News Consumption, Recommender Systems, and Polarization*

Mingduo Zhao

Click here to view the latest version

Abstract

Recommender systems shape how people consume news, possibly reinforcing political polarization. We run two field experiments to identify how user preferences and algorithms interact to amplify partisan news consumption. In the first study, 2,065 U.S. participants use blank Google accounts and a browser extension to track users' activities on Google News. The first-round recommendations are exogenous, allowing us to show that ideologically aligned content draws more clicks. A second experiment uses bots to randomly click on articles, revealing that each click leads to more aligned content. These two pieces of causal evidence establish a feedback loop between user preference and algorithmic recommendations. We also find in the field study that, after interacting with the recommender system, people's level of polarization increases. A structural model combining a discrete choice model (demand side) with a multi-armed bandit algorithm (supply side) confirms this positive-feedback mechanism. The model is then used to simulate a counterfactual "ideology-blind" recommendation policy that ignores political slant when curating content. While this policy reduces polarization, it comes at the cost of likely lower engagement. Overall, the findings provide causal evidence that personalized algorithms reinforce partisan consumption and exacerbate polarization. They also uncover a fundamental trade-off between mitigating polarization and sustaining engagement, which offers important insights for both platform owners and policymakers.

*I am grateful to Ganesh Iyer, Przemyslaw Jeziorski, Steven Tadelis, and Nano Barahona for their invaluable mentorship and advice. I thank Leonardo Arriola, John Canny, Dean Eckles, Frederico Finan, Igal Hendel, Shachar Kariv, Tony Ke, Tai Lam, Gabriel Lenz, Xi Li, Song Lin, Carl Mela, Sridhar Moorthy, Sarah Moshary, Olivia Natan, Yesim Orhun, Ariel Pakes, Navdeep Sahni, Bradley Shapiro, K. Sudhir, Francesco Trebbi, Miguel Villas-Boas, Nils Wernerfelt for helpful comments. The experiments are reviewed and determined to be exempt by UC Berkeley IRB under protocol ID 2023-09-16715. The experiments are pre-registered at AsPredicted (#207062, https://aspredicted.org/9yck-j5j9.pdf and #207064, https://aspredicted.org/f268-4q95.pdf. This work was partially supported by Xlab at UC Berkeley. All errors are my own.

1 Introduction

The issue of political polarization has become increasingly pronounced in modern American society (Pew Research Center, 2014; Allcott et al., 2020). Individuals today appear to find it progressively challenging to find common ground and consensus on contentious political issues. One of the possible explanations could be the influence of social media and digital platforms powered by recommender systems (Rowland, 2011). Unlike traditional media such as newspapers and television which deliver the same content to everyone, today's media landscape is shaped by online platforms and apps utilizing algorithm-based recommender systems that tailor content to individual users. These systems infer user preferences from observed behavior and rapidly adapt the content they deliver to provide personalized experiences (Gillespie, 2014). Nevertheless, there are concerns that such selectively curated content could reinforce users' existing beliefs, limiting their exposure to differing viewpoints (Chaney et al., 2018; Acemoglu et al., 2025). These findings raise two important research questions. First, do recommendation algorithms used by modern media contribute to political polarization, and if so, through what mechanisms? Second, what policy or platform interventions can effectively mitigate algorithmically-driven polarization, and what are the broader implications of such interventions? This paper aims to investigate both questions in depth.

Figure 1 illustrates the core mechanism addressing our first research question—the feedback loop between users' news preferences and the recommender system. Specifically, users' choices—shaped by their underlying preferences—feed into the algorithm and cause the system to generate increasingly slanted recommendations. As a result, as individuals' choice sets become more slanted, they are increasingly likely to encounter and choose ideologically aligned content, thereby further reinforcing the feedback cycle. In this paper, we provide causal evidence for both sides of this loop and demonstrate that this dynamic amplifies individuals' polarization levels.

To unpack the feedback mechanism, we leverage two complementary field experiments. First, a news-browsing study using a Chrome extension records the two-way interaction between

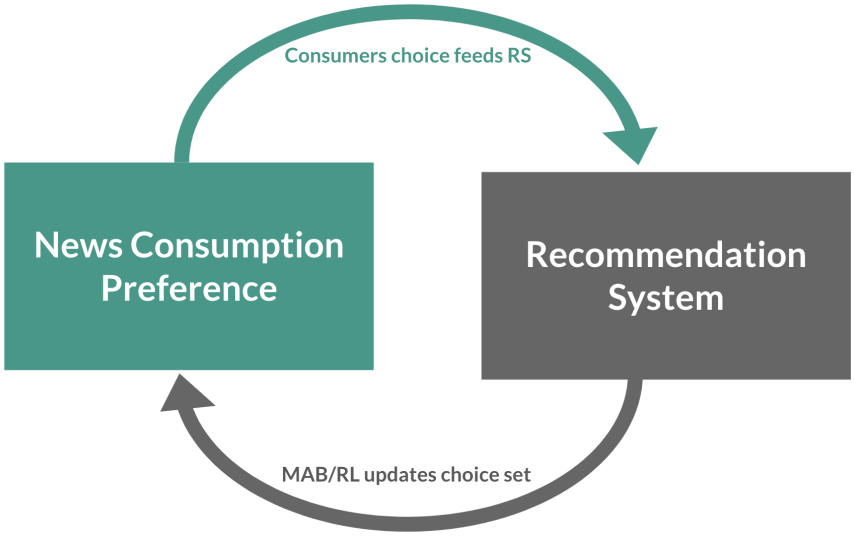


Figure 1: Feedback Loop Between User Preferences and the Recommender System

Note: This figure illustrates the reinforcement cycle between users' news consumption preferences and algorithmic recommender systems (RS). User choices feed into the recommender systems, which in turn updates the news choice set using multi-armed bandit (MAB) or reinforcement learning (RL) methods.

users and the recommender system. Second, a randomized-click experiment with Selenium bots generates exogenous click signals.[1] These designs allow us to isolate how recommendations causally influence later clicks, and how users' clicks shape subsequent recommendations, respectively. We measure the resulting changes in polarization by comparing participants' political survey responses before and after the study. To evaluate a counterfactual recommender system design—specifically, an ideology-blind system that curates content solely based on topic without considering political slant[2]—we develop and estimate a structural model of polarization reinforcement. On the demand side, we show that users consistently prefer content that aligns with their ideological views. The supply-side estimation uncovers strong positive feedback effects, indicating that the system progressively reinforces users' existing content preferences. Using these estimates, our counterfactual analysis shows that implementing an ideology-blind recommendation policy can meaningfully mitigate increases

[1]We use Selenium, an open-source web automation tool, to program bots that simulate user browsing behavior by automatically loading web pages and clicking on news articles according to pre-specified random rules. This allows us to generate exogenous click signals that are not influenced by personal preferences or prior consumption history.

[2]This is discussed in the Algorithmic Accountability Act of 2019, Patankar et al. (2019), and MacCarthy (2020).

in both affective and ideological polarization caused by the recommender system.[3] However, it also produces a notable drop in overall engagement, since users receive fewer articles that match their existing preferences. This highlights a fundamental trade-off: efforts to reduce echo chambers may inherently conflict with platforms' engagement goals.

After providing a general overview of the study, we now describe each field experiment in greater detail and preview some of the results. We begin with the news-browsing field study. In this study, participants first completed a pre-survey designed to measure their baseline levels of political polarization. They were then instructed to log in using a provided blank Google account and install a browser extension built from scratch for this study.[4] The extension recorded detailed, iteration-level data on how users interacted with the recommended content, capturing both their news consumption choices and how the recommendation algorithm adapted in response to clicks.[5]

The data shows a positive correlation between prior recommendations and current user clicks, as well as between current clicks and subsequent recommendations. To identify the causal impact of recommendations on user behavior, we leveraged the exogeneity of the first recommendations. Because the provided new Google accounts had no prior browsing history, the first-round recommendations were free from prior user data, which is confirmed by a balance table. Using variation of the first round recommendations, we find that for the same increase in partisan recommendations, users are 2.3–4.9% more likely to consume politically aligned content compared to misaligned content. These asymmetric responses accumulate over time and interact with the supply side—namely, the behavior of the recommender system, where user clicks influence subsequent recommendations.

[3](Ideological) polarization measures individuals' positions on policy issues, while affective polarization reflects their attitudes toward those with opposing political affiliations. These measures are widely used in the political science and economics literature, including studies such as Pew Research Center (2014); Iyengar et al. (2019); Boxell et al. (2024); Reiljan et al. (2024). For example, ideological polarization is measured through questions like: "How would you evaluate the statement: 'Immigrants today strengthen our country because of their hard work and talents'?" Affective polarization is measured with questions such as: "How comfortable would you feel having a person from the opposing party as a neighbor on your street?"

[4]A detailed description of the Chrome extension we developed is provided in Appendix A.

[5]To ensure compliance, we implemented a back-end machine learning system (MLSys) within the extension. This system verified that participants had logged in using the assigned account and had correctly installed the extension.

Since participants completed surveys both before and after the news-browsing study, we can directly assess how interactions with the recommender system influence their level of polarization. To do this, we administered a survey adapted from Pew Research Center (2014), which measures two key outcomes: ideological polarization, capturing views on policy issues, and affective polarization, reflecting attitudes toward individuals with opposing political affiliations. We find that the interaction between user behavior and algorithmic recommendations intensifies polarization relative to participants' baseline levels prior to system engagement, with statistically significant increases in both measures. On average, participants' polarization levels rose from 0.9 to 2 percentage points after the news-browsing study. These effects were observed across political affiliations, with slightly stronger affective polarization among Democratic participants. This suggests that engagement with personalized recommendations can meaningfully shift political attitudes and exacerbate partisan divisions.

To identify the causal effect of user clicks on subsequent recommendations, we designed a random-click field experiment using automated Selenium bots. Selenium is an open-source tool that enables the automation of web browser actions. These bots mimic user behavior—such as scrolling, clicking, and navigating pages. By programming them to click on news content at random, we created a controlled environment to analyze how the recommender system responds to user behavior. We utilized fifty brand-new Google News accounts with no prior browsing history. Each bot logged into Google News and connected via a VPN configured to simulate a Democratic, Republican, or Battleground region. It then interacted with the recommender system by recording up to 50 news articles before randomly selecting one to click. After each click, the bot immediately recorded the complete set of subsequent recommendations.

The results revealed significant algorithmic adaptation on both extensive and intensive margins. On the extensive margin, engaging with a partisan-leaning article increases the proportion of subsequent recommendations with the same political slant by 4.4%, indicating that clicking leads to increasingly similar content. On the intensive margin, a one-unit increase in a clicked article's stance score produced a 4.1% rise in the mean stance score of subsequent recommendations, demonstrating that more extreme clicks yield more extreme recommenda-

tions.[6] These findings provide compelling evidence that Google's recommendation algorithm dynamically reinforces the ideological preferences signaled by users' clicks.

To evaluate potential solutions to polarization in news consumption and beliefs, we develop a structural model that combines supply and demand. On the demand side, we use a discrete choice model to capture how users select among recommended news articles. On the supply side, we incorporate a multi-armed bandit algorithm that updates news recommendation using Thompson Sampling. Both components are estimated via maximum likelihood. The model's estimates align with the causal evidence: users exhibit strong preferences for politically aligned content, prompting the system to reinforce these preferences by recommending more of the same.

Building on this framework, we simulate counterfactual designs for recommender systems aimed at reducing polarization. Specifically, we model an "ideology-blind" system that selects content based solely on topic, without considering political slant. Although such interventions substantially reduce polarization—bringing users' polarization change close to zero—they likely incur a trade-off in the form of reduced user engagement. Counterfactual analysis indicates that over 60% of users would experience a decline in utility under the ideology-blind policy.[7,8]

All together, these results provide robust causal evidence that the feedback loop between user behavior and algorithmic recommendations amplifies partisan content consumption and polarization. They also provide valuable quantitative estimates to inform the design of recommender systems that more effectively balance the dual objectives of maximizing user engagement and minimizing polarization.

There is a rich literature on targeted advertising and behavioral targeting in Marketing (Iyer et al., 2005; Narayanan and Manchanda, 2009; Dong et al., 2009; Dubé et al., 2017;

[6]The stance score is based on Pew Research Center (2014) and Figure 2, and I rescale the original -10 to 10 scale to a 0 to 1 range, where 0 represents the extreme left and 1 represents the extreme right.

[7]Our news browsing field study did not include an explicit "exit the platform" option. This design choice was made to prevent participants from leaving early and collecting payment without meaningful participation. As a result, we use utility loss as a proxy for engagement loss.

[8]The social welfare implications of an ideology-blind policy depend on how our society values or weighs political polarization.

Athey et al., 2018; Rafieian and Yoganarasimhan, 2021). Previous research has focused on how firms use consumer signals to direct advertisements. Our research builds on this foundation by proposing a framework in which recommender systems leverage user behavior signals to tailor news content delivery. We argue that when these systems are optimized for engagement, they can have unintended consequences on political polarization.

There is also an established strand of literature on polarization spanning Marketing, Economics, and Political Science (Prior, 2007; Levendusky, 2013; Allcott et al., 2020; Iyer and Yoganarasimhan, 2021; Schoenmuller et al., 2022; Boxell et al., 2024; Yoganarasimhan and Iakovetskaia, 2024). Our contribution lies in observing the reinforcement mechanism at the most granular level of platform interaction. At each iteration, users make choices based on their preferences, and the recommender system adjusts accordingly. This iterative feedback loop strengthens ideological alignment over time and contributes to increasing polarization.

Several studies have examined how media outlets adjust their slant to cater to audience preferences and influence voting behavior (Mullainathan and Shleifer, 2005; DellaVigna and Kaplan, 2007; Prior, 2007; Gentzkow and Shapiro, 2010; Levendusky, 2013; Barberá et al., 2015; Athey et al., 2018; Allcott and Gentzkow, 2017; Enikolopov et al., 2020). Much of this work focuses on traditional media, which delivers uniform content to all users and adapts relatively slowly. In contrast, recommender system–based media have the capacity to offer different content to different users at the same time and update recommendations rapidly based on individual behavior. A key contribution of this paper is its causal analysis of adaptive processes at a fine-grained, iteration-level scale, revealing how digital recommender systems can quickly accelerate polarization.

The concept of the 'echo chamber' also appears in the literature, along with alternative designs for recommender systems that might alleviate this problem (Bakshy et al., 2015; Kunaver and Požrl, 2017; Chaney et al., 2018; Jiang et al., 2019; Levy, 2021). Many existing studies in this area measure behavior in a descriptive and correlational manner, often confounded by individuals' past social media activity. Some rely on lab experiments or use custom-built recommender systems rather than those deployed in real-world platforms. In contrast, we contribute by running a field study with users interacting with Google

News's live recommender system. We also causally identify both sides of the feedback loop and further employ economic modeling to evaluate the potential effects of implementing an ideology-blind recommender system, offering quantitative insights for platform designers and policymakers considering such interventions.

The rest of the paper is structured as follows. Section 2 outlines the design of our two field studies. Section 3 presents the main reduced-form results. Section 4 presents the structural model and its estimates for both demand and supply side. Section 5 reports the counterfactual results. Section 6 discusses the managerial and policy implications of our findings. Section 7 concludes.

2 Field Experiments

We implement two field experiments: the news browsing study and the random click experiment. Both are conducted on the Google News platform.[9] Google News is one of the most influential platforms in the digital news ecosystem. Launched in 2002, it serves as a comprehensive aggregator, curating headlines from over 50,000 publishers worldwide and offering users personalized recommendations to current news across more than 35 languages.[10,11] By 2021, Google reported that its Search and News platforms collectively directed users to news publishers' websites over 24 billion times each month, significantly boosting traffic and revenue for media organizations.[12,13] With approximately 119.6 million visits in December 2024, Google News ranks among the top ten most-visited news websites in the United States.[14] Central to its operation is a sophisticated recommender system that personalizes content based on user behavior, aiming to enhance engagement through tailored news experiences (Liu et al., 2010).

[9]Website: https://news.google.com/foryou
[10]Google News: the secret sauce, *The Guardian*, February 25, 2013.
[11]Spreading the News in New Languages, *Google News Blog*, March 4, 2016.
[12]How Google supports journalism and the news industry, *Google*.
[13]Facebook, Google fight bill that would help U.S. news industry, *Reuters*, March 12, 2021.
[14]Top 50 news websites in the US: The Atlantic saw 40% rise in visits amid group chat scoop, *Press Gazette*, April 10, 2025.

2.1 News Browsing Study

This field study aims to explore how algorithmic recommendations and user clicks interact in the real world setting, and to determine whether engagement with Google News's recommender system exacerbates political polarization. It unfolds in three phases—(1) a pre-study survey to measure baseline polarization level, (2) users' direct interaction with the Google News recommendation algorithm, and (3) a post-interaction survey to capture shifts in polarization.[15,16]

We recruited N = 2,065 U.S. participants on Prolific. Each participant was (i) assigned a blank Google account with no prior browsing history, and (ii) instructed to install a custom Chrome extension developed for this project. We assigned blank accounts to participants to eliminate confounding factors such as prior browsing history. This supports a more robust identification of causal relationships. To increase compliance, participants submitted a screenshot for verification, and we cross-referenced it with a backend machine learning system to confirm they logged into the assigned account and installed the extension correctly.

The field study was carried out in three stages:

- **Stage 1: Pre-study survey**

  The study began with a survey. It collected participants' demographic information—age, gender, ethnicity, ZIP code, etc. Most importantly, each participant was also asked to report their political affiliation (Democrat, Republican, or Independent). Participants then completed measures of ideological polarization ("beliefs" regarding social policy issues) and affective polarization ("attitudes" toward members of the opposing party), using Pew Research Center (2014) methodology. These scores served as the baseline for assessing how participants' political beliefs and intergroup attitudes change after engaging with the recommender system.

[15]The detailed study procedure, survey steps, example questions, and survey screenshots are provided in Appendix B.

[16]Polarization in our study is characterized along two dimensions. Behavioral polarization arises from patterns in news consumption and algorithmic recommendation. Our study provides causal evidence of the underlying feedback loop. Stated polarization refers to self-reported ideological and affective polarization measured by surveys.

- **Stage 2: Interaction with the recommender system**

  Following the survey, participants engaged with the Google News platform for about 30-60 minutes in the field using its recommender system. On average, users made around 20 iterations of news selections. The installed extension tracked the articles—headlines, sources, publication times, images, and authors—that participants selected during their browsing sessions and recorded the list of recommendations generated by the platform's algorithm at each iteration.[17]

- **Stage 3: Post-study survey** Immediately after the news browsing, participants completed a post-study survey very similar to the pre-experiment assessment, capturing any shifts in their ideological polarization and affective polarization. By comparing pre- and post-study responses, we can quantify the immediate impact of interacting with the recommender system.[18]

Given the structure of our data, we begin by examining correlations. First, we regress the slant of current recommendations on users' past clicks; next, we regress users' subsequent clicks on the slant of current recommendations. As shown in Table 1, both coefficients are positive and statistically significant. However, these estimates should not be interpreted causally, as they are affected by serial correlation: persistent user preferences over time induce strong autocorrelation in both the click and recommendation sequences.

To identify causal relationships in both directions, we leverage exogenous variation in the independent variables. Specifically, to identify the effect of recommendations on clicks, we take advantage of brand-new Google News accounts with no browsing history. Each

[17] We use the article's source and the news-media classification from Pew Research Center (2014) to identify slant.

[18] Admittedly, this result is best interpreted as the treatment effect of receiving a targeted information intervention followed immediately by a test, which may amplify the measured impact. Additionally, while participants are likely exposed to news in their everyday lives, the level of engagement during the study—typically 30 to 60 minutes of focused interaction—is considerably more intense. This concentrated exposure may produce stronger cognitive or emotional responses than casual or passive media consumption at home. Another key factor is the personalized nature of the content in the experiment. Unlike traditional media sources such as newspapers, television, or radio—which broadcast generalized content to a broad audience and update slowly—the information in this setting is tailored to the individual's interests or prior behaviors. This personalization may make the content feel more relevant, credible, or persuasive, further intensifying its effect.

user's first–round recommendations are determined by two main factors: (1) the platform's population-level priors, which are not user-specific (i.e., what is trendy on the platform now) and (2) the algorithm's randomized exploration mechanism.[19] Population-level priors evolve over time as users enter the study at different points and as trending topics shift. These priors are not driven by individual user preferences. In parallel, the exploration algorithm randomly samples articles to learn which ones resonate with each new user. Because neither component depends on individual prior browsing behavior and their preference, they provide a valid exogenous source of variation to identify how recommendations influence clicks. Conversely, to isolate how clicks shape subsequent recommendations, we run an additional field experiment in which automated bots perform randomized clicks and we then observe the changes in the recommendations they received.

2.2 Random Click Experiment

The objective of this field experiment is to isolate the causal impact of individual clicks on Google News's recommendation by exogenously randomizing clicks. We implemented bots using Selenium to simulate clicking behavior on Google News. Specifically, we wrote a suite of Selenium-based bots that launch a Chrome session for each experimental account, navigate to the recommendation page, and programmatically click on article links. Then we collected article-level metadata for downstream analysis.

In particular, each bot logs into its assigned Google account by following the standard web login process. It is also programmed to handle optional authentication steps, such as backup email verification, if they appear. Once logged in, the bot navigates to the Google News "For You" feed, which is curated by the platform's recommendation system. In each session, the bot uses an IP to randomly connect to one of three predefined location endpoints. Each endpoint corresponds to a different political leaning—Democratic, Battleground, or

[19]Recommender systems typically use Explore-Then-Commit (ETC) and ϵ-Greedy strategies to address the cold start problem in recommender systems (Felício et al., 2017). The ETC strategy divides the process into two phases. The initial phase involves exploration, which corresponds to the period when a user account is newly created. We leverage this period for causal identification, treating it as exogenous. After this phase, the algorithm commits to the option that has performed best during exploration. The ϵ-Greedy strategy incorporates random exploration by choosing a random article with a certain probability.

Republican.[20] Upon landing on the news page for each location, the bot executes a series of interactions to resemble real user engagement. These include randomized scrolling and timed waits to ensure the full page content is rendered. It then identifies up to 50 individual news articles and records essential fields such as the article title, source, link, and publication time. We used fifty brand-new Google News accounts with no prior browsing history. In each click iteration, we select one of the recommended articles at random. Immediately after each click, the bot records the complete set of recommendations generated in response. We repeat this procedure around 100 times for every account so that each click position is drawn independently. Because each account starts without a browsing history and initial clicks are randomly assigned, subsequent changes in recommended content are driven by the experimental clicks. This strategy causally identifies the effect of user clicks on future recommendations.

One potential concern is that the magnitude of the causal effect of clicks may depend on the specific sequence of interactions—that is, path dependence. Random clicks are independent of user preferences and therefore prevent systematic directional reinforcement at each iteration coming from preference. By averaging over randomly generated paths, we aim to minimize the influence of any individual sequence and mitigate bias from path dependence.[21] This approach also yields a valid estimate of our target estimand: the average treatment effect of a click on recommendations across randomly generated paths. We also have a version of this analysis only using the data from the first click in Appendix C1. The results are similar but with larger magnitudes, reflecting the system's heightened sensitivity to click signals when the account has no prior history.

3 Results

In this section, we present the main results of our two experiments. We aim to demonstrate how recommendations and clicks causally reinforce one another and how engaging with the

[20] We are using three locations in our experiment: Sacramento, CA (Democratic-leaning), Saginaw, MI (a battleground area), and Fort Wayne, IN (Republican-leaning). In the 2024 presidential election, the Democratic and Republican vote shares in these locations were approximately 58% vs. 38%, 51% vs. 49%, and 43% vs. 55%, respectively.

[21] We are also currently running a version of the random click experiment using more accounts.

recommender system exacerbates polarization. We begin by presenting correlational evidence of the two-way relationship: how recommendations influence subsequent clicks, and how those clicks, in turn, shape later recommendations. To estimate causal effects, we leverage two sources of exogenous variation embedded in our study design. The first comes from the first-round recommendations shown to newly created accounts with no prior browsing history. The second source is the randomized click behavior generated by automated bots, which simulates clicks independent of user preferences. These allow us to separately identify the causal effect of recommendations on user clicks and how user clicks causally influence subsequent recommendations. Finally, we quantify the impact of interacting with the recommender system on both ideological and affective polarization, conditioning on each user's baseline polarization levels.

3.1 Consumption and Recommendation Are Highly Correlated

This section investigates the correlational feedback loop between user clicks and algorithmic recommendations—specifically, how past user clicks influence future recommendations, and how those recommendations subsequently shape user click behavior. Before estimating these relationships, we first classify each article by political leaning and assign it an ideological stance score. Following Pew Research Center (2014), we assign each recommended news a left-leaning or right-leaning label and apply the corresponding slant score based on its source.[22] Figure 2 presents a visualization of this classification and the associated scores.

Table 1 documents a strong, bidirectional relationship between users' clicking behavior and the algorithm's recommendations: Past user clicks influence the ideological makeup of the next set of recommended articles, and the ideological slant of those recommendations, in turn, affects future clicking behavior. First, Columns (1) and (2) in Panel A indicate a strong association between a user's prior click and the ideological slant of subsequent recommendations. For example, a one-point increase in the stance score of the prior click (ranging from 0 = extreme Democratic to 1 = extreme Republican) is linked to a 3.5 percent increase in the average stance of subsequent recommendations. Additionally, clicking on a

[22] In Pew Research Center (2014), media slant scores range from -10 to 10. A score below 0 indicates the news source is left-leaning, while a score above 0 indicates it is right-leaning.

Panel A	Outcome: Rec_t	
	Mean Stance Score	Share of Republican
	(1)	(2)
$Click_{t-1}(score)$	0.035***	
	(0.001)	
$Click_{t-1}(rep)$		0.033***
		(0.001)
Observations	43,219	43,219
Panel B	Outcome: $Click_t$	
	is Democrat	is Republican
	(1)	(2)
$Rec_t(dem)$	0.942***	
	(0.021)	
$Rec_t(rep)$		0.985***
		(0.026)
Observations	22,788	14,395
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 1: Correlation Between User Clicks and Recommendations

Note: Using data from the news-browsing study, this table reports the correlations between users' clicking behavior and the recommendation outcomes. Panel A shows coefficients from regressing the current recommendations (Rec_t) on the previous iteration's clicks (Click_{t-1}). The two click measures are: $\text{Click}_{t-1}(score)$, the stance score of the prior click (0 = extreme Democratic, 1 = extreme Republican), and $\text{Click}_{t-1}(rep)$, an indicator for whether the prior click was Republican-leaning. Stance scores and partisan classifications follow Pew Research Center (2014) with rescaling. The outcome variables are the average stance score of the recommendations and the proportion that are Republican-leaning. Panel B shows coefficients from regressing the next iteration's clicks (Click_t) on the current recommendations (Rec_t). Here, $\text{Rec}_t(dem)$ is the proportion of recommended articles that are Democrat, and $\text{Rec}_t(rep)$ is the proportion of recommended articles that are Republican. The outcome variables are the indicators for whether the clicked article is Democrat or Republican. The number of observations in Panel A is 43,219, representing the full sample. In Panel B, the regressions are conducted separately for users identified as Democrats (22,788 observations) and Republicans (14,395 observations). The sum of the two partisan groups is less than the full sample because users who identified as independents or other affiliations are not included from this subgroup analysis.

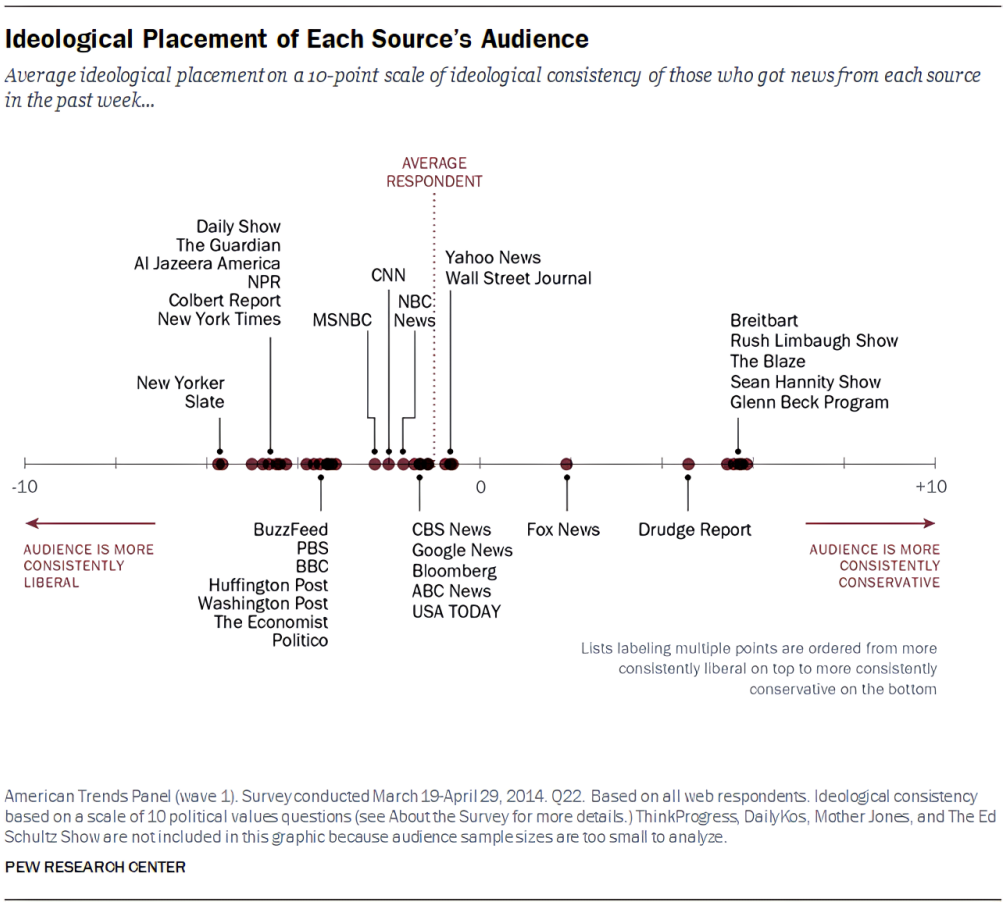


Figure 2: Ideological Slant of U.S. News Audiences

Note: This figure is taken from (Pew Research Center, 2014), section "Media Outlets by the Ideological Composition of Their Audience". It shows the ideological leanings of audiences for major U.S. media outlets. The horizontal axis represents ideological consistency, with negative values indicating more consistently liberal audiences and positive values indicating more consistently conservative audiences.

Republican-leaning article is associated with a 3.3 percentage point increase in the share of Republican recommendations. In short, as users continue clicking, they tend to receive not only more recommendations that are aligned, but ones that are more ideologically extreme. Both relationships are highly statistically significant. Second, Panel B presents results from a reversed specification, examining how the slant of prior recommendations correlates with subsequent clicking behavior. A one-point increase in the share of Democratic-leaning recommendations is associated with a 0.942 increase in the likelihood of clicking on a Democratic article, while a one-point increase in Republican recommendations corresponds to a 0.985 increase in the probability of a Republican click. These statistically significant coefficients suggest that users tend to click on content that is recommended more. Taken together, these

patterns point to a strong correlational feedback loop: Click behavior correlates with more ideologically aligned future recommendations, and exposure to such content is in turn linked to increased future engagement. While these estimates are correlational and may not be precise in magnitude due to autocorrelation, they remain informative. While not causal, they provide a useful benchmark for understanding the overall patterns in the data.

3.2 Impact of Recommendations on Clicks

This section presents causal evidence on how recommendations influence user clicks. We begin by validating our key identifying assumption: That the first-round recommendations presented to accounts without prior browsing history are exogenous to users' pre-existing characteristics.[23] To verify this exogeneity, we conduct a balance test on the ideological slant of those initial recommendations.

Table 2 displays the balance test for our primary source of exogenous variation. We examined the first-round recommendations across various demographic groups, including gender, age, political affiliation, education, income, and ethnicity. The analysis shows no statistically significant differences in first round recommendations across these demographic categories. These results validate that the first-round recommendations are evenly distributed across individuals and not influenced by their pre-existing traits, supporting their exogeneity.

Since the first-round recommendations are exogenous to users' pre-existing characteristics and independent of their prior behavior, we use them to causally identify their impact on participants' immediate clicking behavior—that is, their first click. The specification we estimate is:

$$Click_i = \alpha + \beta Rec_i + \gamma' X_i + \epsilon_i, \tag{1}$$

where $Click_i$ denotes news consumption for individual i at the first iteration. To account for partisan heterogeneity, we run separate regressions for Democrats and Republicans, reported in Rows 1 and 2 of Table 3, respectively. The variable Rec_i denotes the share of recommended

[23] As discussed earlier, the initial content mix likely consists of a combination of trending items and randomly selected content. This approach aligns with the Explore-Then-Commit strategy, which is commonly employed to address the cold-start problem in recommender systems.

Table 2: Recommendation Balance Table

	% Democrat News (P-value)
By Gender	
Female	50.752
Male	50.549
Non-binary	50.476
F-test	(0.971)
By Age	
$18-44$	50.455
$45-59$	51.585
60+	50.149
F-test	(0.558)
By Political Affiliation	
Democrat	51.392
Independent	50.673
Republican	49.417
F-test	(0.125)
By Education	
High school or less	49.445
College or bachelor's degree	51.330
Associates or technical degree	50.026
Graduate or professional degree	49.664
F-test	(0.287)
By Income	
$25,000 - $49,999	50.028
$50,000 - $74,999	50.590
$75,000 - $99,999	51.409
$100,000 - $149,999	50.130
$150,000 and more	51.363
F-test	(0.825)
By Ethnicity (ranked alphabetically)	
American Indian or Alaska Native	57.130
Asian	49.328
Black or African American	49.832
Hispanic or Latino	49.090
Middle Eastern	51.447
Multiracial/Mixed ethnicity	49.414
Native Hawaiian or Other Pacific Islander	65.473
Other/Not specified	45.878
White	51.022
F-test	(0.429)

Note: This table reports the proportion of Democrat-leaning news articles recommended in the first round across various demographic groups, including political stance, age, gender, education, income, and ethnicity (analyzed separately). For each category, the p-value of an F-test of the null hypothesis that all subgroups received the same proportion of Democrat news is reported in parentheses.

content with a Democratic (Row 1) or Republican (Row 2) slant shown to individual i at the first iteration, depending on whether the user is a Democrat (Row 1) or a Republican (Row 2). The indicator variable $Click_i$ equals 1 if the individual clicks on content aligned with their political affiliation—Democratic for Row 1 and Republican for Row 2. Put it simply, Row 1 of Table 3 reports the effect of exposing Democratic users to more Democratic-leaning content, while Row 2 presents the corresponding effect for Republican users exposed to more Republican-leaning content. X_i denotes a vector of control variables capturing the proportion of recommendations falling into each topic—namely politics; economy & business; technology; health; science; sports; entertainment; education; environment; and lifestyle—where topic labels are assigned by an LLM (ChatGPT).

	Outcome: *Click*	
	is Democrat	is Republic
	(1)	(2)
Rec(*dem*)	1.024***	
	(0.094)	
Rec(*rep*)		1.011***
		(0.117)
Covariates	Yes	Yes
Observations	1,062	642
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 3: Causal Effect of Recommendations on Clicks

Note: This table reports how the ideological composition of first-round news recommendations affects users' initial consumption choices. The key independent variables are the proportions of partisan content in the recommendation slate. Corresponding outcomes are the indicators of whether the clicked news is Democrat-leaning or Republican-leaning. Columns 1 and 2 report Democrat and Republican subsample respectively. All models control for the distribution of recommended topics—politics; economy & business; technology; health; science; sports; entertainment; education; environment; and lifestyle—based on LLM classification.

Table 3 reports the estimated causal effects of recommendation slant on user click behavior.[24] For the same increase in partisan recommendations, users are 2.2–4.9% more likely to consume politically aligned content than misaligned content.[25] Specifically, Democrats

[24]For how Table 3 relates to the 2.2–4.9% estimates shown, see the footnote calculations on the next page.

[25]We define a recommendation or click as "aligned" if the political slant of the article matches the user's self-reported party affiliation provided in the pre-survey.

(Row 1) are 4.9% more likely to click on aligned content compared to misaligned content.[26] Similarly, Republicans (Row 2) are 2.2% more likely to engage with aligned content relative to misaligned content.[27] These findings suggest that Democrats exhibit a stronger preference for in-group content, a pattern consistent with the demand-side estimates presented in Section 4.1. In summary, we find causal evidence that a higher share of politically aligned content in the recommendation slate increases the likelihood that users click on content matching their own political affiliation.

3.3 Impact of Clicks on Recommendations

After establishing the causal effect of recommendations on clicks—the first half of the feedback loop—we now turn to the second: estimating how clicks influence subsequent recommendations. To identify this relationship, we conducted a randomized-click field experiment using automated Selenium bots (described in Section 2). Each bot was assigned a blank Google News account, received a list of recommended articles, and, in each iteration, clicked one of the presented articles uniformly at random. Immediately following the click, the bot recorded the next set of recommended articles. By leveraging the randomness of the click as an exogenous source of variation, we estimate how the prior click—measured both as a continuous stance score and a binary partisan label—affects the slant of the following recommendations. The specification we run is:

$$Rec_{it+1} = \alpha + \beta Click_{it} + \epsilon_{it}, \tag{2}$$

where $Click_{it}$ denotes the random click made by bot i at iteration t. The variable Rec_{it+1} captures the recommender system's response in the next iteration—specifically, the average ideological slant of recommended content and the share of Republican-leaning articles. Since all clicks are random, these results identify the causal effect of a single exogenous click on subsequent recommendations. The findings reveal a clear pattern: each additional click

[26] A 1% increase in partisan recommendations leads to approximately a 1.024% increase in clicks on ideologically aligned news for Democrats. In contrast, the increase in clicks on misaligned content is only 0.976%, implying a 4.9% higher likelihood of engaging with aligned content, i.e., $\frac{1.024-0.976}{0.976} \approx 0.049$.

[27] A 1% increase in partisan recommendations results in a 1.011% increase in clicks on aligned content for Republicans, compared to 0.989% for misaligned content, yielding a 2.2% difference, i.e., $\frac{1.011-0.989}{0.989} \approx 0.022$.

leads the system to serve more ideologically aligned content. It also pushes recommendations toward more extreme positions.

	Outcome: Rec_{t+1}	
	Mean Stance Score	Share of Republican
	(1)	(2)
$Click_t(score)$	0.041***	
	(0.003)	
$Click_t(rep)$		0.044***
		(0.003)
Individual Fixed Effect	Yes	Yes
Observations	4,210	4,210
R^2	0.033	0.042
Adjusted R^2	0.033	0.042
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 4: Causal Effect of Clicks on Recommendations

Note: This table reports the causal impact of randomized click events on the political slant of subsequent recommendations. Column (1) shows the effect of the stance score of the prior random click ($\text{Click}_t(score)$) on the mean stance score of the next recommendation set (Rec_{t+1}: Mean Stance Scores); Column (2) shows the effect of clicking a Republican-leaning article ($\text{Click}_t(rep)$) on the proportion of Republican-leaning recommendations (Rec_{t+1}: Proportion of Republican). Stance scores and partisan labels follow the Pew Research Center's 2014 classification of news outlets Pew Research Center (2014).

The results of the random-click experiment appear in Table 4. They reveal significant effects on both the intensive and extensive margins. On the intensive margin, a one-unit increase in the previous click's stance score produces a 4.1% rise in the mean stance score of the next recommendations, showing that more extreme clicks yield more extreme recommendations. On the extensive margin, clicking a Republican-leaning article increases the proportion of Republican-leaning recommendations by 4.4%.[28] Together, these findings demonstrate that the recommendation algorithm dynamically adapts to each user's interactions, reinforcing the ideological preferences signaled by their clicks.[29]

[28]Since outlets are categorized dichotomously following Pew Research Center (2014), a similar symmetric effect is present for Democratic clicks.

[29]First-click analysis in Appendix C1 shows that one click results in approximately a 7–8% increase in subsequent recommendations. We do not interpret this as an average treatment effect, since the recommendation algorithm is especially responsive to click when the account has no prior history.

3.4 Recommender Systems Exacerbate Polarization

Now that we have a clear understanding of how user choices and system recommendations causally reinforce each other, we aim to examine the treatment effect of this interaction on individuals' polarization level. At this stage of the study, we investigate how interacting with the recommender system influences participants' political views, specifically focusing on two outcomes: attitudes toward individuals with opposing views (affective polarization) and participants' political views on social issues (ideological polarization). Changes in these measures are calculated by comparing participants' responses to political surveys, before and after the news-browsing study.

	Treatment			
	Before	After	Difference	P-value
	(1)	(2)	(3)	(4)
Outcome I: Affective Polarization				
Republicans	48.457	49.362	0.905^{***}	< 0.001
Democrats	63.284	65.292	2.008^{***}	< 0.001
Outcome II: Ideological Polarization				
Republicans	57.123	58.163	1.039^{***}	0.001
Democrats	83.161	84.923	1.762^{***}	< 0.001
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 5: Shifts in Affective and Ideological Polarization

Note: This table presents the measures of (1) affective polarization (attitudes toward individuals with opposing views) and (2) ideological polarization (political views on social issues) before and after our study, segmented by two groups: Republican and Democratic. Both outcomes are scaled from 0 to 100, where higher values indicate a greater degree of polarization. Column 3 and 4 reports the differences between pre- and post-study polarization levels, along with the corresponding p-values from a paired t-test. A positive difference implies that individuals became more polarized following the study.

Table 5 shows the results.[30] Both affective polarization and ideological polarization measures are scaled from 0 to 100, where higher values indicate greater levels of polarization. For affective polarization, we observe statistically significant changes in both partisan groups, with different magnitudes. Before the study, the affective polarization score for Republicans was 48.457, which increased to 49.362 after the engaging with the platform, resulting in a significant difference of 0.905. Democrats exhibited a larger and notable change, with an increase from 63.284 to 65.292, resulting in a difference of 2.008. These findings suggest that interaction with the recommender system heightened negative attitudes toward individuals with differing political views across both partisan groups, with Democrat showing a somewhat stronger response. For ideological polarization regarding political beliefs, we observed modest but statistically significant changes for both groups. Among Republicans, the polarization score increased from 57.123 to 58.163, an increase of 1.039. Similarly, Democrats showed a change of comparable magnitude, with scores rising from 83.161 to 84.923, an increase of 1.762.

These findings suggest that 30-60 minutes interaction with the recommender system over approximately 20 iterations has a small but statistically significant impact on both polarization measures.[31] For Republicans, increases in both measures suggest that exposure to ideologically aligned recommendations may intensify negative perceptions of opposing groups and reinforce more rigid policy stances. Democrats exhibited slightly more sensitivity to the treatment in terms of affective polarization but showed a similarly significant shift in ideological polarization. This suggests that recommender systems can indeed shape political attitudes. If individuals engage with the system over a longer and more sustained period,

[30]Since the Google News recommender system cannot be easily disabled, constructing a conventional control group is not feasible. Consequently, the news browsing study is best understood, at least in part, as an observational study. The identification of the recommender system's effect on user clicks relies on the random and exogenous nature of the first-round recommendations. To estimate treatment effects on polarization, we employ a before-and-after comparison of users' responses—effectively using each participant as their own control.

[31]In large part, we interpret the short term treatment in the study affects participants' beliefs rather than their underlying preferences. Changes in attitudes toward certain social issues may result simply from exposure to related content that shifts their beliefs. Notably, the feedback loop can function even in the absence of any change in preferences. Changes in polarization levels over the long term may differ from those observed in the short term. These longer-term effects depend on the degree to which participants continue to engage with similar content after the study concludes. The extent of this continued exposure can further shape their beliefs and, over time, potentially influence their underlying preferences as well.

these effects may persist—and may even intensify—over time. This suggests the value of a future long-term study spanning several months or years. However, implementing such a study poses significant challenges, including high attrition rates and substantial costs associated with maintaining participant participation over extended durations.[32]

4 Structural Model

This section employs a structural model to examine the effects of ideology-blind recommender systems on users and platforms.[33] An ideology-blind recommendation policy curates content without considering political slant. Our research question addresses a critical issue in digital platform design: How would an ideology-blind recommender system affect user polarization and platform engagement? To answer this question, we develop a structural model that integrates both the demand and supply side. To model the demand side, we estimate a discrete choice framework that captures users' news choice at each iteration of engagement with the platform. The results show that both Democrats and Republicans exhibit a clear preference for content that aligns with their political affiliation. The supply side is modeled using Thompson Sampling within a multi-arm bandit framework, which captures how the recommender system selects the articles offered to consumers. We use the model to simulate the counterfactual impact of implementing ideology blindness in a recommender system. Our structural approach allows us to quantify the tradeoffs involved: while ideology-blind recommendations mitigate increases in polarization, they also lead to utility losses for a significant portion of users, ultimately reducing overall platform engagement.

4.1 Demand Side

The dataset records each user's sequence of interactions with the recommender system, including the content presented by the system and the selection made by the user at each iteration. For causal identification on the demand side, we restrict our analysis to the first

[32]We conducted a follow-up political survey 30 days after the experiment concluded, with results presented in Appendix D2. Compared to the immediate post-experiment survey, the longer-term results are more mixed. Ideological polarization tends to return to baseline levels once users stop engaging with Google News. However, there is some persistence in affective polarization.

[33]This is discussed in the "Algorithmic Accountability Act of 2019".

round of interaction—that is, users' responses to the initial recommendations.[34] Consistent with the logic outlined in Section 3.2, these first-round recommendations are exogenous, as they are generated without prior user behavior. This enables us to recover preference parameters without concerns of autocorrelation.

We model the utility that consumer i derives from news article j in the first round as:

$$U_{ij} = \beta_i^d D_{ij} + \beta_i^r R_{ij} + \beta_i^x X_{ij} + \varepsilon_{ij}, \tag{3}$$

where D_{ij} and R_{ij} are binary indicators for whether the article leans Democratic or Republican, and X_{ij} is a vector capturing the article's topic composition.[35] The error term ε_{ij} follows a Type I extreme-value distribution.

In this specification, β_i does not indicate individual-level heterogeneity, as such heterogeneity cannot be identified from cross-sectional data based solely on the first round of interaction. The only source of heterogeneity we incorporate is users' political affiliation, as reported in the pre-survey. Specifically, we define the utility coefficients as:

$$\beta_i^d = S_i \times \beta_r^d + (1 - S_i) \times \beta_d^d, \quad \beta_i^r = S_i \times \beta_r^r + (1 - S_i) \times \beta_d^r, \tag{4}$$

where $S_i = 1$ if individual i identifies as Republican, and $S_i = 0$ if identified as Democrat. Thus, for Democratic users ($S_i = 0$), we have $\beta_i^d = \beta_d^d$, and for Republican users ($S_i = 1$), we have $\beta_i^r = \beta_r^r$. Our primary parameters of interest are β_d^d and β_r^r, which reflect the marginal utility that Democrats derive from Democratic content and Republicans derive from Republican content, respectively.

Since there is no outside option—participants could not exit the study before fulfilling the

[34] We also estimate users' preferences using data from all rounds. Results are reported in the Online Appendix, Section E.

[35] Both D_{ij} and R_{ij} can be included since the model does not have an intercept.

requirements[36,37]—the probability that consumer i selects news item j in the first round follows the standard logit choice model:

$$p_{ij} = \frac{\exp(V_{ij})}{\sum_{k \in \mathscr{C}_i} \exp(V_{ik})}, \tag{5}$$

where V_{ik} represents U_{ik} without the idiosyncratic shock ε_{ij}, and $\mathscr{C}_i$ denotes the choice set in the first round for consumer i.

The likelihood function for an individual consumer is:

$$L_i = \prod_{j \in \mathscr{C}_i} (p_{ij})^{d_{ij}}, \tag{6}$$

where d_{ij} is an indicator variable that equals 1 if consumer i clicked on news j in the first round, and 0 otherwise.

By pooling consumers within each political group (i.e., Democratic or Republican), the likelihood function for group S can be expressed as:

$$L_S = \prod_{i \in S} L_i = \prod_{i \in S} \prod_{j \in \mathscr{C}_i} (p_{ij})^{d_{ij}}, \tag{7}$$

where p_{ij} denotes the predicted probability that consumer i selects item j, and d_{ij} is an indicator equal to 1 if item j was chosen by consumer i, and 0 otherwise.

Finally, we estimate the parameters separately for the two groups—Democratic and Republican—denoted as $\{\beta_s\}_{s \in \{d,r\}} \equiv \{(\beta_s^d, \beta_s^r, \beta_s^x)\}_{s \in \{d,r\}}$, using MLE as follows:

$$\{\hat{\beta}_s\}_{s \in \{d,r\}} = \left\{ \arg\max_{\beta_s} \log L_S(\beta_s) \right\}_{s \in \{d,r\}} = \left\{ \arg\max_{\beta_s} \sum_{i \in S} \sum_{j \in \mathscr{C}_i} d_{ij} \log \left(\frac{\exp(\beta_s' Z_j)}{\sum_{k \in \mathscr{C}_i} \exp(\beta_s' Z_k)} \right) \right\}_{s \in \{d,r\}}. \tag{8}$$

[36]This restriction prevents participants from immediately leaving the study to claim payment without providing meaningful input.

[37]Participants were required to either make 20 effective choices or remain in the study for at least 30 minutes. The effectiveness of their choices was determined by a backend machine learning system that analyzed scrolling behavior, click timing intervals, and related metrics to estimate whether users were engaging attentively. Participants were informed that inattentive or careless clicks would not be counted as participation, which helps explain why they spent, on average, 1–2 minutes per choice.

where Z_k is a vector that includes all information related to news topics (X_k) and news slant (D_k, R_k).

Since there is no outside option in the choice set, we cannot normalize utility relative to an outside alternative. In the results presented in Table 6, we normalize the utility coefficients by setting β_d^r—the indirect utility that a Democratic user derives from consuming Republican-leaning news—as the baseline. All other estimates—β_r^r and β_d^d—are relative to this.[38] The coefficients β_r^r and β_d^d represent the average indirect utility that Republican users derive from right-leaning content and Democratic users derive from left-leaning content, respectively, and serve as the key parameters of interest in our analysis.

Parameters	Estimate
β_r^r	0.312***
	(0.097)
β_d^d	0.405***
	(0.027)
Note:	***p<0.01

Table 6: Demand Side Estimation

Note: This table presents the parameter estimates from the demand-side models, with robust standard errors reported in parentheses. β_r^r represents the marginal utility gain from a piece of Republican news for individuals with a republican stance, relative to β_d^r. The model also controls for topic-specific possibility variables, which account for various news categories, including "politics", "economy & business", "technology", "health", "science", "sports", "entertainment", "education", "environment", and "lifestyle".

Table 6 presents the estimate from our demand-side model, which quantifies the strength of political-alignment preferences among users. The estimates highlight key patterns in partisan news consumption. The coefficient β_r^r, which measures the marginal utility Republican users derive from right-leaning content, is positive and statistically significant, indicating a clear preference for ideologically aligned news for Republicans. Likewise, β_d^d captures the marginal utility Democratic users gain from left-leaning content and is notably larger than β_r^r. This suggests that Democrats, on average, display an even stronger preference for aligned content than Republicans. Since our model already controls topic, these alignment effects likely

[38] Since adding a constant to all utilities does not affect choice probabilities, the absolute levels of the β coefficients are not identified. However, their relative differences remain informative and can meaningfully reflect consumer preferences between partisan news sources.

reflect a taste for ideologically aligned content rather than mere subject-matter preferences.[39] Users—particularly Democrats—tend to prefer news that aligns with their political views. As a result, platforms that account for political slant in their recommendation algorithms may have an incentive to promote aligned content. We confirm this by the supply-side pattern in Section 4.2. Furthermore, these results feed directly into our counterfactual exercise: an ideology-blind recommendation approach would erode user satisfaction and, in turn, may reduce engagement.[40]

4.2 Supply Side

Demand-side estimation shows that users prefer news aligned with their political slant. If a recommender system could directly observe users' political affiliation, it would naturally leverage this information by predominantly serving aligned content. In reality, the platform must infer each user's affiliation from their interactions, creating an exploration–exploitation tradeoff: the algorithm must experiment with different content to learn a user's political leaning (exploration) while also capitalizing on its current belief to maximize engagement (exploitation). We model this problem as a multi-armed bandit problem with Thompson sampling. A large body of research demonstrates that Thompson sampling—widely adopted in various domains, including recommender systems—effectively balances exploration and exploitation under certain regularity conditions (Thompson, 1933; Honda and Takemura, 2014). We implement this framework to evaluate whether—and by how much—the platform's sequential learning process updates its slant and reinforces users' political preferences.

We define a "round" as the moment when a new set of recommendations is generated following a user click; each click and its resulting recommendations constitute one round. As part of our study design, the extension refreshes the page each time a consumer selects a piece of news. The number of news articles that is recommended in round t is denoted as M_t. The slant of the j-th news article recommended to consumer i at time t is represented by $s_{ijt} \in \{0, 1\}$, where $s_{ijt} = 1$ indicates right-wing content, and $s_{ijt} = 0$ indicates left-wing

[39] We assume that the utility derived from a topic and its political slant is additively separable.

[40] As noted earlier, our news browsing field study did not offer an explicit "exit the platform" option. This design choice aimed to discourage participants from receiving compensation without meaningful engagement. Consequently, we use utility loss as a proxy for reduced engagement.

content. For each individual i and round t, we observe the platform's action in terms of the proportion of right-wing news recommended, denoted as $a_{it} \in [0, 1]$. The proportion of left-wing news recommended is then $1 - a_{it}$. By definition, a_{it} is the average slant of the recommended news articles: $a_{it} = \frac{\sum_{j=1}^{M_t} s_{ijt}}{M_t}$. The type of news consumed by consumer i at time t can be represented as a binary variable $c_{it} \in \{0, 1\}$, where 0 indicates left-wing news consumption and 1 indicates right-wing news consumption. The true political type of a consumer is unknown to the platform. When a consumer's political type aligns with the type of news they consume, the reward for that round is higher. The reward can be interpreted as either the time a user spends engaging with the news platform or their willingness to continue interacting with the system. The recommender system, such as Google News, seeks to maximize consumer engagement. The signals sent by the consumers to the platform are a sequence of choices: $\{c_{i1}, c_{i2}, c_{i3}, \dots\}$.

As described above, we frame the recommendation platform's problem as a multi-armed bandit problem and implement Thompson sampling to manage the exploration–exploitation trade-off. While Thompson Sampling is known to be optimal under certain regularity conditions (e.g., Honda and Takemura (2014)), we treat our specification as a reduced-form calibration. The model is intended to capture the fundamental trade-off that underlies any engagement-optimizing behavioral targeting system, regardless of the specific algorithm it employs. Although the true underlying algorithm—such as Google's—may follow a different specification, possibly based on more complex approaches like reinforcement learning, the essential mechanism we model remains representative of the broader optimization logic.[41]

The system's prior belief for user i is initialized as

$$a_{i0} \sim \text{Beta}(\alpha_0, \beta_0). \tag{9}$$

[41] That said, certain recommender systems explicitly implement designs based on Thompson Sampling, as documented in Knyazev and Oosterhuis (2023).

We have the following updates after observing click c_{it}

$$\text{If } c_{it} = 0, \quad a_{i,t+1} \sim \text{Beta}(\alpha_t,\ \beta_t + c_b), \tag{10}$$

$$\text{If } c_{it} = 1, \quad a_{i,t+1} \sim \text{Beta}(\alpha_t + c_a,\ \beta_t). \tag{11}$$

Here, α_0 and β_0 encode the platform's initial prior belief, while c_a and c_b control how strongly each click updates the right- and left-leaning posterior, respectively. Our parameters of interest are $\Theta = \{\alpha_0, \beta_0, c_a, c_b\}$.

The specification we use can be viewed as a reduced-form heuristic. It is designed to capture the fundamental trade-off faced by any recommender system that infers user preferences from observed behavior—namely, the balance between exploitation and exploration. Notably, our updating procedure departs a bit from the classic Bayesian updating of a Beta distribution. In our settings, Bayesian updating does not yield a conjugate posterior. A detailed proof is provided in Appendix G. Our specification approximates the learning process in a tractable and interpretable way. We introduce separate updating parameters, c_a and c_b, to allow for asymmetric responses to Republican and Democratic clicks. This asymmetry reflects the possibility that the mapping from clicks to inferred user preferences is not symmetric, and may result from heuristic-based design choices in the recommendation algorithm—potentially made by engineers aiming to maximize "reward". This perspective aligns with design considerations discussed in the recommender system literature, such as Knyazev and Oosterhuis (2023). Another reason for deviating from standard Bayesian updating is that our model serves as a simplified mapping to approximate a complex reinforcement learning algorithm that the firm might actually use in reality. We adopt this interpretable formulation because we lack access to Google's actual recommendation algorithm, yet aim to capture the core trade-off it likely embodies. We also estimate a nonparametric version of the updating process for α_t and β_t in Appendix Figure I5. The results provide additional validation for the specification used in this study. They show that it closely mirrors the behavior of the actual system, particularly in capturing the asymmetric responses to partisan clicks. Moreover, the updating rule appears to be approximately linear, consistent with observed patterns.

To estimate these parameters, we employ maximum likelihood. For each user i with browsing history $\{c_{i1}, c_{i2}, ..., c_{iT}\}$, let

$$rep_{it} = \sum_{k=1}^{t} c_{ik}, \quad dem_{it} = t - rep_{it} \tag{12}$$

denote the cumulative counts of right- and left-leaning clicks up to round t. The posterior after t observations is

$$\text{Beta}(\alpha_0 + c_a\, rep_{it},\ \beta_0 + c_b\, dem_{it}), \tag{13}$$

The log-likelihood over all users and iterations is

$$\log L(\Theta; dem_{it}, rep_{it}) = \sum_{i=1}^{N} \sum_{t=1}^{T} \log \left(\frac{x^{\alpha_0 + c_a\, rep_{it} - 1} (1-x)^{\beta_0 + c_b\, dem_{it} - 1}}{\int_0^1 t^{\alpha_0 + c_a\, rep_{it} - 1} (1-t)^{\beta_0 + c_b\, dem_{it} - 1}\, dt} \right), \tag{14}$$

and we estimate

$$\hat{\Theta} \ = \ \arg\max_{\Theta} \log L(\Theta; dem_{it}, rep_{it}). \tag{15}$$

Our supply-side calibration results, presented in Table 7, provide insights into how recommender systems dynamically adapt to user behaviors and reinforce content preferences, which could contribute to polarization. The initial parameters of the Beta distribution (α_0, β_0) are of similar magnitude, which reveals that without any user interaction history, the recommender system begins with an initial recommendation that is pretty balanced with a slight rightward slant. That is, the system maintains reasonable political diversity in its initial recommendations.

More importantly, the parameters that govern the updating process (c_a, c_b) demonstrate the system's learning and adaptation process. These positive and highly significant coefficients confirm that the recommender system actively reinforces user preferences through a feedback loop: when users select right-leaning content, the system increases the probability of

Parameter	Estimate
α_0	4.949^{***}
	(0.049)
β_0	4.298^{***}
	(0.045)
c_a	0.067^{***}
	(0.005)
c_b	0.090^{***}
	(0.006)
Note:	***p<0.01

Table 7: Supply Side Calibration

Note: The table presents the parameter estimates from the supply-side models, with robust standard errors reported in parentheses. α_0 and β_0 represent the initial parameter of the beta distribution that characterizes the proportion of Republican-leaning news among the recommended news. The parameter c_α (c_β) denotes the change applied to α_0 (β_0) after observing the political stance of clicked news. This adjustment reflects the reinforcement of the political stance recommendations based on consumer interactions, capturing the dynamic nature of news recommendations as they adapt to consumer preferences.

recommending similar content in subsequent rounds, and likewise for left-leaning content.[42] Interestingly, the larger magnitude of $\hat{c}_b$ compared to $\hat{c}_a$ suggests that the system responds more strongly to signals of left-leaning content than right-leaning content, potentially accelerating preference reinforcement for left-leaning users.

The finding of positive and statistically significant updating parameters (c_a and c_b) on the supply side carries important implications when considered together with our demand side results. Our demand-side estimates reveal that users exhibit strong preferences for politically aligned content. When these preferences are reinforced by a recommender system that adjusts its recommendations accordingly, it constitutes a powerful mechanism for potential polarization. Each click on politically aligned content not only gives the user's immediate utility for consumption but also increases the likelihood of receiving similar content in the future, potentially creating increasingly homogeneous information environments. This dynamic

[42]We also estimate the supply-side model using the random click data, as reported in Appendix F4. In that specification, both (c_a, c_b) remain positive and statistically significant. The point estimates differ slightly. We do not use that specific specification here because the experiment was conducted during an active political campaign period, which caused the initial recommendations to deviate substantially from a neutral 50% baseline. Moreover, if we conceptualize the clicking sequence as branching paths, bots—due to their randomized clicking behavior—do not systematically follow one side of the branches as humans tend to do. Instead, their behavior resembles a random walk, which leads to supply-side estimates that diverge from the patterns observed in actual human trajectories.

feedback process helps explain why our counterfactual analysis finds that an ideology-blind recommender system would mitigate polarization while decreasing overall engagement. By disrupting the reinforcement mechanism captured here, such a system would no longer cater to the strong alignment preferences revealed in our demand-side estimation.

5 Counterfactual

We have shown that user preferences and the recommender system causally reinforce one another, thereby contributing to increased polarization. In addition, we have estimated and calibrated both the demand-side and supply-side models. With these components established, we proceed to evaluate the impact of an ideology-blind recommendation policy, as referenced in the Algorithmic Accountability Act of 2019. In this counterfactual, the system would recommend articles solely by topics, without regard to political slant, allowing us to assess whether such an approach would mitigate the polarization effects we observe and to what level it also harms engagement.

5.1 Polarization Level

We implement the ideology-blind counterfactual and investigate its impact in three steps. First, we randomly reassign the political slant of each recommended article in the news browsing study dataset, while keeping the topic composition fixed. We then use our estimated demand model to compute the choice probabilities under the modified recommendation slants. We then simulate each individual's click sequence d_{ijt} multiple times based on the counterfactual choice probabilities. The variable d_{ijt} is a dummy equal to 1 if consumer i clicks on news article j in round t, and 0 otherwise. Finally, using the empirically estimated linear mapping from news choices to polarization change from our original data, we translate each simulated click sequence into a predicted shift in polarization, and average these shifts across all simulated draws. Table 8 outlines the implementation using the notation introduced in Section 4. A more detailed description is provided in Appendix H. The corresponding results are reported in Table 9.

Counterfactual Simulation

Compute the overall share of Republican news (p) across all recommendations.
for each consumer $i \in \{1, \ldots, N\}$ do:
  for each recommendation round $t \in \{1, \ldots, T\}$ do:
    for each piece of news $j \in \mathscr{C}_{it}$ do:
      Reassign R_j^c such that round t contains a proportion p of Republican news
      Compute counterfactual utility using the original topic and reassigned stance:
        $\hat{V}_{ijt}^c = \hat{\beta}_i Z_{ijt}^c = \hat{\beta}_i^d D_{ijt}^c + \hat{\beta}_i^r R_{ijt}^c + \hat{\beta}_i^x X_{ijt}$
      Compute the counterfactual choice probability for item j:
        $p_j^c = \frac{\hat{V}_{ijt}^c}{\sum_{k \in \mathscr{C}_{it}} \hat{V}_{ijt}^c}$
        Simulate the counterfactual news choice dummy $\hat{d}_{ijt}$ based on $\{p_{ijt}^c\}_{j \in \mathscr{C}_{it}}$
    end for
  end for
  Count the number of counterfactual Democratic (dem_i^c) and Republican (rep_i^c) choices.
  Compute counterfactual polarization change using estimated coefficients:
    $\Delta\text{Polar}_i^c = \hat{\gamma}_1 rep_i^c + \hat{\gamma}_2 dem_i^c$.
end for

Table 8: Counterfactual Simulation Steps

Note: This table outlines the steps used in the counterfactual simulation. The set $\mathscr{C}_{it}$ represents the recommended news articles for consumer i at round t in the news browsing study. The utility coefficients β_i are estimated via Maximum Likelihood Estimation. The political slant of each news article is reassigned using the overall Republican share p, computed from the union of all recommendation sets (assumed to reflect the platform's full recommendation pool). The counterfactual polarization change is computed based on article-level consumption, using coefficients $\hat{\gamma}_1$ and $\hat{\gamma}_2$, which are estimated via OLS by regressing observed changes in polarization on the number of Republican and Democratic articles selected during the news browsing study. Superscripts c are used to indicate counterfactual objects.

The counterfactual results presented in Table 9 provide compelling evidence of mitigation on polarization level under the ideology-blind design.[43] These findings reveal several insights with implications for platform design and regulatory considerations. In the observed data in Table 5, we see substantial increases in polarization across both partisan groups and both ideological polarization measures. The counterfactual scenario (Table 9) tells a different story. By implementing an ideology-blind recommender system, we observe substantially different outcomes. In the counterfactual scenario, the changes in both polarization and affective polarization are nearly zero—and even negative in some cases. This likely reflects users' more frequent exposure to counter-attitudinal content, which substantially reduces their ideological polarization and affective polarization levels.

Specifically, we find small reductions in ideological polarization under the counterfactual scenario: 0.060 points for Democrats and 0.054 points for Republicans. These changes contrast sharply with the increases under the standard recommender system, suggesting that an ideology-blind approach can slow—and potentially reverse—polarization on policy-related dimensions. For affective polarization, Republicans exhibit a modest increase of 0.291 points, which represents a 68% decline from the 0.905-point increase seen under the regular system that reinforces existing beliefs. In contrast, Democrats experience a statistically significant decrease of 0.286 points, effectively undoing the earlier 2.008-point increase. This asymmetric pattern indicates that Democrats may be more responsive to ideologically diverse content than Republicans. Taken together, these findings indicate that an ideology-blind recommender system could not only avoid contributing to polarization but may also foster modest depolarization in both attitudinal and policy-related domains.

Additionally, these findings also underscore the substantial role that political alignment in standard recommender systems plays in driving polarization. Specifically, removing ideological targeting significantly slows—or even reverses—the growth in polarization. This

[43]Based on the procedure outlined in Table 8 and Appendix H, we compute the counterfactual change in polarization using a linear model trained on data from our field study. Specifically, we use individuals' news consumption choices as explanatory variables and estimate a linear model where the outcome variable is the observed change in polarization from the news browsing study. We then apply this estimated model to the counterfactual news choices in order to predict the corresponding change in polarization under the counterfactual scenario.

	After-Before	
	Democrat	Republican
	(1)	(2)
Affective Polarization	-0.286***	0.291***
	(0.008)	(0.007)
Ideological Polarization	-0.060***	-0.054***
	(0.006)	(0.002)
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 9: Simulated Polarization Change

Note: This table reports changes in polarization levels following the study, for the counterfactual scenario. Standard errors are reported in parentheses. Measures include Democratic/Republican Affective Polarization and Ideological Polarization. Column 1 reports the polarization change for Democrats. Column 2 reports the polarization change for Republicans.

finding provides empirical support for longstanding concerns about algorithmic filter bubbles and echo chambers, quantifying their precise impact on polarization dynamics. They suggest that even some adjustments to recommendation algorithms' treatment of political content could yield significant reductions in polarization without completely eliminating the personalization features on topics that drive user engagement. However, the polarization-reducing benefits of an ideology-blind recommender system must be considered alongside its potential costs. Our demand-side estimates suggest that, given users' strong preferences for ideologically aligned content, such an approach would likely lead to lower user satisfaction and decreased platform engagement. This highlights the fundamental tension between business incentives that favor engagement optimization and societal concerns about political polarization. The next section examines the impact of ideology blindness on platform engagement.

5.2 Platform Engagement

Having demonstrated that ideology-blind recommender systems can mitigate polarization, we now turn to assessing their potential cost—specifically, a reduction in user engagement. To quantify this trade-off, we calculate the percentage of users whose expected utility falls under

the counterfactual recommendation policy. Specifically, we compare each user's expected utility under the Google News current recommender system with that under the ideology-blind design.

Following the procedure described in Table 8, we simulate news choices B times under both the current recommender system and a counterfactual ideology-blind system. For each simulation, we compute the associated utility, allowing us to estimate the individual's expected utility under each of the two recommender systems design. We then compute each user's expected utility under the current recommendation policy as $\hat{V}_i = \frac{1}{B}\sum_b \sum_t \sum_{j\in\mathscr{C}_{it}} \hat{d}_{ijt}\hat{V}_{ijt}$, and under the ideology-blind counterfactual as $\hat{V}_i^c = \frac{1}{B}\sum_b \sum_t \sum_{j\in\mathscr{C}_{it}} \hat{d}^c_{ijt}\hat{V}^c_{ijt}$.[44] The utility difference between the two policies is then given by $\Delta\hat{V}_i = \hat{V}_i^c - \hat{V}_i$, which serves as a proxy for changes in user engagement.

	Group		
	Democrats	Republicans	Overall
	(1)	(2)	(3)
$(\Delta\hat{V}_i < 0)$ %	67.2***	48.0***	60.3***
	(1.88)	(2.67)	(1.57)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 10: Percent of Consumer Experiencing Utility Loss

Note: This table reports the percentage of consumers who experience a utility loss in the counterfactual scenario, broken down by group (Democrat, Republican, and Overall). Standard errors are reported in parentheses.

The results in Table 10 show the percentage of users who would experience utility loss under an ideology-blind recommender system compared to the regular personalized system. This measurement helps us approximate the potential engagement cost of implementing this policy. Overall, 60.3% of users would experience some degree of utility loss when exposed to ideology-blind recommendations. This finding is statistically significant at the 1% level.

[44] Recall that utility is computed as $\hat{V}_{ijt} = \hat{\beta}_i^d D_{ijt} + \hat{\beta}_i^r R_{ijt} + \hat{\beta}_i^x X_{ijt}$, and under the counterfactual as $\hat{V}^c_{ijt} = \hat{\beta}_i^d D^c_{ijt} + \hat{\beta}_i^r R^c_{ijt} + \hat{\beta}_i^x X_{ijt}$, where the superscript c denotes counterfactual quantities. The simulation of choice $\hat{d}_{ijt}$ is based on probability estimation as $\hat{p}_{ijt} = \frac{exp(\hat{V}_{ijt})}{\sum_{k\in\mathscr{C}_{it}} exp(\hat{V}_{ikt})}$. Under the counterfactual, the simulation of $\hat{d}^c_{ijt}$ is based on $\hat{p}^c_{ijt} = \frac{exp(\hat{V}^c_{ijt})}{\sum_{k\in\mathscr{C}_{it}} exp(\hat{V}^c_{ikt})}$.

When broken down by political affiliation, we see that Democrats would be slightly more impacted than Republicans, with 67.2% of Democrats experiencing utility loss compared to 48.0% of Republicans. This outcome is expected, given that our demand estimates indicate Democrats exhibit a stronger preference for ideologically aligned content.

These findings in Section 5 quantify the trade-off between user engagement and polarization that both platform designers and policymakers must consider. While implementing ideology-blind recommendations reduces polarization, it comes at a cost: 60% of users experience a decline in utility. While this approach provides societal benefits by reducing polarization, it comes at the likely cost of reduced user engagement. Moreover, it underscores important short- and long-term trade-offs that platform designers must carefully evaluate. The key insight is that although polarization may boost engagement in the short run, it can undermine long-term engagement by fostering a more divided and fragmented user base. We discuss this in the next section.

6 Managerial and Policy Implication

Our paper reveals a clear trade-off platform that owners and policymakers face when considering recommender systems. Ideology-blind recommender systems reduce polarization but at a likely cost of reduced user engagement. This presents them with competing objectives. Platform owners and policymakers must trade off between the societal benefits of reduced polarization and engagement losses. Policymakers could consider regulating platforms to reduce slant-based targeting. If necessary, regulators might also explore compensating platforms for adopting ideology-blind recommender systems, enabling them to internalize the societal costs of polarization rather than treating it as an externality.

Additionally, platform owners face another trade-off—one not formally examined in this paper but potentially highly relevant: the tension between short-term and long-term user engagement. They must balance the boost in short-term engagement from polarizing content against its long-term risks. Although stories predominantly from one side can drive immediate clicks and time on site, polarization on content recommendation in the long run

could foster a hostile environment and alienate users seeking diverse perspectives (Cinelli et al., 2021). This could ultimately erode long-term engagement. To achieve sustainable growth, platform owners might want to explore hybrid approaches, striking an optimal balance between short-term and long-term user engagement. For instance, platforms could adopt moderate diversity constraints that guarantee some exposure to cross-cutting content, while still permitting a degree of slant-based personalization to mitigate the impact on short-term user engagement.

For policymakers, our findings provide quantitative evidence for potential regulatory approaches. Instead of requiring fully ideology-blind systems, regulators could implement minimum diversity standards that balance the goals of reducing polarization with maintaining platform viability and user satisfaction. Regulators may also need to consider compensating platforms that adopt ideology-blind recommender systems. These more nuanced policy approaches recognize that while such systems can meaningfully enhance social cohesion by reducing polarization, their implementation must be carefully designed to avoid substantial engagement losses that could undermine platform sustainability.

7 Conclusion and Future Work

In this paper, we provide causal evidence of both sides of the feedback loop between users' news preferences and the recommender system. We also show that this dynamic exacerbate individuals' polarization level. We leverage two field experiments to unpack this feedback loop. First, a news-browsing field study conducted via a Chrome extension captures the two-way interaction between users and the recommender system. Importantly, the first recommendation is exogenous to user behavior and preference. Second, a randomized-click field experiment with Selenium bots generates exogenous click signals. These designs allow us to isolate how recommendations causally influence later clicks, and how users' clicks shape subsequent recommendations, respectively. Next, we quantify the resulting changes in polarization by comparing participants' pre- and post-study survey responses. We then estimate two complementary models: On the demand side, Maximum Likelihood Estimation of the choice model confirms that individuals tend to favor ideologically aligned news. On

the supply side, a multi-armed bandit model calibrated using Thompson Sampling reveals strong positive feedback effects that amplify users' content preferences. Building on these estimates, our counterfactual analysis shows that an ideology-blind recommendation policy—one that ignores political slant and curates contents solely by topic—can substantially mitigate increases in both affective and ideological polarization content. However, it also produces a notable drop in overall engagement, since users receive fewer articles that match their existing preferences. This tradeoff underscores a fundamental dilemma: initiatives to reduce echo chambers may conflict with platforms' short-term engagement goals. To address this tension, regulators may need to consider compensating platforms for efforts that reduce polarization.

Looking forward, we see three promising avenues for further exploration. First, platforms can test implementation of ideology-blind features on a small subset of users to validate our counterfactual estimates in real-world settings. Second, researchers could explore alternative algorithmic designs—beyond pure blindness—that strike an optimal balance between minimizing polarization and preserving user engagement, framing this as a multi-objective optimization problem. Finally, building richer longitudinal datasets that follow users over several months or years—an effort that could cost millions of dollars—would allow for the measurement of the long-term effects of recommendation-driven polarization.

References

Acemoglu, D., Ozdaglar, A., and Siderius, J. (2025). AI and social media: A political economy perspective. 2

Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., and Yang, D. (2020). Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. *Journal of Public Economics*, 191:104254. 2, 7

Allcott, H. and Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–236. 7

Athey, S., Calvano, E., and Gans, J. S. (2018). The impact of consumer multi-homing on advertising markets and media competition. *Management Science*, 64(4):1574–1590. 7

Bakshy, E., Messing, S., and Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132. 7

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., and Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, 26(10):1531–1542. 7

Boxell, L., Gentzkow, M., and Shapiro, J. M. (2024). Cross-country trends in affective polarization. *Review of Economics and Statistics*, 106(2):557–565. 4, 7

Chaney, A. J., Stewart, B. M., and Engelhardt, B. E. (2018). How algorithmic confounding in recommendation systems increases homogeneity and decreases utility. In *Proceedings of the 12th ACM Conference on Recommender Systems*, pages 224–232. 2, 7

Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., and Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences of the United States of America*, 118(9):e2023301118. 38

DellaVigna, S. and Kaplan, E. (2007). The fox news effect: Media bias and voting. *Quarterly Journal of Economics*, 122(3):1187–1234. 7

Dong, X., Manchanda, P., and Chintagunta, P. K. (2009). Quantifying the benefits of individual-level targeting in the presence of firm strategic behavior. *Journal of Marketing Research*, 46(2):207–221. 6

Dubé, J.-P., Fang, Z., Fong, N., and Luo, X. (2017). Competitive price targeting with smartphone coupons. *Marketing Science*, 36(6):944–975. 6

Enikolopov, R., Makarin, A., and Petrova, M. (2020). Social media and protest participation: Evidence from Russia. *Econometrica*, 88(4):1479–1514. 7

Felício, C. Z., Paixão, K. V., Barcelos, C. A., and Preux, P. (2017). A multi-armed bandit model selection for cold-start user recommendation. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, UMAP '17, page 32–40, New York, NY, USA. Association for Computing Machinery. 11

Gentzkow, M. and Shapiro, J. M. (2010). What drives media slant? evidence from U.S. daily newspapers. *Econometrica*, 78(1):35–71. 7

Gillespie, T. (2014). The relevance of algorithms. *Media Technologies: Essays on Communication, Materiality, and Society*, 167(2014):167. 2

Honda, J. and Takemura, A. (2014). Optimality of Thompson sampling for Gaussian bandits depends on priors. In *Artificial Intelligence and Statistics*, pages 375–383. PMLR. 27, 28

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., and Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, 22(1):129–146. 4

Iyer, G., Soberman, D., and Villas-Boas, J. M. (2005). The targeting of advertising. *Marketing Science*, 24(3):461–476. 6

Iyer, G. and Yoganarasimhan, H. (2021). Strategic polarization in group interactions. *Journal of Marketing Research*, 58(4):782–800. 7

Jiang, R., Chiappa, S., Lattimore, T., György, A., and Kohli, P. (2019). Degenerate feedback loops in recommender systems. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pages 383–390. 7

Knyazev, N. and Oosterhuis, H. (2023). A lightweight method for modeling confidence in recommendations with learned Beta distributions. In *Proceedings of the 17th ACM Conference on Recommender Systems*, RecSys '23, page 306–317, New York, NY, USA. Association for Computing Machinery. 28, 29

Kunaver, M. and Požrl, T. (2017). Diversity in recommender systems–A survey. *Knowledge-based Systems*, 123:154–162. 7

Levendusky, M. S. (2013). Why do partisan media polarize viewers? *American Journal of Political Science*, 57(3):611–623. 7

Levy, R. (2021). Social media, news consumption, and polarization: Evidence from a field experiment. *American Economic Review*, 111(3):831–870. 7

Liu, J., Dolan, P., and Pedersen, E. R. (2010). Personalized news recommendation based on click behavior. In *Proceedings of the 15th international conference on Intelligent user interfaces*, pages 31–40. 8

MacCarthy, M. (2020). *An examination of the Algorithmic Accountability Act of 2019*. SSRN. 3

Mullainathan, S. and Shleifer, A. (2005). The market for news. *American Economic Review*, 95(4):1031–1053. 7

Narayanan, S. and Manchanda, P. (2009). Heterogeneous learning and the targeting of marketing communication for new products. *Marketing Science*, 28(3):424–441. 6

Patankar, A., Bose, J., and Khanna, H. (2019). A bias aware news recommendation system. In *2019 IEEE 13th International Conference on Semantic Computing (ICSC)*, pages 232–238. IEEE. 3

Pew Research Center (2014). Political polarization in the american public. https://www.pewresearch.org/journalism/2014/10/21/political-polarization-media-habits/. 2, 4, 5, 6, 9, 10, 13, 14, 15, 20, 48

Prior, M. (2007). *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge University Press. 7

Rafieian, O. and Yoganarasimhan, H. (2021). Targeting and privacy in mobile advertising. *Marketing Science*, 40(2):193–218. 7

Reiljan, A., Garzia, D., Da Silva, F. F., and Trechsel, A. H. (2024). Patterns of affective polarization toward parties and leaders across the democratic world. *American Political Science Review*, 118(2):654–670. 4

Rowland, F. (2011). The filter bubble: What the internet is hiding from you. *portal: Libraries and the Academy*, 11(4):1009–1011. 2

Schoenmuller, V., Netzer, O., and Stahl, F. (2022). Polarized America: From political polarization to preference polarization. *Marketing Science Frontiers*. 7

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294. 27

Yoganarasimhan, H. and Iakovetskaia, I. (2024). From feeds to inboxes: A comparative study of polarization in facebook and email news sharing. *Management Science*, 70(9):6461–6472. 7

Appendix

Table of Contents

A Google News Extension Overview 45

B Procedure of the News-browsing Study 46

C Effect of Random Clicks on Recommendations Using First-Round Data . . . 48

D Longer-Term Change in Polarization 49

E Demand Side Estimation for Full Sample 50

F Supply-Side Calibration Based on Random Click Data 51

G Supply-Side Model with Bayesian Updating 52

H Detailed Steps for Counterfactual 55

I Other Summary Statistics 57

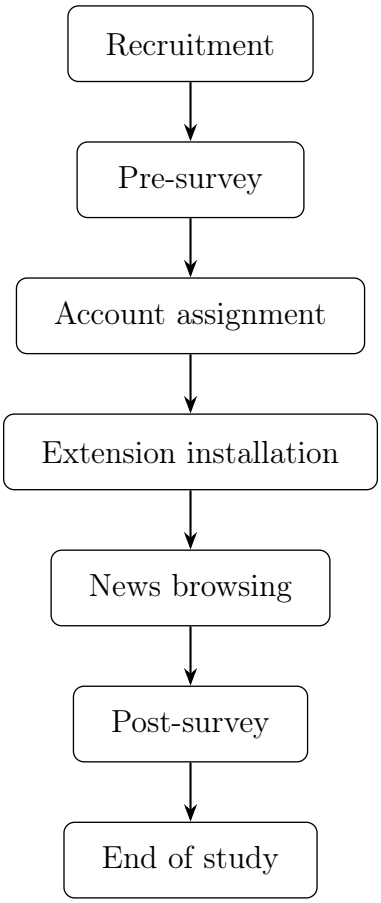
A Google News Extension Overview

For both field studies presented in this paper—the news browsing study and the random click study—we developed a custom Google News extension from scratch. Below is a summary of its functionality.

The Chrome extension is designed to operate passively in the background as users interact with the Google News platform. Upon installation, it requests permission to access a predefined set of URL patterns, primarily targeting domains associated with Google News (e.g., news.google.com and relevant subpaths). When a user navigates to one of these pages, the extension automatically injects a content script into the rendered DOM. This script monitors page-level activity and establishes a persistent WebSocket connection to a remote logging server built on an Express backend. The connection is authenticated using a short-lived JWT token, which the extension retrieves either from a configuration file or through manual set-up on the options page.

Each time a supported page is loaded, the extension waits for the main content feed to fully render, ensuring all asynchronous elements—such as lazy-loaded article modules—are present. It then queries the DOM using CSS selectors and heuristic rules to identify discrete article blocks. For each block, the extension extracts key metadata, including the headline, publication source, URL slug, category tag (when available), and visible timestamp. If an explicit publication time is not present, the extension infers timing based on relative labels (e.g., "2 hours ago") and the current page load time. To capture a broader range of engagement signals, the extension instruments several low-level user behaviors. These include scroll depth (tracked through scroll listeners with threshold-based triggers), dwell time (monitored via focus and blur events), and article clicks (captured using delegated click handlers on anchor elements). Each interaction is timestamped and linked to a persistent session ID. Events are accumulated in a local queue. Once the threshold is reached, the events are serialized into JSON and transmitted to the backend via WebSocket. If the WebSocket connection fails, the extension falls back to RESTful POST requests using Fetch or XHR, depending on the client version.

B Procedure of the News-browsing Study



We appreciate your participation in this study. Our research examines patterns in news consumption. We want to emphasize that it is required for participants to use Google Chrome, as the study is optimized for this browser. We are excited to have you on board and look forward to your participation.

Instructions

Your participation in the experiments will require the following five steps:

1. Please carry out a pre-survey using Qualtrics.
2. Enter your Prolific ID to receive your unique Gmail Account, Password, and backup recovery email address.
3. Please install the following extension. This extension is only guaranteed to work optimally on Google Chrome, hence we mandate its usage for participation in our study. If Google Chrome is not installed on your device, you can download it using this link: https://www.google.com/chrome. Please note, failure to use Google Chrome may prevent you from completing the experiments.
4. Browse through Google News, selecting the articles you wish to read. Please note: any clicks within a short time interval will not be considered as valid.
5. Please conduct a final survey on Qualtrics.

Upon completing the survey, you will receive a completion code. Please enter this identifier accurately into the text box below to proceed to the subsequent sections.
Should you have any questions or concerns, feel free to reach out to us at help@haasresearch.org.

Consent

By selecting the button below, you are consenting to participate in this study. You are also affirming that you are a minimum of 18 years of age and that you have comprehended the information provided in the consent form. Upon clicking the button, you will be directed to a pre-survey designed to gather basic information about you, conducted via the Qualtrics platform. Upon survey completion, you will receive a completion code. Kindly return to this page and input the code into the text box provided below.

Figure B1: Procedure of News-browsing Study

Note: This figure presents the procedure of our news-browsing study. The procedure of post-survey is similar to that of pre-survey described in Figure B2. The instructions participants received is shown on the right side.

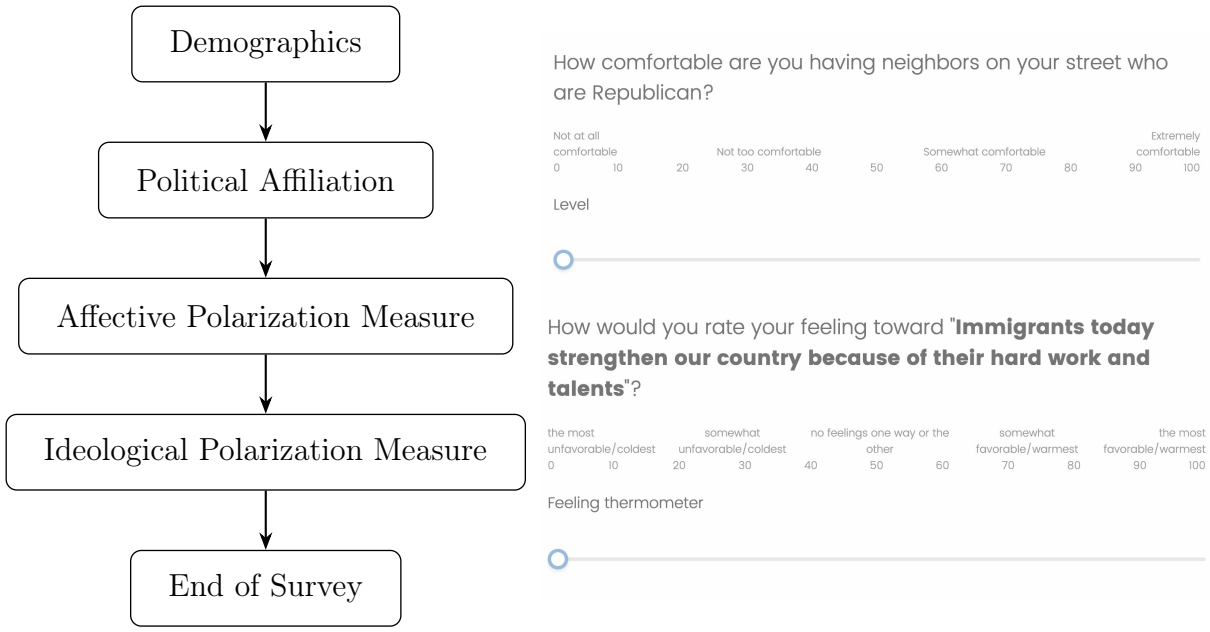


Figure B2: Procedure of Surveys

Note: This figure presents the procedure of pre-survey and post-survey in our study. Affective polarization is measured with questions such as: "How comfortable would you feel having a person from the opposing party as a neighbor on your street?". Ideological polarization is measured through questions like: "How would you evaluate the statement: 'Immigrants today strengthen our country because of their hard work and talents'?". The screenshots of example questions are shown on the right.

C Effect of Random Clicks on Recommendations Using First-Round Data

	Outcome: Rec_2	
	Mean Stance Score	Share of Republican
	(1)	(2)
$Click_1(score)$	0.071^{***}	
	(0.018)	
$Click_1(rep)$		0.079^{**}
		(0.034)
Observations	50	50
R^2	0.240	0.102
Adjusted R^2	0.225	0.083
Note:		*p<0.1; **p<0.05; ***p<0.01

Table C1: Random Click Effect in The First Round

Note: This table reports the causal impact of randomized click events on the political slant of subsequent recommendations in the first round. Column (1) shows how the stance score of the prior random click ($\text{Click}_1(score)$) affects mean stance score of the next recommendation set (Rec_2: Mean Stance Score); Column (2) shows the how a Republican-leaning article ($\text{Click}_1(rep)$) affects the proportion of Republican-leaning recommendations (Rec_2: Proportion of Republican). Stance scores and partisan labels follow the Pew Research Center's 2014 classification of news outlets Pew Research Center (2014).

D Longer-Term Change in Polarization

	Treatment			
	Before	Follow-up	Difference	P-value
	(1)	(2)	(3)	(4)
Outcome I: Affective Polarization				
Republicans	49.030	50.851	1.821**	0.019
Democrats	63.358	64.473	1.115***	0.004
Outcome II: Ideological Polarization				
Republicans	54.936	55.065	0.129	0.859
Democrats	82.576	82.886	0.310	0.473
Note:				*p<0.1; **p<0.05; ***p<0.01

Table D2: Long-term Change in Polarization

Note: This table presents the measures of (1) affective polarization (attitudes toward individuals with opposing views) and (2) ideological polarization (political views on social issues) before our study and 30 days after the study, segmented by two groups: Republican and Democratic. Both outcomes are scaled from 0 to 100, where higher values indicate a greater degree of polarization. The table also reports the differences between pre-study and 30 days after the study polarization levels, along with the corresponding p-values from paired t-test. A positive difference implies that individuals became more polarized following the study.

E Demand Side Estimation for Full Sample

Table E3: Demand Side Estimation

coefficient	value
β_r^r	0.891^{***}
	(0.100)
β_d^d	1.857^{***}
	(0.495)
baseline	β_{rd}

Note: This table presents the parameter estimates from the demand-side models, using data of all iterations of Democrat and Republican. β_r^r represents the marginal utility gain from a piece of Republican news for individuals with a Republican stance, relative to β_d^r. The model also controls for topic-specific possibility variables, which account for various news categories, including "politics", "economy & business", "technology", "health", "science", "sports", "entertainment", "education", "environment", and "lifestyle".

F Supply-Side Calibration Based on Random Click Data

Parameter	Estimate
α_0	19.084^{***}
	(0.202)
β_0	15.253^{***}
	(0.157)
c_a	0.357^{***}
	(0.026)
c_b	0.360^{***}
	(0.026)
Note:	***p<0.01

Table F4: Supply Side Calibration for Random Click Experiment

Note: The table presents the parameter estimates from the supply-side models for the random click experiment. α_0 and β_0 represent the initial parameter of the beta distribution that characterizes the proportion of Republican-leaning news among the recommended news. The parameter c_α (c_β) denotes the change applied to α_0 (β_0) after observing the political stance of clicked news. This adjustment reflects the reinforcement of the political stance recommendations based on participants' interaction with the random-clickling bots, capturing the dynamic nature of news recommendations as they adapt to consumer consumption behavior.

G Supply-Side Model with Bayesian Updating

Proposition: *The Bayesian updating for the proportion of Republican news, a_{it}, gives a non-conjugate posterior.*

Proof:

In this framework, we summarize the assumptions and notations as follows. The platform determines the share of Republican-leaning recommendations based on its current belief about the user's political orientation. Specifically, the proportion a_{it} of Republican content shown to user i in round t is set equal to the platform's posterior belief that the user is Republican, denoted $p_{it}(\theta_i = 1)$, where $\theta_i = 1$ indicates a Republican-leaning user. The user's latent type θ_i is drawn from a Bernoulli distribution, $\theta_i \sim \text{Bernoulli}(w)$, with the prior on w following a $\text{Beta}(\alpha, \beta)$ distribution. After each round, the platform updates its belief based on the political slant of the article clicked by the user, denoted by c_{it}, where $c_{it} = 1$ if the clicked article is Republican and $c_{it} = 0$ otherwise. This process allows the platform to continuously refine its estimate of the user's political stance.

The proportion of Republican news in recommendation is

$$a_{it} = p_{it}(\theta) = w_{it}|c_{i,t-1}$$

The posterior probability of w given c is

$$P(w|c) = \frac{P(c|w)P(w)}{\int_w P(c|w)P(w)dw}$$

where

$$\begin{aligned} P(c|w) &= P(c, \theta = 1|w) + P(c, \theta = 0|w) \\ &= P(c|\theta = 1, w)P(\theta = 1|w) + P(c|\theta = 0, w)P(\theta = 0|w) \\ &= wP(c|\theta = 1, w) + (1 - w)P(c|\theta = 0, w) \end{aligned}$$

Let n denote the number of news articles. Define $P_r(\theta)$ as the probability that a user clicks on a Republican news article given preference parameter θ, i.e., $P_r(\theta) = P(d_j = 1 \mid \theta, s_j = r)$. Similarly, let $P_d(\theta)$ denote the probability of clicking on a Democratic article given θ, that is, $P_d(\theta) = P(d_j = 1 \mid \theta, s_j = d)$.

$$P(c = 1|\theta) = \frac{anP_r(\theta)}{anP_r(\theta) + (1-a)nP_d(\theta)} = \frac{aP_r(\theta)}{aP_r(\theta) + (1-a)P_d(\theta)}$$

Denote $t(\theta) = \frac{P_r(\theta)}{P_d(\theta)}$ as the relative probability of clicking Republican news compared to Democrat news. It is natural to assume that $t(1) \geq 1$ and $t(0) \leq 1$. Thus,

$$P(c = 1|\theta) = \frac{at(\theta)}{at(\theta) + 1 - a}$$
$$P(c = 0|\theta) = \frac{1-a}{at(\theta) + 1 - a}$$

And with $a|w = w$, we have

$$P(c = 1|\theta, w) = \frac{wt(\theta)}{wt(\theta) + 1 - w}$$
$$P(c = 0|\theta, w) = \frac{1-w}{wt(\theta) + 1 - w}$$

Therefore, we have

$$P(c = 1|w) = w\frac{wt(1)}{wt(1) + 1 - w} + (1-w)\frac{wt(0)}{wt(0) + 1 - w}$$

And

$$P(w|c = 1) \propto w^{\alpha-1}(1-w)^{\beta-1}\left[w\frac{wt(1)}{wt(1) + 1 - w} + (1-w)\frac{wt(0)}{wt(0) + 1 - w}\right]$$

For the general case, however, $p(w \mid c = 1)$ does not follow the functional form $w^a(1-w)^b$, and thus is not Beta-distributed.

Similarly, we can prove that

$$P(c=0|w) = w\frac{1-w}{wt(1)+1-w} + (1-w)\frac{1-w}{wt(0)+1-w}$$

and

$$P(w|c=0) \propto w^{\alpha-1}(1-w)^{\beta-1}\left[w\frac{1-w}{wt(1)+1-w} + (1-w)\frac{1-w}{wt(0)+1-w}\right]$$

which means $w|c=0$ is not Beta-distributed either.

For special cases:

1. $t(1) = t(0) = 1$, meaning people with different stances have the same preference for news of different stances. In this case, we infer that $P(w|c=1) \propto w^{\alpha}(1-w)^{\beta-1}$, implying that $w|c=1 \sim \text{Beta}(\alpha+1, \beta)$. The recommender system will be updated in one step.

2. $t(1) = +\infty$, $t(0) = 0$, meaning people will only consume aligned content. In this case, we infer that $P(w|c=1) \propto w^{\alpha}(1-w)^{\beta-1}$, implying that $w|c=1 \sim \text{Beta}(\alpha+1, \beta)$. The recommender system will also be updated in one step.

In conclusion, a_{it} is not Beta-distributed after Bayesian updating, unless the preference index $t(\theta) = 1$, meaning that individuals with different political stances equally prefer Republican and Democratic news; or unless $t(1) = +\infty$ and $t(0) = 0$, meaning that individuals only view content aligned with their own stance.

H Detailed Steps for Counterfactual

We implement the ideology-blind counterfactual and simulate the new polarization level in three steps. As outlined in Section 4, individuals' expected utility from news consumption is modeled as

$$V_{ijt} = \beta_i^d D_j + \beta_i^r R_j + \beta_i^x X_j,$$

where D_j and R_j are binary indicators for whether article j is Democrat-leaning or Republican-leaning, and X_j denotes the topics of news.

Step 1: Random reassignment of political stance. In each round t, for each user i, we randomly reassign the political stance of articles to balance the Democrat recommendations and Republican recommendations. Specifically, we define the average share of Republican articles across all recommendations as:

$$p = \frac{1}{NT} \sum_i \sum_t \frac{\sum_{j \in \mathscr{C}_{it}} R_j}{|\mathscr{C}_{it}|},$$

where N is the number of users, T is the number of rounds, and $\mathscr{C}_{it}$ is the set of articles recommended to user i in round t. p can reflect the total share of Republican news in the whole recommendation pool.

To balance the recommendations of Democrat and Republican news, we re-assigned the political slant of news R_j^c for all news j and ensured that the post-replacement Republican share in each round matches the overall average,

$$\frac{\sum_{j \in \mathscr{C}_{it}} R_j^c}{|\mathscr{C}_{it}|} = p \qquad \forall \mathscr{C}_{it}.$$

Using the reassigned stances, while holding the topic composition constant, we compute counterfactual utilities as:

$$\hat{V}_{ijt}^c = \hat{\beta}_i^d D_j^c + \hat{\beta}_i^r R_j^c + \hat{\beta}_i^x X_j.$$

where $\hat{\beta}_i^d$, $\hat{\beta}_i^r$, $\hat{\beta}_i^x$ are estimated using observed data.

Step 2: Simulate counterfactual news consumption. For each individual, we simulate

their counterfactual choice using the Multinomial Logit Model:

$$P^c_{ijt} := \Pr(d^c_{ijt} = 1) = \frac{exp(V^c_{ijt})}{\sum_{j \in C_{it}} exp(V^c_{ijt})},$$

where d^c_{ijt} is a dummy equal to 1 if consumer i clicks on news article j in round t, in the counterfactual scenario, and 0 otherwise. We draw d^c_{ijt} from a multinomial distribution with probability P^c_{ijt}.

We then obtain the counterfactual click sequence for each consumer, denoted by $\{c^c_{it}\}$. $c^c_{it} = 1$ represents the user i clickes a Republican news article in the round t, in the counterfactual scenario. Following the approach in Section 4, we construct the accumulated counterfactual consumption of Republican-leaning and Democrat-leaning news as $rep^c_{it} = \sum_{k=1}^{t} c^c_{ik}$ and $dem^c_{it} = t - rep^c_{it}$, respectively.

Step 3: Estimate counterfactual polarization change. We then use a linear model to map news consumption to changes in polarization:

$$\Delta\text{Polar}_i = \gamma_1 rep_{iT} + \gamma_2 dem_{iT} + \sigma_i,$$

where rep_{iT} and dem_{iT} are the total number of Republican- and Democrat-leaning articles consumed by user i, and σ_i is an i.i.d. error term.

After estimating γ_1 and γ_2 from observed data, we apply them to the counterfactual consumption totals rep^c_{iT} and dem^c_{iT} to compute the predicted change in polarization under ideology-blind exposure. The aggregate results are reported in Table 9.

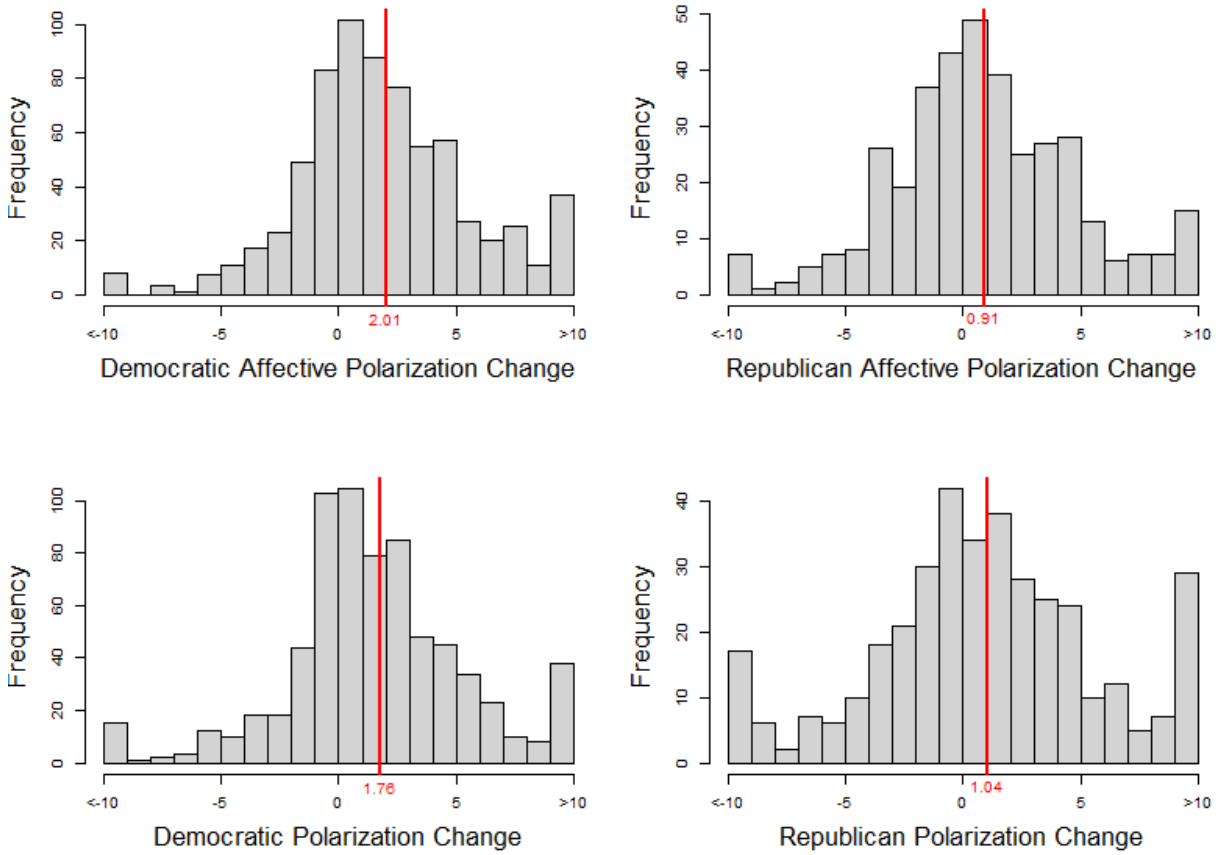
I Other Summary Statistics

Table I5: Share of Political Stance by Topic

	Summary Statistics
	% of Republican news
Topic Group	(1)
Politics	0.41
Economy & Business	0.77
Technology	0.71
Health	0.47
Science	0.43
Sports	0.73
Entertainment	0.54
Education	0.50
Environment	0.35
Lifestyle	0.52
Total	0.56

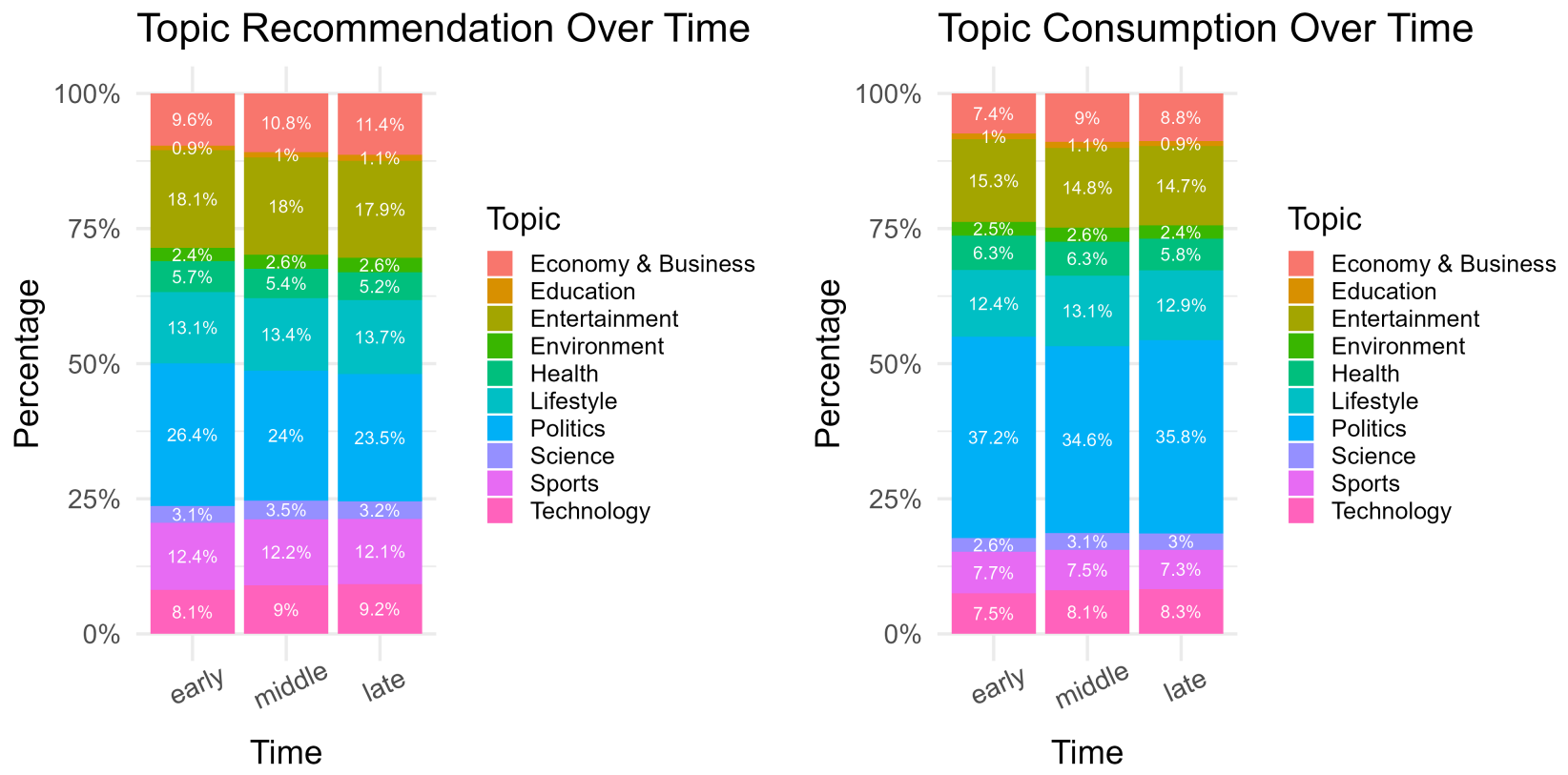
Note: This table presents the political slant of news across each topic category. Specifically, we report the proportion of Republican news of each topic in our full dataset. For example, there are 606,100 news articles about politics in our dataset, of which 41% are Republican-leaning.

Figure I3: Distribution of Polarization Change



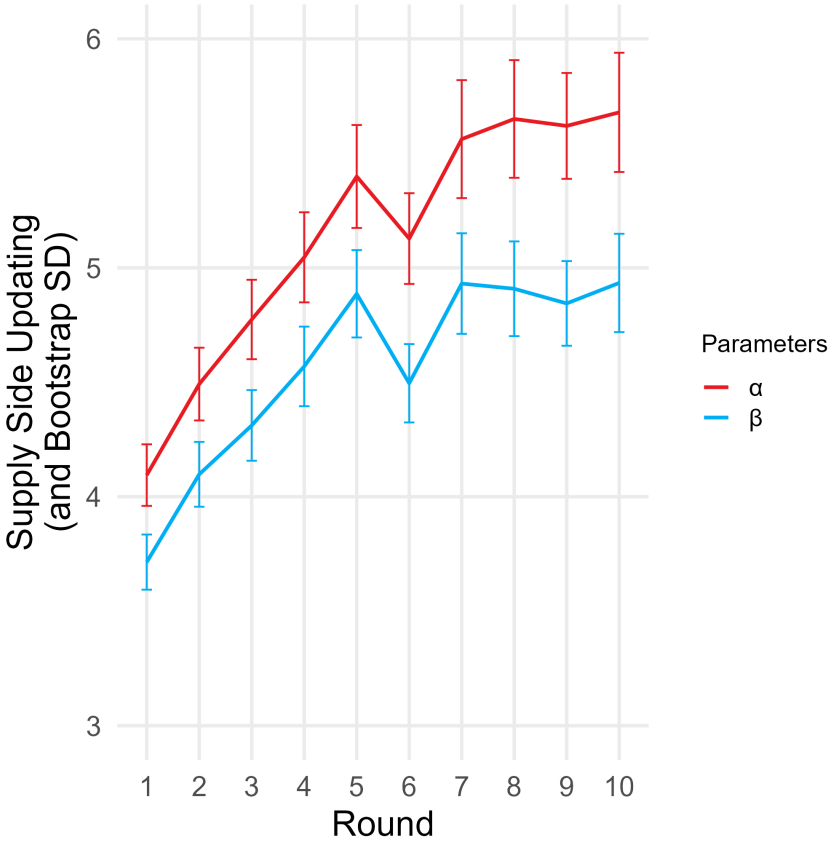
Note: This figure presents the distribution of individual changes in affective and (ideological) polarization for Democratic and Republican participants. Changes below –10 or above 10 are binned into the first/last bars to display outliers. Vertical red lines mark the mean change in each panel.

Figure I4: Topic Distribution Over Time



Note: This figure presents the distribution of recommended and consumed news topics across different stages of the process. Left panel shows recommended topics and right panel shows consumed topics. Specifically, topics are categorized as "Politics," "Economy & Business," "Technology," "Health," "Science," "Sports," "Entertainment," "Education," "Environment," and "Lifestyle." We divide each user's sequence into three equal phases: early (first third of the rounds), middle (second third), and late (final third), based on when the news was recommended.

Figure I5: Supply Side Estimation for the First 10 Rounds



Note: This figure illustrates the supply-side estimation, modeling the proportion of Republican news as $a_{it} \sim \text{Beta}(\alpha_t, \beta_t)$ for each of the first 10 rounds. For example, in round $t = 5$, we assume the proportion of Republican news recommended to individual i follows a Beta distribution with parameters α_5 and β_5, which we estimate. The red line represents the estimated α_t, and the blue line represents β_t. The vertical bars show bootstrap confidence intervals.