.016	0.038	[-0.052, 0.128]	[-0.037, 0.114]	0.046	0.404	4965
	Yes	0.175	[0.046, 0.304]	[0.067, 0.284]	0.066	0.008	0.056	[-0.016, 0.127]	[-0.004, 0.116]	0.037	0.126	
Beliefs: Investigating Trump unacceptable (all)	No	0.152	[0.022, 0.282]	[0.043, 0.261]	0.066	0.022	-0.030	[-0.106, 0.047]	[-0.094, 0.035]	0.039	0.452	4965
	Yes	0.121	[0.011, 0.231]	[0.029, 0.213]	0.056	0.031	0.003	[-0.058, 0.065]	[-0.048, 0.055]	0.031	0.913	
Pro-Kremlin attitudes on Ukraine War (PCA)	No	0.222	[0.063, 0.381]	[0.088, 0.356]	0.081	0.006	-0.027	[-0.121, 0.067]	[-0.106, 0.052]	0.048	0.569	4596
	Yes	0.187	[0.060, 0.313]	[0.080, 0.293]	0.065	0.004	0.013	[-0.061, 0.087]	[-0.049, 0.075]	0.038	0.727	
All attitudes on policies and current news (PCA)	No	0.239	[0.075, 0.404]	[0.101, 0.378]	0.084	0.004	0.010	[-0.083, 0.103]	[-0.068, 0.088]	0.047	0.837	4596
	Yes	0.208	[0.086, 0.330]	[0.106, 0.310]	0.062	0.001	0.051	[-0.016, 0.119]	[-0.005, 0.108]	0.034	0.136	
Low well-being (PCA)	No	0.032	[-0.062, 0.125]	[-0.047, 0.110]	0.048	0.506	0.006	[-0.052, 0.063]	[-0.042, 0.054]	0.029	0.845	4850
	Yes	0.043	[-0.056, 0.141]	[-0.040, 0.126]	0.050	0.394	0.016	[-0.044, 0.076]	[-0.034, 0.066]	0.031	0.598	
Panel D: Engagement and Political Attitudes. ITT: Republicans and Independents vs. Democrats (Extended Data Fig. 4)												
Republicans and Independents												
User engagement (PCA)	No	0.108	[-0.045, 0.262]	[-0.021, 0.237]	0.078	0.167	-0.000	[-0.087, 0.086]	[-0.073, 0.072]	0.044	0.992	2659
	Yes	0.124	[-0.025, 0.274]	[-0.001, 0.250]	0.076	0.104	0.004	[-0.081, 0.088]	[-0.067, 0.075]	0.043	0.930	
Partisanship and affective polarisation (PCA)	No	-0.038	[-0.137, 0.060]	[-0.121, 0.044]	0.050	0.445	-0.015	[-0.061, 0.032]	[-0.054, 0.025]	0.024	0.538	1059
	Yes	-0.036	[-0.134, 0.062]	[-0.118, 0.047]	0.050	0.476	-0.012	[-0.059, 0.034]	[-0.051, 0.027]	0.024	0.610	
Conservative policy priorities (PCA)	No	0.134	[-0.026, 0.294]	[-0.000, 0.269]	0.082	0.101	0.038	[-0.044, 0.120]	[-0.031, 0.107]	0.042	0.361	2659

	Yes	0.142	[0.000, 0.284]	[0.023, 0.262]	0.072	0.050	0.029	[-0.044, 0.102]	[-0.032, 0.090]	0.037	0.438	
Beliefs: Investigating Trump unacceptable (all)	No	0.162	[0.009, 0.315]	[0.033, 0.290]	0.078	0.038	-0.021	[-0.107, 0.065]	[-0.094, 0.051]	0.044	0.630	2659
	Yes	0.171	[0.037, 0.306]	[0.058, 0.285]	0.069	0.013	-0.020	[-0.094, 0.053]	[-0.082, 0.042]	0.038	0.590	
Pro-Kremlin attitudes on Ukraine War (PCA)	No	0.244	[0.085, 0.404]	[0.110, 0.378]	0.081	0.003	0.004	[-0.086, 0.093]	[-0.072, 0.079]	0.046	0.938	2452
	Yes	0.224	[0.078, 0.370]	[0.101, 0.346]	0.074	0.003	0.015	[-0.067, 0.096]	[-0.054, 0.083]	0.042	0.723	
All attitudes on policies and current news (PCA)	No	0.226	[0.059, 0.394]	[0.085, 0.367]	0.086	0.008	0.012	[-0.077, 0.102]	[-0.063, 0.088]	0.046	0.787	2452
	Yes	0.206	[0.061, 0.352]	[0.084, 0.328]	0.074	0.006	0.012	[-0.064, 0.089]	[-0.052, 0.076]	0.039	0.749	
Low well-being (PCA)	No	0.077	[-0.020, 0.175]	[-0.005, 0.159]	0.050	0.120	0.024	[-0.032, 0.081]	[-0.023, 0.071]	0.029	0.397	2595
	Yes	0.091	[-0.006, 0.189]	[0.010, 0.173]	0.050	0.067	0.035	[-0.022, 0.092]	[-0.013, 0.083]	0.029	0.232	
Democrats												
User engagement (PCA)	No	0.175	[0.007, 0.342]	[0.034, 0.315]	0.086	0.041	-0.116	[-0.210, -0.021]	[-0.195, -0.036]	0.048	0.017	2306
	Yes	0.157	[-0.010, 0.324]	[0.017, 0.297]	0.085	0.065	-0.125	[-0.216, -0.033]	[-0.202, -0.047]	0.047	0.008	
Partisanship and affective polarisation (PCA)	No	0.004	[-0.030, 0.039]	[-0.024, 0.033]	0.017	0.799	0.007	[-0.019, 0.032]	[-0.015, 0.028]	0.013	0.604	2278
	Yes	0.009	[-0.024, 0.041]	[-0.019, 0.036]	0.017	0.600	0.008	[-0.016, 0.033]	[-0.012, 0.029]	0.013	0.507	
Conservative policy priorities (PCA)	No	0.077	[-0.035, 0.189]	[-0.017, 0.171]	0.057	0.179	0.062	[-0.006, 0.129]	[0.005, 0.118]	0.034	0.072	2306
	Yes	0.075	[-0.031, 0.182]	[-0.014, 0.165]	0.054	0.164	0.043	[-0.018, 0.104]	[-0.008, 0.094]	0.031	0.164	
Beliefs: Investigating Trump unacceptable (all)	No	-0.002	[-0.043, 0.039]	[-0.037, 0.032]	0.021	0.909	0.018	[-0.004, 0.040]	[-0.000, 0.037]	0.011	0.103	2306
	Yes	-0.015	[-0.052, 0.023]	[-0.046, 0.017]	0.019	0.440	0.023	[0.002, 0.044]	[0.006, 0.041]	0.011	0.029	
Pro-Kremlin attitudes on Ukraine War (PCA)	No	-0.004	[-0.110, 0.102]	[-0.093, 0.085]	0.054	0.936	0.005	[-0.061, 0.072]	[-0.051, 0.061]	0.034	0.873	2144
	Yes	0.009	[-0.082, 0.101]	[-0.068, 0.086]	0.047	0.842	-0.001	[-0.055, 0.053]	[-0.047, 0.044]	0.028	0.969	
All attitudes on policies and current news (PCA)	No	0.043	[-0.057, 0.144]	[-0.041, 0.128]	0.051	0.397	0.037	[-0.025, 0.100]	[-0.015, 0.090]	0.032	0.242	2144
	Yes	0.053	[-0.035, 0.140]	[-0.021, 0.126]	0.045	0.238	0.027	[-0.022, 0.077]	[-0.014, 0.069]	0.025	0.277	
Low well-being (PCA)	No	-0.045	[-0.137, 0.048]	[-0.123, 0.033]	0.047	0.346	-0.021	[-0.079, 0.036]	[-0.070, 0.027]	0.029	0.466	2255
	Yes	-0.032	[-0.127, 0.062]	[-0.112, 0.047]	0.048	0.504	-0.019	[-0.077, 0.039]	[-0.067, 0.030]	0.029	0.520	
Panel E: Engagement and Political Attitudes. ITT: Re-weighted to Address Selection into Initial Feed (Extended Data Fig. 5)												
User engagement (PCA)	No	0.110	[-0.010, 0.229]	[0.009, 0.210]	0.061	0.072	-0.043	[-0.110, 0.023]	[-0.100, 0.013]	0.034	0.202	4965
	Yes	0.108	[-0.011, 0.226]	[0.008, 0.207]	0.060	0.074	-0.049	[-0.114, 0.017]	[-0.104, 0.006]	0.033	0.143	
Partisanship and affective polarisation (PCA)	No	-0.012	[-0.054, 0.031]	[-0.047, 0.024]	0.022	0.590	0.006	[-0.016, 0.029]	[-0.013, 0.025]	0.012	0.579	3337
	Yes	-0.010	[-0.051, 0.031]	[-0.044, 0.024]	0.021	0.637	0.010	[-0.012, 0.032]	[-0.009, 0.028]	0.011	0.386	
Conservative policy priorities (PCA)	No	0.162	[0.041, 0.283]	[0.061, 0.264]	0.062	0.009	0.033	[-0.035, 0.102]	[-0.024, 0.091]	0.035	0.340	4965
	Yes	0.134	[0.040, 0.228]	[0.055, 0.213]	0.048	0.005	0.048	[-0.004, 0.101]	[0.004, 0.092]	0.027	0.071	
Beliefs: Investigating Trump unacceptable (all)	No	0.129	[0.025, 0.234]	[0.041, 0.218]	0.054	0.016	-0.014	[-0.068, 0.041]	[-0.060, 0.032]	0.028	0.623	4965

	Yes	0.103	[0.023, 0.184]	[0.036, 0.171]	0.041	0.012	0.008	[-0.035, 0.050]	[-0.028, 0.043]	0.022	0.728	
Pro-Kremlin attitudes on Ukraine War (PCA)	No	0.156	[0.031, 0.280]	[0.051, 0.260]	0.064	0.015	-0.011	[-0.081, 0.058]	[-0.070, 0.047]	0.036	0.747	4596
	Yes	0.091	[-0.005, 0.187]	[0.010, 0.172]	0.049	0.064	0.018	[-0.035, 0.070]	[-0.026, 0.062]	0.027	0.506	
All attitudes on policies and current news (PCA)	No	0.191	[0.065, 0.318]	[0.085, 0.298]	0.065	0.003	0.002	[-0.068, 0.073]	[-0.056, 0.061]	0.036	0.947	4596
	Yes	0.128	[0.041, 0.215]	[0.055, 0.201]	0.044	0.004	0.028	[-0.020, 0.076]	[-0.012, 0.069]	0.025	0.250	
Low well-being (PCA)	No	0.029	[-0.048, 0.106]	[-0.036, 0.094]	0.039	0.461	0.011	[-0.031, 0.052]	[-0.024, 0.045]	0.021	0.616	4850
	Yes	0.038	[-0.041, 0.116]	[-0.028, 0.103]	0.040	0.345	0.012	[-0.030, 0.053]	[-0.023, 0.046]	0.021	0.587	
Panel F: Engagement and Political Attitudes. Selective Attrition Does Not Drive Results: Lee Bounds of ITT Estimates by Initial Feed Setting (Extended Data Fig. 6)												
User engagement (PCA)	No	0.142	[-0.086, 0.287]	[-0.056, 0.264]	–	–	-0.053	[-0.169, 0.019]	[-0.152, 0.006]	–	–	4965
Partisanship and affective polarisation (PCA)	No	0.069	[-0.127, 0.378]	[-0.098, 0.340]	–	–	-0.043	[-0.138, 0.094]	[-0.122, 0.075]	–	–	3337
Conservative policy priorities (PCA)	No	0.145	[0.007, 0.298]	[0.029, 0.274]	–	–	0.027	[-0.050, 0.107]	[-0.038, 0.094]	–	–	4965
Beliefs: Investigating Trump unacceptable (all)	No	0.140	[0.010, 0.293]	[0.027, 0.270]	–	–	0.004	[-0.077, 0.067]	[-0.067, 0.055]	–	–	4965
Pro-Kremlin attitudes on Ukraine War (PCA)	No	0.163	[0.003, 0.368]	[0.028, 0.340]	–	–	-0.019	[-0.154, 0.058]	[-0.140, 0.042]	–	–	4596
All attitudes on policies and current news (PCA)	No	0.176	[0.025, 0.416]	[0.049, 0.385]	–	–	-0.006	[-0.090, 0.109]	[-0.076, 0.092]	–	–	4596
Low well-being (PCA)	No	0.096	[-0.164, 0.326]	[-0.138, 0.297]	–	–	0.028	[-0.045, 0.174]	[-0.030, 0.154]	–	–	4850
Panel G: Content that Users See: ITT Estimates by Initial Feed Setting (Extended Data Fig. 7)												
Conservative share	No	0.379	[0.014, 0.744]	[0.073, 0.685]	0.186	0.042	0.007	[-0.170, 0.184]	[-0.142, 0.156]	0.090	0.938	599
	Yes	0.351	[0.064, 0.639]	[0.110, 0.593]	0.147	0.018	-0.044	[-0.179, 0.090]	[-0.157, 0.068]	0.068	0.517	
Liberal share	No	0.098	[-0.226, 0.421]	[-0.174, 0.369]	0.165	0.554	-0.093	[-0.277, 0.092]	[-0.248, 0.062]	0.094	0.324	599
	Yes	0.110	[-0.152, 0.372]	[-0.110, 0.330]	0.134	0.412	-0.053	[-0.204, 0.098]	[-0.180, 0.074]	0.077	0.493	
Any political activist share	No	0.503	[0.164, 0.841]	[0.218, 0.787]	0.173	0.004	-0.233	[-0.410, -0.055]	[-0.382, -0.083]	0.091	0.011	599
	Yes	0.373	[0.115, 0.630]	[0.157, 0.589]	0.131	0.005	-0.277	[-0.423, -0.131]	[-0.399, -0.155]	0.074	0.000	
Conservative activist share	No	0.411	[0.029, 0.792]	[0.091, 0.730]	0.194	0.035	-0.038	[-0.211, 0.134]	[-0.184, 0.107]	0.088	0.663	599
	Yes	0.369	[0.054, 0.683]	[0.105, 0.632]	0.160	0.024	-0.089	[-0.225, 0.046]	[-0.203, 0.024]	0.069	0.197	
Liberal activist share	No	0.224	[-0.091, 0.539]	[-0.040, 0.488]	0.161	0.164	-0.264	[-0.445, -0.082]	[-0.416, -0.112]	0.093	0.005	599
	Yes	0.203	[-0.066, 0.472]	[-0.022, 0.429]	0.137	0.141	-0.227	[-0.378, -0.075]	[-0.354, -0.100]	0.077	0.004	
Any news outlet share	No	-0.503	[-0.782, -0.223]	[-0.737, -0.268]	0.143	0.000	0.794	[0.612, 0.976]	[0.641, 0.947]	0.093	0.000	599
	Yes	-0.513	[-0.785, -0.241]	[-0.741, -0.285]	0.139	0.000	0.774	[0.604, 0.944]	[0.632, 0.917]	0.087	0.000	
Conservative news share	No	-0.029	[-0.277, 0.219]	[-0.237, 0.179]	0.126	0.819	0.329	[0.126, 0.531]	[0.159, 0.498]	0.103	0.002	599
	Yes	-0.058	[-0.292, 0.176]	[-0.254, 0.139]	0.120	0.630	0.282	[0.101, 0.463]	[0.130, 0.434]	0.092	0.002	
Liberal news share	No	-0.423	[-0.717, -0.129]	[-0.670, -0.176]	0.150	0.005	0.481	[0.290, 0.671]	[0.321, 0.641]	0.097	0.000	599

	Yes	-0.433	[-0.707, -0.160]	[-0.663, -0.204]	0.139	0.002	0.498	[0.322, 0.673]	[0.350, 0.645]	0.089	0.000	
Entertainment share	No	-0.055	[-0.397, 0.288]	[-0.342, 0.233]	0.175	0.755	-0.310	[-0.490, -0.130]	[-0.461, -0.159]	0.092	0.001	599
	Yes	0.085	[-0.161, 0.332]	[-0.122, 0.293]	0.126	0.499	-0.272	[-0.421, -0.123]	[-0.398, -0.147]	0.076	0.000	
Official share	No	0.034	[-0.187, 0.255]	[-0.152, 0.219]	0.113	0.766	0.039	[-0.162, 0.241]	[-0.130, 0.209]	0.103	0.703	599
	Yes	-0.008	[-0.230, 0.214]	[-0.194, 0.178]	0.113	0.943	0.043	[-0.149, 0.235]	[-0.118, 0.204]	0.098	0.663	
Panel H: Followed Accounts. ITT Estimates by Initial Feed Setting: Lee Bounds Estimates (Fig. S1.15)												
Conservative share	No	0.176	[0.013, 0.662]	[0.047, 0.609]	–	–	0.018	[-0.087, 0.265]	[-0.070, 0.229]	–	–	2387
Liberal share	No	-0.065	[-0.350, 0.305]	[-0.311, 0.256]	–	–	-0.043	[-0.179, 0.139]	[-0.160, 0.110]	–	–	2387
Any political activist share	No	0.119	[-0.058, 0.491]	[-0.024, 0.452]	–	–	-0.002	[-0.154, 0.196]	[-0.136, 0.161]	–	–	2387
Conservative activist share	No	0.166	[0.047, 0.599]	[0.079, 0.557]	–	–	0.016	[-0.100, 0.306]	[-0.083, 0.264]	–	–	2387
Liberal activist share	No	0.008	[-0.192, 0.379]	[-0.160, 0.336]	–	–	-0.009	[-0.143, 0.244]	[-0.126, 0.204]	–	–	2387
Any news outlet share	No	-0.034	[-0.292, 0.296]	[-0.254, 0.254]	–	–	0.014	[-0.069, 0.267]	[-0.051, 0.236]	–	–	2387
Conservative news share	No	0.160	[-0.065, 0.543]	[-0.029, 0.503]	–	–	0.029	[-0.052, 0.326]	[-0.036, 0.287]	–	–	2387
Liberal news share	No	-0.084	[-0.302, 0.385]	[-0.264, 0.332]	–	–	-0.036	[-0.116, 0.223]	[-0.099, 0.190]	–	–	2387
Entertainment share	No	0.015	[-0.290, 0.307]	[-0.249, 0.265]	–	–	0.014	[-0.099, 0.210]	[-0.076, 0.186]	–	–	2387
Official share	No	-0.041	[-0.297, 0.300]	[-0.261, 0.262]	–	–	-0.009	[-0.142, 0.231]	[-0.123, 0.195]	–	–	2387
Panel I: All Survey-Based Outcomes. ITT Estimates by Initial Feed Setting (Figure S2.9)												
User engagement (PCA)	No	0.142	[0.029, 0.256]	[0.047, 0.237]	0.058	0.014	-0.053	[-0.117, 0.011]	[-0.107, 0.000]	0.033	0.103	4965
	Yes	0.140	[0.028, 0.251]	[0.046, 0.233]	0.057	0.014	-0.055	[-0.118, 0.007]	[-0.108, -0.003]	0.032	0.081	
Use X no less than before	No	0.130	[0.017, 0.243]	[0.035, 0.225]	0.058	0.024	-0.039	[-0.103, 0.025]	[-0.093, 0.014]	0.033	0.230	4965
	Yes	0.123	[0.010, 0.235]	[0.028, 0.217]	0.057	0.033	-0.037	[-0.100, 0.026]	[-0.090, 0.016]	0.032	0.250	
Post on X no less than before	No	0.092	[-0.022, 0.205]	[-0.003, 0.187]	0.058	0.112	-0.044	[-0.108, 0.020]	[-0.097, 0.010]	0.033	0.181	4965
	Yes	0.091	[-0.016, 0.199]	[0.001, 0.182]	0.055	0.096	-0.045	[-0.105, 0.015]	[-0.095, 0.006]	0.031	0.143	
Partisanship and affective polarisation (PCA)	No	-0.007	[-0.045, 0.031]	[-0.039, 0.025]	0.019	0.706	0.002	[-0.021, 0.025]	[-0.017, 0.021]	0.012	0.874	3337
	Yes	-0.004	[-0.041, 0.032]	[-0.035, 0.027]	0.019	0.824	0.002	[-0.021, 0.024]	[-0.017, 0.021]	0.011	0.883	
Affective polarization	No	0.014	[-0.048, 0.077]	[-0.038, 0.067]	0.032	0.652	0.002	[-0.034, 0.037]	[-0.028, 0.031]	0.018	0.933	3337
	Yes	0.023	[-0.034, 0.079]	[-0.025, 0.070]	0.029	0.432	0.006	[-0.027, 0.038]	[-0.022, 0.033]	0.017	0.733	
Self-declared Republican	No	-0.024	[-0.060, 0.013]	[-0.054, 0.007]	0.019	0.202	-0.003	[-0.026, 0.020]	[-0.022, 0.017]	0.012	0.820	4965
	Yes	-0.025	[-0.061, 0.012]	[-0.055, 0.006]	0.018	0.183	-0.003	[-0.026, 0.020]	[-0.023, 0.016]	0.012	0.779	
Conservative policy priorities (PCA)	No	0.145	[0.027, 0.263]	[0.046, 0.244]	0.060	0.016	0.027	[-0.036, 0.090]	[-0.026, 0.080]	0.032	0.404	4965
	Yes	0.111	[0.021, 0.201]	[0.035, 0.186]	0.046	0.016	0.035	[-0.013, 0.084]	[-0.005, 0.076]	0.025	0.150	

Shares Republican priorities	No	0.137	[0.019, 0.254]	[0.038, 0.235]	0.060	0.023	0.027	[-0.036, 0.090]	[-0.026, 0.080]	0.032	0.400	4965
	Yes	0.102	[0.011, 0.193]	[0.026, 0.179]	0.046	0.028	0.041	[-0.008, 0.091]	[-0.000, 0.083]	0.025	0.101	
Beliefs: Investigating Trump unacceptable (all)	No	0.111	[0.016, 0.205]	[0.032, 0.190]	0.048	0.022	-0.021	[-0.075, 0.033]	[-0.066, 0.025]	0.027	0.452	4965
	Yes	0.083	[0.010, 0.157]	[0.022, 0.145]	0.037	0.026	-0.000	[-0.041, 0.041]	[-0.034, 0.034]	0.021	0.992	
Investigations are against the rule of law	No	0.132	[0.020, 0.244]	[0.038, 0.226]	0.057	0.021	0.010	[-0.054, 0.074]	[-0.044, 0.064]	0.033	0.763	4965
	Yes	0.094	[0.008, 0.180]	[0.022, 0.166]	0.044	0.032	0.035	[-0.015, 0.086]	[-0.007, 0.077]	0.026	0.172	
Investigations undermine democracy	No	0.108	[-0.004, 0.220]	[0.014, 0.202]	0.057	0.058	-0.031	[-0.095, 0.033]	[-0.085, 0.023]	0.033	0.340	4965
	Yes	0.070	[-0.010, 0.151]	[0.002, 0.138]	0.041	0.088	-0.005	[-0.052, 0.043]	[-0.044, 0.035]	0.024	0.852	
Investigations are an attack on people like me	No	0.098	[-0.012, 0.208]	[0.006, 0.191]	0.056	0.080	-0.027	[-0.092, 0.037]	[-0.081, 0.027]	0.033	0.405	4965
	Yes	0.071	[-0.014, 0.157]	[-0.001, 0.143]	0.044	0.103	-0.008	[-0.058, 0.043]	[-0.050, 0.035]	0.026	0.768	
Investigations are an attempt to stop campaign	No	0.075	[-0.036, 0.186]	[-0.018, 0.168]	0.057	0.184	-0.036	[-0.101, 0.028]	[-0.090, 0.018]	0.033	0.267	4965
	Yes	0.041	[-0.041, 0.123]	[-0.028, 0.110]	0.042	0.330	-0.022	[-0.071, 0.027]	[-0.063, 0.019]	0.025	0.375	
Pro-Kremlin attitudes on Ukraine War (PCA)	No	0.163	[0.046, 0.279]	[0.065, 0.261]	0.059	0.006	-0.019	[-0.086, 0.047]	[-0.075, 0.037]	0.034	0.569	4596
	Yes	0.123	[0.034, 0.211]	[0.048, 0.197]	0.045	0.007	0.007	[-0.043, 0.058]	[-0.035, 0.050]	0.026	0.773	
Anti-Zelensky sentiment	No	0.151	[0.038, 0.265]	[0.056, 0.247]	0.058	0.009	-0.017	[-0.083, 0.049]	[-0.073, 0.038]	0.034	0.612	4700
	Yes	0.109	[0.013, 0.206]	[0.028, 0.190]	0.049	0.027	0.003	[-0.054, 0.059]	[-0.045, 0.050]	0.029	0.927	
Pro-Putin sentiment	No	0.108	[0.001, 0.215]	[0.018, 0.198]	0.054	0.047	0.012	[-0.054, 0.077]	[-0.043, 0.067]	0.034	0.726	4873
	Yes	0.090	[-0.007, 0.188]	[0.008, 0.172]	0.050	0.070	0.011	[-0.047, 0.069]	[-0.038, 0.060]	0.030	0.712	
Republicans handle the Ukraine crisis better	No	0.118	[0.013, 0.224]	[0.030, 0.207]	0.054	0.027	-0.023	[-0.088, 0.042]	[-0.078, 0.032]	0.033	0.486	4965
	Yes	0.088	[0.007, 0.170]	[0.020, 0.157]	0.042	0.034	0.003	[-0.047, 0.054]	[-0.039, 0.046]	0.026	0.894	
Disapproval of Biden government in Ukraine	No	0.130	[0.016, 0.243]	[0.034, 0.225]	0.058	0.026	-0.014	[-0.080, 0.051]	[-0.069, 0.040]	0.033	0.663	4807
	Yes	0.094	[0.001, 0.188]	[0.016, 0.173]	0.048	0.049	0.012	[-0.042, 0.066]	[-0.033, 0.057]	0.028	0.661	
Ukraine measures	No	0.098	[-0.015, 0.210]	[0.003, 0.192]	0.057	0.089	-0.019	[-0.083, 0.045]	[-0.073, 0.035]	0.033	0.562	4965
	Yes	0.077	[-0.020, 0.175]	[-0.005, 0.159]	0.050	0.122	0.005	[-0.051, 0.061]	[-0.042, 0.052]	0.029	0.849	
Low well-being (PCA)	No	0.023	[-0.045, 0.091]	[-0.034, 0.080]	0.035	0.506	0.004	[-0.036, 0.044]	[-0.030, 0.038]	0.021	0.845	4850
	Yes	0.033	[-0.035, 0.101]	[-0.024, 0.090]	0.035	0.344	0.010	[-0.031, 0.051]	[-0.024, 0.044]	0.021	0.632	
Life dissatisfaction	No	0.060	[-0.015, 0.135]	[-0.003, 0.123]	0.038	0.116	-0.009	[-0.054, 0.035]	[-0.046, 0.028]	0.023	0.679	4911
	Yes	0.057	[-0.015, 0.129]	[-0.004, 0.118]	0.037	0.122	0.003	[-0.040, 0.046]	[-0.033, 0.039]	0.022	0.892	
Low well-being	No	-0.009	[-0.085, 0.068]	[-0.073, 0.055]	0.039	0.822	0.018	[-0.028, 0.065]	[-0.021, 0.058]	0.024	0.437	4886
	Yes	-0.007	[-0.083, 0.069]	[-0.070, 0.057]	0.039	0.857	0.014	[-0.032, 0.060]	[-0.024, 0.052]	0.023	0.550	
Negative X experience	No	0.066	[-0.050, 0.182]	[-0.031, 0.164]	0.059	0.264	0.053	[-0.010, 0.116]	[-0.000, 0.106]	0.032	0.101	4965
	Yes	0.071	[-0.035, 0.177]	[-0.018, 0.160]	0.054	0.191	0.054	[-0.001, 0.109]	[0.008, 0.101]	0.028	0.053	
Anti-NATO sentiment	No	0.026	[-0.088, 0.140]	[-0.069, 0.121]	0.058	0.654	-0.017	[-0.082, 0.049]	[-0.072, 0.038]	0.033	0.614	4789

	Yes	0.009	[-0.090, 0.108]	[-0.075, 0.092]	0.051	0.862	0.007	[-0.051, 0.065]	[-0.041, 0.056]	0.029	0.809	
Importance of the war in Ukraine	No	0.028	[-0.086, 0.141]	[-0.068, 0.123]	0.058	0.633	-0.011	[-0.076, 0.053]	[-0.066, 0.043]	0.033	0.727	4965
	Yes	0.053	[-0.053, 0.159]	[-0.036, 0.142]	0.054	0.326	-0.022	[-0.083, 0.038]	[-0.073, 0.028]	0.031	0.466	
Did not solve puzzle for Ukraine	No	0.009	[-0.101, 0.118]	[-0.083, 0.101]	0.056	0.874	-0.038	[-0.102, 0.027]	[-0.092, 0.016]	0.033	0.251	4965
	Yes	-0.001	[-0.104, 0.102]	[-0.087, 0.086]	0.053	0.991	-0.028	[-0.090, 0.033]	[-0.080, 0.023]	0.031	0.367	
Panel J: Engagement and Political Attitudes. ITT: Reweighted for Twitter/𝕏 population (ANES) (Figure S2.10)												
User engagement (PCA)	No	0.166	[0.027, 0.305]	[0.050, 0.283]	0.071	0.019	-0.097	[-0.177, -0.017]	[-0.164, -0.030]	0.041	0.017	4965
	Yes	0.151	[0.014, 0.288]	[0.036, 0.266]	0.070	0.031	-0.092	[-0.168, -0.015]	[-0.156, -0.027]	0.039	0.019	
Partisanship and affective polarisation (PCA)	No	-0.007	[-0.059, 0.045]	[-0.051, 0.036]	0.027	0.782	-0.009	[-0.039, 0.022]	[-0.034, 0.017]	0.015	0.579	3337
	Yes	-0.002	[-0.053, 0.050]	[-0.045, 0.041]	0.026	0.941	-0.007	[-0.037, 0.023]	[-0.032, 0.018]	0.015	0.648	
Conservative policy priorities (PCA)	No	0.124	[-0.008, 0.255]	[0.013, 0.234]	0.067	0.066	0.000	[-0.066, 0.066]	[-0.056, 0.056]	0.034	0.997	4965
	Yes	0.091	[-0.009, 0.191]	[0.007, 0.175]	0.051	0.076	0.022	[-0.031, 0.074]	[-0.022, 0.066]	0.027	0.413	
Beliefs: Investigating Trump unacceptable (all)	No	0.137	[0.024, 0.249]	[0.042, 0.231]	0.057	0.017	-0.020	[-0.083, 0.044]	[-0.073, 0.033]	0.032	0.543	4965
	Yes	0.097	[0.017, 0.178]	[0.030, 0.165]	0.041	0.018	0.016	[-0.032, 0.064]	[-0.024, 0.056]	0.024	0.509	
Pro-Kremlin attitudes on Ukraine War (PCA)	No	0.114	[-0.028, 0.256]	[-0.005, 0.233]	0.072	0.117	-0.022	[-0.100, 0.055]	[-0.088, 0.043]	0.040	0.572	4596
	Yes	0.074	[-0.031, 0.178]	[-0.014, 0.161]	0.053	0.167	0.013	[-0.048, 0.073]	[-0.038, 0.063]	0.031	0.679	
All attitudes on policies and current news (PCA)	No	0.167	[0.030, 0.305]	[0.052, 0.283]	0.070	0.017	-0.023	[-0.093, 0.046]	[-0.082, 0.035]	0.036	0.510	4596
	Yes	0.115	[0.029, 0.202]	[0.043, 0.188]	0.044	0.009	0.020	[-0.026, 0.066]	[-0.019, 0.059]	0.024	0.399	
Low well-being (PCA)	No	0.034	[-0.051, 0.118]	[-0.037, 0.105]	0.043	0.436	0.003	[-0.052, 0.058]	[-0.043, 0.049]	0.028	0.924	4850
	Yes	0.048	[-0.039, 0.135]	[-0.025, 0.121]	0.044	0.278	0.018	[-0.038, 0.074]	[-0.029, 0.065]	0.029	0.530	
Panel K: Engagement and Political Attitudes. ITT: Republicans vs. Independents (Figure S2.11)												
Republicans												
User engagement (PCA)	No	0.028	[-0.223, 0.280]	[-0.183, 0.240]	0.128	0.824	-0.040	[-0.177, 0.096]	[-0.155, 0.074]	0.070	0.562	1065
	Yes	0.044	[-0.204, 0.292]	[-0.164, 0.252]	0.126	0.728	-0.035	[-0.167, 0.098]	[-0.146, 0.077]	0.068	0.610	
Partisanship and affective polarisation (PCA)	No	-0.038	[-0.137, 0.060]	[-0.121, 0.044]	0.050	0.445	-0.015	[-0.061, 0.032]	[-0.054, 0.025]	0.024	0.538	1059
	Yes	-0.036	[-0.134, 0.062]	[-0.118, 0.047]	0.050	0.476	-0.012	[-0.059, 0.034]	[-0.051, 0.027]	0.024	0.610	
Conservative policy priorities (PCA)	No	0.177	[-0.021, 0.375]	[0.011, 0.343]	0.101	0.081	0.045	[-0.061, 0.150]	[-0.044, 0.133]	0.054	0.409	1065
	Yes	0.179	[0.006, 0.352]	[0.034, 0.324]	0.088	0.044	0.015	[-0.074, 0.103]	[-0.060, 0.089]	0.045	0.747	
Beliefs: Investigating Trump unacceptable (all)	No	0.180	[-0.067, 0.426]	[-0.028, 0.387]	0.126	0.154	0.040	[-0.093, 0.172]	[-0.071, 0.151]	0.068	0.555	1065
	Yes	0.140	[-0.086, 0.365]	[-0.050, 0.329]	0.115	0.227	0.022	[-0.090, 0.135]	[-0.072, 0.117]	0.058	0.697	
Pro-Kremlin attitudes on Ukraine War (PCA)	No	0.349	[0.132, 0.565]	[0.167, 0.531]	0.110	0.002	0.029	[-0.090, 0.148]	[-0.071, 0.129]	0.061	0.635	995

	Yes	0.328	[0.134, 0.523]	[0.165, 0.491]	0.099	0.001	0.036	[-0.071, 0.143]	[-0.053, 0.126]	0.054	0.507	
All attitudes on policies and current news (PCA)	No	0.273	[0.062, 0.483]	[0.096, 0.449]	0.107	0.011	0.034	[-0.078, 0.146]	[-0.060, 0.128]	0.057	0.553	995
	Yes	0.246	[0.071, 0.422]	[0.099, 0.393]	0.089	0.006	0.018	[-0.070, 0.105]	[-0.056, 0.091]	0.045	0.693	
Low well-being (PCA)	No	0.155	[-0.014, 0.325]	[0.013, 0.298]	0.087	0.073	0.054	[-0.038, 0.147]	[-0.023, 0.132]	0.047	0.251	1048
	Yes	0.180	[0.015, 0.345]	[0.042, 0.319]	0.084	0.034	0.064	[-0.029, 0.158]	[-0.014, 0.142]	0.048	0.179	
Independents												
User engagement (PCA)	No	0.154	[-0.040, 0.348]	[-0.009, 0.316]	0.099	0.121	0.028	[-0.084, 0.139]	[-0.066, 0.121]	0.057	0.627	1594
	Yes	0.170	[-0.018, 0.357]	[0.012, 0.327]	0.096	0.077	0.031	[-0.079, 0.141]	[-0.062, 0.123]	0.056	0.584	
Partisanship and affective polarisation (PCA)	No	0.132	[-0.055, 0.318]	[-0.025, 0.288]	0.095	0.166	0.087	[-0.023, 0.198]	[-0.006, 0.180]	0.056	0.123	0
	Yes	0.133	[-0.053, 0.320]	[-0.023, 0.290]	0.095	0.163	0.088	[-0.022, 0.198]	[-0.004, 0.181]	0.056	0.118	
Conservative policy priorities (PCA)	No	0.112	[-0.092, 0.317]	[-0.059, 0.284]	0.104	0.281	0.041	[-0.069, 0.151]	[-0.051, 0.133]	0.056	0.466	1594
	Yes	0.122	[-0.078, 0.322]	[-0.046, 0.289]	0.102	0.233	0.039	[-0.069, 0.147]	[-0.052, 0.129]	0.055	0.479	
Beliefs: Investigating Trump unacceptable (all)	No	0.154	[-0.020, 0.327]	[0.008, 0.299]	0.089	0.083	-0.056	[-0.158, 0.045]	[-0.141, 0.029]	0.052	0.278	1594
	Yes	0.190	[0.021, 0.358]	[0.048, 0.331]	0.086	0.028	-0.050	[-0.147, 0.047]	[-0.132, 0.031]	0.050	0.311	
Pro-Kremlin attitudes on Ukraine War (PCA)	No	0.176	[-0.025, 0.378]	[0.007, 0.346]	0.103	0.087	-0.009	[-0.129, 0.111]	[-0.110, 0.092]	0.061	0.884	1457
	Yes	0.164	[-0.036, 0.365]	[-0.004, 0.332]	0.102	0.109	-0.001	[-0.118, 0.117]	[-0.099, 0.098]	0.060	0.990	
All attitudes on policies and current news (PCA)	No	0.189	[-0.019, 0.396]	[0.014, 0.363]	0.106	0.076	0.005	[-0.113, 0.122]	[-0.094, 0.103]	0.060	0.936	1457
	Yes	0.184	[-0.021, 0.389]	[0.012, 0.356]	0.104	0.080	0.009	[-0.107, 0.124]	[-0.088, 0.105]	0.059	0.883	
Low well-being (PCA)	No	0.033	[-0.086, 0.153]	[-0.067, 0.133]	0.061	0.587	0.004	[-0.066, 0.074]	[-0.055, 0.063]	0.036	0.912	1547
	Yes	0.041	[-0.079, 0.161]	[-0.060, 0.142]	0.061	0.505	0.014	[-0.058, 0.085]	[-0.046, 0.074]	0.037	0.705	
Panel L: Engagement and Political Attitudes. ITT: Observed Compliers (Figure S2.12)												
User engagement (PCA)	No	0.134	[-0.204, 0.473]	[-0.150, 0.418]	0.173	0.437	0.005	[-0.195, 0.204]	[-0.163, 0.172]	0.102	0.965	534
	Yes	0.164	[-0.174, 0.502]	[-0.119, 0.448]	0.172	0.343	-0.014	[-0.206, 0.178]	[-0.175, 0.147]	0.098	0.886	
Partisanship and affective polarisation (PCA)	No	0.036	[-0.044, 0.116]	[-0.031, 0.103]	0.041	0.382	0.015	[-0.026, 0.057]	[-0.020, 0.050]	0.021	0.473	324
	Yes	0.040	[-0.045, 0.124]	[-0.031, 0.110]	0.043	0.363	0.029	[-0.015, 0.073]	[-0.008, 0.066]	0.023	0.200	
Conservative policy priorities (PCA)	No	0.252	[-0.122, 0.625]	[-0.062, 0.565]	0.191	0.187	0.148	[-0.046, 0.341]	[-0.015, 0.310]	0.099	0.135	534
	Yes	0.217	[-0.071, 0.506]	[-0.025, 0.459]	0.147	0.143	0.073	[-0.079, 0.224]	[-0.054, 0.200]	0.077	0.347	
Beliefs: Investigating Trump unacceptable (all)	No	0.315	[0.057, 0.573]	[0.099, 0.532]	0.132	0.017	0.020	[-0.143, 0.183]	[-0.117, 0.157]	0.083	0.812	534
	Yes	0.235	[0.024, 0.445]	[0.058, 0.411]	0.107	0.031	-0.020	[-0.145, 0.105]	[-0.125, 0.085]	0.064	0.752	
Pro-Kremlin attitudes on Ukraine War (PCA)	No	0.242	[-0.102, 0.587]	[-0.047, 0.532]	0.176	0.169	0.072	[-0.133, 0.277]	[-0.100, 0.244]	0.105	0.490	496
	Yes	0.216	[-0.060, 0.491]	[-0.015, 0.447]	0.140	0.128	0.059	[-0.106, 0.225]	[-0.080, 0.198]	0.084	0.483	
All attitudes on policies and current news (PCA)	No	0.273	[-0.090, 0.636]	[-0.032, 0.578]	0.185	0.141	0.128	[-0.073, 0.328]	[-0.040, 0.296]	0.102	0.212	496

	Yes	0.213	[-0.056, 0.482]	[-0.012, 0.439]	0.137	0.124	0.087	[-0.060, 0.234]	[-0.037, 0.210]	0.075	0.248	
Low well-being (PCA)	No	0.035	[-0.139, 0.209]	[-0.111, 0.181]	0.089	0.692	0.112	[-0.002, 0.226]	[0.017, 0.208]	0.058	0.054	524
	Yes	0.041	[-0.153, 0.236]	[-0.122, 0.205]	0.099	0.677	0.115	[0.001, 0.230]	[0.019, 0.211]	0.058	0.049	
Panel M: Engagement and Political Attitudes. LATE: Robustness (Figure S2.13)												
User engagement (PCA)	No	0.313	[0.061, 0.565]	[0.102, 0.524]	0.128	0.015	-0.125	[-0.274, 0.025]	[-0.250, 0.001]	0.076	0.103	4965
	Yes	0.286	[0.021, 0.551]	[0.064, 0.508]	0.135	0.034	-0.116	[-0.270, 0.038]	[-0.245, 0.013]	0.079	0.140	
Partisanship and affective polarisation (PCA)	No	-0.017	[-0.102, 0.068]	[-0.088, 0.055]	0.043	0.703	0.000	[-0.054, 0.055]	[-0.045, 0.046]	0.028	0.987	3337
	Yes	-0.032	[-0.125, 0.062]	[-0.110, 0.047]	0.048	0.507	-0.002	[-0.060, 0.057]	[-0.051, 0.048]	0.030	0.960	
Conservative policy priorities (PCA)	No	0.319	[0.057, 0.580]	[0.099, 0.538]	0.133	0.017	0.063	[-0.085, 0.210]	[-0.061, 0.187]	0.075	0.404	4965
	Yes	0.326	[0.110, 0.542]	[0.145, 0.508]	0.110	0.003	0.090	[-0.031, 0.210]	[-0.011, 0.191]	0.061	0.144	
Beliefs: Investigating Trump unacceptable (all)	No	0.244	[0.035, 0.453]	[0.068, 0.419]	0.107	0.022	-0.048	[-0.175, 0.078]	[-0.155, 0.058]	0.064	0.452	4965
	Yes	0.197	[0.010, 0.385]	[0.040, 0.355]	0.096	0.039	0.017	[-0.087, 0.120]	[-0.070, 0.104]	0.053	0.753	
Pro-Kremlin attitudes on Ukraine War (PCA)	No	0.350	[0.098, 0.603]	[0.138, 0.563]	0.129	0.007	-0.044	[-0.197, 0.109]	[-0.173, 0.084]	0.078	0.569	4596
	Yes	0.318	[0.109, 0.528]	[0.143, 0.494]	0.107	0.003	0.032	[-0.089, 0.154]	[-0.070, 0.134]	0.062	0.602	
All attitudes on policies and current news (PCA)	No	0.378	[0.117, 0.639]	[0.159, 0.597]	0.133	0.005	0.016	[-0.135, 0.167]	[-0.111, 0.143]	0.077	0.837	4596
	Yes	0.356	[0.155, 0.556]	[0.188, 0.524]	0.102	0.001	0.095	[-0.017, 0.206]	[0.001, 0.188]	0.057	0.095	
Low well-being (PCA)	No	0.051	[-0.099, 0.201]	[-0.075, 0.177]	0.077	0.505	0.009	[-0.085, 0.103]	[-0.070, 0.088]	0.048	0.845	4850
	Yes	0.065	[-0.101, 0.231]	[-0.074, 0.205]	0.085	0.442	0.025	[-0.075, 0.125]	[-0.059, 0.109]	0.051	0.623	
Panel N: Followed Accounts. ITT: Reweighted to Resemble the Full Sample (Figure S2.14)												
Conservative share	No	0.212	[0.037, 0.387]	[0.065, 0.358]	0.089	0.018	0.015	[-0.076, 0.106]	[-0.061, 0.091]	0.046	0.746	2387
	Yes	0.181	[0.045, 0.316]	[0.067, 0.294]	0.069	0.009	0.003	[-0.066, 0.072]	[-0.055, 0.061]	0.035	0.930	
Liberal share	No	-0.101	[-0.267, 0.065]	[-0.240, 0.039]	0.085	0.235	-0.073	[-0.161, 0.016]	[-0.147, 0.002]	0.045	0.109	2387
	Yes	-0.047	[-0.169, 0.075]	[-0.149, 0.056]	0.062	0.452	-0.073	[-0.143, -0.004]	[-0.131, -0.015]	0.035	0.037	
Any political activist share	No	0.143	[-0.008, 0.294]	[0.016, 0.270]	0.077	0.064	-0.045	[-0.134, 0.044]	[-0.120, 0.030]	0.046	0.321	2387
	Yes	0.134	[0.018, 0.250]	[0.037, 0.231]	0.059	0.024	-0.036	[-0.106, 0.033]	[-0.095, 0.022]	0.036	0.307	
Conservative activist share	No	0.218	[0.054, 0.382]	[0.080, 0.356]	0.084	0.009	0.002	[-0.087, 0.091]	[-0.073, 0.077]	0.046	0.963	2387
	Yes	0.189	[0.053, 0.324]	[0.075, 0.303]	0.069	0.007	0.000	[-0.071, 0.071]	[-0.059, 0.060]	0.036	0.999	
Liberal activist share	No	-0.008	[-0.156, 0.139]	[-0.132, 0.115]	0.075	0.911	-0.050	[-0.138, 0.037]	[-0.124, 0.023]	0.045	0.258	2387
	Yes	-0.004	[-0.107, 0.099]	[-0.091, 0.082]	0.053	0.937	-0.044	[-0.112, 0.024]	[-0.101, 0.013]	0.035	0.200	
Any news outlet share	No	-0.032	[-0.201, 0.137]	[-0.174, 0.110]	0.086	0.711	0.042	[-0.048, 0.132]	[-0.033, 0.118]	0.046	0.358	2387
	Yes	0.011	[-0.143, 0.164]	[-0.118, 0.140]	0.078	0.890	0.022	[-0.061, 0.105]	[-0.048, 0.091]	0.042	0.607	

Conservative news share	No	0.143	[-0.046, 0.332]	[-0.016, 0.302]	0.097	0.139	0.035	[-0.053, 0.123]	[-0.039, 0.108]	0.045	0.439	2387
	Yes	0.134	[-0.026, 0.294]	[-0.000, 0.269]	0.082	0.101	0.016	[-0.057, 0.089]	[-0.046, 0.077]	0.037	0.670	
Liberal news share	No	-0.074	[-0.251, 0.103]	[-0.222, 0.075]	0.090	0.415	-0.023	[-0.109, 0.062]	[-0.096, 0.049]	0.044	0.592	2387
	Yes	-0.018	[-0.164, 0.127]	[-0.140, 0.104]	0.074	0.805	-0.024	[-0.100, 0.052]	[-0.087, 0.040]	0.039	0.540	
Entertainment share	No	-0.002	[-0.171, 0.167]	[-0.144, 0.140]	0.086	0.980	0.042	[-0.051, 0.136]	[-0.036, 0.121]	0.048	0.377	2387
	Yes	-0.026	[-0.171, 0.120]	[-0.148, 0.096]	0.074	0.728	0.052	[-0.029, 0.134]	[-0.016, 0.121]	0.042	0.209	
Official share	No	-0.078	[-0.248, 0.092]	[-0.220, 0.065]	0.087	0.371	-0.040	[-0.141, 0.061]	[-0.125, 0.045]	0.052	0.436	2387
	Yes	-0.063	[-0.220, 0.093]	[-0.195, 0.068]	0.080	0.429	-0.055	[-0.147, 0.037]	[-0.132, 0.022]	0.047	0.240	
Panel O: Followed Accounts. ITT: Reweighted for Twitter/𝕏 population (ANES) (Figure S2.15)												
Conservative share	No	0.211	[0.005, 0.417]	[0.038, 0.384]	0.105	0.045	-0.005	[-0.105, 0.095]	[-0.089, 0.079]	0.051	0.919	2387
	Yes	0.179	[0.014, 0.344]	[0.040, 0.318]	0.084	0.034	-0.000	[-0.080, 0.079]	[-0.067, 0.067]	0.041	0.999	
Liberal share	No	-0.065	[-0.228, 0.099]	[-0.202, 0.072]	0.083	0.437	-0.042	[-0.131, 0.047]	[-0.116, 0.033]	0.045	0.355	2387
	Yes	-0.052	[-0.180, 0.077]	[-0.160, 0.056]	0.066	0.432	-0.052	[-0.126, 0.022]	[-0.114, 0.010]	0.038	0.171	
Any political activist share	No	0.157	[-0.011, 0.326]	[0.016, 0.299]	0.086	0.067	-0.022	[-0.117, 0.074]	[-0.102, 0.058]	0.049	0.652	2387
	Yes	0.143	[0.004, 0.283]	[0.027, 0.260]	0.071	0.044	-0.021	[-0.098, 0.056]	[-0.085, 0.043]	0.039	0.590	
Conservative activist share	No	0.194	[-0.011, 0.398]	[0.022, 0.365]	0.104	0.063	-0.007	[-0.112, 0.097]	[-0.095, 0.080]	0.053	0.891	2387
	Yes	0.177	[0.003, 0.351]	[0.031, 0.323]	0.089	0.046	-0.001	[-0.085, 0.084]	[-0.071, 0.070]	0.043	0.990	
Liberal activist share	No	0.026	[-0.101, 0.153]	[-0.081, 0.133]	0.065	0.687	-0.013	[-0.090, 0.064]	[-0.078, 0.051]	0.039	0.737	2387
	Yes	0.026	[-0.071, 0.124]	[-0.056, 0.108]	0.050	0.596	-0.015	[-0.080, 0.050]	[-0.070, 0.040]	0.033	0.655	
Any news outlet share	No	-0.047	[-0.236, 0.143]	[-0.206, 0.112]	0.097	0.630	-0.008	[-0.114, 0.097]	[-0.097, 0.080]	0.054	0.876	2387
	Yes	-0.035	[-0.206, 0.137]	[-0.179, 0.109]	0.087	0.690	-0.014	[-0.115, 0.087]	[-0.099, 0.071]	0.051	0.785	
Conservative news share	No	0.142	[-0.037, 0.321]	[-0.009, 0.292]	0.091	0.121	-0.002	[-0.091, 0.087]	[-0.076, 0.073]	0.045	0.969	2387
	Yes	0.134	[-0.015, 0.283]	[0.009, 0.259]	0.076	0.078	-0.006	[-0.086, 0.074]	[-0.074, 0.061]	0.041	0.876	
Liberal news share	No	-0.073	[-0.230, 0.083]	[-0.204, 0.058]	0.080	0.358	-0.031	[-0.124, 0.062]	[-0.109, 0.047]	0.048	0.514	2387
	Yes	-0.061	[-0.195, 0.074]	[-0.174, 0.052]	0.069	0.375	-0.028	[-0.118, 0.062]	[-0.104, 0.048]	0.046	0.543	
Entertainment share	No	0.060	[-0.162, 0.282]	[-0.126, 0.246]	0.113	0.596	0.074	[-0.052, 0.201]	[-0.032, 0.180]	0.064	0.248	2387
	Yes	0.065	[-0.141, 0.270]	[-0.108, 0.237]	0.105	0.539	0.077	[-0.037, 0.191]	[-0.019, 0.173]	0.058	0.185	
Official share	No	-0.173	[-0.449, 0.103]	[-0.404, 0.058]	0.141	0.219	-0.056	[-0.181, 0.068]	[-0.161, 0.048]	0.064	0.377	2387
	Yes	-0.185	[-0.449, 0.079]	[-0.406, 0.037]	0.135	0.171	-0.078	[-0.196, 0.039]	[-0.177, 0.020]	0.060	0.190	
Panel P: Followed Accounts. LATE Estimates by Initial Feed Setting (Figure S2.16)												
Conservative share	No	0.254	[0.032, 0.475]	[0.067, 0.440]	0.113	0.025	0.011	[-0.108, 0.130]	[-0.089, 0.111]	0.061	0.855	2387

	Yes	0.228	[0.045, 0.410]	[0.074, 0.381]	0.093	0.015	0.006	[-0.089, 0.101]	[-0.073, 0.086]	0.048	0.894	
Liberal share	No	-0.117	[-0.333, 0.099]	[-0.299, 0.064]	0.110	0.288	-0.088	[-0.208, 0.032]	[-0.189, 0.012]	0.061	0.149	2387
	Yes	-0.050	[-0.209, 0.109]	[-0.184, 0.083]	0.081	0.537	-0.085	[-0.178, 0.009]	[-0.163, -0.007]	0.048	0.075	
Any political activist share	No	0.191	[-0.009, 0.391]	[0.023, 0.359]	0.102	0.061	-0.060	[-0.183, 0.062]	[-0.163, 0.043]	0.062	0.336	2387
	Yes	0.167	[0.014, 0.321]	[0.039, 0.296]	0.078	0.033	-0.057	[-0.150, 0.037]	[-0.135, 0.022]	0.048	0.235	
Conservative activist share	No	0.274	[0.062, 0.486]	[0.096, 0.452]	0.108	0.011	-0.010	[-0.130, 0.111]	[-0.111, 0.092]	0.062	0.875	2387
	Yes	0.242	[0.060, 0.423]	[0.090, 0.394]	0.092	0.009	-0.015	[-0.112, 0.081]	[-0.097, 0.066]	0.049	0.755	
Liberal activist share	No	0.004	[-0.198, 0.206]	[-0.165, 0.174]	0.103	0.967	-0.056	[-0.180, 0.067]	[-0.160, 0.047]	0.063	0.370	2387
	Yes	0.015	[-0.127, 0.156]	[-0.104, 0.133]	0.072	0.840	-0.057	[-0.149, 0.035]	[-0.135, 0.020]	0.047	0.224	
Any news outlet share	No	-0.053	[-0.264, 0.157]	[-0.230, 0.123]	0.107	0.619	0.047	[-0.069, 0.162]	[-0.050, 0.144]	0.059	0.429	2387
	Yes	-0.001	[-0.196, 0.193]	[-0.165, 0.162]	0.099	0.989	0.018	[-0.092, 0.129]	[-0.074, 0.111]	0.056	0.745	
Conservative news share	No	0.157	[-0.077, 0.391]	[-0.039, 0.354]	0.120	0.188	0.045	[-0.071, 0.161]	[-0.052, 0.142]	0.059	0.444	2387
	Yes	0.162	[-0.048, 0.372]	[-0.015, 0.338]	0.107	0.132	0.021	[-0.083, 0.125]	[-0.066, 0.109]	0.053	0.690	
Liberal news share	No	-0.092	[-0.323, 0.139]	[-0.286, 0.102]	0.118	0.436	-0.033	[-0.145, 0.080]	[-0.127, 0.062]	0.057	0.566	2387
	Yes	-0.019	[-0.215, 0.177]	[-0.184, 0.146]	0.100	0.851	-0.032	[-0.139, 0.075]	[-0.122, 0.058]	0.055	0.556	
Entertainment share	No	-0.001	[-0.208, 0.207]	[-0.175, 0.173]	0.106	0.995	0.064	[-0.055, 0.184]	[-0.036, 0.164]	0.061	0.291	2387
	Yes	-0.033	[-0.216, 0.151]	[-0.187, 0.122]	0.094	0.728	0.085	[-0.021, 0.192]	[-0.004, 0.175]	0.054	0.117	
Official share	No	-0.087	[-0.290, 0.115]	[-0.257, 0.082]	0.103	0.396	-0.042	[-0.165, 0.081]	[-0.145, 0.061]	0.063	0.502	2387
	Yes	-0.046	[-0.237, 0.144]	[-0.206, 0.113]	0.097	0.633	-0.065	[-0.184, 0.054]	[-0.165, 0.035]	0.061	0.282	
Panel Q: Followed Accounts. LATE: Robustness (Figure S2.17)												
Conservative share	No	0.385	[0.048, 0.721]	[0.102, 0.667]	0.172	0.025	0.017	[-0.165, 0.198]	[-0.135, 0.169]	0.093	0.855	2387
	Yes	0.342	[0.061, 0.622]	[0.106, 0.577]	0.143	0.017	0.015	[-0.133, 0.162]	[-0.109, 0.138]	0.075	0.846	
Liberal share	No	-0.178	[-0.506, 0.150]	[-0.453, 0.097]	0.167	0.288	-0.134	[-0.316, 0.048]	[-0.287, 0.019]	0.093	0.149	2387
	Yes	-0.066	[-0.314, 0.181]	[-0.274, 0.141]	0.126	0.599	-0.143	[-0.287, 0.000]	[-0.264, -0.023]	0.073	0.051	
Any political activist share	No	0.290	[-0.014, 0.593]	[0.035, 0.544]	0.155	0.062	-0.092	[-0.278, 0.095]	[-0.248, 0.065]	0.095	0.336	2387
	Yes	0.263	[0.027, 0.499]	[0.064, 0.461]	0.120	0.030	-0.086	[-0.230, 0.058]	[-0.207, 0.035]	0.074	0.243	
Conservative activist share	No	0.416	[0.093, 0.739]	[0.145, 0.687]	0.165	0.012	-0.015	[-0.199, 0.169]	[-0.169, 0.140]	0.094	0.875	2387
	Yes	0.367	[0.092, 0.642]	[0.137, 0.598]	0.140	0.009	-0.016	[-0.166, 0.134]	[-0.142, 0.110]	0.077	0.836	
Liberal activist share	No	0.007	[-0.300, 0.313]	[-0.251, 0.264]	0.156	0.967	-0.086	[-0.274, 0.102]	[-0.243, 0.072]	0.096	0.370	2387
	Yes	0.029	[-0.189, 0.247]	[-0.154, 0.212]	0.111	0.794	-0.093	[-0.235, 0.049]	[-0.213, 0.026]	0.073	0.199	
Any news outlet share	No	-0.081	[-0.400, 0.238]	[-0.349, 0.187]	0.163	0.619	0.071	[-0.105, 0.247]	[-0.077, 0.219]	0.090	0.429	2387
	Yes	0.012	[-0.295, 0.320]	[-0.246, 0.270]	0.157	0.938	0.016	[-0.155, 0.187]	[-0.128, 0.159]	0.087	0.857	

Conservative news share	No	0.239	[-0.116, 0.593]	[-0.059, 0.536]	0.181	0.188	0.069	[-0.107, 0.245]	[-0.079, 0.217]	0.090	0.444	2387
	Yes	0.243	[-0.087, 0.573]	[-0.034, 0.520]	0.168	0.149	0.042	[-0.120, 0.204]	[-0.094, 0.178]	0.083	0.613	
Liberal news share	No	-0.139	[-0.489, 0.211]	[-0.433, 0.155]	0.179	0.436	-0.050	[-0.221, 0.121]	[-0.194, 0.094]	0.087	0.566	2387
	Yes	-0.023	[-0.333, 0.288]	[-0.283, 0.238]	0.158	0.887	-0.070	[-0.234, 0.094]	[-0.207, 0.068]	0.084	0.405	
Entertainment share	No	-0.001	[-0.316, 0.313]	[-0.265, 0.263]	0.160	0.995	0.098	[-0.084, 0.280]	[-0.054, 0.250]	0.093	0.291	2387
	Yes	-0.064	[-0.347, 0.220]	[-0.302, 0.174]	0.145	0.659	0.154	[-0.011, 0.318]	[0.016, 0.291]	0.084	0.067	
Official share	No	-0.133	[-0.440, 0.175]	[-0.391, 0.125]	0.157	0.398	-0.064	[-0.251, 0.123]	[-0.221, 0.093]	0.095	0.502	2387
	Yes	-0.081	[-0.379, 0.218]	[-0.331, 0.170]	0.152	0.597	-0.124	[-0.308, 0.060]	[-0.278, 0.030]	0.094	0.187	
Panel R: Followed Accounts. ITT: Republicans and Independents vs. Democrats (Figure S2.18)												
Republicans and Independents												
Conservative share	No	0.273	[-0.033, 0.579]	[0.016, 0.530]	0.156	0.081	0.008	[-0.143, 0.160]	[-0.119, 0.136]	0.077	0.914	1288
	Yes	0.301	[0.036, 0.566]	[0.078, 0.523]	0.135	0.027	0.010	[-0.118, 0.138]	[-0.097, 0.117]	0.065	0.878	
Liberal share	No	-0.012	[-0.187, 0.164]	[-0.159, 0.136]	0.090	0.896	-0.018	[-0.106, 0.071]	[-0.092, 0.057]	0.045	0.698	1288
	Yes	0.006	[-0.145, 0.158]	[-0.121, 0.134]	0.077	0.933	-0.013	[-0.092, 0.067]	[-0.079, 0.054]	0.041	0.757	
Any political activist share	No	0.211	[-0.010, 0.432]	[0.025, 0.396]	0.113	0.062	-0.040	[-0.163, 0.082]	[-0.143, 0.063]	0.063	0.519	1288
	Yes	0.236	[0.049, 0.424]	[0.079, 0.394]	0.096	0.014	-0.019	[-0.117, 0.080]	[-0.102, 0.064]	0.050	0.712	
Conservative activist share	No	0.304	[0.010, 0.599]	[0.057, 0.551]	0.150	0.043	-0.028	[-0.186, 0.130]	[-0.160, 0.104]	0.080	0.728	1288
	Yes	0.316	[0.052, 0.580]	[0.094, 0.537]	0.135	0.020	-0.030	[-0.164, 0.105]	[-0.143, 0.083]	0.069	0.665	
Liberal activist share	No	0.003	[-0.155, 0.160]	[-0.130, 0.135]	0.080	0.975	-0.013	[-0.099, 0.072]	[-0.085, 0.058]	0.044	0.759	1288
	Yes	0.020	[-0.123, 0.162]	[-0.100, 0.139]	0.073	0.786	0.004	[-0.073, 0.082]	[-0.061, 0.070]	0.040	0.912	
Any news outlet share	No	0.099	[-0.140, 0.338]	[-0.102, 0.299]	0.122	0.419	0.069	[-0.043, 0.180]	[-0.025, 0.162]	0.057	0.227	1288
	Yes	0.159	[-0.068, 0.387]	[-0.031, 0.350]	0.116	0.170	0.036	[-0.067, 0.139]	[-0.050, 0.122]	0.053	0.493	
Conservative news share	No	0.161	[-0.171, 0.492]	[-0.118, 0.439]	0.169	0.343	0.039	[-0.114, 0.192]	[-0.089, 0.167]	0.078	0.619	1288
	Yes	0.217	[-0.076, 0.511]	[-0.029, 0.464]	0.150	0.148	0.013	[-0.122, 0.148]	[-0.100, 0.126]	0.069	0.853	
Liberal news share	No	0.057	[-0.100, 0.213]	[-0.075, 0.188]	0.080	0.479	-0.026	[-0.112, 0.059]	[-0.098, 0.045]	0.044	0.547	1288
	Yes	0.087	[-0.059, 0.233]	[-0.035, 0.209]	0.074	0.243	-0.032	[-0.114, 0.050]	[-0.101, 0.037]	0.042	0.441	
Entertainment share	No	-0.002	[-0.237, 0.233]	[-0.200, 0.195]	0.120	0.986	0.011	[-0.115, 0.137]	[-0.095, 0.117]	0.064	0.864	1288
	Yes	-0.053	[-0.263, 0.157]	[-0.230, 0.123]	0.107	0.618	0.017	[-0.092, 0.125]	[-0.075, 0.108]	0.056	0.765	
Official share	No	0.050	[-0.177, 0.277]	[-0.141, 0.240]	0.116	0.667	-0.007	[-0.121, 0.106]	[-0.103, 0.088]	0.058	0.900	1288
	Yes	0.089	[-0.119, 0.298]	[-0.086, 0.264]	0.106	0.401	-0.021	[-0.122, 0.081]	[-0.106, 0.065]	0.052	0.692	
Democrats												
Conservative share	No	0.047	[-0.030, 0.125]	[-0.018, 0.112]	0.040	0.233	-0.004	[-0.029, 0.022]	[-0.025, 0.018]	0.013	0.784	1099

	Yes	0.040	[-0.030, 0.110]	[-0.019, 0.099]	0.036	0.266	0.003	[-0.023, 0.028]	[-0.019, 0.024]	0.013	0.841	
Liberal share	No	-0.088	[-0.340, 0.165]	[-0.299, 0.124]	0.129	0.496	-0.116	[-0.268, 0.035]	[-0.243, 0.011]	0.077	0.133	1099
	Yes	-0.090	[-0.288, 0.109]	[-0.256, 0.077]	0.101	0.376	-0.125	[-0.247, -0.003]	[-0.227, -0.023]	0.062	0.045	
Any political activist share	No	0.092	[-0.133, 0.317]	[-0.097, 0.281]	0.115	0.425	-0.053	[-0.199, 0.092]	[-0.176, 0.069]	0.074	0.472	1099
	Yes	0.019	[-0.129, 0.168]	[-0.105, 0.144]	0.076	0.797	-0.059	[-0.163, 0.045]	[-0.146, 0.028]	0.053	0.265	
Conservative activist share	No	0.060	[-0.029, 0.150]	[-0.015, 0.135]	0.046	0.186	0.007	[-0.007, 0.020]	[-0.005, 0.018]	0.007	0.334	1099
	Yes	0.051	[-0.028, 0.130]	[-0.015, 0.118]	0.040	0.205	0.014	[-0.000, 0.028]	[0.002, 0.026]	0.007	0.054	
Liberal activist share	No	0.062	[-0.200, 0.325]	[-0.158, 0.283]	0.134	0.642	-0.071	[-0.244, 0.102]	[-0.216, 0.074]	0.088	0.422	1099
	Yes	0.003	[-0.172, 0.178]	[-0.144, 0.150]	0.089	0.971	-0.086	[-0.212, 0.041]	[-0.192, 0.020]	0.065	0.184	
Any news outlet share	No	-0.175	[-0.400, 0.049]	[-0.364, 0.013]	0.115	0.127	-0.002	[-0.145, 0.140]	[-0.122, 0.117]	0.073	0.973	1099
	Yes	-0.171	[-0.371, 0.029]	[-0.339, -0.003]	0.102	0.095	-0.022	[-0.153, 0.109]	[-0.132, 0.088]	0.067	0.742	
Conservative news share	No	0.022	[-0.088, 0.131]	[-0.070, 0.114]	0.056	0.697	0.021	[-0.005, 0.047]	[-0.001, 0.042]	0.013	0.112	1099
	Yes	0.012	[-0.092, 0.116]	[-0.075, 0.099]	0.053	0.817	0.026	[-0.000, 0.051]	[0.004, 0.047]	0.013	0.051	
Liberal news share	No	-0.134	[-0.446, 0.178]	[-0.396, 0.128]	0.159	0.399	-0.016	[-0.169, 0.137]	[-0.144, 0.112]	0.078	0.836	1099
	Yes	-0.134	[-0.397, 0.129]	[-0.355, 0.087]	0.134	0.319	-0.015	[-0.152, 0.121]	[-0.130, 0.099]	0.070	0.826	
Entertainment share	No	-0.002	[-0.229, 0.225]	[-0.192, 0.188]	0.116	0.986	0.096	[-0.038, 0.230]	[-0.017, 0.208]	0.068	0.161	1099
	Yes	0.007	[-0.184, 0.199]	[-0.154, 0.168]	0.098	0.941	0.114	[-0.001, 0.230]	[0.017, 0.211]	0.059	0.053	
Official share	No	-0.187	[-0.410, 0.035]	[-0.374, -0.001]	0.113	0.099	-0.062	[-0.219, 0.096]	[-0.194, 0.071]	0.080	0.443	1099
	Yes	-0.197	[-0.402, 0.008]	[-0.369, -0.025]	0.105	0.060	-0.077	[-0.222, 0.067]	[-0.198, 0.044]	0.074	0.293	
Panel S: Followed Accounts. ITT: Observed Compliers (Figure S2.19)												
Conservative share	No	0.531	[0.040, 1.021]	[0.119, 0.942]	0.250	0.035	0.062	[-0.137, 0.262]	[-0.105, 0.230]	0.102	0.540	359
	Yes	0.402	[0.010, 0.793]	[0.073, 0.730]	0.200	0.050	0.032	[-0.121, 0.185]	[-0.096, 0.161]	0.078	0.679	
Liberal share	No	-0.064	[-0.478, 0.350]	[-0.412, 0.284]	0.211	0.762	0.021	[-0.220, 0.262]	[-0.182, 0.223]	0.123	0.865	359
	Yes	-0.079	[-0.392, 0.235]	[-0.342, 0.185]	0.160	0.626	0.044	[-0.146, 0.235]	[-0.115, 0.204]	0.097	0.647	
Any political activist share	No	0.468	[0.054, 0.882]	[0.121, 0.816]	0.211	0.027	0.048	[-0.196, 0.292]	[-0.157, 0.252]	0.124	0.701	359
	Yes	0.296	[-0.033, 0.625]	[0.020, 0.572]	0.168	0.084	0.023	[-0.156, 0.201]	[-0.127, 0.173]	0.091	0.805	
Conservative activist share	No	0.604	[0.073, 1.135]	[0.158, 1.050]	0.271	0.027	0.066	[-0.149, 0.281]	[-0.114, 0.246]	0.110	0.548	359
	Yes	0.451	[0.022, 0.880]	[0.091, 0.811]	0.219	0.045	0.027	[-0.142, 0.196]	[-0.115, 0.169]	0.086	0.754	
Liberal activist share	No	0.070	[-0.276, 0.417]	[-0.220, 0.361]	0.177	0.690	0.000	[-0.256, 0.257]	[-0.214, 0.215]	0.131	0.997	359
	Yes	0.085	[-0.232, 0.402]	[-0.181, 0.351]	0.162	0.601	-0.006	[-0.210, 0.198]	[-0.177, 0.165]	0.104	0.954	
Any news outlet share	No	-0.065	[-0.475, 0.344]	[-0.409, 0.278]	0.209	0.754	0.151	[-0.062, 0.365]	[-0.028, 0.331]	0.109	0.166	359
	Yes	-0.164	[-0.525, 0.198]	[-0.467, 0.140]	0.184	0.380	0.164	[-0.022, 0.351]	[0.008, 0.321]	0.095	0.086	

Conservative news share	No	0.315	[-0.036, 0.666]	[0.021, 0.610]	0.179	0.079	0.022	[-0.130, 0.175]	[-0.106, 0.150]	0.078	0.775	359
	Yes	0.224	[-0.084, 0.532]	[-0.035, 0.482]	0.157	0.162	0.007	[-0.132, 0.145]	[-0.110, 0.123]	0.071	0.925	
Liberal news share	No	-0.139	[-0.589, 0.311]	[-0.517, 0.238]	0.230	0.544	0.094	[-0.102, 0.291]	[-0.071, 0.259]	0.100	0.349	359
	Yes	-0.175	[-0.515, 0.165]	[-0.460, 0.110]	0.173	0.318	0.146	[-0.020, 0.313]	[0.007, 0.286]	0.085	0.085	
Entertainment share	No	-0.288	[-0.776, 0.200]	[-0.697, 0.122]	0.249	0.248	-0.057	[-0.293, 0.180]	[-0.255, 0.142]	0.121	0.639	359
	Yes	-0.141	[-0.554, 0.271]	[-0.488, 0.205]	0.210	0.505	-0.037	[-0.231, 0.157]	[-0.200, 0.126]	0.099	0.708	
Official share	No	0.070	[-0.275, 0.415]	[-0.219, 0.360]	0.176	0.691	-0.034	[-0.260, 0.192]	[-0.224, 0.156]	0.115	0.770	359
	Yes	0.044	[-0.251, 0.339]	[-0.204, 0.291]	0.150	0.773	-0.012	[-0.213, 0.190]	[-0.181, 0.158]	0.103	0.911	

Table 2.16. Treatment Effects of Switching Feed Settings: Detailed Results

Estimates of switching the algorithm on (left panel) and switching it off (right panel). Left panel: the effect of moving from the chronological to the algorithmic feed for users initially on the chronological feed. Right panel: the effect of moving in the opposite direction for users initially on the algorithmic feed. “Chrono” and “Algorithm” refer to the chronological and algorithmic feeds, respectively. 90% and 95% CIs are reported, as well as standard errors. For each outcome, the results of two specifications are reported. No controls refers to the unconditional estimates with robust standard errors, controlling only for the initial feed setting and, where applicable, pre-treatment outcome levels. Controls ”Yes” refers to the conditional estimates, controlling for pre-treatment covariates using Generalized Random Forests (GRFs). All outcomes are standardized.

3 Survey Questions in Original Layout

3.1 Pre-treatment Survey Questions

YouGov

How often do you use Twitter?

- ◯ **Several times a day**
- ◯ Once a day
- ◯ Several times a week
- ◯ Once a week
- ◯ Several times a month
- ◯ Once a month or less
- ◯ N/A - I do not have a twitter account

YouGov

You are about to participate in a research project on the Twitter newsfeed. This research is being conducted in conjunction with the University of St. Gallen, the Federal Institute of Technology of Zurich, and Paris School of Economics.

As part of this project we would like you to participate in two surveys: this current survey, and another one in six weeks. Both surveys ask you about your political opinions and your Twitter use.

Today's survey will ask you to **switch to a specific setting in your Twitter newsfeed** (if you do not already have this setting on). Please stick with this assigned setting for six weeks, until you are contacted for the follow-up survey.

You will receive **500 points** for taking part in this survey and it should take 8-10 minutes to complete.

If you also stick with the Twitter setting assigned later in this survey and partipicate in the survey in six weeks time, you will receive an **additional 2500 in points**.

Please tick the below if you agree to take part in this study:

- ○ By ticking this box you are giving your consent to participate in this academic study
- ○ I do not wish to continue with this survey

YouGov

Firstly, some questions about **your political opinions**.

YouGov

Generally speaking, do you usually think of yourself as a Democrat, a Republican, an Independent, or what?

- ○ Republican
- ○ Democrat
- ○ Independent
- ○ Some other party

YouGov

Would you call yourself a strong Republican or a not very strong Republican?

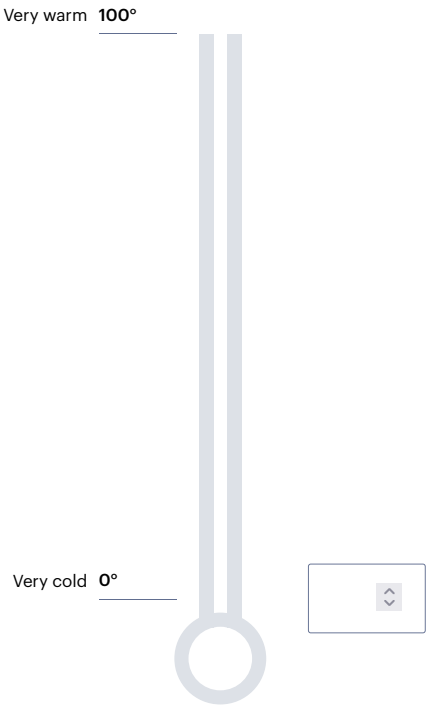
- ○ Strong
- ○ Not very strong
- ○ Not applicable

YouGov

In the list that follows, rate that party using something we call the feeling thermometer. Ratings between 50 degrees and 100 degrees mean that you feel favourable and warm toward the party. Ratings between 0 and 50 degrees mean that you don't feel favourable toward the party and that you don't care too much for that party. You would rate the party at the 50 degree mark if you don't feel particularly warm or cold toward the party.

Democratic Party

Click on the thermometer to give a rating.

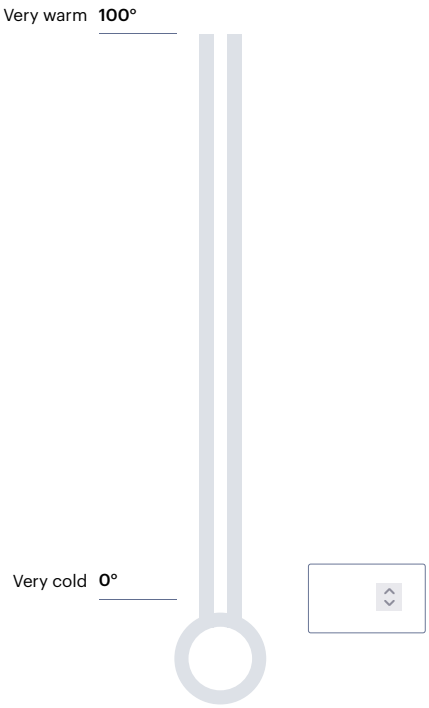


YouGov

In the list that follows, rate that party using something we call the feeling thermometer. Ratings between 50 degrees and 100 degrees mean that you feel favourable and warm toward the party. Ratings between 0 and 50 degrees mean that you don't feel favourable toward the party and that you don't care too much for that party. You would rate the party at the 50 degree mark if you don't feel particularly warm or cold toward the party.

Republican Party

Click on the thermometer to give a rating.



Don't know

YouGov

We now turn to questions about **your life satisfaction and happiness**

YouGov

Please imagine a ladder, with steps numbered from 0 at the bottom to 10 at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time?

Bottom of the ladder **Top of the ladder**

Not sure

YouGov

Taking all things together, would you say you are...

- ○ Very happy
- ○ Rather happy
- ○ Not very happy
- ○ Not at all happy
- ○ Don't know

YouGov

We now turn to questions about **your use of Twitter**

YouGov

What do you use Twitter for? You can select multiple options.

- [ ] Engaging in professional activities
- [ ] Reading news on politics
- [ ] Reading news other than politics (e.g., sports, fashion, people, games)
- [ ] Being entertained (e.g., memes, jokes)
- [ ] Engaging in political activities
- [ ] Connecting with friends and family
- [ ] Other

YouGov

How often do you tweet?

- ○ Several times a day
- ○ Once a day
- ○ Several times a week
- ○ Once a week
- ○ Several times a month
- ○ Once a month or less
- ○ Never

YouGov

As part of this survey, you will now take some action on Twitter.

You must **keep open the browser tab with this survey** while performing the required action on Twitter in a **separate browser tab**.

*As a reminder, you will receive an additional **2500 points** for sticking with the Twitter setting assigned later in this survey and taking a follow up survey in 6 weeks time.*

- ○ I understand I must keep open the browser tab with this survey until I have completed the entire survey
- ○ I do not agree to opening Twitter in a seperate browser and do not wish to participate in this part of the study

YouGov

Our research team studies the **Twitter news feed**.

On Twitter, you can change how other users' content is displayed to you on the Twitter Home Page.

There are two settings.

In a separate browser tab, please check which setting you are currently using. For now, **do not change your current setting.**

To see your setting, go to the Twitter Home Page in this new tab: https://twitter.com/home (https://twitter.com/home) At the top of the screen, you will see **"For You"** and **"Following"** or the corresponding terms in your language.

If the newsfeed currently shows you the tweets Twitter recommends to you, you will see a blue line underneath "For you".

If the newsfeed currently shows you the latest tweets as they appear, you will see a blue line underneath "Following".

Which of the two options is your blue line currently under?

- ○ My blue line is underneath "For you"
- ○ My blue line is underneath "Following"

YouGov

As part of this research project, we will now randomly assign you a newsfeed setting.

Your assigned setting is **"Following"**.

If your current setting is not "Following", please change it to "Following". That is, in your browser tab with the Twitter Home Page, click on "Following" at the top of your newsfeed so it has a blue line underneath it.

Please confirm that your current newsfeed setting is now **"Following"**

*As a reminder, you will receive an additional **2500 points** for sticking with the Twitter setting assigned later in this survey and taking a follow up survey in 6 weeks time.*

- ○ I confirm that my setting is now **"Following"**
- ○ I do not agree to participate in this part of the study

YouGov

Please stick to the **"Following"** setting for the next six weeks -- until we recontact you for the follow-up.

*As a reminder, you will receive an additional **2500 points** for doing this and participating in the survey in six weeks time.*

- ○ I understand that I must not change my newsfeed setting for the next six weeks
- ○ I do not agree to keeping my newsfeed setting the same for the next six weeks

YouGov

Thank you for answering the questions so far! Next we will ask you about sharing additional data for this study. This is optional.

YouGov

What device are you currently useing to complete this survey?

- ○ Desktop/Laptop
- ○ Tablet
- ○ Smart phone
- ○ Other

YouGov

This research is being conducted in conjunction with the University of St. Gallen, the Federal Institute of Technology of Zurich, and Paris School of Economics. They are interested in understanding more about what people see in their Twitter newsfeeds. **To do this, the University of St. Gallen would like to collect 200 tweets from your Twitter newsfeed using Google Chrome browser extension.**

A browser extension is a small software application that can be used to add functionality to a web browser. If you want to take part, you would first need to install the browser extension to your device, and log into Twitter. You would then click on the extension and click on 'Start Parsing' for the extension to collect and automatically download the first 100 tweets shown in your Twitter "For You" and "Following". Once you have downloaded the file of tweets, you would simply upload it within our survey.

Once we have the file, we'll link it to your survey responses and pass the combined information to the University of St. Gallen.

The process should be simple and easy to follow and should not take any longer than 8-10 minutes. If you do wish to participate in this part of the study, you will also receive an additional incentive of **2000 points**, for uploading the file containing the tweets.

The University of St. Gallen will only use the Twitter data for the purposes described above, and you will not be contacted directly as a result of sharing your data with them . You can find out more about how the University of St. Gallen will use your personal data here **privacy notice (https://d2krm5etor5n5r.cloudfront.net/private/95587268ff8af103d64630ef9241ef02/649ece7f49773d7e9f623a57.pdf?Expires=1699464095&Signature=cGn4b~wupf9Dd9aaeXTWmSyktnqz2TZzuow0daYkg85a1QRpumi5BOmCj-5UeaNoQd7ZpmBde~uHObF6wYgpKtXEZqbZTzqPF7zMuFzdNcp4qDmsOLac7fXJGUvP58M9oe1dZhmkygWMkgXmfJgV5pOgYHFGgK4LU192Key-Pair-Id=APKAJVD327QNO34KQAZA)**.

If you decide to participate, you can also opt out of this activity up to two months after the survey by contacting roland.hodler@unisg.ch. After this, the data will be fully anonymised and unidentifiable. If you do not wish to share this data with the University of St. Gallen, that's fine, and your decision will not in any way impact the points you have earned for completing this survey, your eligibility to participate in the follow-up survey for this research project, or your eligibility to participate in further YouGov surveys.

Your consent
By participating in this survey you are consenting to:
• Downloading the Chrome browser extension;
• Uploading a file of tweets to our survey;
• YouGov sharing this data along with your survey responses with the University of St. Gallen; and
• The University of St. Gallen using this data for the purposes outlined above.

Are you happy to install the chrome extension and upload a file containing tweets from your newsfeed?

- ○ Yes, I'm happy to install the chrome extension and upload a file containing tweets from your newsfeed
- ○ No, I'd prefer not to participate in this part of the study

YouGov

Thank you for agreeing to participate in this part of the study. To install the chrome extension please click on the link below:

Chrome Extension (https://chrome.google.com/webstore/detail/haiolgiaiecogbfdciicioeldobccgmh?authuser=0&hl=en)

As a reminder you will need to do the following:
• Download the Chrome extension
• Login to your **Twitter** account
• Go to the Twitter Home Page
• Access the Chrome extension by clicking on the puzzle icon in the upper right corner of your browser
• Click 'Start Parsing'
• Allow 8-10 minutes for the process to take place
• You must leave the tab open, and we recommend not engaging in other activities on your computer while the extension is running (otherwise, the extension might throw an error and have to be relaunched) • A csv file will automatically download once completed
• Upload the csv file below

Once you have the csv file containing the first 200 tweets from your newsfeed, please upload the file here.

Choose file to upload

or drag and drop
them here

YouGov

Thanks for agreeing to share this information.

Please provide your Twitter 'handle' (this is the name that appears at the top of your profile).

Twitter handle

(e.g. “@YouGov”)

3.2 Post-treatment Survey Questions

YouGov

You are about to participate in the follow-up survey for the research project on the Twitter/X newsfeed.

Six weeks ago, you filled out a first survey. Today's survey should take you max. 8-10 minutes.

It asks you about your political opinions and your Twitter/X use. Twitter has been rebranded as X in the meantime. The rebranding does not impact your experiment participation in any way.

You will receive **2500 points** for completing this survey.

Your participation is an invaluable contribution to our research on the societal effects of social media. Thank you!

Should you have any questions, please contact the project leader at: roland.hodler@unisg.ch.

Please tick the below if you agree to take part in this study:

- ○ By ticking this box you are giving your consent to participate in this academic study
- ○ I do not wish to continue with this survey

YouGov

Firstly, some questions about **your political opinions**.

YouGov

Generally speaking, do you usually think of yourself as a Democrat, a Republican, an Independent, or what?

- ○ Republican
- ○ Democrat
- ○ Independent
- ○ Some other party

YouGov

Would you call yourself a strong Democrat or a not very strong Democrat?

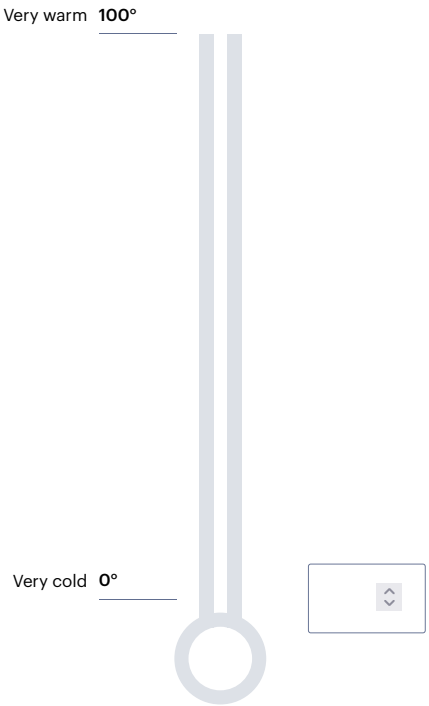
- ○ Strong
- ○ Not very strong
- ○ Not applicable

YouGov

In the list that follows, rate that party using something we call the feeling thermometer. Ratings between 50 degrees and 100 degrees mean that you feel favourable and warm toward the party. Ratings between 0 and 50 degrees mean that you don't feel favourable toward the party and that you don't care too much for that party. You would rate the party at the 50 degree mark if you don't feel particularly warm or cold toward the party.

Democratic Party

Click on the thermometer to give a rating.



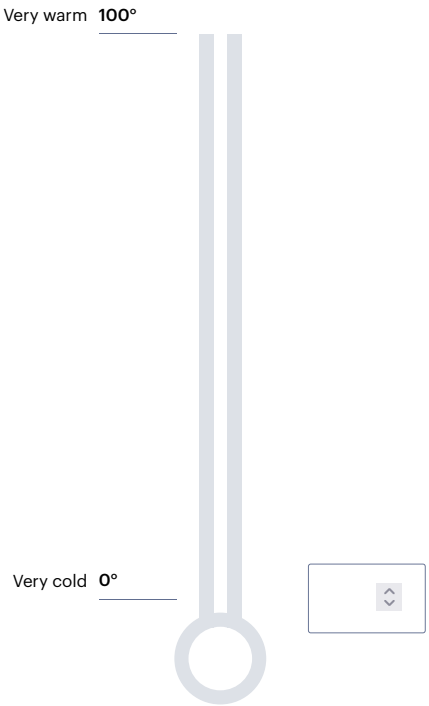
Don't know

YouGov

In the list that follows, rate that party using something we call the feeling thermometer. Ratings between 50 degrees and 100 degrees mean that you feel favourable and warm toward the party. Ratings between 0 and 50 degrees mean that you don't feel favourable toward the party and that you don't care too much for that party. You would rate the party at the 50 degree mark if you don't feel particularly warm or cold toward the party.

Republican Party

Click on the thermometer to give a rating.



Don't know

YouGov

We now turn to questions about **your life satisfaction and happiness**.

YouGov

Please imagine a ladder, with steps numbered from 0 at the bottom to 10 at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time?

Bottom of the ladder **Top of the ladder**

Not sure

YouGov

Taking all things together, would you say you are...

- ○ Very happy
- ○ Rather happy
- ○ Not very happy
- ○ Not at all happy
- ○ Don't know

YouGov

We now turn to questions about **your use of Twitter/X**.

YouGov

For this question, please note that however you answer will not affect the number of points you receive. Being as honest as possible will be incredibly useful to the research.

In the previous survey you were asked to keep your newsfeed setting as "Following".

While participating in this experiment, would you say that you used Twitter/X with the "Following" newsfeed setting (please consider your Twitter/X use on any device, such as computer, tablet, or phone)...

- ○ Always
- ○ Most of the time
- ○ Sometimes
- ○ Rarely
- ○ Never

YouGov

What have you used Twitter/X for over the past few weeks? You can select multiple options.

- [ ] Connecting with friends and family
- [ ] Engaging in professional activities
- [ ] Reading news other than politics (e.g., sports, fashion, people, games)
- [ ] Being entertained (e.g., memes, jokes)
- [ ] Reading news on politics
- [ ] Engaging in political activities
- [ ] Other

YouGov

How often have you used Twitter/X over the past few weeks?

- ○ Several times a day
- ○ Once a day
- ○ Several times a week
- ○ Once a week
- ○ Several times a month
- ○ Less than once a month

YouGov

How often have you tweeted over the past few weeks?

- ○ Several times a day
- ○ Once a day
- ○ Several times a week
- ○ Once a week
- ○ Several times a month
- ○ Once a month or less
- ○ Never

YouGov

Recently, have you found your time spent on Twitter/X a positive or negative experience?

- ○ Very positive
- ○ Somewhat positive
- ○ Neutral
- ○ Somewhat negative
- ○ Very Negative

YouGov

We now turn to questions about some **current political events**.

YouGov

Of the following, what are the most important issues that you want your government to address? Please rank your top 3 issues below:

Drag your choices onto the numbered boxes on the left to rank each of the characteristics below.

1

2

3

- Abortion
- Education
- Crime
- Gun control
- Crisis in Ukraine
- Health care
- Immigration
- Inflation

☐ Don't know

YouGov

Regarding issues that you want the President and U.S. Congress to address, how important do you consider the crisis in Ukraine?

- ○ Not important at all
- ○ One of the least important topics
- ○ A moderately important topic
- ○ One of the most important topics
- ○ The most important topic
- ○ Don't know

YouGov

Now, we are providing you with a list of names and organizations. Please tell us if you have a favorable or unfavorable opinion of each person or organization?

	Strongly unfavorable	Somewhat unfavorable	Neutral	Somewhat favorable	Very favorable	Don't know
Vladimir Putin	○	○	○	○	○	○
NATO	○	○	○	○	○	○
Volodymyr Zelensky	○	○	○	○	○	○

YouGov

Do you approve or disapprove of the job the U.S. government is doing on handling the crisis in Ukraine?

- Strongly disapprove
- Somewhat disapprove
- Neither approve nor disapprove
- Somewhat approve
- Strongly approve
- Don't know

YouGov

Between the Democratic Party and the Republican Party, who in your opinion is best able to handle the crisis in Ukraine?

- ○ Republican
- ○ Democrat
- ○ Both equally
- ○ Neither

YouGov

Which actions should the U.S. take or keep taking to help Ukraine? You can select multiple options.

- [ ] Sending more U.S. troops to European countries other than Ukraine
- [ ] Creating a no-fly zone over Ukraine
- [ ] Placing economic sanctions on Russia
- [ ] Taking Ukrainian refugees
- [ ] Sending military aid to Ukraine, such as weapons and equipment
- [ ] Sending U.S. troops to Ukraine
- [ ] None of these
- [ ] Don't know

YouGov

Thinking about the current indictments and investigations into Donald Trump's actions while in office...

Do you think these indictments and investigations are or are not a deliberate attempt to stop the Trump campaign?

- ○ Are a deliberate attempt to stop the Trump campaign
- ○ Are not a deliberate attempt to stop the Trump campaign

Do you think these indictments and investigations do or do not uphold the rule of law?

- ○ Do uphold the rule of law
- ○ Do not uphold the rule of law

Do you think these indictments and investigations protect or undermine democracy?

- ○ Protect democracy
- ○ Undermine democracy

Do you think these indictments and investigations are or are not an attack on people like you?

- ○ Are an attack on people like me
- ○ Are not an attack on people like me

YouGov

This question is optional.

Now, you can solve a math puzzle to make a donation to the Red Cross Ukraine. That is, if you correctly solve the puzzle, we will donate 1 USD to the Red Cross Ukraine. Whether or not you solve the puzzle will not affect your own remuneration.
If you want to solve it, please write the solution in the field. Otherwise, just leave the field blank.

How many **odd numbers are higher than 12 but smaller than 30?**

YouGov

What device are you currently useing to complete this survey?

- ○ Desktop/Laptop
- ○ Tablet
- ○ Smart phone
- ○ Other

YouGov

Thank you for answering the questions so far! Next we will ask you about sharing additional data for this study. This is optional and you will receive extra points if you chose to participate.

YouGov

This research is being conducted in conjunction with the University of St. Gallen, the Federal Institute of Technology Zurich, and Paris School of Economics. They are interested in understanding more about what people see in their Twitter/X newsfeeds. **To do this, the University of St. Gallen would like to collect 200 tweets from your Twitter/X newsfeed using a Google Chrome browser extension.**

Even if you have already completed this task in an earlier survey, it is very useful for us if you do it a second time.

A browser extension is a small software application that can be used to add functionality to a web browser. If you want to take part, you would first need to install the browser extension to your device, and log into Twitter/X. You would then click on the extension and click on 'Start Parsing' for the extension to collect and automatically download the first 100 tweets shown in your Twitter/X "For You" and "Following". Once you have downloaded the file of tweets, you would simply upload it within our survey.

Once we have the file, we'll link it to your survey responses and pass the combined information to the University of St. Gallen.

The process should be simple and easy to follow and should not take any longer than 8-10 minutes. If you do wish to participate in this part of the study, you will also receive an additional incentive of **2500 points**, for uploading the file containing the tweets.

The University of St. Gallen will only use the Twitter/X data for the purposes described above, and you will not be contacted directly as a result of sharing your data with them . You can find out more about how the University of St. Gallen will use your personal data here **privacy notice (https://d2krm5etor5n5r.cloudfront.net/private/95587268ff8af103d64630ef9241ef02/649ece7f49773d7e9f623a57.pdf?Expires=1699463232&Signature=D9h8FEx2AQfOq2pcBcbM-0S7yMQAiiEow~SyHFh6awnU5H-U37xibfdKrCd23vIGNkIOUQA9gie9jkdHJs5oasFtjjp2hvyVnzclrKtHXFL2KLTdIcUTMXOOQvrZShWO0ZhgxRK7x6FdnCzF1A0h~5DHLze83DLBv5Key-Pair-Id=APKAJVD327QNO34KQAZA)**.

If you decide to participate, you can also opt out of this activity up to two months after the survey by contacting roland.hodler@unisg.ch. After this, the data will be fully anonymised and unidentifiable. If you do not wish to share this data with the University of St. Gallen, that's fine, and your decision will not in any way impact the points you have earned for completing this survey, or your eligibility to participate in further YouGov surveys.

Your consent
By participating in this survey you are consenting to:
• Downloading the Chrome browser extension;
• Uploading a file of tweets to our survey;
• YouGov sharing this data along with your survey responses with the University of St. Gallen; and
• The University of St. Gallen using this data for the purposes outlined above.

Are you happy to install the Chrome extension and upload a file containing tweets from your newsfeed?

- ○ Yes, I'm happy to install the Chrome extension and upload a file containing tweets from my newsfeed
- ○ No, I'd prefer not to participate in this part of the study

>

YouGov

Thank you for agreeing to participate in this part of the study. To install the Chrome extension please click on the link below:

Chrome Extension (https://chrome.google.com/webstore/detail/haiolgiaiecogbfdciicioeldobccgmh?authuser=0&hl=en)

As a reminder you will need to do the following:
• Download the Chrome extension
• Login to your **Twitter/X** account
• Go to the Twitter/X Home Page
• Access the Chrome extension by clicking on the puzzle icon in the upper right corner of your browser
• Click 'Start Parsing'
• Allow 8-10 minutes for the process to take place
• You must leave the tab open, and we recommend not engaging in other activities on your computer while the extension is running (otherwise, the extension might throw an error and have to be relaunched)
• A csv file will automatically download once completed
• Upload the csv file below

Once you have the csv file containing the first 200 tweets from your newsfeed, please upload the file here.

Choose file to upload

or drag and drop
them here

3.3 Reminder Emails

Dear YouGov Member,

In July you participated in a survey on the Twitter newsfeed. As part of this project you were asked to keep your Twitter newfeed setting as **"For You"**.

This is a reminder to keep your Twitter newfeed setting as **"For you"** for the next **four weeks** until we invite you to a second survey as part of this project.

After completing the second survey and keeping your newsfeed setting as **"For you"** , you will receive **2500 points**.

Thank you for participating in this study.

Best wishes,
YouGov Research Team

Fig. 3.22. First Reminder Email

Dear YouGov member,

In July you participated in a survey on the Twitter newsfeed. This is a short reminder that you were assigned to the **Following** newsfeed setting for the duration of the experiment (approximately six weeks). Please use only this newsfeed setting until we re-contact you for a follow-up survey in the next couple of weeks.

Importantly, please stick to the **Following** setting when using Twitter on any device (on desktops, tablets, and smartphones).

As you may have noticed, Twitter has rebranded itself as X. Our experiment remains unaffected by this change.

Best wishes,

YouGov Research Team

Fig. 3.23. Second Reminder Email

4 Instructions for Human Annotators of the Content Data

4.1 User Feed Data

Annotation Instructions: Political Content Labeling

Overview

You will receive a file with the following columns:

- Account_Info
- tweet_index
- account_slant_human
- account_type_human

Your task is to read the account information with specific tweets, then provide two annotations: account political slant and account type. Please ignore the column "tweet_index"; this is just for us to process the data.

Annotation Tasks

1. Account-level Political Slant

- **Column to review**: Account_Info
- **Column for your annotation**: account_slant_human
- **Options**: Conservative, Liberal, Cannot say
- **Instructions**:
    - Consider the account name, example posts, and any knowledge you have about the account (e.g., if it is a famous account)
    - Determine the political leaning based on the US political context
    - Look for subtle cues in language and content that indicate political leaning
    - Choose "Cannot say" if you're uncertain or the determination would be overly speculative

2. Account Type

- **Column to review**: Account_Info
- **Column for your annotation**: account_type_human
- **Options**: News, Political Activist, Entertainment, Official, Other
- **Instructions**:
    - Determine what category best fits the account
    - "News" is for news outlets and other accounts that primarily post news articles
    - "Official" includes government accounts (e.g., POTUS, elected officials)
    - "Political Activist" is for accounts that primarily post political content

- "Entertainment" is for accounts that post memes, jokes, or other recreational content (e.g., games, movies, sports)
- If you are hesitating, it can help to think in terms of topics (e.g., an account can be a news account if it primarily shares and discusses news without necessarily having to be an official newspaper)
- "Cannot say" is NOT an option here -- choose "Other" if unsure

Important Guidelines

No external input

Since you are hired as a human annotator to evaluate NLP-based annotations (including large language models), you MUST NOT use any LLM such as ChatGPT! Also, please do not google or search for the subjects on Twitter. Just annotate based on your current knowledge of the US political context (the tweets are from 2023) and political intuition. If you do not know, choose the option for this case (as described above) but do not consult ANY other sources. **Complying with this rule will ensure that scientific principles are upheld. Given the project scope, we expect public scrutiny of our data; hence, we rely on your utmost integrity.**

Temporary access to the file and no copies allowed

These are data from a survey experiment with human subjects. Therefore, you are not allowed to share this data with anyone. After submission of your annotations, your access to the file will be removed. By accepting this job, you agree not to save copies of the file anywhere else.

Offensive content

These are actual posts from humans. They may cover the entire political spectrum, including extreme viewpoints. They can also include offensive content (e.g., pornography or violence) since such content is not always filtered out on Twitter.

Technical issues

If you spot anything irregular in the data, please contact the research team as soon as possible. For example, if the same account appears twice in the annotation file (this should not happen), please contact us. We are also happy to help with any technical issues you may have.

Submission Format

- Enter annotations directly in the designated columns
- Do not use quotation marks (write Liberal not "Liberal")
- Be consistent with capitalization as shown in the instructions

4.2 Followed Accounts

Annotation Instructions: Political Content Labeling

Overview

You will receive a file with the following columns:

- Account_Info
- account_index
- account_slant_human
- account_type_human

Your task is to read the account information, including the account name and description. Then, provide two annotations: account political slant and account type. Please ignore the column "account_index"; this is just for us to process the data.

Annotation Tasks

1. Account-level Political Slant

- **Column to review**: Account_Info
- **Column for your annotation**: account_slant_human
- **Options**: Conservative, Liberal, Cannot say
- **Instructions**:
    - Consider the account name, account description, and any knowledge you have about the account (e.g., if it is a famous account)
    - Determine the political leaning based on the US political context
    - Look for subtle cues in language and content that indicate political leaning
    - Choose "Cannot say" if you're uncertain or the determination would be overly speculative

2. Account Type

- **Column to review**: Account_Info
- **Column for your annotation**: account_type_human
- **Options**: News, Political Activist, Entertainment, Official, Other
- **Instructions**:
    - Determine what category best fits the account
    - "News" is for news outlets and other accounts that primarily post news articles
    - "Political Activist" is for accounts that primarily post political content
    - "Entertainment" is for accounts that post memes, jokes, or other recreational content (e.g., games, movies, sports)

- If you are hesitating, it can help to think in terms of topics (e.g., an account can be a news account if it primarily shares and discusses news without necessarily having to be an official newspaper)
- "Cannot say" is NOT an option here -- choose "Other" if unsure

Important Guidelines

No external input

Since you are hired as a human annotator to evaluate NLP-based annotations (including large language models), you MUST NOT use any LLM such as ChatGPT! Also, please do not google or search for the subjects on Twitter. Just annotate based on your current knowledge of the US political context (the tweets are from 2023) and political intuition. If you do not know, choose the option for this case (as described above), but do not consult ANY other sources. **Complying with this rule will ensure that scientific principles are upheld. Given the project scope, we expect public scrutiny of our data; hence, we rely on your utmost integrity.**

Temporary access to the file and no copies allowed

These are data from a survey experiment with human subjects. Therefore, you are not allowed to share this data with anyone. After submission of your annotations, your access to the file will be removed. By accepting this job, you agree not to save copies of the file anywhere else.

Offensive content

These are actual posts from humans. They may cover the entire political spectrum, including extreme viewpoints. They can also include offensive content (e.g., pornography or violence) since such content is not always filtered out on Twitter.

Technical issues

If you spot anything irregular in the data, please contact the research team as soon as possible. For example, if the same account appears twice in the annotation file (this should not happen), please contact us. We are also happy to help with any technical issues you may have.

Submission Format

- Enter annotations directly in the designated columns
- Do not use quotation marks (write Liberal not "Liberal")
- Be consistent with capitalization as shown in the instructions

References and Notes

[47] D. M. Blei, A. Y. Ng, M. I. Jordan, Latent Dirichlet Allocation. *Journal of Machine Learning Research* **3** (Jan), 993–1022 (2003).

[48] U. Matter, P. Widmer, *Who Owns the Online Media?*, Tech. rep., SSRN (2021).

[49] R. E. Robertson, *et al.*, Auditing partisan audience bias within Google Search. *Proceedings of the ACM on Human-Computer Interaction* **2** (CSCW), 148 (2018), doi:10.1145/3274417.

[50] D. S. Lee, Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies* **76** (3), 1071–1102 (2009), doi:10.1111/j.1467-937X.2009.00536.x.

[51] H. Tauchmann, Lee (2009) treatment-effect bounds for nonrandom sample selection. *The Stata Journal* **14** (4), 884–894 (2014), doi:10.1177/1536867X1401400411.

[52] S. Athey, J. Tibshirani, S. Wager, Generalized Random Forests. *The Annals of Statistics* **47** (2), 1148–1178 (2019).