News Customization with AI

Felix Chopra Ingar Haaland Fabian Roeben
Christopher Roth Vanessa Sticher

November 8, 2025

Abstract

We introduce a new method to study news preferences by unbundling presentation from coverage. In our AI-powered news app, users can customize article characteristics, such as complexity or extent of opinion, while holding the underlying news event constant. In large-scale field experiments, we find that customization leads to better matching between the news consumed and stated preferences, increasing news satisfaction. While a significant fraction of users demand politically aligned news, the majority of users display high and persistent demand for less opinionated and more fact-driven news. By contrast, users not offered customization keep consuming news misaligned with their stated preferences.

Keywords: News Consumption, Customization, Artificial Intelligence, Matching.
JEL codes: C93, D83, L82, P00

We thank Elliott Ash, Luca Braghieri, Simon Cordes, Thomas Graeber, Paul Grass, Rafael Jiménez-Durán, Aaron Leonard, Filip Milojević, Sendhil Mullainathan, Aakaash Rao, Frank Schilbach, Carlo Schwarz, Jesse Shapiro, Adam Szeidl, David Yanagizawa-Drott and Tom Zimmermann for very constructive comments on our paper. We thank seminar audiences at the Berlin Applied Micro and Behavioral Economics seminars, BI Norwegian Business School, CEPR Bocconi 2025 Media Workshop, the CEPR Text-as-Data Workshop, Copenhagen University, the DFG Workshop on Consumer Preferences Consumer Mistakes, the Economics of AI Conference at IESE, the FAIR Media Bias and Political Polarization conference, the 2024 iOG-BFI fall workshop at Chicago, the IZA Belief Workshop, LMU Munich, the Max Planck Institute for Research on Collective Goods, MIT, the Political Economy of Ideas & Ideology Conference at Northwestern, the University of Bern, the Workshop in Behavioral Public Economics and Political Economy in Vienna, and the 3rd Zurich Workshop in AI+Economics for excellent comments. We thank Pietro Ducco, Maximilian Fell, Miguel Camacho Horvitz, Milena Jessen, Luca Michels, Max Müller, and Till Stange for excellent research assistance. The experiments were all pre-registered on the AEA RCT registry (#14561). The experimental instructions are available here: `https://raw.githubusercontent.com/cproth/papers/master/tmt_instructions_appendix.pdf`. Roth and Roeben acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1-390838866. Haaland acknowledges financial support from the Research Council of Norway (RCN) through the project "Media Bias and Political Polarization," under Grant Number 344979. Haaland and Roth acknowledge financial support from the RCN through its Centre of Excellence Scheme (FAIR project No 262675). We received IRB approval from MIT, the Frankfurt School of Finance & Management, the University of Cologne, and the NHH Norwegian School of Economics. Chopra: Frankfurt School of Finance & Management, f.chopra@fs.de; Haaland: NHH Norwegian School of Economics, Ingar.Haaland@nhh.no; Roeben: University of Cologne and ECONtribute, f.roeben@wiso.uni-koeln.de; Roth: University of Cologne and ECONtribute, NHH Norwegian School of Economics, Max Planck Institute for Research on Collective Goods, IZA, CEPR, roth@wiso.uni-koeln.de; Sticher: MIT, vsticher@mit.edu.

1 Introduction

A majority of Americans report dissatisfaction with the news media (Gottfried, 2020), a trend linked to declining news engagement (Forman-Katz, 2023) and broader negative social and political consequences (Prior, 2007; Sunstein, 2018; Zhuravskaya et al., 2020). This dissatisfaction is often attributed to how news is presented, including its negativity, political slant, technicality, and opinionated tone, suggesting a substantial mismatch between consumer preferences and the current supply of news.

Although markets often align supply with consumer preferences, there are several reasons to expect a persistent mismatch in news markets. News articles are multidimensional experience goods: fit is only revealed after consumption, and key attributes such as factual accuracy are costly to verify even ex-post. Because consumers make frequent, low-stakes choices over many potential articles, identifying content that matches preferences is very difficult (Gentzkow, 2018), making matching frictions especially pronounced in online news markets. On the supply side, increasing reliance on advertising revenue and referrals from social media networks create incentives to optimize for short-term engagement rather than consumer satisfaction, leading to clickbait content that attracts clicks but might cause regret (Kleinberg et al., 2024).

These market dynamics highlight the potential for a better alignment between the supply of news and consumer preferences. A precondition for improving the alignment is understanding consumers' news preferences: How do they actually want the news to be presented, and how much heterogeneity is there between consumers? To better understand these preferences, we need a setting where people actively choose characteristics of news articles, such as their level of opinion or complexity, while the underlying news events remain fixed.

To create such a news environment—where presentation is unbundled from coverage—we custom-built our own AI-powered news app, *TailorMadeTimes*. This app enables users to tailor news articles along *one* of the following dimensions at a time: extent of opinion, entertainment, sentiment, political perspective, and complexity.[1] The customization dimensions included in our app are directly motivated by the key sources of dissatisfaction about the prevailing news supply expressed by consumers (Reuters, 2023). As users enter their preferences with a set of sliders, the app dynamically presents AI-created rewrites, customized in real time, while holding constant the news event based on articles from a highly credible source, the Associated Press (see Figure 1 for an illustration of the app).[2]

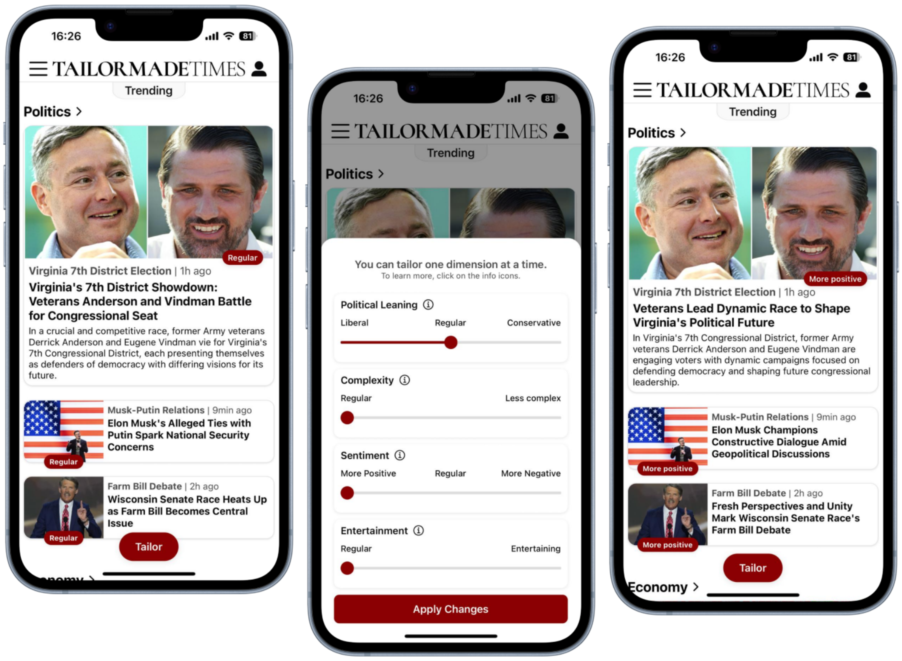
By fixing the underlying news event and allowing users to vary one dimension at a time, the app cleanly separates presentation from coverage, an option largely absent in today's news

[1]To verify that the underlying news event is held constant across article rewrites, we develop a custom fact-level alignment procedure. Furthermore, we use a human validation survey to confirm that the AI rewrites meaningfully vary along the intended dimensions.

[2]In particular, the core factual elements (actors, outcomes, statistics, etc.) stay identical across versions.

market.[3] This unbundling may lower search costs and expand the product space, enabling combinations such as fact-driven reporting on highly politicized topics or non-technical treatments of economic news. The active choice environment, where all article attributes are transparently labeled, ensures that consumers know in advance what they will receive, reducing the likelihood of ex post regret.

Figure 1: The AI News App



Note: This figure presents a mockup of the AI news app. The left screenshot shows the home screen, which comprises the top three trending articles per section ('Politics', 'Economy', 'US', 'World', 'Lifestyle', 'Sports') and is set to the *Regular* version. The middle screenshot shows the menu that opens when clicking the *Tailor* button, where users can customize articles via sliders. The right screen is the same as the first, but shows the articles in the *More Positive* version. The screenshots were taken on October 25, 2024, from an account for which we enabled all customizations except for *Facts Only*.

Leveraging our AI-powered news app in a large-scale field experiment, we examine the following main research questions: How much heterogeneity is there in news preferences across consumers? And does news customization help consumers improve the alignment between their news consumption and their stated preferences?

Latent demand for customization Before turning to our main experiment, we first conduct a pre-registered experiment with 5,005 U.S. respondents recruited from Prolific to test whether there is latent demand for language customization.[4] Participants were randomly assigned to learn

[3]Naturally, news customization at scale has only very recently become technically feasible through recent advances in Generative AI.

[4]While there is a concern about LLM-powered bots in online surveys, we take several steps to mitigate this issue in our experiments. Most importantly, our main outcome—verifiable app downloads—requires a device with access to either the App Store or Google Play and is thus robust to even highly sophisticated bots.

about an AI news app either with or without the customization feature. Offering customization increases download rates by around 40 percent ($p < 0.01$), suggesting that users value the ability to tailor language features. We find similar download rates irrespective of whether the additional customization feature allows for political rewrites or the removal of opinions.

Measuring news preferences Having uncovered latent demand for customization, we next characterize heterogeneity in people's news preferences and study whether people use customization to align their news consumption with their preferences. This requires a different design with the following features: Experience with the customization feature, a naturalistic setting, holding the set of news events covered in the app constant, and comparable engagement over time across treatment arms with and without customization. We implement such a design in our main experiment.

By holding the underlying news events constant and allowing deliberate variation in article features—such as complexity, extent of opinion, or political slant—we isolate specific dimensions of news preferences. At the same time, our approach of measuring preferences using real consumption data in an app directly captures the real world behavior of interest and mitigates concerns about social desirability bias, which confound self-reported data on news preferences.

Our experiment proceeds across multiple survey waves with U.S. respondents recruited from Prolific. In the initial recruitment survey, respondents are asked whether they are interested in downloading an AI-powered news app. Those with an interest are later invited to an app download survey in which they can download our app. Participants then use the app for 10 days with usage incentives, followed by midline and endline surveys measuring downstream outcomes.

We employ a between-subjects design, randomly assigning respondents to one of two customization treatments or to a control group without customization. Respondents only learn about customization after app login, ensuring no differential selection into app download by treatment status. In both customization groups, respondents can customize news articles along four dimensions. Three of the four dimensions are identical in both customization treatments, while the fourth dimension differs across treatments. Specifically, respondents in both customization treatments can choose between a regular default version or customize the articles in terms of entertainment (*Entertaining*), sentiment (*More Positive*, *More Negative*), and complexity (*Less Complex*), while holding the underlying news event constant. For the fourth dimension, we cross-randomize whether respondents can customize the political perspective of the articles (*Liberal*, *Conservative*) or remove opinion content (*Facts Only*).

To isolate the effect of being assigned different article versions without being offered customization, respondents in the control group are randomized into receiving only one of the eight article versions without customization. Articles are not labeled on the app for control group respondents, who are thus unaware of which version they receive.[5]

How people use customization A precondition for successfully characterizing people's news preferences using this experiment is that respondents actually use the news app. We therefore start by verifying compliance with the usage incentives. Respondents across all different treatment conditions spend a similar amount of time on the app. On average, they spend about four minutes per day and use the app on between 6 and 7 days during the 10-day incentive period. Taken together, our usage data confirm that people complied with the incentives. This ensures that users have experience with the customization tool and news articles from different rewrite versions.

Having established that usage incentives successfully encouraged respondents to use the app, we now turn to our first main research question, drawing on data from the incentive period:[6] How much heterogeneity is there in news preferences across consumers? In the customization treatment with political rewrites, the most frequently used option is the politically more neutral *Regular* rewrite (37% of the time), but substantial time shares are also devoted to politically aligned news: Democrats spend 24% of their time in the *Liberal* setting and Republicans spend 25% in the *Conservative* setting. The *More Negative* option is rarely selected (2%), consistent with users making choices in the moment that are inconsistent with what they actually want (Kleinberg et al., 2024). These patterns indicate that when political rewrites are available, a significant fraction of our respondents use customization to align news with their ideological views. Our evidence is thus consistent with prior work showing that a subset of consumers have a preference for politically aligned news (Gentzkow and Shapiro, 2010). However, an important difference in our setting is that users make a deliberate choice to align the news with their own ideological views. Furthermore, while we find that a subset of consumers demand slanted news, we see that the large majority, in fact, prefer news without politically slanted content.

To assess whether the observed demand for the *Regular* setting reflects a genuine preference for politically balanced reporting or a default effect, we turn to the second customization treatment, which replaces the political rewrite options with a *Facts Only* option explicitly framed as "less opinionated and focused on the facts." In this treatment, which keeps the *Regular* version as the default, respondents spend 46% of their time in *Facts Only*—more than in any other setting—with only 18% of the time devoted to the *Regular* default. The pattern persists across partisan lines, is particularly pronounced among those with low baseline news satisfaction, and is equally pronounced among respondents with high and low baseline news consumption. This evidence suggests that stated preferences for more fact-driven and less opinionated content reflect people's true preferences, and not just social desirability bias.

Customization and alignment with stated preferences Having established significant heterogeneity in news preferences across consumers, we next examine our second main research

[5]Control group respondents are also unaware that other respondents can customize the news.

[6]As we discuss later, we obtain very similar results in the post-incentive period.

question: To what extent news customization improves alignment between people's news consumption and stated preferences? We leverage pre-treatment questions from the recruitment survey in which respondents reported whether they find the news too complex, negative, positive, liberal, conservative, opinionated, or insufficiently entertaining. We then relate these assessments to subsequent consumption patterns during both the incentive and post-incentive periods. The results reveal that individuals systematically select news that corrects the perceived mismatch between existing news content and their preferences. For instance, respondents who think the news is too opinionated are significantly more likely to consume the *Facts Only* version whereas those who think the news is too complex spend more time using the *Less Complex* setting.

To examine whether the improved alignment between news consumption and preferences is unique to the customization treatments, we next study similar correlations between the perceived mismatch and news consumption among our control group respondents who are randomly assigned a fixed version throughout the whole period. For this exercise, we focus on the post-incentive period when control group respondents can freely adjust their consumption. While control respondents who believe the news is too opinionated spend significantly more time using the app if they are assigned the *Facts Only* version, we do not observe significant correlations for any of the other dimensions. These findings demonstrate that customization strongly helps increase alignment with news preferences compared to being assigned a fixed version at random. Furthermore, the notable exception for the *Facts Only* version reinforces our earlier finding that there is a strong demand for more fact-driven and less opinionated news.

We next examine whether there are systematic group differences between treatment and control group respondents in their overall news consumption during the post-incentive period. While all respondents reduce their consumption in the post-incentive period, there are no treatment differences in the time spent using the app. Any treatment differences in news consumption will thus be driven by different time shares across article versions. The results reveal a striking *customization wedge* in news consumption: Respondents offered customization overwhelmingly consume fact-driven and less opinionated news and consume virtually no negative rewrites. By contrast, control group respondents spend very similar amounts of time across the different fixed article versions, making them much more likely to consume negative news in particular.

Customization and news satisfaction This customization wedge raises the question of whether a higher alignment with news preferences among respondents offered customization also raises news satisfaction. Consistent with an increased preference alignment, we observe significantly higher news satisfaction among respondents offered customization. This effect is particularly pronounced among respondents offered customization with the *Facts Only* option, consistent with the high demand for more fact-driven content. Looking into potential mechanisms, we find that the lower news satisfaction among control group respondents is largely driven by respondents assigned to more negative or politically slanted news. The fact that control group respondents consume the different article versions in roughly similar amounts, yet report lower satisfaction, implies that observed choices absent customization may be a poor proxy for welfare—consistent with the theory in Kleinberg et al. (2024) and echoing empirical evidence from social media on the consumption of toxic content (Beknazar-Yuzbashev et al., 2025).

Although customization improves alignment with preferences and positively affects news satisfaction, we find precisely estimated null effects on political polarization and factual knowledge. These null results are in line with a broader literature documenting muted downstream effects from news interventions, including those that alter ideological alignment or total consumption. Given that we hold the coverage of news events constant, it is perhaps unsurprising that our intervention does not meaningfully affect political polarization or factual knowledge.

Related literature We contribute to a growing empirical literature on people's news preferences (Braghieri et al., 2025a; Chopra et al., 2022, 2024; Thaler, 2024). Robertson et al. (2023) show that negative sentiment drives online news engagement using large-scale field experiments, and Beknazar-Yuzbashev et al. (2025) find that removing toxic content from social media feeds reduces overall consumption. Our main contribution is to introduce a new approach for studying news preferences that unbundles content from coverage. Leveraging this approach, we show that most consumers—when offered an active choice over transparent product attributes—prefer fact-driven over opinionated news. These preferences are challenging to identify from market data because news articles are experience goods and outlets maximize engagement rather than user satisfaction (Kleinberg et al., 2024). Our finding that consumers show a large preference for *Facts Only* rewrites is consistent with models emphasizing the role of accuracy concerns when there is uncertainty about news quality (Gentzkow and Shapiro, 2006). Moreover, the significant wedge we document between news preferences under customization and those observed without customization aligns with recent evidence from social media suggesting that making more deliberate choices may substantially increase the reliability and reduce the slant of consumed news (Braghieri et al., 2025a). Our evidence on the substantial demand for customization with *Facts Only* rewrites contrasts with the well-documented popularity of opinion shows in the US. This apparent discrepancy is consistent with recent evidence suggesting that consumers perceive opinion-oriented content as highly informative (Bursztyn et al., 2023a), especially when it aligns with their own political views (Mitchell et al., 2018a).

Our study contributes to a large literature on the economics of media markets (Demsetz and Lehn, 1985; Mullainathan and Shleifer, 2005), media bias (Caprini, 2023; DellaVigna and Kaplan, 2006; Gentzkow and Shapiro, 2010; Groseclose and Milyo, 2005), and recent work that studies how artificial intelligence technologies and algorithms impact the market for news (Acemoglu, 2021; Acemoglu et al., 2023; Aridor et al., 2024, 2025; Cheng et al., 2025). Previous research has highlighted that AI technologies may increase the spread of misinformation (Acemoglu et al., 2025; Campante et al., 2025; Chen and Shu, 2023; Simon et al., 2023; Zhou et al., 2023). We highlight how AI technology can be used to address a market failure by increasing the match between consumer preferences and the supply of news.[7]

More broadly, our paper contributes to an emerging literature on how platform design choices shape digital news markets. Prior studies have focused on content moderation policies (Jiménez-Durán, 2023; Jiménez-Durán et al., 2024) and recommendation algorithms on social media platforms (Acemoglu et al., 2025; Beknazar-Yuzbashev et al., 2024; Kalra, 2025; Levy, 2021). Levy (2021) shows that Facebook's algorithm systematically limits exposure to counter-attitudinal news, creating mismatches between users' potential interests and the content they are actually shown, and thereby reinforcing polarization. By contrast, we highlight how deliberate news customization in a news app affects consumption patterns in a very different context—steering readers toward more fact-driven and less opinionated content.

2 The AI News App

2.1 App

For the purposes of this paper, we developed a mobile news app called *TailorMadeTimes* (see Figure 1 for a mockup). The design and structure of our news app are informed by common best practices of other mobile news apps. Specifically, news articles are previewed in the format of a scrollable *home feed*, which is organized by news categories ("Politics", "Economy", "US", "World", "Lifestyle", "Sports"; in this order). Article previews include the article title, timestamp (e.g., "1 hour ago"), a short AI-generated topic label (e.g., "US Presidential Election"), and an image for most articles (during the study period, 82% of articles had an image). For the leading article in each section, we also display the article lead already in the feed. Every news category in the feed includes the top three trending articles, and users can switch to dedicated category sections with additional articles. The app is updated with new articles every hour, resulting in an average of 112 new articles per day, where articles remain in the app for three days. Older articles can still be accessed via the search and bookmark features. Furthermore, users with push notifications enabled receive four notifications throughout the day, each containing the most trending news story since the previous notification (see Section 2.2 for a definition of "trending").

The distinguishing feature from other news apps is that users of our app can *customize* news articles in terms of entertainment, sentiment, political perspective, complexity, and the extent of opinions—while holding the underlying news event constant. As shown in Figure 1, users can indicate their preferences by clicking the "Tailor" button, which activates a pop-up menu with a set of sliders. As users enter their preferences by changing the slider settings, the app dynamically updates *all* articles. Users can make changes to the sliders at any point in time. Importantly, the set of news articles remains identical: customization only affects the presentation

[7]Methodologically, we relate to a growing literature that uses LLMs, e.g., in the context of the collection of qualitative data (Chopra and Haaland, 2023) or the classification of text (Ash and Hansen, 2023; Braghieri et al., 2025b; Graeber et al., 2024; Habibi et al., 2024).

of news articles along one dimension at a time, not which articles are shown in the app.

2.2 Source of Articles

The news app relies on news articles sourced from the Associated Press (AP), which is widely regarded as a highly credible news organization.[8] Established in 1846, AP has a long-standing reputation for its strict adherence to journalistic ethics and its commitment to unbiased news coverage. It operates on a not-for-profit model, which allows it to focus solely on news gathering and reporting. As a result, its news is trusted by a vast array of subscribers and outlets around the world. For our app, we select English-language news articles covering politics, economy, the US, the world, lifestyle, and sports via a news API subscription (`www.newsapi.ai`), focusing on articles that have between 100 and 1,000 words. Then, for each article, we create a *trending score* to capture its popularity. Specifically, we retrieve a score from the news API that indicates how often entities within an article (e.g., people, events, countries) have been mentioned across all available news sources within the past two days before the article was published. We combine this score with a measure of recency and sort articles by this combined score, such that articles that contain trending entities and that are more recent appear higher up within the home and category feeds of the app.

2.3 Article Versions

This section motivates and characterizes the different article versions we offer users as part of the customization feature of the app (via the "Tailor" button). The baseline version is the *Regular* version, which is a summary of around 300 words of the original AP article. The customization options on top of *Regular* directly correspond to common news reader complaints emphasizing the dryness, negativity, political bias, and technicality of mainstream reporting (Reuters, 2023). The rewrite dimensions we focus on (extent of opinion, political perspective, entertainment, sentiment, and complexity) directly reflect these key sources of dissatisfaction. Political and sentiment rewrites address concerns about political bias and negativity. Entertainment and complexity rewrites respond to complaints about dryness and technicality. *Facts Only* rewrites reduce opinionated content, offering an option for more neutral reporting. All rewrites vary the presentation while covering the same core news event (e.g., events, actors, dates, outcomes, statistics). Table 1 provides an overview of the rewrite dimensions with examples, illustrating how the same news event is expressed differently across these dimensions. Below, we describe each rewrite in more detail, explicitly referring to and directly quoting the examples presented in the table (the technical implementation of the rewriting process is described in Section 2.4).

[8]The Associated Press is consistently rated as one of the news agencies with the highest reliability scores (Bhadani et al., 2022) and is also among the most trusted news sources in the US (see, e.g., `https://today.yougov.com/politics/articles/52272-trust-in-media-2025-which-news-sources-americans-use-and-trust`.

Table 1: Rewrite Examples

Dimension	Title and first sentence of body	Explanation
Regular	*Title*: Republican Lawmakers Push Back on Key Aspects of Trump Administration's Efficiency Reforms *Body*: Republican lawmakers are voicing concerns about President Trump's sweeping government reforms under the Department of Government Efficiency (DOGE), led by billionaire Elon Musk. [...]	(Baseline Associated Press Summary)
Conservative	*Title*: Republicans Balance Support for Trump Reforms with Local Concerns *Body*: Republican lawmakers are expressing thoughtful concerns about President Trump's ambitious government reforms spearheaded by the Department of Government Efficiency (DOGE), under the leadership of prominent entrepreneur Elon Musk. [...]	Downplaying issues, positive framing of Republican government plans
Liberal	*Title*: Trump's Government Reform Plan Faces Bipartisan Criticism for Overreach *Body*: Republican lawmakers are voicing concerns about President Trump's sweeping government reforms under the Department of Government Efficiency (DOGE), an initiative spearheaded by billionaire Elon Musk that has faced significant criticism for its potential overreach. [...]	Highlights criticism and concerns associated with Republican government plans
More Positive	*Title*: Republicans Collaborate on Ambitious Government Reforms for Greater Efficiency *Body*: Republican lawmakers are engaging in constructive discussions about President Trump's ambitious government reforms under the Department of Government Efficiency (DOGE), spearheaded by visionary entrepreneur Elon Musk. [...]	Optimistic framing, positive attributes
More Negative	*Title*: Trump's Sweeping Reforms Spark Chaos and GOP Backlash *Body*: Republican lawmakers are increasingly alarmed by President Trump's sweeping government reforms under the Department of Government Efficiency (DOGE), headed by billionaire Elon Musk. [...]	Cautionary and alarmist framing
Entertaining	*Title*: Musk-Led Government Overhaul Sparks GOP Drama Over Cuts *Body*: Republican lawmakers are raising eyebrows at President Trump's ambitious government overhaul helmed by the Department of Government Efficiency (DOGE), which, in a plot twist worthy of Silicon Valley, is being led by billionaire entrepreneur Elon Musk. [...]	Uses dramatic and vivid language
Less Complex	*Title*: Republicans Worry Musk-Led Government Changes May Hurt Jobs and Programs *Body*: Some Republican lawmakers are worried about President Trump's major changes to the government. These changes are being managed by the Department of Government Efficiency, or DOGE, which is led by billionaire Elon Musk. [...]	Simpler wording, shorter sentences, easier syntax
Facts Only	*Title*: Republican Lawmakers Raise Concerns Over DOGE Government Reforms *Body*: Republican lawmakers have expressed concerns about President Trump's government reforms under the Department of Government Efficiency (DOGE), led by Elon Musk. [...]	Objective and neutral phrasing

Note: This table provides an overview of the rewrite dimensions with examples of how article title and first sentence of the body vary across dimensions. The example shown is from an article published on February 12, 2025. Text highlighted in yellow indicates examples of dimension-specific rewrites. The full article body along with title and lead are available on the following link: `https://raw.githubusercontent.com/cproth/papers/master/TMT_examplerewrites.pdf`.

Political Political rewrites adjust wording and framing to reflect liberal or conservative perspectives. For example, as shown in Table 1, the *Conservative* version frames the reforms positively by highlighting lawmakers expressing "thoughtful concerns" about President Trump's "ambitious" government reforms, led by the "prominent entrepreneur Elon Musk."[9] Conversely,

[9]When this article appeared, Elon Musk worked for the government and was looked upon favorably by Trump.

the *Liberal* rewrite describes the same initiative skeptically, noting criticism for its potential overreach, reflected in its title emphasizing “Bipartisan Criticism.” As another example, inflation data under a *Conservative* rewrite might be framed as a concern, such as “Inflation still above its target under the current Biden administration.” By contrast, *Liberal* rewrites highlight progress and successes, such as “Inflation significantly down from its peak under the current Biden administration.” More generally, political rewrites vary the use of partisan terminology, as in Gentzkow and Shapiro (2010), where, for example, the *Liberal* rewrite would refer to “estate tax,” whereas the *Conservative* rewrite would say “death tax.” These rewrites aim to reflect differing political perspectives without introducing unsupported causal claims.

Sentiment Sentiment rewrites shift the article’s tone toward optimism or caution. As shown in Table 1, the positive rewrite emphasizes lawmakers engaging in “constructive discussions,” framing the reforms as “ambitious,” with the title highlighting “Collaboration” and “Greater Efficiency.” In contrast, the negative rewrite describes lawmakers as “increasingly alarmed,” emphasizing “chaos” and “backlash,” underscored in the title as well.

Entertainment *Entertaining* rewrites employ dynamic and engaging language, using vivid descriptions, creative analogies, and storytelling techniques. For instance, Table 1 presents an engaging narrative by characterizing the situation as sparking GOP drama and describing Elon Musk’s leadership as a “plot twist worthy of Silicon Valley.” The title itself emphasizes drama to enhance reader interest. As another illustration, a straightforward phrase like “Persistent inflation” might be transformed into “Inflation refuses to leave the room,” framing economic issues with a touch of humor or drama. Storytelling techniques may frame developments as dramatic arcs, such as likening market fluctuations to a “roller-coaster ride.” Relatable metaphors or analogies, such as comparing inflation to “an uninvited guest overstaying its welcome,” further enhance reader engagement.

Complexity *Complexity* rewrites simplify language and structure. These rewrites adopt simpler vocabulary and shorter sentences, similar to the style of the *Simple English Wikipedia*. Table 1 provides a direct example: “Some Republican lawmakers are worried about President Trump’s major changes to the government. These changes are being managed by the Department of Government Efficiency, or DOGE, which is led by billionaire Elon Musk.” This example demonstrates how a long sentence can be effectively split into two shorter sentences, each using simple sentence structures and accessible vocabulary such as “worried” and “changes” instead of more complex terms. As another example, the sentence “Signs of decreasing inflation and a cooling job market have led to expectations of monetary easing” might be rewritten as “Because inflation is going down and the job market is slowing, people think the central bank will lower rates.” Additionally, technical terms are either replaced with simpler terms or explained in simple language, such as defining “monetary easing” as “making it easier for businesses and people to borrow money.”

Facts Only *Facts Only* rewrites reduce opinionated content by removing subjective assessments. Table 1 exemplifies this approach with a neutral description: "Republican lawmakers have expressed concerns about President Trump's government reforms under the Department of Government Efficiency (DOGE), led by Elon Musk." The title also neutrally states that lawmakers "Raise Concerns" without subjective framing. As another example, the sentence "The government's reckless spending has led to an unsustainable increase in national debt, putting the economy at risk" might be rewritten as "The government's spending has increased the national debt." These rewrites eliminate subjective descriptors (e.g., "reckless" or "unsustainable"), present the facts objectively, and simplify the statements. Such rewrites present the news objectively, fostering unbiased reporting by removing evaluative language.

2.4 Multi-Agent AI System for Rewrites

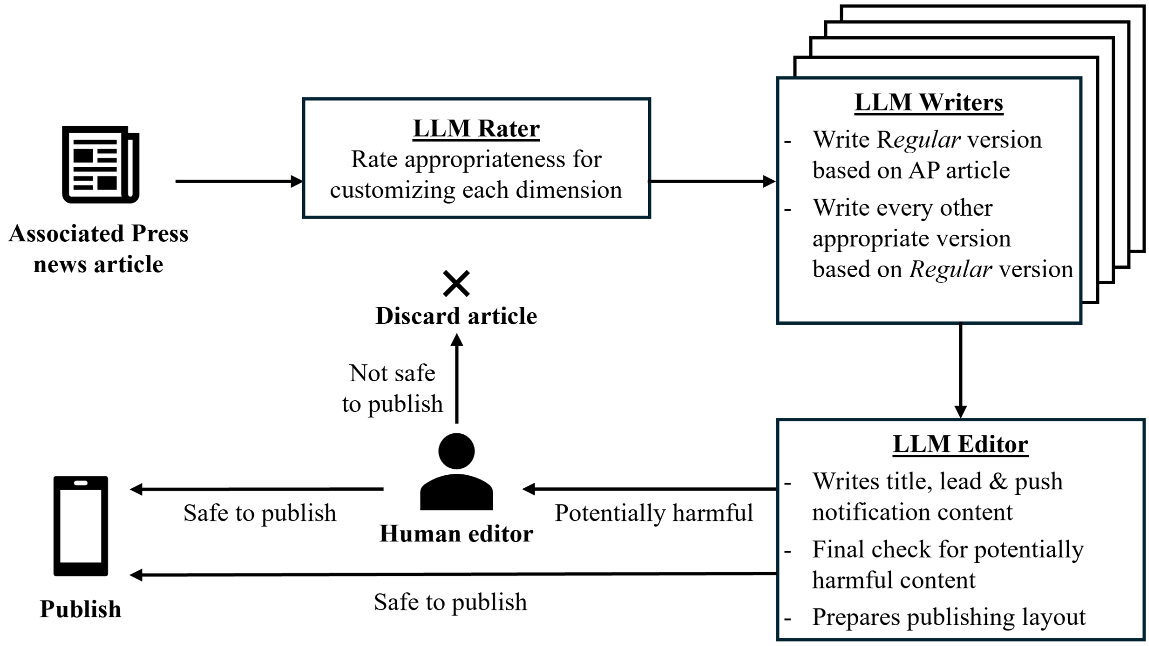
To systematically vary article characteristics, we built a multi-agent AI system. The objective of this system is to generate alternative versions of each AP news article that hold the underlying news event constant and uphold journalistic integrity across versions by not omitting or fabricating facts. To achieve this, we use a variety of AI agents for specific generative and evaluative tasks. Our system is built on the Generative Pre-trained Transformer 4o model family ("GPT-4o" and "GPT4o-mini") by OpenAI, which was state-of-the-art at the time of the experiments. Figure 2 illustrates this workflow, which consists of three components outlined below.

Preprocessing Starting with a given AP news article (sourced as described in Section 2.2), the first agent assesses the appropriateness of customizing the article. Specifically, for every dimension we set out a rule for when it should not be customizable. For example, articles that contain sensitive topics, such as death or terror, cannot be made more entertaining or more positive. Further, it also flags articles categorized as "Sports" or "Lifestyle" to not be customizable on the political dimension, given the limited scope for political rewrites for this type of content.[10]

Writing Given the original AP article, an agent first writes the *Regular* version, which is a shorter, summarized version of the AP article of around 300 words. Based on this version, a first draft of all other appropriate versions is written. All rewrite agents are instructed to (i) maintain accuracy by holding the news event constant, (ii) maintain article length and logical flow, and (iii) not mislead, omit, or fabricate even in light of changing emphasis and framing. Additionally, we equip every rewrite agent with relevant context beyond the model's cutoff date (e.g., who the current US president is, what today's date is, etc.). On top of these general guidelines, every agent is provided with a description of the desired characteristics of the rewrite and a version-specific curated set of input/output examples (*few-shot prompting*). By showing representative example rewrites and explanations of rewritten sentences, we effectively teach

[10]In the app, users receive an explanation via a tooltip on why a specific article-version combination is unavailable.

Figure 2: Multi-Agent AI System for Rewrites



Note: This figure illustrates the multi-agent AI system used to generate the rewrites of the articles. First, an agent assesses whether customizing the article is appropriate (preprocessing). Next, a writing agent generates the *Regular* version by summarizing the article and drafts the rewrites. An editor agent then creates titles, leads, and push notification content. Finally, the rewrites are screened through OpenAI's Moderation API and, if flagged, sent for manual review before publication.

the agent how to implement a rewrite that corresponds to the desired characteristics. Lastly, we ask the agents to conduct the rewrite in a two-step process, where they first identify suitable sentences to be rewritten and provide a justification for each suggested sentence rewrite. Second, they are instructed to reflect on the suggested changes and implement those that are in line with the general rewrite guidelines and the desired dimension characteristics. This reflective procedure aims at maximizing output quality while adhering to the constraints we set.

Editing Lastly, an editor agent, instructed similarly to the rewrite agents for the main body, writes a version-specific title and lead. Further, it generates push notification content for articles that, based on the trending score, are chosen to be sent as a push notification to users. The push notifications correspond to the customization setting of the user when the notification is sent. As a final check, all content that has been generated by AI in the process is sent to OpenAI's Moderation API to rule out publishing unsafe content. If any content is flagged by the Moderation API, the entire article with all its versions is not published, but sent to us for review. In addition, all articles that received the lowest appropriateness score are also sent to us for review. Unflagged articles are automatically and immediately published on the app.

2.5 Validation of News Event Consistency

Methodology To validate that rewrites do not change the underlying news event, we developed a custom fact-level alignment procedure. We begin by extracting factual claims, defined as single, objectively verifiable statements following Ni et al. (2024), from every article version. Next, we classify these extracted facts into core and supplementary using an LLM classifier (GPT-4.1). Core facts are defined as those that (i) directly convey the article's main event, "who did what", and its primary significance ("why it matters"), and (ii) form the minimal sufficient set to reconstruct a faithful account of the core events. Supplementary facts provide context, background, or stylistic detail and are allowed to vary across rewrites (e.g., explanations added in the *Less Complex* version). We restrict validation to core facts because variation in supplementary details is expected by construction and does not change the underlying news event. This means checking whether all core facts from the *Regular* version are contained in the rewrites (to detect omissions) and vice versa (to detect hallucinations). To operationalize this, we align facts via embeddings: for each core fact in the *Regular* version (and vice versa), we take the maximum cosine similarity to any fact in the rewrite (core and supplementary) and declare a match if the maximum similarity $\geq \tau$, where τ is calibrated against human judgments on a held-out set (τ=0.66). As a safeguard against missed matches, we then conduct a final LLM-based consistency check to reduce false negatives. In addition to the embeddings-based checks, we separately check the consistency of extracted numbers across versions (e.g., dates, statistics, percentages). A detailed account of the technical implementation of this three-step validation, as well as example outputs of all steps, is provided in Appendix A2.

Results Table 2 reports directional coverage of core facts across rewrites across 59,092 article versions in total. We present two measures: $R \to V$ (share of *Regular* core facts covered by the rewrite; 100% minus the omission rate) and $V \to R$ (share of *Version* core facts supported by the *Regular* article; 100% minus the hallucination rate). The "Embeddings" block uses cosine similarity at the calibrated threshold of 0.66, while "+ LLM Adjudication" reevaluates the unmatched remainder of the embeddings-based analysis against the full target article to avoid potential false negatives. Lastly, "Numeric Alignment" assesses equality of extracted numbers (dates, quantities, percentages) across versions.

Across all comparisons, embedding-only alignment yields coverage of 82.4% (R→V) and 82.0% (V→R), indicating substantial baseline alignment. Under embeddings alone, coverage is lowest for *Entertaining* and *Less Complex* (≈ 72–78%), consistent with heavier paraphrasing and structural changes, and highest for political rewrites and *Facts Only* (≈ 86–90%). After LLM adjudication, coverage rises to 96.5% (R→V) and 96.3% (V→R). Under this adjudicated measure, our metrics are all above 90% in both directions, and five out of seven versions exceed 95% in both directions. Furthermore, quantitative content is preserved: numeric alignment is close to 100% in both directions for every rewrite, providing an orthogonal validation that dates,

Table 2: Alignment of Core Facts

	Embeddings		+ LLM Adjudication		Numeric Alignment	
Version	*R→V*	*V→R*	*R→V*	*V→R*	*R→V*	*V→R*
Conservative	88.4%	87.5%	96.8%	95.9%	99.2%	100%
Liberal	90.0%	87.7%	98.0%	95.8%	99.6%	100%
Entertaining	72.0%	73.5%	97.5%	97.7%	98.3%	99.7%
Less Complex	76.7%	78.3%	98.2%	99.4%	98.8%	99.0%
More Negative	83.5%	79.9%	96.2%	91.8%	99.3%	99.9%
More Positive	82.3%	83.2%	91.3%	95.3%	97.7%	100%
Facts Only	86.4%	87.4%	96.2%	95.8%	99.9%	100%
All	82.4%	82.0%	96.5%	96.3%	98.9%	99.7%

Note: R=Regular, V=Version. "R→V" is the share of *Regular* core facts covered by the rewrite (1–omissions). "V→R" is the share of *Version* core facts covered by *Regular* (1–hallucinations). "Embeddings" uses `all-mpnet-base-v2` at the calibrated cosine similarity threshold (τ=0.66). "+ LLM Adjudication" checks unmatched pairs with an LLM to rule out false negatives. "Numeric Alignment" reports consistency of extracted numeric tokens across versions. Examples of this alignment procedure can be found in Appendix A2.

statistics, and percentages are consistent. Overall, the evidence indicates that rewrites retain the underlying news event: embedding similarity captures most matches; the LLM adjudication reliably recovers hard paraphrases; and numerical details remain consistent. For more details and examples, see Appendix A2.

2.6 Human Validation of Rewrites

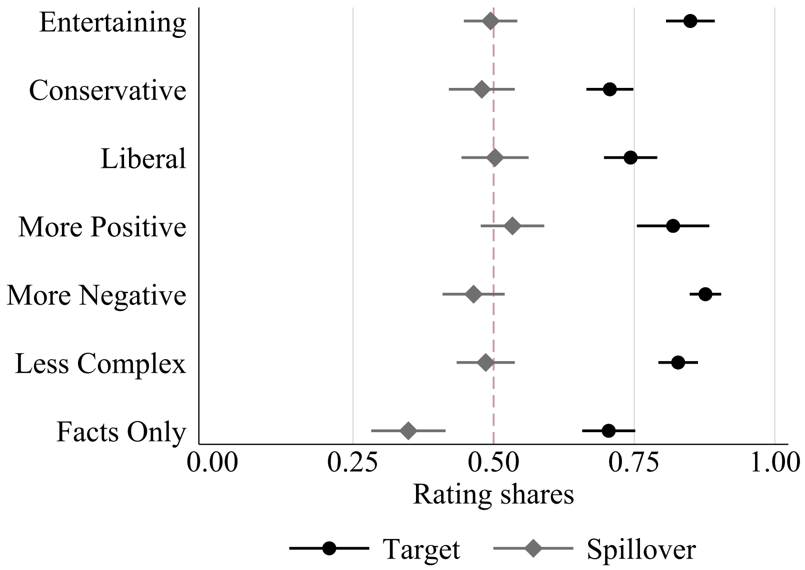
Design To validate whether and how the rewrites affected the articles, we recruited 1,169 respondents from Prolific (see Appendix Table A1 for an overview of studies).[11] To ensure political balance in our sample, we stratified the collection to recruit as many Republicans as Democrats. The data collection based on articles from our wave 1 field experiment took place in December 2024, while the one for wave 2 took place in August 2025.[12] For each wave, we randomly selected 21 news articles among the trending articles from the politics section, which appear at the top of the app during the study period and are the most commonly read on the app. For these news articles, we then construct within-article version pairs, with each version pair consisting of the *Regular* version and a rewritten version.[13] Each respondent then rated six randomly selected version pairs along different dimensions. Version pairs are presented in randomized order. We provide respondents with the title, the article's lead, and the full article. To measure whether our rewrites affected the target dimensions as expected, we ask half of the

[11]The experimental instructions are available here: `https://raw.githubusercontent.com/cproth/papers/master/tmt_instructions_appendix.pdf`.

[12]Given increasing concerns about bots on online survey platforms (Rilla et al., 2025), we took additional measures in the wave 2 survey. In particular, participation in our survey required authentication with a mobile phone, where each physical device could unlock the study only once.

[13]In wave 1 of the human validation, we did not include the *Facts Only* rewrite. To ensure overall similar sample size across the waves, we over-sampled *Facts Only* rewrites in wave 2 of the human validation by a factor of 2.

Figure 3: Human Validation of Article Rewrites



Note: This figure shows the share of respondents rating the rewritten version of the article as being more entertaining, conservative, liberal, positive, negative, or less complex compared to the regular version of the article in a blinded binary comparison. The black dots show the fraction of respondents correctly identifying the article version that was varied along the target dimension (e.g., *Entertaining* rewrites being rated as more entertaining). The gray squares show the fraction of respondents classifying an article as belonging to a spillover dimension (e.g., non-*Entertaining* rewrites being rated as more entertaining). Error bars represent 95% confidence intervals.

respondents to rate the article pairs in the target dimension of interest. For example, if the target dimension is complexity, we ask our respondents "which article is more complex?" To assess whether our rewrites systematically induced spillovers to other dimensions, we assign the other half of respondents to rate the article pair along a randomly chosen non-target dimension (e.g., "which article is more positive?"). This design allows us to examine whether we effectively changed articles only along the target dimension or also along other dimensions.

Results The black circles in Figure 3 separately present the fraction of respondents that correctly classified the article for each of the seven different rewrite dimensions. The figure highlights that the fraction of human raters who correctly classify the articles is quite high on average. Around 80% of raters correctly identify the *Entertaining*, *More Positive*, *More Negative*, and *Less Complex* rewrites. For the political rewrites (*Liberal* and *Conservative*) and *Facts Only*, the fraction of correct classifications is slightly lower at around 70%. We document similar human ratings of the article pairs for raters who identify as Democrats and Republicans (Appendix Figure A1).

The gray squares in Figure 3 show the fraction of respondents who classified the article as belonging to a given spillover dimension. For example, the gray square for *Entertaining* means that articles rewritten in all dimensions other than *Entertaining* are, on average, classified as the *Entertaining* version in 49% of the time. If non-target dimensions were not systematically affected by rewrites, we would expect this number to be close to 50% if respondents select a version at random in the absence of a perceptible difference. We find similar patterns in all other spillover dimensions and cannot reject the null hypothesis that rewrites did not systematically affect individual non-target dimensions. We observe similar human classification patterns in wave 1 and wave 2 of the survey.

Appendix Figure A2 shows the spillovers to other rewrite dimensions separately for each target dimension. This figure reveals overall muted spillovers for each of our target dimensions, except for the *Facts Only* rewrite. In particular, *Facts Only* articles are rated as less negative and less liberal, consistent with these rewrites being more fact-driven and less opinionated. As such, being rated as less political and less likely to be negative is a natural byproduct of more neutral and fact-driven reporting. Appendix Figure A30 provides additional evidence that respondents update their beliefs about the article characteristics after using the news app for 10 days in all dimensions except for entertainment.

2.7 Ethical Considerations

We took several measures to mitigate potential concerns about AI-generated news content.[14] First, trained research assistants vetted the factual accuracy of rewritten versions compared to the baseline AP article during our study period. Second, our multi-agent AI system incorporates multiple quality checks, including screening for harmful content. Third, our AI writing assistants are explicitly instructed to uphold rigorous standards of journalistic integrity and to neither omit nor fabricate information. Fourth, our rewrites for conservative and liberal perspectives do not introduce unsupported causal claims and do not affect the underlying facts mentioned in the rewritten articles. Fifth, our news app only relies on articles from a high-credibility source, the Associated Press. Sixth, we employ a conservative rewriting system where an AI agent first reviews and then flags articles that would be potentially inappropriate to customize along a specific dimension (e.g., stories surrounding death or terror cannot be made more entertaining or more positive).

3 The Demand for Customization

Before addressing our main research questions, we first examine whether there is demand for news customization, as this would suggest that there is a mismatch between their preferences and the supply of news, which they think news customization can help address.

Examining whether people value news customization requires a design that isolates this demand from people's demand for the underlying news articles, since customization can only be offered as a bundle with the articles to which it is applied. To address this, we offer participants

[14] We received IRB approval from MIT, the Frankfurt School of Finance & Management, the University of Cologne, and the NHH Norwegian School of Economics.

access to different versions of the app in a between-subjects design: one version with the option of customization and an identical version without customization. We then infer the demand for customization from differential download rates of different app versions. This revealed-preference measure captures the first stage of product adoption: whether people are willing to incur the cost of downloading a new app in order to gain access to news customization.

3.1 Experimental Design

Sample On December 13 and 14, 2024, we recruited 5,005 Android and iPhone users via Prolific, a survey platform commonly used in social science research (Peer et al., 2022). Appendix Table A4 presents summary statistics of our sample and Appendix Table A5 examines the integrity of randomization. Our sample matches the US population in terms of age and gender, but includes a higher share of individuals with at least some college education (86% in our sample vs 63% in the general population), which is a common feature of online panels (Armantier et al., 2017). 42% identify as Democrats, 33% as Independents, and 25% as Republicans. 84% primarily consume news online, which is very close to the 86% of Americans who report consuming news online.[15]

Design Appendix Figure A3 provides an overview of the design. Participants are randomly assigned in equal proportions to one of three conditions: the control condition or one of the two customization conditions. All participants are informed about our AI news app with AI-written articles based on content from the Associated Press, a high-quality news provider. The instructions emphasize the factual accuracy and human review of AI-written articles to mitigate potential participant concerns.

Participants in the customization conditions *additionally* learn that the news app includes the option of customizing news articles in terms of three target dimensions (complexity, sentiment, and entertainment) without changing the facts included in the news article. Specifically, participants receive the following instructions:

> The key feature of our news app is that you can tailor the news stories according to your preferences. Click the Tailor button and move the sliders to choose how you want the news to be reported. [. . .] You can change slider settings at any point in time, and the app will update the writing of all articles in real-time. Irrespective of how you tailor the news, the underlying facts remain constant. Tailoring also does not affect the selection of articles – it only changes how they are presented.

We explain the effect of each customization dimension to participants and demonstrate how customization would work in practice using a short video.

To assess whether demand is sensitive to the range of adjustable features, we vary the fourth

[15] `https://www.pewresearch.org/journalism/fact-sheet/news-platform-fact-sheet`

dimension offered in the treatment conditions. In one variant, participants can additionally select the political orientation of articles (conservative versus liberal). In the other, they may opt to reduce opinion content and emphasize factual reporting.

> Changes the political perspective of the article by using language and terms that are more commonly used by either conservatives or liberals. Also changes the focus of the article to highlight themes and issues that matter most to each political group.

The rewrite dimension that makes articles less opinionated (*Facts Only*) is described as follows:

> Makes the article less opinionated by focusing on facts only. Also avoids subjective assessments and judgments, and uses more neutral language.

We then ask all participants whether they are interested in downloading the app. For those who express interest, we provide detailed download instructions, including a link to the app and a QR code for scanning.[16] We explicitly communicate to participants that payment for survey participation will not depend on downloading the app. Our main outcome of interest is a dummy variable for whether participants actually download and log into the app.

3.2 Results

Figure 4 and Appendix Table A6 show that news customization increases the download rates and logins into the news app compared to an otherwise identical news app. In particular, Figure 4 shows that the download rate in the customization group is 9.6% compared to 6.8% in the control group without customization ($p < 0.01$). The magnitude of this effect is economically large, corresponding to a 40% increase in download rates compared to the control group.

The customization conditions with political rewrites and *Facts Only* rewrites allow us to see whether there is a higher demand for customization that adds a political perspective or removes opinions. Interestingly, we find similar download rates across the two customization conditions ($p = 0.747$): 9.4% download the app when the customization allows for political rewrites, compared to 9.7% when the customization allows for less opinionated rewrites (as shown in Figure 4). This finding suggests either heterogeneity in demand — where some users prefer political alignment and others prefer more fact-driven content — or that both customization options are viewed as equally attractive.

In addition to measuring latent demand for the app, we can also track how time spent in the app differs between different treatments.[17] While app downloads are endogenous to the

[16]Android users can download the app directly from the Google Play Store, while iPhone users need to first download *TestFlight*—an app for beta testing—before being able to download our app. Respondents take between three to four minutes on average to download and log into the app.

[17]We do not directly observe if the app is open on a user's phone. Instead, we infer time spent from user actions (e.g., clicks, scrolls). Actions separated by less than ten minutes are grouped into a session and the session duration is defined as the time between the first and last action. Session durations are then aggregated (e.g., to the daily level).

treatments, it is nonetheless interesting to examine whether those in the customization treatments spend more time in the app. As shown in Appendix Figure A4, on the day of the app download, users spend significantly more time in the app when offered the *Facts Only* customization compared to no customization (p = 0.067). They also spend somewhat more time when offered political rewrites, although this difference is not statistically significant (p = 0.371). While the overall usage is rather low (users across conditions spend on average 125 seconds in the app on the first day, conditional on downloading it, and usage quickly drops thereafter), these results suggest that customization in general and the *Facts Only* option in particular lead to higher initial news engagement.[18] Given our discussion above, our first main result follows.

Result 1. *We document a latent demand for news customization: respondents offered the app with customization are 40% more likely to download it than those offered the same app without the customization option. We find similar demand for customization with political rewrites and less opinionated rewrites, suggesting an underlying demand both for amplifying and minimizing opinion in the news.*

While this data underscores that there is latent demand for news customization, it only allows us to characterize user preferences for a small subset of the population and it does not allow us to study downstream outcomes with sufficient statistical power. Measuring news consumption with customization in a broader population, and assessing the implications of such customization, requires an experimental design with two additional features: prior user experience with the customization tool and comparable levels of engagement across the conditions with and without customization. We turn to such a design in the next section.

4 News Consumption With Customization

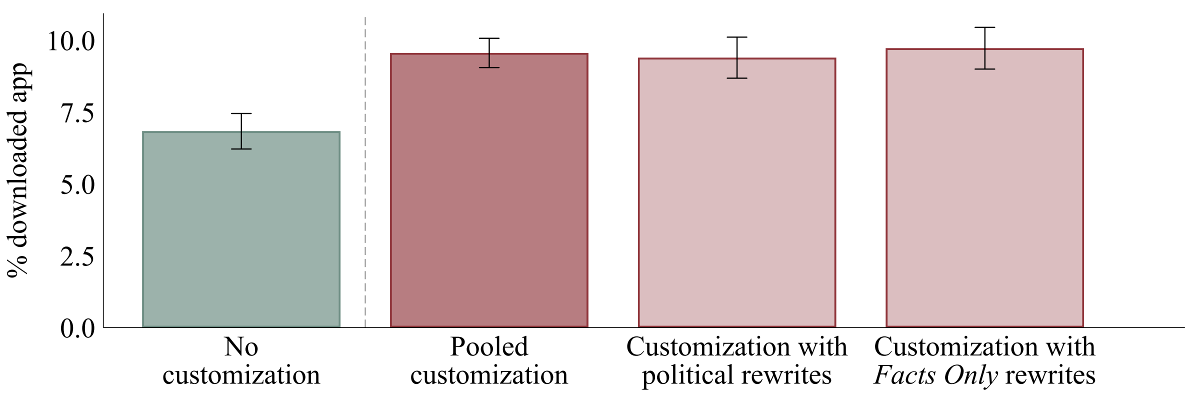
Having established latent demand for news customization, we now turn to our main research question: measuring news preferences. In our main experiment, users are randomly assigned to customization or control, ensuring comparable exposure and enabling us to identify preferences.

4.1 Measuring News Preferences with Customization

Measuring consumers' news preferences is difficult for a variety of reasons. First, news articles vary across multiple dimensions, even when covering the same event. This multi-dimensionality complicates revealed-preference approaches that infer preferences from observed choices. If we conceptualize news articles as a vector of product attributes, then news consumption involves simultaneous trade-offs over multiple correlated attributes. If, for example, slanted news tends to be more vivid or sensational, then disentangling a preference for slant from a preference for a more vivid language becomes more challenging.

[18]Appendix Figure A5 presents average app usage frequency and time during the first seven days after download, confirming the results from day 1.

Figure 4: Demand for News Customization



Note: This figure shows the share of respondents who downloaded the AI news app and logged in at least once across conditions. The first bar shows the download rate in the no customization condition. The second bar shows pooled download rates for both customization conditions. The last two bars show download rates separately for each customization condition (with political rewrites and with *Facts Only* rewrites). Error bars represent standard errors of the mean.

Second, news articles are *experience goods*—their quality and fit can only be assessed after consumption (Gentzkow, 2018). Since news outlets typically optimize for engagement rather than consumer satisfaction (Kleinberg et al., 2024), "clickbait" stories are supplied even though consumers might regret reading them ex post. While readers use past experience with outlets to form expectations (Jo, 2019), these expectations might be systematically miscalibrated, especially with clickbait stories where headlines deliberately misrepresent actual coverage. As a result, the news supply may not fully span the full support of actual consumer preferences.

Our descriptive evidence on how much time people spend consuming news in different settings offers a new approach for measuring and characterizing heterogeneity in news preferences. Inferring consumer preferences from news consumption under customization has several advantages. First, choices are made in a deliberate way that should align with people's expectations of what they are getting. For example, by choosing the *Facts Only* customization, consumers directly reveal their preference for a less opinionated article. Second, as our customization aims to vary individual attributes while minimizing "spillovers" to other dimensions, we at least partly address the issue of correlated product attributes. Third, by measuring actual consumption patterns over a prolonged period of time, our evidence is less plagued by concerns about social desirability bias compared to survey evidence (Bursztyn et al., 2025).[19]

[19] Although most Americans report that newspapers should not favor one side (Mitchell et al., 2018b), social desirability bias may deter readers from admitting they prefer news aligned with their views.

4.2 Experimental Design

The ideal design for measuring news preferences with customization requires four main features. First, users need a period of time to get sufficient experience with customization, actively trying different settings. Second, this learning should occur in a naturalistic, real-world news app environment rather than in a stylized lab setting. Third, to isolate the effects of customization from the content of the underlying news stories, we need a control group that reads the same underlying *distribution* of news stories without being offered customization. Finally, the design must induce exposure to customization while ensuring that all treatment groups exhibit comparable overall app use, so that any differences in downstream outcomes reflect customization rather than different usage levels or selection into the treatments.

We address these requirements by conducting a field experiment over the course of three weeks. The experiment includes four survey waves: a recruitment survey, an app download survey with randomization into treatments, a midline survey, and an endline survey. In the recruitment survey, we measure baseline characteristics and ask for the respondents' interest in getting "free early access" to an AI-tailored news app. A few days later, we invite respondents who expressed an interest in the app to a follow-up survey, where respondents who still express an interest in the app get a link to download it. Importantly, respondents only learn about their treatment status *after* they have downloaded and logged in to the app. Respondents subsequently use the app for 10 days and are paid monetary incentives for daily usage. We then administer a midline survey where we measure downstream outcomes and app perceptions. After a further 10 days, we administer the endline survey where we again elicit downstream outcomes and app perceptions. Figure 5 provides an overview of the experiment, which we describe in more detail below.

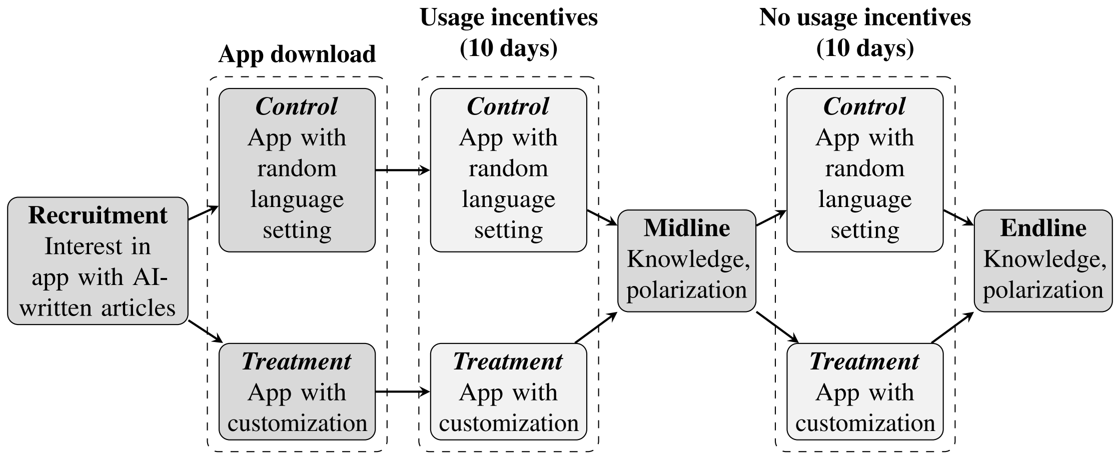
4.2.1 Recruitment Survey

The purpose of the recruitment survey is twofold. First, we want to recruit a relevant group of people interested in receiving access to an AI-based news app and who have an Android or iOS smartphone. Second, we want to measure demographics and other baseline characteristics.

The survey starts with a standard attention screener and questions about their current smartphone. Respondents who fail the attention check or do not have an iPhone or Android phone are screened out of the survey. We then proceed with questions on news consumption and attitudes towards the media, including questions about which news outlets they had consumed in the prior 12 months as well as questions about news satisfaction, news fatigue, interest and trust in the news, and views on current news reporting. We also ask questions on views about AI, political attitudes (including questions on affective and issue political polarization), questions on mental health, and a rich battery of demographic questions.

To elicit interest in our AI-based news app in a natural way, we frame it as an opportunity to

Figure 5: Overview of the Usage Incentive Experiment



Note: This figure presents an overview of the experiment with usage incentives to study news consumption with customization. The experiment starts with a recruitment survey. Participants who complete the recruitment survey and express interest in an app with AI-written articles are randomized into treatment (app with customization and cross-randomized to political or *Facts Only* rewrites) and control (app without customization and cross-randomized to a fixed rewrite version), and invited to the app download survey. Participants who download the app receive incentives for using the app for the next 10 days. After that, they are invited to a midline survey. Then, they can use the app for another 10 days but do not receive incentives. After that, they are invited to an endline survey.

get "free early access" to the app before its official launch. We inform our respondents that the app offers frequent updates with relevant news articles. We highlight that the content is based on trending stories from the Associated Press and that the stories are rewritten using AI technology "to deliver essential information quickly and clearly." Finally, we emphasize that the core facts remain unchanged from the AP sources, "guaranteeing the same level of trust and credibility you expect from seasoned journalists." Importantly, we do not mention customization in this description. Respondents are told that those who express an interest in the app will be invited for a follow-up survey with instructions on how to get access.

4.2.2 App Download Survey

A few days after the recruitment survey, respondents who expressed an interest in the app are invited to the app download survey. We first ask respondents whether they are still interested in downloading the app. For those who say yes, we then give them more information about the app. Respondents are told that they will receive a $2 bonus if they download the app, enter their log-in credentials, complete a short onboarding process in the app, and enable push notifications. If they agree with this, they receive login credentials and download instructions.

To avoid differential selection into downloading the app across treatment arms, respondents only learn about the details of the app during the onboarding process *after* they have downloaded it. Our pre-registered analysis sample includes respondents who download the app, log in, and complete the onboarding process.

Treatment assignment We randomly assign respondents to two different customization treatments and a control group. In both customization groups, respondents can customize news articles along four dimensions in real time. Three of the four dimensions are identical in both customization treatments, while the fourth dimension differs across treatments. Respondents in both customization treatments can choose between a regular default version or customize the articles in terms of entertainment, sentiment (positive or negative), and complexity, while holding the underlying news event constant. For the fourth dimension, we cross-randomize whether respondents can customize the political leaning of the articles (liberal or conservative) or remove opinion content from the articles. This cross-randomization allows us to make a treatment comparison between adding a political perspective or removing opinion content. As shown in Figure 1, respondents can, at any time, customize articles through changes in the slider settings. The order in which sliders are shown in the app is randomized between respondents.

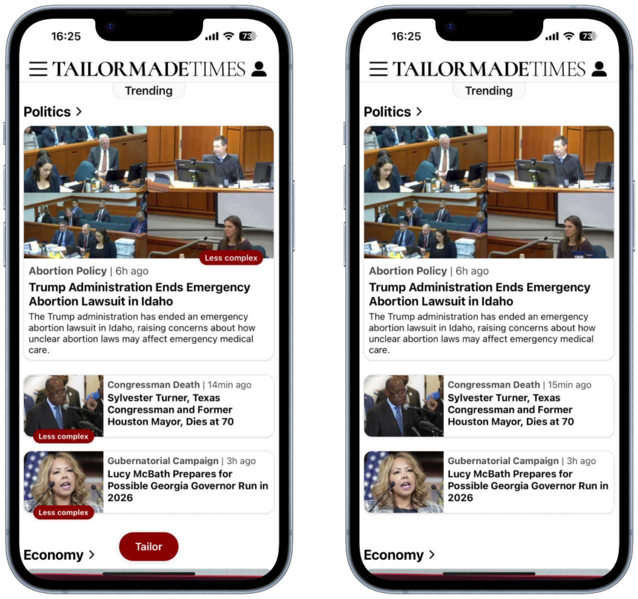
Respondents in the control group are not offered customization and are also not aware that multiple different rewrites of the same underlying regular version exist. Instead, they are randomly assigned one of the versions for all articles they read in the app (*Regular*, *Facts Only*, *More Positive*, *More Negative*, *Less Complex*, *Liberal*, *Conservative*, or *Entertaining*). As shown in Figure 6, the feed looks identical for control group respondents and treatment group respondents, except that control group respondents are not shown version labels and also do not have access to the "Tailor" button. Since control group respondents are not informed about which rewrite version they receive and are also unaware that multiple versions of the articles exist, they must rely on their experience with the app to infer anything about the articles.

App usage incentives To provide respondents with enough experience using the app and experimenting with the different article versions and to measure the consequences of consuming customizable versus non-customizable news, it is essential to ensure high news consumption in the app and roughly equal consumption between the different treatment groups. To achieve this, we implement a series of usage incentives. First, we make the $2 bonus for downloading the app conditional on enabling push notifications. Second, after people have completed the onboarding process and enabled push notifications in the app, we inform them about a second bonus opportunity in which they earn $1 for each day they read at least three articles in the app over a 10-day period. Note that we do *not* incentivize using the customization slider. Third, we offer participants an additional $2 bonus if they keep push notifications enabled for the following 21 days. Since we observe all activities within the app, we can monitor compliance without drawing on any self-reported data.

4.2.3 Midline and Endline Surveys

Ten days after the app download survey, respondents are invited to the midline survey. In this survey, we measure news satisfaction, self-reported news consumption outside of the app, time substitution, and general app perceptions. In addition, we measure affective political polarization,

Figure 6: The AI News App: Customization vs. Control Version



Note: This figure shows screenshots of the app in the customization and control group. The screenshot on the left displays the app as shown to respondents in the customization treatment who have actively chosen the *Less Complex* rewrite. The screenshot on the right displays the app as shown to control group respondents cross-randomized into the *Less Complex* rewrite. Labels are only shown for respondents in the customization group. Respondents in the customization group can customize using the "Tailor" button.

issue polarization, factual knowledge, and mental health. Ten days after the midline survey, respondents are invited to the endline survey. The endline survey follows a similar structure to the midline survey.

4.2.4 Sample

Setting As in the demand experiment, we recruit respondents through Prolific, a survey platform commonly used in social science research with high data quality (Peer et al., 2022). Prolific makes it easy to pay out monetary bonuses and re-contact respondents for multi-wave studies. 17,854 participants completed the recruitment survey.[20] Of these, 53% expressed interest in early access to the app and were invited to the app download survey. Among those invited to the download survey, 53% started the survey, and 70% of those also downloaded the app. This

[20] We collected the data in two separate waves: one in October/November 2024 and another in January/February 2025. While we aimed to recruit 2,500 respondents in the first wave, as indicated in the pre-analysis plan, we only managed to recruit 1,439 app users. We therefore decided to conduct a second data collection in February 2025. In this second wave, we also included the *Facts Only* customization and control arms, which were not part of the initial collection. We oversampled the *Facts Only* arm in the second wave to achieve statistical power comparable to that of the other treatment comparisons. Otherwise, the waves were identical except for minor adjustments in some of the questions due to the new presidential administration.

resulted in a final sample of 3,478 respondents who downloaded the app and were assigned to treatment (Appendix Figure A6). To guard against bots, our experiment required participants to pass multiple attention checks, complete CAPTCHAs, and download an app on their mobile phone.[21]

We assign close to 1,500 respondents (43%) to the customization treatments (21.5% in customization with *Facts Only* and 21.5% in customization with political rewrites) and close to 2,000 respondents (the remaining 57%) to the control group (who are cross-randomized in equal proportion to each of the rewrite versions, see Appendix Table A7). Among the 3,478 respondents who download the app and are assigned to treatment, 76.5% (2,661) participated in the midline survey and 71.7% (2,495) participated in the endline survey (Appendix Figure A6). There is no selective attrition by treatment status across survey waves (Appendix Table A8).

Summary statistics Appendix Table A9 presents summary statistics of our final sample and Appendix Table A10 provides evidence on the integrity of randomization. The average age is 37.7 years, with respondents ranging from 18 to 78, skewing younger than the national median of 38.9 years. About 56.4% identify as female, higher than the US average of 50.8%, and 69.1% are White, broadly comparable to the national figure of 61.6%. 87.8% of respondents report having at least some college education, exceeding the national rate of 63%. Politically, the sample leans more Democratic (42.8%) than the general population (30–35%), with 33.2% identifying as Independent and 24.0% as Republican, slightly underrepresenting Republicans. Overall, as is common for online data collections, our sample is not fully representative of the broader US population, skewing younger, more female, more educated, and more digitally oriented.

Baseline news consumption 58% of our respondents read the news at least somewhat frequently, 90.5% primarily access news online, significantly higher than the national average of 58%, and respondents report using an average of 1.8 news apps. To examine whether our sample has a news consumption that is broadly representative of the US online population, we map each outlet they reported reading at baseline onto the ideological alignment scale developed by Bakshy et al. (2015). This broadly used measure of slant assigns each outlet a score ranging from -2 (very liberal) to +2 (very conservative), based on the average self-reported ideology of Facebook users who shared articles from that outlet. Averaging these scores across our sample yields a slightly left-of-center, unimodal distribution with a mean of -0.123 and a Democrat–Republican gap of 0.167, as shown in Appendix Figure A7. Most respondents cluster around centrist outlets, spanning from center-left outlets like *The New York Times* to center-right outlets like *The Wall Street Journal*. This distribution closely resembles the one observed in the nationally weighted browsing panel analyzed by Guess (2021), whose data exhibits a comparable mean (-0.116) and

[21] Downloading an app on a mobile device serves as a particularly strong anti-bot measure. Unlike attention checks or CAPTCHAs, which sophisticated scripts can sometimes bypass, app installation requires access to a real device, interaction with an app store, and user authentication. This additional friction makes it highly unlikely that automated agents could participate in the study.

spread. Taken together, these findings suggest that our sample's baseline news diet is broadly in line with the broader US online audience.

Selection into app downloads While our final sample exhibits a baseline news diet broadly comparable to that of the wider US online population, there could still be some selection into who chooses to download the app. We therefore examine patterns of selection into app downloads using a broad set of baseline characteristics. For this analysis, Table A11 contrasts the full sample of about 17,000 respondents who completed the baseline survey with our final sample that ultimately installed our app using a linear probability model. Given the high statistical power of this comparison, many coefficients are expected to be statistically significant; we thus focus on those of economically meaningful magnitude. As expected, respondents who consume the news at least somewhat frequently and those who primarily consume news online are 2.5 to 4.6 percentage points more willing to download the app ($p < 0.01$, Column 1). Download probability is also increasing in the number of news apps participants have on their phones. We also find that generalized trust in AI positively predicts the chance of downloading our AI news app. Regarding demographic characteristics, individuals with college education are 4 percentage points less likely to download the app ($p < 0.01$). Reassuringly, the ideological composition of respondents' baseline news consumption is unrelated to app downloads (Column 2).

4.2.5 Compliance with Usage Incentives

In this experiment, respondents were paid monetary incentives to engage with the news app. As a first step, we examine the extent to which respondents complied with the usage incentives. Respondents across the customization and the no-customization groups, on average, spend approximately four minutes a day in the app (Appendix Figure A8, Panels (a) and (b)).[22] This corresponds to about 43% of the average daily online news consumption in the United States, which is 9.7 minutes (Allen et al., 2020). Respondents in both groups log in on between 6 and 7 days on average during the 10-day incentive period (Appendix Figure A8, Panels (c) and (d)).[23] Taken together, our usage data confirms that people complied with the incentives. Moreover, there is no differential app usage across treatment arms during the incentive period—as intended by our design.

4.3 How Do People Consume News with Customization?

In this subsection, we provide descriptive evidence on people's news consumption in different customization groups that allow them to express their news preferences. For this exercise, we first derive the share of time spent in each customization setting at the individual level, and then average individual time shares across all respondents, weighting the consumption shares by the

[22]Similar to the demand experiment, we infer time spent from actions (e.g., clicks, scrolls) because we do not directly observe if the app is open on a user's phone.

[23]Appendix Figure A9 shows that there is also no differential downloading across the customization and no customization groups, confirming that the treatment assignment worked as expected.

amount of time spent consuming news in the app.[24] We calculate this share for all respondents with non-zero consumption during the respective period, i.e., those without consumption are excluded. We thus characterize how much time the *average* respondent spends reading news in different customization settings.

We first analyze usage data from the app usage incentive period (days 1–10) to ensure that differences in news consumption patterns across customization groups reflect relative news preferences rather than changes in the sample composition across customization groups.

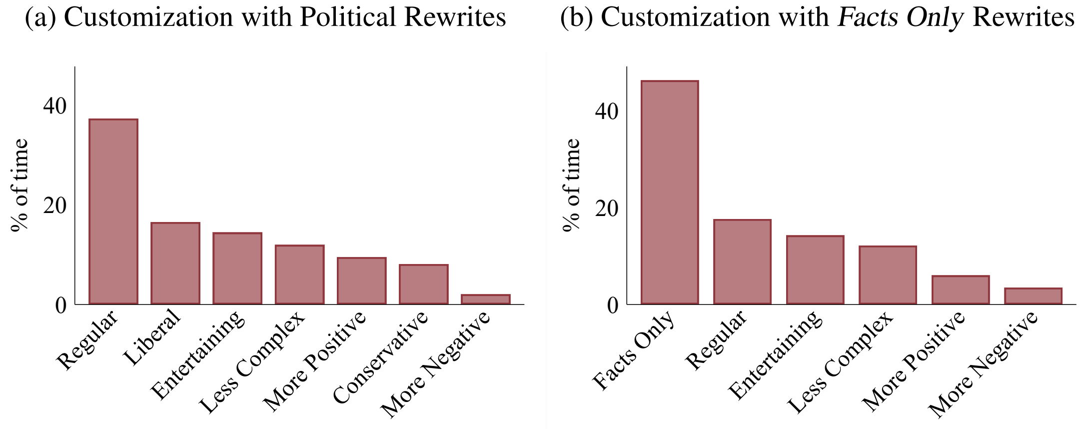
4.3.1 Customization with Political Rewrites

In the political rewrites treatment (Panel (a) of Figure 7), respondents spend 37% of their time in the *Regular* setting, the neutral rewrite closest to the original AP style. Despite being the default, *Regular* accounts for just over one-third of time; almost two-thirds is spent on other options. 17% of time is spent in the *Liberal* setting, followed by *Entertaining* (14%), *Less Complex* (12%), and *More Positive* (10%) settings. Interestingly, despite experimental studies showing that more negative news leads to more engagement and higher consumption rates (Robertson et al., 2023), the only customization option that is virtually never used is the *More Negative* setting (2%). The non-use of this setting suggests that the "demand" for negativity observed with non-customizable news is likely to reflect users making choices in the moment that are inconsistent with what they actually want (Kleinberg et al., 2024). By contrast, when users are aware of the article attributes and can make deliberate choices about how the news should be reported, they actively avoid negative news.

Political heterogeneity The *Liberal* option is used more often than the *Conservative* (17% vs. 8%), partly because the sample has more Democrats than Republicans (Table A9). Analyzing consumption separately by party reveals strong demand for politically aligned news among both Democrats and Republicans (as shown in Figure A11). More specifically, Democrats spend 24% of their time using the *Liberal* setting, making it second most popular after the *Regular* default among Democrats (38%). Republicans spend 25% of their time using the *Conservative* setting (compared to 39% in the *Regular* default). Democrats and Republicans are, thus, equally likely to demand ideologically aligned news. Furthermore, while the political rewrites could have been used to "see both perspectives" by switching between the *Liberal* and *Conservative* options, we find that both Democrats and Republicans use them almost exclusively to align the news with their own ideology. Our evidence is thus consistent with prior work showing that a subset of consumers have a revealed-preference for politically aligned news (Gentzkow and Shapiro, 2010). However, an important difference in our setting is that our users make a *deliberate* choice to align the news with their own ideological views. Nonetheless, the vast majority of consumers

[24] As a robustness check, we weight the consumption shares across article versions equally for all respondents irrespective of the time they spend in the app. Appendix Figure A10 shows that our findings are robust to this weighting procedure.

Figure 7: News Customization in the Incentive Period



Note: This figure shows the share of time spent consuming news in different customization options during the incentive period, weighting the individual consumption shares by the amount of time spent consuming news in the app. Panel (a) shows the shares for the respondents in the customization group with political rewrites (N = 747). Panel (b) shows the shares for the respondents in the customization group with *Facts Only* rewrites (N = 718).

favor politically neutral reporting.

Differences in news preferences also emerge between voters and non-voters (as shown in Figure A12). Non-voters are somewhat less likely than voters to consume the news in the default *Regular* option (29% vs. 38%) and much less likely to consume the political rewrites (8% vs 27%). Instead, non-voters spend more time reading the news in the *Entertaining* version (25% vs 13%) and *Less Complex* version (23% vs 11%). In other words, while a sizable fraction of politically active users have a demand for ideologically aligned news, non-voters who are disengaged from the political system instead largely prefer the news to be more entertaining and less complex.

4.3.2 Customization with *Facts Only* rewrites

Surveys and interviews suggest Americans prefer 'just the facts' reporting without opinions or commentary (Eddy et al., 2025; Mitchell et al., 2018b). However, such self-reports may reflect cheap talk or social desirability bias (Bursztyn et al., 2025).

Consumption data from the political customization setting reveal significant demand for politically aligned news, yet the most frequently consumed option remains the politically neutral *Regular* setting. While this could be taken as evidence of a high demand for politically balanced news, a potential concern about this interpretation is that *Regular* is the default option. Thus, we cannot disentangle a demand for balanced news from default effects. Furthermore, *Regular* is not explicitly framed as a politically neutral setting.

To assess whether a large fraction of consumers actually value politically neutral and less opinionated news, we turn to our second customization treatment, which features the *Facts Only* option instead of the political rewrites option. *Facts Only* is explicitly framed to users as a neutral option that "makes the article less opinionated by focusing only on the facts," and it is also *not* the default option. This allows us to obtain a revealed-preference measure for politically neutral and non-opinionated content.

As shown in Panel (b) of Figure 7, the customization treatment including the *Facts Only* option demonstrates a striking and *deliberate* demand for more fact-driven and less opinionated content: Respondents spend on average 46% of their time using the *Facts Only* setting during the incentive period—a share that exceeds the time spent in any other setting by a large margin. Compared to the treatment with political rewrites, we also see a sizable decrease in the fraction of time spent in the *Regular* default option (18%) in the *Facts Only* treatment.

The time shares of the customization options that are *shared* in both customization treatments (*More Positive*, *More Negative*, *Less Complex*, and *Entertaining*) remain largely unaffected. The time share of *Facts Only* (46%) can thus almost entirely be accounted for by the time shares of *Liberal* (17%) and *Conservative* (8%) in the customization with political rewrites (which are no longer available) and the differential time use of *Regular* across the two treatment arms (37% - 18% = 19%). These patterns underscore that the high use of *Regular* in the customization treatment with political rewrites likely reflects a demand for fact-driven and neutral news rather than default effects. Furthermore, the popularity of the *Facts Only* option suggests that people's stated desire for less biased and more fact-driven news reflects a genuine preference rather than social desirability bias.

These findings raise the important question of why consumers frequently select ideologically aligned news, often presented through highly partisan opinion programs, despite our evidence indicating a substantial preference for neutral, fact-driven content. One plausible resolution is offered by theories positing that consumers perceive aligned news as inherently more accurate (Gentzkow and Shapiro, 2006). Empirical support for this explanation comes from recent studies documenting that Americans regard opinion-based programming as informative (Bursztyn et al., 2023a) and perceive politically congruent news statements—even those explicitly presented as opinions—as more factual (Mitchell et al., 2018a). This phenomenon helps explain our observation of substantial demand for politically aligned content in the initial customization treatment, where the *Facts Only* option is not included, as well as the large time shares spent in the *Facts Only* rewrite.

Heterogeneity We only see minor differences between Republicans and Democrats in their use of the *Facts Only* option—which is by far the most popular rewrite option across both groups—demonstrating that the revealed demand for more fact-driven and less opinionated content holds across the political aisle.

One concern around our results is that they are driven by diversification motives in which respondents only prefer *Facts Only* because their existing news portfolio is politically biased.

Interestingly, the demand for *Facts Only* is consistently high irrespective of whether respondents have a high or low baseline news consumption (Figure A13). Furthermore, it is also consistently high irrespective of the slant of people's prior news consumption (Figure A14). These results suggest that the consumption of *Facts Only* news reflects a genuine preference for more factual and less opinionated content that is not mainly driven by diversification motives.

One of the major complaints among those with low news satisfaction is that the news is too opinionated and politically biased (Reuters, 2023). If this is a genuine concern, we should expect those with low baseline news satisfaction to demand more fact-driven and less opinionated news. This is precisely what we find: Those who express greater dissatisfaction with the news at baseline are significantly more likely to use the *Facts Only* option (as shown in Figure A15). This finding speaks to our broader motivation that news customization allows for a better match between consumer preferences and their news consumption, especially for attributes that are less covered among mainstream news outlets. The *Facts Only* option, which allows people to *transparently* avoid opinionated content, might be especially relevant in this context, as consumers have a difficult time separating facts from opinion in regular news (Mitchell et al., 2018a). Our second main result is as follows:

Result 2. *We document substantial heterogeneity in news preferences. While a nontrivial share of respondents demand politically slanted news aligned with their views, the majority favor politically neutral reporting. Moreover, demand for less opinionated news is large and consistent across the political spectrum.*

4.3.3 Robustness and Use of Customization Across Time and Topics

We examine robustness across several dimensions and explore heterogeneity in customization usage. Order randomization of sliders does not influence time shares spent in the different settings (Table A12). News preferences are consistent across topics, with similar usage distributions indicating a stable preference for particular settings (Appendix Figures A16, A17). After initial experimentation, users rapidly settle into stable news preferences, reflected by a sharp decline in slider adjustments and persistent usage patterns throughout the observation period (Appendix Table A13 and Appendix Figures A18 and A19). Users with above-median slider use in the *Facts Only* group spend somewhat less time in the default *Regular* setting, instead consuming more news in the *Less Complex* and *More Negative* settings. By contrast, within the political customization group, we observe only minor quantitative heterogeneity based on above-median slider use (Appendix Table A14). Finally, baseline trust in AI does not systematically affect consumption patterns under customization (Figure A20).

4.4 Does Customization Improve Matching with Preferences?

Having documented substantial heterogeneity in individuals' use of customization, we next examine the extent to which it enhances the alignment between respondents' news consumption choices and their stated preferences for news reporting. We then investigate how consumption patterns differ between respondents who have access to customization and those in the control group, who are randomly assigned to a fixed setting.

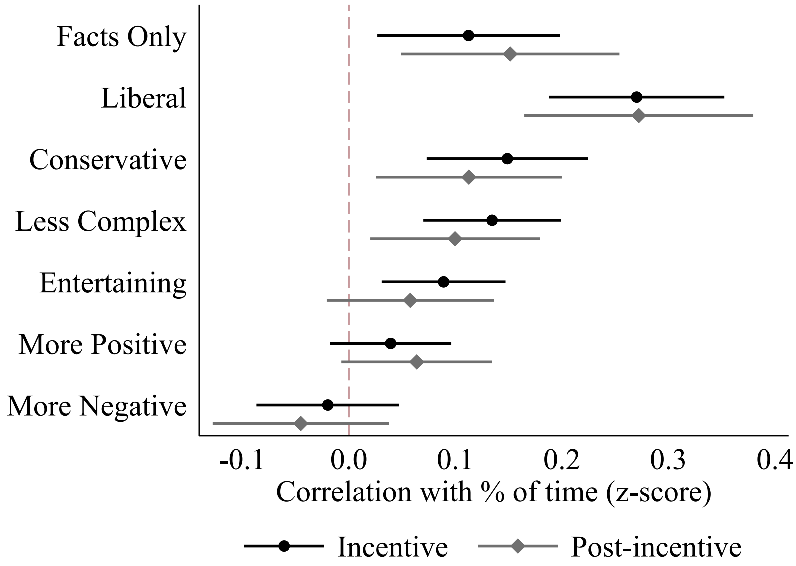
4.4.1 Perceived Mismatch and News Consumption

To understand whether customization improves the alignment between people's news consumption and their news preferences, we exploit a series of baseline pre-treatment questions in which we ask respondents whether they think news articles are too complex, too negative, too positive, too liberal, too conservative, too opinionated, or not entertaining enough. Figure 8 reports results from multivariate regressions that correlate answers to these questions with the time shares spent in different settings both during the incentive period and the post-incentive period.

The results reveal that people use customization to address the perceived mismatch between their existing supply of news and their news preferences. During both the incentive and post-incentive periods, respondents who believe news articles are too opinionated spend around 10-15% of a standard deviation more time using the *Facts Only* setting ($p < 0.01$). Similarly, respondents who believe the news is too conservative spend around 30% of a standard deviation more time using the *Liberal* setting ($p < 0.01$), whereas those who think the news is too liberal spend around 10-15% of a standard deviation more time using the *Conservative* setting ($p < 0.01$). Those who think the news is too complex spend significantly more time using the *Less Complex* setting ($p < 0.01$), while those who think the news is not entertaining enough spend significantly more time using the *Entertaining* setting, though this result is only significant during the incentive period ($p < 0.01$). We also observe positive correlations between thinking the news is too negative and using the *More Positive* setting, though the point estimates are not statistically significant at conventional levels. The only dimension for which we do not observe correlations in the expected direction is the *More Negative* setting, likely reflecting that few people think the news is too positive and very few users end up using the *More Negative* setting in practice, leading to point estimates close to zero.

We can also study similar correlations between the perceived mismatch and news consumption for our control group respondents who were randomly assigned one setting for the whole period (as shown in Figure A21). During the incentive period, when incentives induce high news consumption among all respondents, we see no systematic patterns among control group respondents. Interestingly, for the post-incentive phase, we see that respondents who believe news articles are too opinionated spend around 20% of a standard deviation more time in the app if they are assigned the *Facts Only* rewrite ($p < 0.01$). For the remaining dimensions, however, we see no systematic correlations between perceived mismatch and time use. These results suggest that customization strongly helps improve alignment with preferences relative to being assigned a setting at random. The continued predictive power of preference mismatches for control-group respondents assigned to the *Facts Only* version reinforces our earlier finding

Figure 8: Perceived News Mismatch and Time Spent in Different Customizations



Note: This figure shows correlations between the perceived mismatch of the news and time spent in the corresponding version. Coefficients are estimated from multivariate regressions of the share of time spent with a given customization setting (e.g., *Facts Only*) on the full vector of mismatch dimensions as measured in the recruitment survey. For the mismatch dimensions, we elicit respondents' agreement with the statements "News articles are too [complex/negative/positive/liberal/conservative/opinionated]" and "News articles are not entertaining enough". We encode agreement such that larger values represent a stronger match with the customization option and standardize the values. Results are shown separately for the incentive (black circles) and post-incentive (gray diamonds) periods. All specifications include fixed effects for the day on which the app download survey was completed. Error bars represent 95% confidence intervals.

that individuals strongly prefer factual over opinion-based content. Summarizing the previous findings, our third main result is as follows:

Result 3. *Customization helps users address the perceived mismatch between their preferences for how the news should be presented and their actual news consumption. Those who think the existing news coverage is too opinionated, too conservative, too liberal, too complex, or not entertaining enough actively use the customization to make their news consumption more aligned with their preferences.*

4.4.2 The customization wedge in news consumption

To what extent does customization change people's news consumption towards their stated preferences, such as more fact-driven and less negative content? We answer this question by comparing news consumption in the treatment group with the control groups who were randomly assigned one version. For this exercise, we focus on the post-incentive period to capture people's news consumption in the absence of incentives once they are familiar with the app.

Differences in news consumption between the treatment and control group could emerge either through different *total* news consumption between treatments or different *time shares* spent in each rewrite version. While we observe a natural drop in consumption in the post-incentive period (as shown in Figure A22), there are no treatment differences in time spent on the app between respondents in the customization treatments and the control group (as further discussed in Section A5.4 of the online appendix).[25] Any post-incentive treatment differences in news consumption thus arise from different time shares in the different settings. To benchmark news consumption patterns with customization, we therefore compare the time spent in each setting across the treatments.

To create a time share measure for control group respondents, we first scale each control arm's total time to account for differences in group sizes and then divide this by the total time of all control groups. For the treatment groups with customization, we weight the consumption shares of the two treatment groups by the amount of time spent consuming news in the app.[26] Figure 9 shows the consumption shares across the different article versions for respondents in the control group and compares them to the consumption shares of the treatment groups with customization. To facilitate direct comparisons between the control group and each customization treatment, Panel (a) displays outcomes for all rewrite options in the customization treatment with political rewrites, disaggregated by treatment and corresponding control respondents. Panel (b) presents analogous results for the customization treatment with the *Facts Only* rewrites.

The figure reveals a striking *customization wedge* in time shares between the control groups (gray bars) and the treatments with customization (red bars). As shown in Panel (a), respondents offered customization with the political rewrites spend most of their time in the *Regular* setting (40%) and only 1% of their time using the *More Negative* setting. By contrast, the distribution among control group respondents is remarkably uniform across the different rewrite versions. Most strikingly, the *More Negative* setting has almost the same time share as the *Regular* setting. The only rewrite option that seems considerably less popular than the average among control group respondents is the *Conservative* option, likely driven by the lower fraction of conservatives than liberals in our sample. While the lower time share spent in the *Conservative* control demonstrates that news consumption can be responsive to different rewrite versions, the overall muted response across the control condition demonstrates a surprising non-responsiveness to different versions when customization is not available.

We see an even more striking customization wedge in Panel (b), where we compare respondents offered customization with the *Facts Only* option to the corresponding control groups. Respondents in the customization treatment spend about 50% of their time using the *Facts Only* option, dominating all other customization options. By contrast, the distribution among control group respondents is remarkably uniform, ranging from a time share of 12% in the *More Positive*

[25]We also find no spillovers on self-reported news consumption outside the app (see Appendix Figure A23).

[26]As a robustness check, we also show results for how much time the *average* user spends in the respective customization setting. Appendix Figure A25 shows that our findings are not sensitive to how we aggregate the consumption shares.

setting to 21% in the *Facts Only* setting. All of the control group shares are close to what would be implied by an equal time spent across the settings (the dashed horizontal line in the figure).

Overall, we observe a strong difference in news consumption depending on whether they can customize. While respondents in our customization treatments overwhelmingly prefer fact-driven and less opinionated news when given the choice, our control group respondents show a remarkably uniform response to the different rewrite versions. The contrast is particularly large for *More Negative* news, which is almost never consumed among respondents offered customization but is almost as popular as *Facts Only* among control group respondents. This "customization wedge" is consistent with recent evidence of *inertia* in news consumption in which people keep consuming news misaligned with their stated preferences unless they are prompted to make an active choice about which news outlets to follow (Braghieri et al., 2025a).

Our results demonstrate that customization allows for more sorting and a better match between underlying preferences and news consumption. Furthermore, they raise the question of whether this increased congruence leads to higher news satisfaction among respondents offered customization, which we examine in the next subsection. Given the discussion above, our fourth main result can be stated as follows:

Result 4. *We uncover a substantial customization wedge in news consumption depending on whether respondents are able to customize the news. When respondents customize, they overwhelmingly seek fact-driven and less opinionated news. By contrast, respondents not offered customization spend similar amounts of time in the different versions and are much more likely to consume the negative version in particular.*

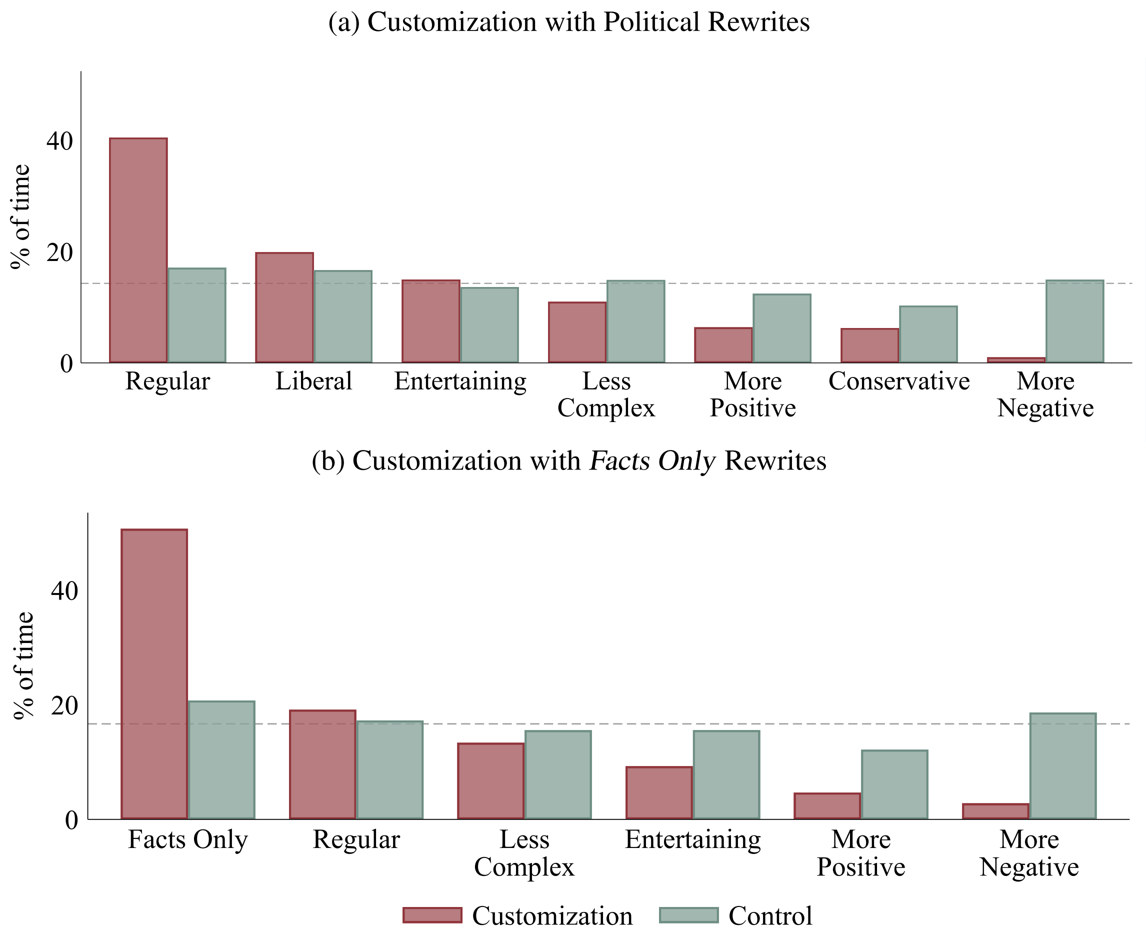
4.5 Customization, News Satisfaction and App Perceptions

The previous section establishes a *customization wedge*, demonstrating that respondents offered customization consume more fact-driven and less opinionated news, whereas those not offered customization are much more likely to consume more negative news. This observation raises the question of whether the absence of customization—and thus a diminished alignment between news preferences and the provided news—also adversely affects consumer satisfaction with news consumption. Moreover, since control group respondents exhibit similar levels of news consumption across rewrite versions, evidence of reduced satisfaction in this group would suggest that observed consumption decisions absent customization poorly capture true underlying preferences.

Empirical strategy To examine the effects of customization on overall news satisfaction, we estimate the following pre-registered specification:

$$Y_i = \alpha + \beta\,\text{CustomizationFacts}_i + \gamma\text{CustomizationPol}_i + \delta X_i + \tau_t + \varepsilon_i, \tag{1}$$

Figure 9: News Customization in the Post-Incentive Period



Note: This figure shows the share of time spent using the different customization options during the post-incentive period weighted by time spent using the app. Panel (a) shows the customization treatment with political rewrites and the corresponding control arms. Panel (b) shows the customization treatment with *Facts Only* rewrites and the corresponding control arms. Panel (a) uses data from both waves, while Panel (b) uses only data from wave 2 given that we only introduced *Facts Only* rewrites in wave 2. The dashed horizontal lines show the shares that would be implied by equal time spent across settings.

where Y_i is respondent i's news satisfaction, X_i is a vector of pre-treatment covariates,[27] τ_t is a fixed effect for the day on which the app download survey was completed, and ε_i is the error term. We use robust standard errors for inference. Our key coefficients of interest are β and γ, which capture, respectively, the effects of being offered customization with political rewrites and customization with *Facts Only*. The omitted group contains control group respondents who are randomized into different language settings without any customization options. β and γ thus capture the change in news satisfaction compared to average satisfaction in the control group.

[27] X_i includes demographics (age, gender, race, income, some college, full-time employed), political leaning (party affiliation and leaning, past voting), news consumption patterns (primarily consuming news online, having a paid subscription to a news platform, consuming news at least somewhat frequently, number of news apps on the phone, primary device for news consumption is mobile), trust in news and AI, and baseline news satisfaction.

To measure news satisfaction, we rely on the following survey item that was included in both our midline and endline collections: "How satisfied are you with the news that you have been reading over the past week?" Respondents answered this question on a five-point Likert scale ranging from 1 (*Very dissatisfied*) to 5 (*Very satisfied*). To ease the interpretation of this measure, we z-score the variable using the mean and standard deviation in the control group.

Results on news satisfaction As illustrated in Figure 10, respondents in the customization treatments report significantly higher satisfaction with news consumption compared to those in the control conditions. The customization treatment with *Facts Only* rewrites increases news satisfaction by 0.19 of a standard deviation ($p < 0.01$), while the customization treatment with political rewrites increases news satisfaction by 0.09 of a standard deviation ($p = 0.032$). Although the difference between the two customization treatments is not statistically significant at conventional levels ($p = 0.10$), the point estimates suggest that customization with *Facts Only* leads to considerably higher news satisfaction than customization with political rewrites. This is consistent with *Facts Only* being the most popular choice among the different settings.[28]

Results on app perceptions To better understand why customization increases news satisfaction, we next study a rich set of outcomes on perceptions of the news app. As shown in Figure 10, both customization treatments increase perceptions of accuracy, trust, and quality. Effect sizes on these outcomes range from 7% to 20% of a standard deviation. Consistent with the results on news satisfaction, effect sizes are consistently larger for the customization treatment with *Facts Only* rewrites. Respondents assigned customization with *Facts Only* perceive the news to be 14% of a standard deviation more fact-driven than control group respondents ($p < 0.01$), whereas respondents assigned customization with political rewrites find the news, if anything, to be somewhat *less* fact-driven. We do not observe any significant or systematic effects of customization on perceived sentiment, political bias, or entertainment.

While customization generally leads to more positive app perceptions, it also increases the perceived complexity of the app. Respondents offered customization report a roughly 20% of a standard deviation higher perceived complexity of the app than control group respondents not offered customization. We thus document a quality-complexity trade-off: Customization leads to higher quality perceptions but at the same time the user experience is more involved. While the net result is still higher news satisfaction on average, the perceived higher complexity of customization could partly explain why our respondents largely prefer to stick to one specific customization setting after a short initial phase of experimentation.

Decomposing average effects In our main specification (Equation (1)), we pool all control group respondents into one group irrespective of their assigned version. To better understand the drivers behind our average treatment effects on news satisfaction, we next examine disaggregated results for all control groups, making the customization treatment with political rewrites the

[28] Figure A26 shows the effects on satisfaction at endline.

omitted group in the regression. As shown in Figure A27, we find that the positive average effect of customization on news satisfaction is driven by lower news satisfaction among respondents receiving the *Conservative*, *More Negative*, *More Positive*, and *Entertaining* rewrites. All of these rewrites result in news satisfaction scores approximately 20 percent of a standard deviation below those from the political customization, and roughly 30 percent below those from the *Facts Only* customization.

Interestingly, the neutral rewrites—*Facts Only*, *Regular*, and *Less Complex*—yield an overall satisfaction that is not significantly different from respondents in the customization treatments. These findings again support the notion that respondents have a latent demand for more fact-driven and politically balanced news. Furthermore, they demonstrate that consumer choices in the absence of customization are a poor measure of consumer preferences: While *More Negative* rewrites are almost never consumed with customization, they are—if anything—over-proportionally consumed among control group respondents. The fact that this consumption of negative news without customization also leads to significantly lower news satisfaction demonstrates that consumption data can lead to misleading conclusions about consumer welfare. This finding is consistent with recent evidence that overly negative or toxic content can be addictive even if it has adverse effects on news satisfaction (Beknazar-Yuzbashev et al., 2025).

Who benefits most from fact-driven news? A consistent finding from our experiment is that readers want *Facts Only* content: It is, by a large margin, the most popular customization choice and the customization treatment with *Facts Only* leads to an increase of 20% of a standard deviation in news satisfaction. One of our main motivations for studying customization is the high general dissatisfaction with the news media and the potential for customization to increase satisfaction through a closer alignment with underlying preferences. Given that the share of Americans closely following the news now is only 38% (down from 51% in 2016)—with potentially large negative externalities for society at large—it is particularly important from a policy perspective to increase news satisfaction among those who currently display high levels of news dissatisfaction.

To shed light on this issue, we estimate treatment effects of the customization with the *Facts Only* treatment separately for each level of baseline news satisfaction (measured on a 5-point Likert scale from 1: *Very dissatisfied* to 5: *Very satisfied*). As shown in Figure A28, we see that the effect is largely driven by those with low baseline satisfaction. Specifically, we observe treatment effects of 23% to 34% among those who are *very* or *somewhat* dissatisfied with the news at baseline, demonstrating that customization with *Facts Only*—where respondents knowingly can select more fact-driven and less opinionated news—is an especially powerful lever for increasing news satisfaction among those who are most dissatisfied with the current news landscape for economic and political news. Given the results discussed in this section, our fifth main result follows:

Result 5. *Respondents offered customization report significantly higher news satisfaction, par-*

ticularly those assigned the customization treatment with the Facts Only option.

4.6 Customization, Political Polarization, and Factual Knowledge

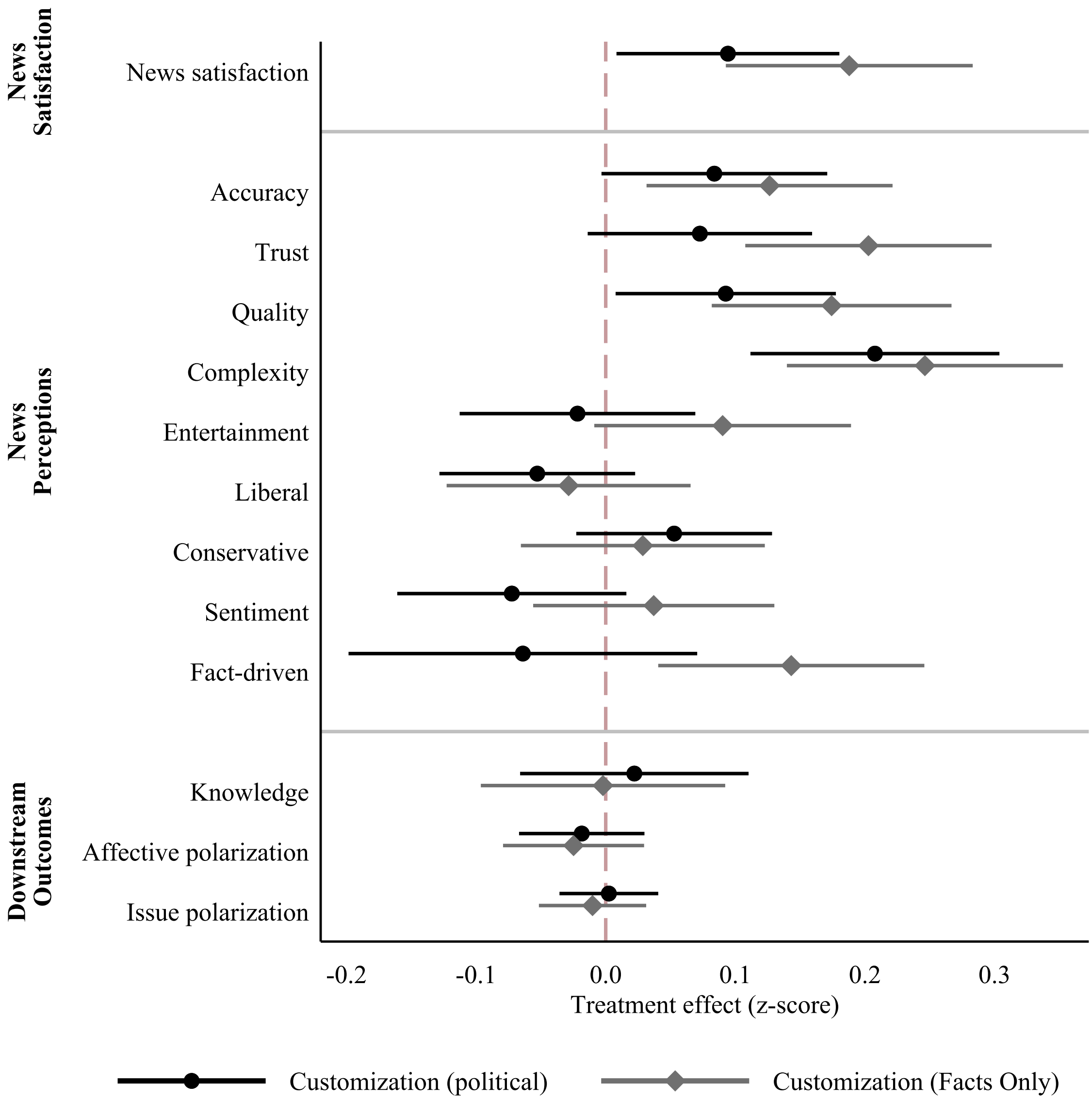
Having established that customization increases news satisfaction, we next examine its impact on political polarization and factual knowledge. Intuitively, customization involving political rewrites could have substantial effects on polarization, though the direction is theoretically ambiguous. While exposure to diverse political viewpoints within a news event may reduce polarization, the prevalent use of political customization to align content with personal ideology may exacerbate polarization by reinforcing echo chambers. Conversely, given its popularity and politically neutral framing, customization with the *Facts Only* option may reduce polarization by promoting exposure to fact-driven rather than opinionated content.

It is also unclear how customization should affect factual knowledge. On the one hand, a better match between underlying preferences and the news might increase the potential for knowledge acquisition; for instance, respondents who think the regular news is too complicated might benefit from using the *Less Complex* option. On the other hand, as discussed in Section 4.5, customization also increases the perceived complexity of the news, which could hinder knowledge acquisition.

Results on affective polarization We begin by examining *affective polarization*, defined as the extent to which individuals express more negative feelings toward those with opposing political views than toward those who share their political orientation (Iyengar et al., 2019). To measure affective polarization, we construct an index from two underlying outcomes (a standard feeling thermometer question and a hypothetical donation task). To estimate treatment effects, we rely on a specification analogous to Equation (1) used to estimate treatment effects on news satisfaction. As shown in Figure 10, we see muted treatment effects on political polarization. The point estimates are close to zero and relatively precisely estimated. Furthermore, the null effect on affective polarization holds across the different control conditions (as shown in Figure A29), emphasizing that the overall muted effect is not driven by potentially polarizing effects in opposite directions of different control conditions.

Results on issue polarization We next examine results on *issue polarization*, namely the extent to which Republicans and Democrats disagree about relevant policy issues. To measure issue polarization, we construct an index based on our respondents' positions on five salient political topics: (i) free trade, (ii) police violence, (iii) sexual harassment, (iv) immigration, and (v) gun control. The index captures the extent to which individuals adopt consistently liberal or conservative positions across these domains. As shown in Figure 10, we also see muted treatment effects on issue polarization, with precisely estimated null effects very close to zero. This null effect also holds across the different control conditions (Figure A29).

Figure 10: Effects on News Satisfaction, Perceptions, and Downstream Outcomes



Note: This figure displays the effects of customization with political rewrites and with *Facts Only* rewrites on user satisfaction and perceptions of the app, as well as downstream outcomes, relative to all control conditions. The first block shows news satisfaction. The second block shows perceptions of accuracy, trustworthiness, quality, complexity, entertainment, political bias (liberal and conservative), sentiment, and whether the content is viewed as fact-driven. The last block shows effects on news knowledge, affective polarization, and issue polarization. All effects are shown in standardized z-scores. All specifications control for demographics (age, gender, race, income, some college, full-time employed), political leaning (party affiliation and leaning, past voting), news consumption patterns (primarily consuming news online, having a paid subscription to a news platform, consuming news at least somewhat frequently, number of news apps on the phone, primary device for news consumption is mobile), and trust in news and AI. All specifications include fixed effects for the day on which the app download survey was completed. Error bars represent 95% confidence intervals.

Results on factual knowledge We measure factual knowledge with a quiz of 15 questions drawn from stories featured in the app, awarding $0.20 per correct answer if completed in under 20 seconds. On average, respondents answered 8 questions correctly in the control condition. Treatment effects on factual knowledge are small and statistically insignificant (Figure 10), though confidence intervals are slightly wider than for polarization outcomes due to the absence of baseline controls. Our final main result follows:

Result 6. *Offering respondents customization does not affect their knowledge, nor does it affect political polarization.*

Discussion Despite significant improvements in news satisfaction, customization has precisely estimated but negligible effects on affective polarization, issue polarization, and factual knowledge. We benchmark our treatment effects using a meta-analysis of randomized controlled trials manipulating overall news exposure or ideological content alignment. This meta-analysis reveals muted effects on polarization and factual knowledge from most interventions (details are presented in Section A5.7 of the online appendix). Given that these studies altered overall news consumption time or ideological content alignment, their potential impact on polarization and knowledge should arguably exceed ours. Nonetheless, as shown in Figure A31, the pooled estimates on affective and issue polarization remain close to zero, and the effects on factual knowledge are also muted and statistically insignificant despite low standard errors. In light of these results, it is not surprising that our intervention, which holds the coverage margin constant, yields muted effects on these outcomes.

5 Conclusion

We study people's news preferences using an AI-powered app that enables users to customize the characteristics of news articles while holding the news event fixed. This design overcomes a core identification challenge: Because news articles are multi-dimensional experience goods and news outlets optimize for short-term engagement rather than long-term satisfaction, market data cannot reliably reveal consumers' underlying preferences (Kleinberg et al., 2024). By allowing respondents to customize article characteristics while holding the underlying news event constant, we observe choices that directly reveal news preferences for specific attributes, such as less complex or more fact-driven content. Furthermore, as customization increases the available product space, we can observe preferences over choices not offered in the existing news market.

Using several large-scale experiments, we document six main findings. First, there is substantial demand for customization, suggesting a mismatch between people's preferences and the existing supply of news that they think customization can help address. Second, we observe substantial heterogeneity in how people use customization. While a significant minority demand politically aligned news, the large majority demand more fact-driven and less opinionated news. Third, customization helps users address the perceived mismatch between their preferences for how news should be presented and their actual news consumption. Fourth, we uncover a "customization wedge" in which respondents not offered customization end up consuming news misaligned with their stated preferences. Fifth, customization leads to significantly higher news satisfaction, particularly when it includes the option to make the news more fact-driven and less opinionated. Finally, while customization increases news satisfaction, it has no significant effects on political polarization or factual knowledge—consistent with evidence from other studies showing that media interventions more generally have muted impacts on these outcomes.

Given the inherent challenges in identifying news preferences from existing market data, tools that allow consumers to make active and deliberate choices about how news is presented are essential for understanding people's underlying preferences. Our results that most people prefer fact-driven news without opinions are consistent with the idea that the observed demand for slanted news in existing media markets to a large extent reflects quality considerations, as in Gentzkow and Shapiro (2006), rather than a strict preference for ideologically aligned content. Consistent with this interpretation, Braghieri et al. (2025a) find that information about bias makes people follow less slanted news outlets on social media.

More broadly, our intervention not only advances understanding of news preferences but also informs policy discussions on the implications of customization as a platform feature. A natural question is why leading news outlets have not already implemented customization more broadly. Given that real-time news customization is only made possible by recent advances in Generative AI, the lack of mainstream adoption could reflect a natural time-lag in adopting new technology. However, it is also worth asking whether it would be profitable for news outlets to offer news customization. While classic models of product differentiation suggest that catering to heterogeneous consumer preferences should increase profitability, it is not obvious that this is the case in the market for news, where many news outlets optimize for engagement rather than consumer satisfaction. Given the importance of social media networks in driving the traffic of many news outlets, it might still be optimal from a revenue perspective to supply "clickbait stories" rather than to offer people more control through customization. Furthermore, while customization increases news satisfaction, it also makes the news seem more complex, potentially further limiting its appeal to news outlets. More work is needed to understand these dynamics further.

References

Acemoglu, Daron, "Harms of AI," Technical Report, National Bureau of Economic Research 2021.

__ **, Asuman Ozdaglar, and James Siderius**, "A model of online misinformation," *Review of Economic Studies*, 2023, p. rdad111.

__ **,** __ **, and** __ , "AI and Social Media: A Political Economy Perspective," Technical Report, National Bureau of Economic Research 2025.

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow, "The Welfare Effects of Social Media," *American Economic Review*, March 2020, *110* (3), 629–676.

__, **Matthew Gentzkow, Winter Mason, Arjun Wilkins, Pablo Barberá, Taylor Brown, Juan Carlos Cisneros, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Sandra González-Bailón, Andrew M. Guess, Young Mie Kim, David Lazer, Neil Malhotra, Devra Moehler, Sameer Nair-Desai, Houda Nait El Barj, Brendan Nyhan, Ana Carolina Paixao De Queiroz, Jennifer Pan, Jaime Settle, Emily Thorson, Rebekah Tromble, Carlos Velasco Rivera, Benjamin Wittenbrink, Magdalena Wojcieszak, Saam Zahedian, Annie Franco, Chad Kiewiet De Jonge, Natalie Jomini Stroud, and Joshua A. Tucker**, "The effects of Facebook and Instagram on the 2020 election: A deactivation experiment," *Proceedings of the National Academy of Sciences*, May 2024, *121* (21), e2321584121.

Allen, Jennifer, Baird Howland, Markus Mobius, David Rothschild, and Duncan J. Watts, "Evaluating the fake news problem at the scale of the information ecosystem," *Science Advances*, 2020, *6* (14), eaay3539.

Aridor, Guy, Rafael Jiménez-Durán, Ro'ee Levy, and Lena Song, "The economics of social media," *Journal of Economic Literature*, 2024.

__, **Tevel Dekel, Rafael Jiménez Durán, Ro'ee Levy, and Lena Song**, "Digital News Consumption: Evidence from Smartphone Content in the 2024 US Elections," *Available at SSRN 5364054*, 2025.

Armantier, Olivier, Giorgio Topa, Wilbert van der Klaauw, and Basit Zafar, "An Overview of the Survey of Consumer Expectations," *Economic Policy Journal*, 2017.

Ash, Elliott and Stephen Hansen, "Text algorithms in economics," *Annual Review of Economics*, 2023, *15* (1), 659–688.

Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky, "Exposure to opposing views on social media can increase political polarization," *Proceedings of the National Academy of Sciences*, September 2018, *115* (37), 9216–9221. Publisher: Proceedings of the National Academy of Sciences.

Bakshy, Eytan, Solomon Messing, and Lada A Adamic, "Exposure to ideologically diverse news and opinion on Facebook," *Science*, 2015, *348* (6239), 1130–1132.

Beknazar-Yuzbashev, George, Rafael Jiménez-Durán, and Mateusz Stalinski, "A model of harmful yet engaging content on social media," in "AEA Papers and Proceedings," Vol. 114 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2024, pp. 678–683.

__, **Rafael Jiménez Durán, Jesse McCrosky, and Mateusz Stalinski**, "Toxic content and user engagement on social media: Evidence from a field experiment," *Available at SSRN 4307346*, 2025.

Bhadani, Saumya, Shun Yamaya, Alessandro Flammini, Filippo Menczer, Giovanni Luca Ciampaglia, and Brendan Nyhan, "Political audience diversity and news reliability in algorithmic ranking," *Nature Human Behaviour*, 2022, *6* (4), 495–505.

Braghieri, Luca, Ro'ee Levy, and Hannah Trachtman, "Frictions in News Consumption: Evidence from Social Media," 2025. Manuscript, May 2025.

__, **Sarah Eichmeyer, Ro'ee Levy, Markus Mobius, Jacob Steinhardt, and Ruiqi Zhong**, "Article-Level Slant and Polarization of News Consumption on Social Media," *Working Paper*, 2025.

Broockman, David E. and Joshua L. Kalla, "Consuming Cross-Cutting Media Causes Learning and Moderates Attitudes: A Field Experiment with Fox News Viewers," *The Journal of Politics*, January 2025, *87* (1), 246–261.

Bursztyn, Leonardo, Aakaash Rao, Christopher Roth, and David Yanagizawa-Drott, "Opinions As Facts," *The Review of Economic Studies*, 2023, *90* (4), 1832–1864.

__, **Benjamin R Handel, Rafael Jimenez, and Christopher Roth**, "When product markets become collective traps: The case of social media," Technical Report, National Bureau of Economic Research 2023.

__, **Ingar Haaland, Nicolas Röver, and Christopher Roth**, "The Social Desirability Atlas," 2025. Manuscript, May 26, 2025.

Campante, Filipe, Ruben Durante, Felix Hagemeister, and Ananya Sen, "GenAI Misinformation, Trust, and News Consumption: Evidence from a Field Experiment," June 2025. Working paper.

Caprini, Giulia, "Visual Bias," *Working Paper*, 2023.

Chen, Canyu and Kai Shu, "Can llm-generated misinformation be detected?," *arXiv preprint arXiv:2309.13788*, 2023.

Cheng, Mengjie (Magie), Elie Ofek, and Hema Yoganarasimhan, "Balancing Engagement and Polarization: Multi-Objective Alignment of News Content Using LLMs," 2025. Working paper.

Chopra, Felix and Ingar Haaland, "Conducting Qualitative Interviews with AI," *CESifo Working Paper*, 2023.

__ , __ , and Christopher Roth, "Do People Demand Fact-checked News? Evidence from U.S. Democrats," *Journal of Public Economics*, 2022, *205*, 104549.

__ , __ , and __ , "The Demand for News: Accuracy Concerns versus Belief Confirmation Motives," *Economic Journal*, 2024.

DellaVigna, Stefano and Ethan Kaplan, "The Fox News effect: Media bias and voting," *Quarterly Journal of Economics*, 2006.

Demsetz, Harold and Kenneth Lehn, "The Structure of Corporate Ownership: Causes and Consequences," *Journal of Political Economy*, 1985, *93* (6), 1155–1177.

Eddy, Kirsten, Katerina Eva Matsa, Michael Lipka, Naomi Forman-Katz, Christopher St. Aubin, Luxuan Wang, and Kaitlyn Radde, "What Is News?," *Pew Research Center*, May 2025. Accessed June 4, 2025.

Forman-Katz, Naomi, "Americans are following the news less closely than they used to," *Pew Research Center*, October 24 2023.

Gentzkow, Matthew, "Media and artificial intelligence," *Toulouse Network for Information Technology Whitepaper*, 2018.

Gentzkow, Matthew A. and Jesse M. Shapiro, "Media Bias and Reputation," *Journal of Political Economy*, 2006, *114* (2), 280–316.

__ and __ , "What Drives Media Slant? Evidence From U.S. Daily Newspapers," *Econometrica*, 2010, *78* (1), 35–71.

Gerber, Alan S., Dean Karlan, and Daniel Bergan, "Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions," *American Economic Journal: Applied Economics*, April 2009, *1* (2), 35–52.

Gottfried, Jeffrey, "Americans' news fatigue isn't going away – about two-thirds still feel worn out," *Pew Research Center*, 2020.

Graeber, Thomas, Christopher Roth, and Constantin Schesch, "Explanations," Technical Report, CESifo Working Paper 2024.

Groseclose, Tim and Jeffrey Milyo, "A measure of media bias," *Quarterly Journal of Economics*, 2005, *120* (4), 1191–1237.

Guess, Andrew M., "(Almost) everything in moderation: new evidence on Americans' online media diets," *American Journal of Political Science*, 2021, *65* (4).

__ , Neil Malhotra, Jennifer Pan, Pablo Barberá, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow, Sandra González-Bailón, Edward Kennedy, Young Mie Kim, David Lazer, Devra Moehler, Brendan Nyhan, Carlos Velasco Rivera, Jaime Settle, Daniel Robert Thomas, Emily Thorson, Rebekah Tromble, Arjun Wilkins, Magdalena Wojcieszak, Beixian Xiong, Chad Kiewiet de Jonge, Annie Franco, Winter Mason, Natalie Jomini Stroud, and Joshua A. Tucker, "How do social media feed algorithms affect attitudes and behavior in an election campaign?," *Science*, July 2023, *381* (6656), 398–404. Publisher: American Association for the Advancement of Science.

— , **Chad Kiewiet De Jonge, Annie Franco, Winter Mason, Natalie Jomini Stroud, and Joshua A. Tucker**, "Reshares on social media amplify political news but do not detectably affect beliefs or opinions," *Science*, July 2023, *381* (6656), 404–408.

_ , Pablo Barberá, Simon Munzert, and JungHwan Yang, "The consequences of online partisan media," *Proceedings of the National Academy of Sciences*, April 2021, *118* (14), e2013464118. Publisher: Proceedings of the National Academy of Sciences.

Habibi, Mahyar, Dirk Hovy, and Carlo Schwarz, "The Content Moderator's Dilemma: Removal of Toxic Content and Distortions to Online Discourse," *arXiv preprint arXiv:2412.16114*, 2024.

Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J. Westwood, "The Origins and Consequences of Affective Polarization in the United States," *Annual Review of Political Science*, 2019, *22.*

Jiménez-Durán, Rafael, "The economics of content moderation: Theory and experimental evidence from hate speech on Twitter," *George J. Stigler Center for the Study of the Economy & the State Working Paper 324*, 2023.

_ , Karsten Müller, and Carlo Schwarz, "The effect of content moderation on online and offline hate: Evidence from Germany's NetzDG," *Available at SSRN 4230296*, 2024.

Jo, Donghee, "Better the Devil You Know: Selective Exposure Alleviates Polarization in an Online Field Experiment," Technical Report, Working Paper 2019.

Kalra, Aarushi, "It's the Algorithm, Stupid! Evidence on the Drivers of Harmful Behavior from a Large-Scale RC," 2025. Working paper, Brown University.

Kleinberg, Jon, Sendhil Mullainathan, and Manish Raghavan, "The challenge of understanding what users want: Inconsistent preferences and engagement optimization," *Management science*, 2024, *70* (9), 6336–6355.

Levy, Ro'ee, "Social Media, News Consumption, and Polarization: Evidence from a Field Experiment," *American Economic Review*, March 2021, *111* (3), 831–70.

Liu, Naijia, Xinlan Emily Hu, Yasemin Savas, Matthew A. Baum, Adam J. Berinsky, Allison J. B. Chaney, Christopher Lucas, Rei Mariman, Justin de Benedictis-Kessner, Andrew M. Guess, Dean Knox, and Brandon M. Stewart, "Short-term exposure to filter-bubble recommendation systems has limited polarization effects: Naturalistic experiments on YouTube," *Proceedings of the National Academy of Sciences*, February 2025, *122* (8), e2318127122. Publisher: Proceedings of the National Academy of Sciences.

Min, Sewon, Kalpesh Krishna, Xinxi Lyu, Mike Lewis, Wen tau Yih, Pang Wei Koh, Mohit Iyyer, Luke Zettlemoyer, and Hannaneh Hajishirzi, "FActScore: Fine-grained Atomic Evaluation of Factual Precision in Long Form Text Generation," *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2023.

Mitchell, Amy, Jeffrey Gottfried, Michael Barthel, and Nami Sumida, "Distinguishing Between Factual and Opinion Statements in the News," Technical Report, Pew Research Center June 2018.

_ , Katie Simmons, Katerina Eva Matsa, and Laura Silver, "Publics Globally Want Unbiased News Coverage, but Are Divided on Whether Their News Media Deliver," `https://www.pewresearch.org/journalism/2018/01/11/publics-globally-want-unbiased-news-coverage-but-are-divided-on-whether-their-news-media-deliver/` 2018. Pew Research Center.

Mullainathan, Sendhil and Andrei Shleifer, "The Market for News," *American Economic Review*, 2005, *95* (4), 1031–1053.

Ni, Jingwei, Minjing Shi, Dominik Stammbach, Mrinmaya Sachan, Elliott Ash, and Markus Leippold, "AFaCTA: Assisting the Annotation of Factual Claim Detection with Reliable LLM Annotators," *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL)*, 2024.

Nyhan, Brendan, Jaime Settle, Emily Thorson, Magdalena Wojcieszak, Pablo Barberá, Annie Y. Chen, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow, Sandra González-Bailón, Andrew M. Guess, Edward Kennedy, Young Mie Kim, David Lazer, Neil Malhotra, Devra Moehler, Jennifer Pan, Daniel Robert Thomas, Rebekah Tromble, Carlos Velasco Rivera, Arjun Wilkins, Beixian Xiong, Chad Kiewiet de Jonge, Annie Franco, Winter Mason, Natalie Jomini Stroud, and Joshua A. Tucker, "Like-minded sources on Facebook are prevalent but not polarizing," *Nature*, August 2023, *620* (7972), 137–144. Publisher: Nature Publishing Group.

Peer, Eyal, David Rothschild, Ariel Gordon, Zak Evernden, and Ekaterina Damer, "Data quality of platforms and panels for online behavioral research," *Behavior Research Methods*, 2022, *54*, 1643–1662.

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

Reimers, Nils and Iryna Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," in "Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing" Association for Computational Linguistics 2019.

Reuters, "Digital News Report 2023: Executive Summary," Technical Report, Reuters Institute for the Study of Journalism 2023. Accessed: December 1, 2024.

Rilla, Raluca, Tobias Werner, Hiromu Yakura, Iyad Rahwan, and Anne-Marie Nussberger, "Recognising, Anticipating, and Mitigating LLM Pollution of Online Behavioural Research," *arXiv preprint arXiv:2508.01390*, 2025.

Robertson, Claire E, Nicolas Pröllochs, Kaoru Schwarzenegger, Philip Pärnamets, Jay J Van Bavel, and Stefan Feuerriegel, "Negativity drives online news consumption," *Nature Human Behaviour*, 2023, *7* (5), 812–822.

Simon, Felix M, Sacha Altay, and Hugo Mercier, "Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown," *Harvard Kennedy School Misinformation Review*, 2023, *4* (5).

Sunstein, Cass R, *# Republic: Divided Democracy in the Age of Social Media*, Princeton University Press, 2018.

Thaler, Michael, "The fake news effect: Experimentally identifying motivated reasoning using trust in news," *American Economic Journal: Microeconomics*, 2024, *16* (2), 1–38.

Wei, Jerry, Chengrun Yang, Xinying Song, Yifeng Lu, Nathan Hu, Jie Huang, Dustin Tran, Daiyi Peng, Ruibo Liu, Da Huang, Cosmo Du, and Quoc V. Le, "Long-form factuality in large language models," *Proceedings of the 38th Conference on Neural Information Processing Systems (NeurIPS)*, 2024.

Zhou, Jiawei, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury, "Synthetic lies: Understanding ai-generated misinformation and evaluating algorithmic and human solutions," in "Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems" 2023, pp. 1–20.

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov, "Political effects of the internet and social media," *Annual review of economics*, 2020, *12* (1), 415–438.

For online publication only:

News Customization with AI

Felix Chopra Ingar Haaland Fabian Roeben

Christopher Roth Vanessa Sticher

Summary of the Online Appendix

Section A1 provides information about the data collections. Section A2 provides details about our validation procedures of our rewrite system. Section A3 provides information about the human validation study. Section A4 provides information about the demand experiment. Section A5 provides information about the usage incentive experiment.

A1 Summary of Data Collections

Table A1: Overview of Data Collections

Data collection	Sample	Treatment arms	Main outcomes
Human Validation Wave 1 (December 2, 2024)	Prolific ($n = 500$)	No treatments	Rating of articles in terms of complexity, entertainment, political leaning and sentiment
Human Validation Wave 2 (August 19-21, 2025)	Prolific ($n = 669$)	No treatments	Rating of articles in terms of facts, complexity, entertainment, political leaning, sentiment
Latent Demand Experiment (December 13 and 14 2024)	Prolific ($n = 5,005$)	No Customization (Regular), Customization with political rewrites, Customization with *Facts Only* rewrites	Download and log in to the app
Usage Incentive Experiment Wave 1 (October and November, 2024)	Prolific ($n = 1,439$)	Customization with political rewrites, No Customization (Regular, Less Complex, Entertaining, Liberal, Conservative, More Positive, More Negative)	Time spent on app (in different versions), knowledge acquisition, political polarization
Usage Incentive Experiment Wave 2 (January and February, 2025)	Prolific ($n = 2,039$)	Customization with political rewrites, Customization with Facts Only rewrites, No Customization (Regular, Less Complex, Entertaining, Liberal, Conservative, More Positive, More Negative, Facts Only)	Time spent on app (in different versions), knowledge acquisition, political polarization

Note: Table A1 displays the overview of the data collections. All data collections were pre-registered on the AEA RCT registry (#14561). The experimental instructions for each collection can be found on the following link: `https://raw.githubusercontent.com/cproth/papers/master/tmt_instructions_appendix.pdf`.

A2 Validation of News Event Consistency

We validate that rewrites preserve the underlying news event using a custom three-step semantic fact-level alignment procedure applied to the full corpus of 59,092 article versions across both waves of the experiment. For all LLM usage in this pipeline, we ensure reproducible outputs by disabling sampling (temperature = 0, top_p = 1) and setting both frequency and presence penalties to zero.[1]

Extraction We define a factual claim as a single, objectively verifiable statement, following Ni et al. (2024). This definition is consistent with recent advances in automatic factuality evaluation, which assess content at the atomic level (Min et al., 2023; Wei et al., 2024). Facts were extracted sentence by sentence from each article version using GPT-4.1 (nano). To ensure high-quality extraction for every sentence, we provided the model with the three most relevant examples from a hand-annotated library of 96 sentences, following the approach of Min et al. (2023). Relevance was determined using a standard text-matching algorithm (BM25), which ensures the model receives highly contextual demonstrations. This retrieval-augmented generation (RAG) setup grounds the model in closely related examples, reduces variance across domains, and achieves high extraction accuracy. Further, outputs were parsed and schema-validated as well as near-duplicates removed. As a final quality filter, we retained only those extractions that formed complete, standalone statements (i.e., finite clauses). This step ensures that every fact is a grammatically whole proposition that can be independently verified, discarding sentence fragments or incomplete thoughts. Lastly, numeric metadata were preserved for later validation checks. An example extraction for a *Regular* version appears in Table A2.

Classification The extracted facts are classified into core or supplementary facts using GPT-4.1. Core facts capture the "who did what and why" of the main event as well as its primary intent and consist of the set sufficient to form a 1-2 sentence, faithful account of the article. Supplementary facts cover background, context, reactions, sourcing/meta information, or other details not essential for understanding the main event. To improve consistency of the fact extraction procedure, the model first resolved pronouns using the article text before classification, and its output was constrained to a structured JSON schema, ensuring every fact was assigned to exactly one category. Two safeguards were applied: ambiguous or uncertain cases were conservatively assigned to supplementary, and duplicate facts were reduced to the single strongest version in core. Table A2 illustrates the resulting split. Outputs were also periodically inspected during development to ensure rule adherence and to refine prompts.

[1] *Temperature* rescales token logits, where lower values reduce stochasticity. *Top_p* restricts draws to the smallest set of tokens whose cumulative probability exceeds p. Setting temperature = 0 and top_p = 1 yields greedy decoding in practice. *Frequency* and *presence* penalties down-weight previously generated tokens to discourage repetition; both were set to zero to avoid altering base probabilities.

Table A2: Example of Fact Extraction and Classification

Core Facts	
1. Republican lawmakers are voicing concerns about President Trump's government reforms.	
2. Specific measures are drawing pushback from Republican lawmakers.	
3. The Department of Government Efficiency (DOGE) is led by Elon Musk.	
4. The government reforms are under the Department of Government Efficiency (DOGE).	
5. The pushback is due to potential negative impacts on home-state industries, jobs, and programs.	
Supplementary Facts	
6. Constituents in the Research Triangle expressed concerns about potential damage to scientific progress.	
7. Critics argued that these measures exceed presidential authority.	
8. House Speaker Mike Johnson praised the administration's efforts as necessary reforms.	
9. Idaho Representative Mike Simpson highlighted concerns about reduced hiring at national parks ahead of the summer tourism season.	
10. Other Republicans, including Senator Susan Collins, detailed how funding caps could undermine critical biomedical research.	
11. Representative Carlos Gimenez is advocating for Venezuelans under Temporary Protected Status in Florida.	
12. Representative Carlos Gimenez is urging the administration to evaluate individual cases to prevent deportations.	
13. Senator Jerry Moran is working on legislation to transfer food aid management to the Department of Agriculture.	
14. Senator Jerry Moran of Kansas raised alarms about disruptions to the U.S. Agency for International Development's food aid.	
15. Senator Katie Britt warned that these reductions could harm essential research at institutions like the University of Alabama.	
16. Senator Katie Britt of Alabama has criticized cuts to National Institutes of Health (NIH) grants.	
17. Senator Ted Budd is a senator representing North Carolina.	
18. Senator Ted Budd reported hearing from constituents in the Research Triangle.	
19. Senate Minority Leader Chuck Schumer accused the administration of bypassing congressional debate.	
20. Some Republican leaders have defended the administration's efforts.	
21. The disruptions to USAID's food aid risk spoiling American-grown produce intended for global distribution.	
22. The executive actions by Musk's DOGE team have sparked legal challenges across the country.	

Note: This table illustrates fact extraction and classification for the *Regular* article version of the example article.

Consistency We then tested whether core facts were consistent across article versions. For each comparison and direction, we embedded all facts using the sentence-transformer model `all-mpnet-base-v2` (unit-normalized) and computed cosine similarities. A source core fact is counted as *covered* if *any* target-side fact (core $\cup$ supplementary) attains similarity $\geq \tau$ with the source (τ=0.66). This threshold was selected by maximizing F1 on 400 hand-labeled fact pairs, following Reimers and Gurevych (2019). This procedure yielded three sets per comparison: matched pairs, unmatched facts from the *Regular* version, and unmatched facts from the respective rewritten version. Table A3 visualizes the embedding step on our running example: five of six source core facts meet the threshold and are matched; the final row (similarity $= 0.59$) falls below τ and remains unmatched under embeddings alone.

As a final validation step to reduce false negatives, each unmatched source core fact is presented together with the full target article to a state-of-the-art LLM (GPT-5 at the time of writing) under a strict JSON schema to assess whether the fact is nonetheless substantively covered. In the example from Table A3, the LLM adjudication marks the last unmatched pair as

Table A3: Example Facts Alignment

Facts from *Regular* version	**Facts from *Less Complex* version**	**Similarity**	**Match**
The Department of Government Efficiency (DOGE) is led by Elon Musk.	The Department of Government Efficiency is led by billionaire Elon Musk.	0.92	✓
Republican lawmakers are voicing concerns about President Trump's government reforms.	Some Republican lawmakers are worried about President Trump's major changes to the government.	0.88	✓
The executive actions by Musk's DOGE team have sparked legal challenges across the country.	The actions taken by Elon Musk's Department of Government Efficiency team led to legal challenges across the country.	0.86	✓
The government reforms are under the Department of Government Efficiency (DOGE).	The changes are being managed by the Department of Government Efficiency.	0.76	✓
Specific measures are drawing pushback from Republican lawmakers.	*Some Republican lawmakers are worried about how certain changes might hurt people and important programs.*	0.68	✓
The pushback is due to potential negative impacts on home-state industries, jobs, and programs.	*Some Republican lawmakers are concerned that President Trump's major changes to the government could hurt industries in their states.*	0.59	×

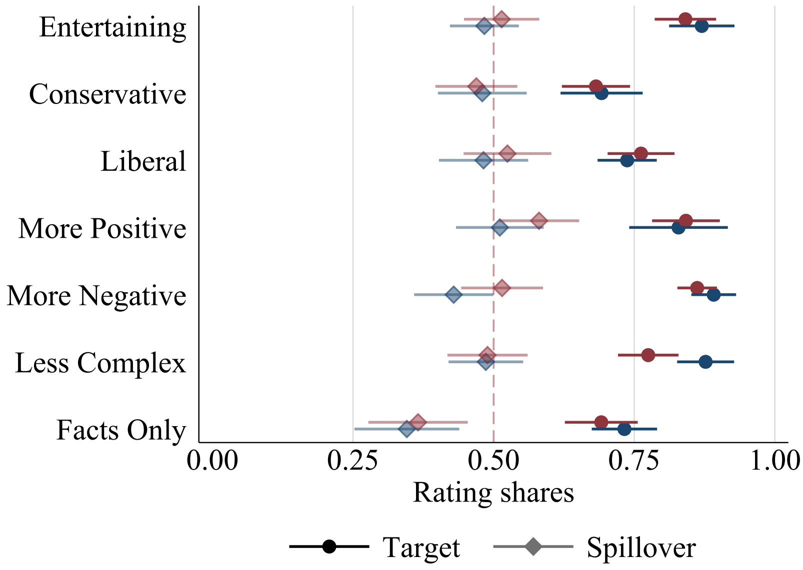
Note: This table shows the matching of core facts for the *Regular* and *Less Complex* version of the example article. Facts in italics represent supplementary facts from the respective article version. If a pair of facts has cosine similarity ≥ 0.66, it is considered a match.

covered with the justification: "Article explicitly says concerns stem from impacts on home-state industries, jobs, and programs."

As an additional diagnostic, we directly compared the number metadata preserved in the extraction step across article versions. This check focused on numerical values (e.g., dates, statistics, percentages). Consistency across these dimensions provides assurance that rewrites did not alter, add or remove quantitative information, complementing the semantic alignment procedure.

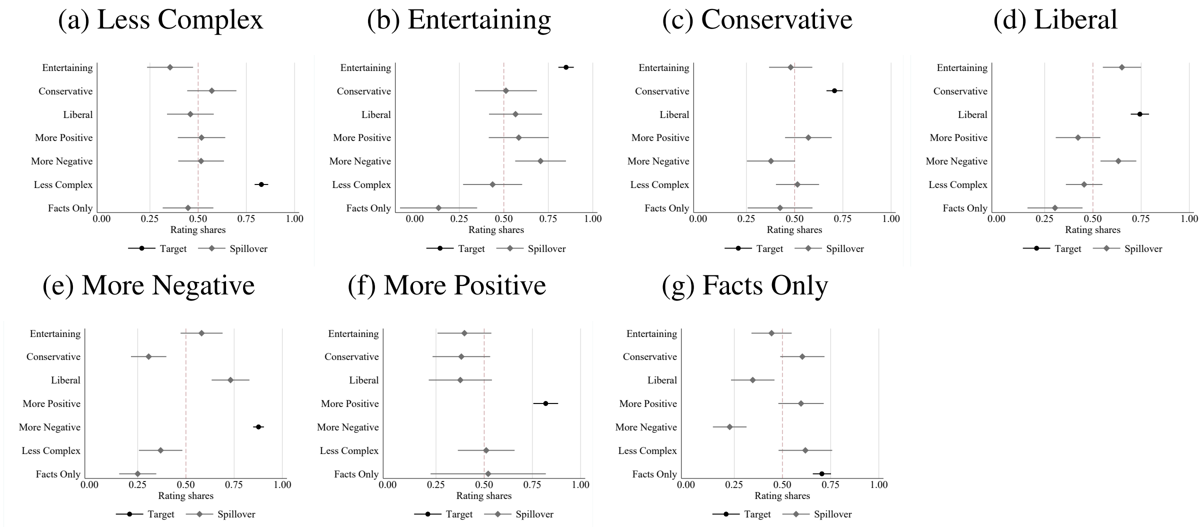
A3 Human Validation Study

Figure A1: Human Validation of Article Rewrites by Party Affiliation



Note: This figure shows the share of respondents rating the rewritten article version as more entertaining, conservative, liberal, positive, negative, or less complex than the regular version in a blinded binary comparison, by party affiliation. Red denotes ratings by Republicans, while blue denotes ratings by Democrats. Dots show the share correctly identifying the version varied along the target dimension (e.g., *Entertaining* rewrites rated as more entertaining). Squares show spillovers (e.g., non-*Entertaining* rewrites rated as more entertaining). Error bars represent 95% confidence intervals.

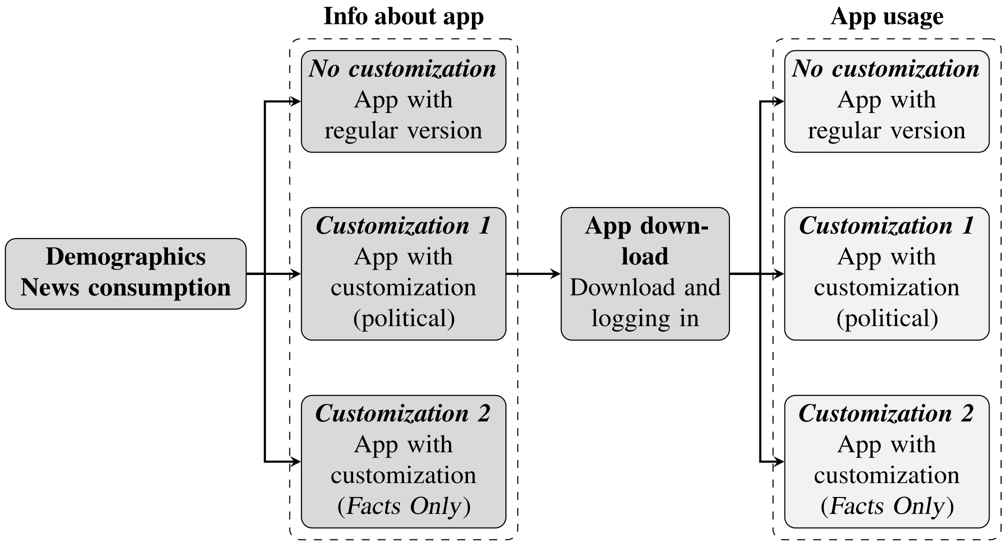
Figure A2: Human Validation by Rewrite Dimension



Note: This figure shows the share of respondents rating the rewritten article version as more entertaining, conservative, liberal, positive, negative, or less complex compared to the regular version in a blinded binary comparison separately by rewrite version. Each panel reports ratings across all dimensions for one version. Black dots indicate correctly identifying the varied dimension (e.g., *entertaining* rewrites rated as more entertaining). Gray squares indicate spillovers (e.g., *Less Complex* rewrites rated as more entertaining). Error bars represent 95% confidence intervals.

A4 Demand Experiment

Figure A3: Overview of the Demand Experiment



Note: This figure presents an overview of the experiment to study the latent demand for customization. The experiment starts with a survey eliciting demographics and baseline news consumption. Participants are then randomized into three conditions and provided information about the app (with or without customization, depending on the condition). Next, they can download the app. Finally, we observe their app usage.

Table A4: Summary Statistics for the Demand Experiment

	Mean	SD	Min	Max
Demographics				
Age	40.04975	13.38931	18	83
Female	.5428571	.4982097	0	1
White	.6973027	.4594712	0	1
At least some college	.8587413	.3483231	0	1
Political leaning				
Democrat	.4207792	.4937335	0	1
Independent	.3302697	.4703571	0	1
Republican	.248951	.4324486	0	1
News consumption				
News primarily online	.8435564	.3633116	0	1
Somewhat frequent news	.5524476	.4972913	0	1
News apps	1.417582	1.588154	0	6
Observations	5005			

Note: This table presents summary statistics for the demand experiment. “Age” is the respondent’s age in years. “Female”, “White”, and “At least some college” are binary indicators for gender, race, and highest level of education, respectively. Party affiliation is captured by the mutually exclusive binary indicators “Democrat”, “Independent”, and “Republican”. “News primarily online” is an indicator that equals one if the respondent consumes news primarily online; “Somewhat frequent news” equals one if the respondent consumes news at least somewhat frequently. “News apps” is the number of news apps on a respondent’s phone.

Table A6: App Downloads by Customization Condition

	(1) Downloaded	(2) Downloaded	(3) Downloaded	(4) Downloaded
Customization	0.0272***	0.0259***		
	(0.00801)	(0.00798)		
Customization (political)			0.0256***	0.0246***
			(0.00944)	(0.00942)
Customization (Facts Only)			0.0289***	0.0272***
			(0.00954)	(0.00951)
Mean with no customization	0.0683	0.0683	0.0683	0.0683
P-value: pol=Facts Only			0.747	0.798
N	5005	5005	5005	5005
Controls	No	Yes	No	Yes

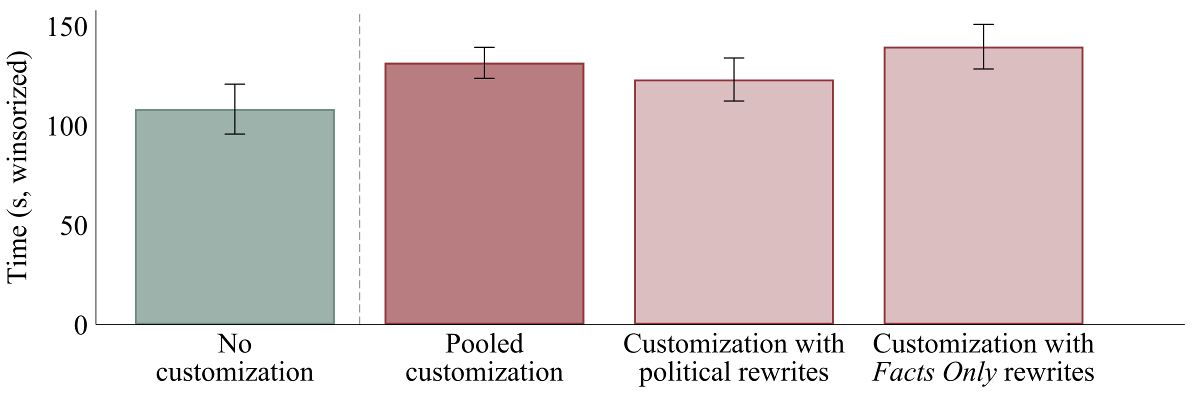
Note: This table presents regression coefficients for the effect of customization on app downloads. "Customization" equals one for respondents in any customization condition (pooled). Columns (1)-(2) show pooled effects. Columns (3)-(4) separate the two customization conditions: political and *Facts Only* rewrites. Columns (2) and (4) control for demographics (age, gender, race, income, some college, full-time employed, party affiliation) and news consumption (primarily consuming news online, number of news apps on the phone). Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Balance Table for the Demand Experiment

	(1) No customization	(2) Customization (political)	(3) Customization (Facts Only)	(4) p-val (2) vs. (1)	(5) p-val (3) vs. (1)
Demographics					
Age	39.707	40.166	40.276	0.323	0.222
	(13.519)	(13.297)	(13.352)		
Female	0.540	0.552	0.536	0.479	0.810
	(0.499)	(0.497)	(0.499)		
White	0.699	0.686	0.707	0.408	0.611
	(0.459)	(0.464)	(0.455)		
Some college	0.873	0.852	0.851	0.082	0.069
	(0.333)	(0.355)	(0.356)		
Political leaning					
Democrat	0.415	0.423	0.424	0.655	0.627
	(0.493)	(0.494)	(0.494)		
Independent	0.329	0.336	0.326	0.686	0.844
	(0.470)	(0.472)	(0.469)		
Republican	0.255	0.241	0.250	0.342	0.735
	(0.436)	(0.428)	(0.433)		
News consumption					
News primarily online	0.853	0.837	0.842	0.205	0.379
	(0.355)	(0.370)	(0.365)		
Somewhat frequent news	0.550	0.545	0.562	0.791	0.462
	(0.498)	(0.498)	(0.496)		
News apps	1.405	1.412	1.436	0.889	0.572
	(1.578)	(1.586)	(1.601)		
Observations	1668	1671	1666		

Note: This table reports summary statistics for the demand experiment by customization condition. "Age" is in years. "Female", "White", and "At least some college" are binary indicators for gender, race, and highest level of education, respectively. Party affiliation is captured by mutually exclusive indicators for "Democrat", "Independent", and "Republican". "News primarily online" equals one if the respondent consumes news primarily online; "Somewhat frequent news" equals one if they consumes news at least somewhat frequently. "News apps" counts the number of news apps on a respondent's phone. Columns (1)-(3) show means with standard deviations in parentheses for the no-customization condition and the two customization conditions. Columns (4) and (5) report p-values testing equality of no-customization and each customization conditions,.

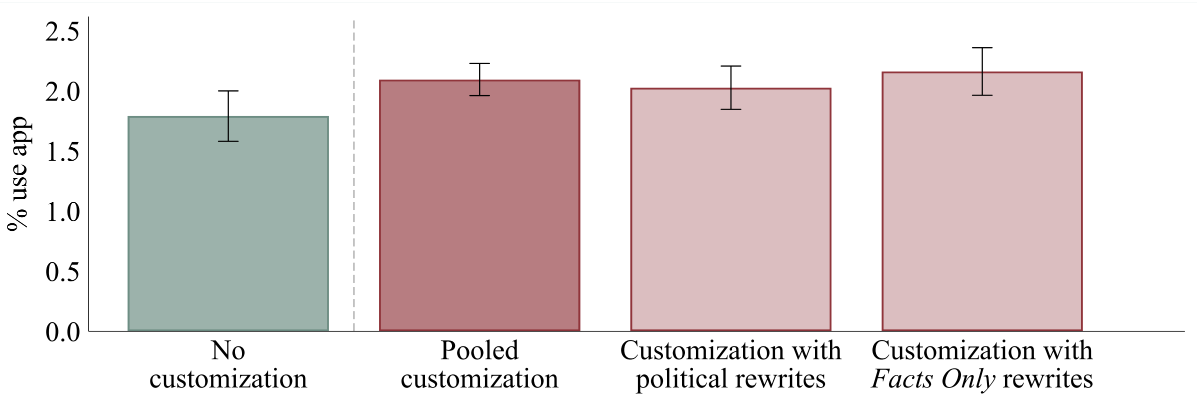
Figure A4: App Usage Time on Day 1 by Customization Condition



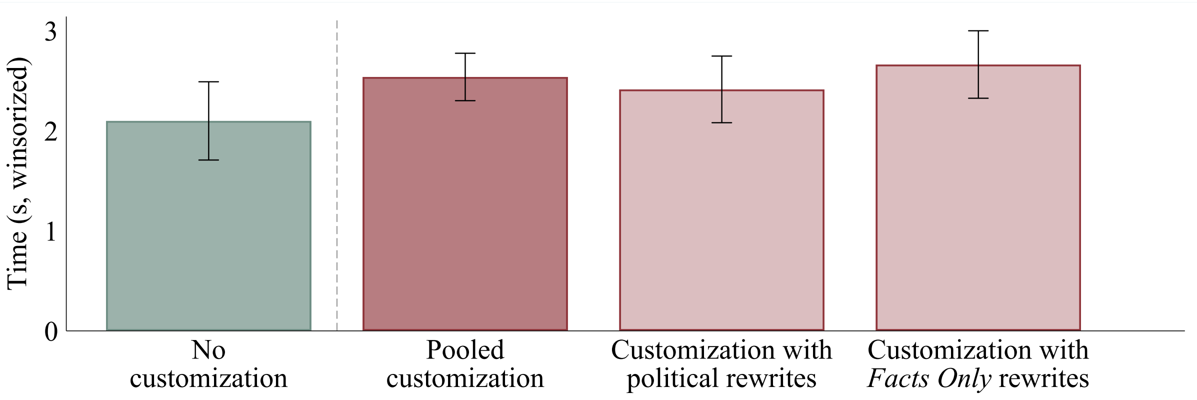
Note: This figure shows app usage time on day 1 by customization, conditional on app download. Time is in seconds and winsorized at the 98th percentile. The first bar shows the no customization condition, the second bar shows both customization conditions pooled, the last two bars show customization conditions separately (political and *Facts Only*). Error bars represent standard errors of the mean.

Figure A5: App Usage Frequency and Time by Customization Condition

(a) App Usage Frequency



(b) Average Time Spent

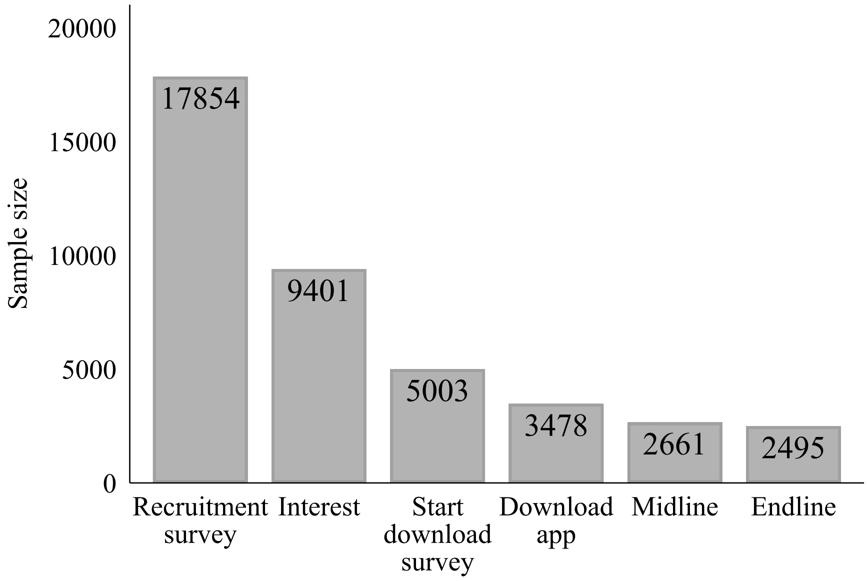


Note: This figure shows daily app usage for the first seven days after the download survey by customization condition. Panel (a) shows the average daily share of users. Panel (b) shows average daily usage time in seconds and winsorized at the 98th percentile. In each panel, the first bar shows no customization, the second bar pools both customization conditions, and the last two bars show political and *Facts Only* separately. Error bars represent standard errors of the mean.

A5 Usage Incentive Experiment

A5.1 Sample

Figure A6: Sample Size in the Usage Incentive Experiment



Note: This figure shows the sample sizes at the different stages of the usage incentive experiment. The bars represent the number of participants who completed the recruitment survey, expressed interest in early access to the app, started the download survey, downloaded the app, and completed the midline and endline surveys.

Table A7: Treatment Arms in the Usage Incentive Experiment

	Total	Wave 1	Wave 2
Customization (political)	747	430	317
Customization (Facts Only)	718		718
Regular	261	140	121
Liberal	237	135	102
Conservative	254	151	103
Less Complex	238	131	107
More Negative	264	158	106
More Positive	254	146	108
Entertaining	263	148	115
Facts Only	242		242
N	3478	1439	2039

Note: This table reports the number of respondents in each treatment arm, both in total (Column 1) and separately for Wave 1 and Wave 2 (Columns 2 and 3, respectively). The first two rows show to the two customization conditions. The remaining rows show to the control conditions, where respondents were shown one randomly assigned rewrite version. The *Facts Only* conditions were only introduced in Wave 2 and therefore over-sampled to achieve comparable final sample sizes. Target sample sizes were 750 participants per customization condition and 250 per control condition. While exact targets could not be achieved due to technical constraints, the final sample sizes closely approximate the targets.

Table A8: Attrition in the Usage Incentive Experiment

	(1) Midline	(2) Endline
Customization (political)	-0.00834	-0.00873
	(0.0177)	(0.0190)
Customization (Facts Only)	0.00343	0.00306
	(0.0202)	(0.0211)
Constant	0.766***	0.719***
	(0.00939)	(0.0100)
Download day FE	Yes	Yes
N	3478	3478

Note: This table reports regression estimates of attrition by treatment arm in the usage incentive experiment. The dependent variable is a binary indicator for completing the midline survey (Column 1) or the endline survey (Column 2). Regressions include fixed effects for the day on which the app download survey was completed. Coefficients for both customization conditions are close to zero, indicating no evidence of differential attrition across groups. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Summary Statistics for the Usage Incentive Experiment

	Mean	SD	Min	Max
Demographics				
Age	37.71794	12.14871	18	78
Female	.5644048	.495906	0	1
White	.6912018	.4620641	0	1
At least some college	.8780909	.3272279	0	1
Political leaning				
Democrat	.4278321	.4948355	0	1
Independent	.3323749	.4711323	0	1
Republican	.239793	.4270184	0	1
News consumption				
News primarily online	.9045428	.2938876	0	1
Somewhat frequent news	.5802185	.493594	0	1
News apps	1.780046	1.611072	0	6
Observations	3478			

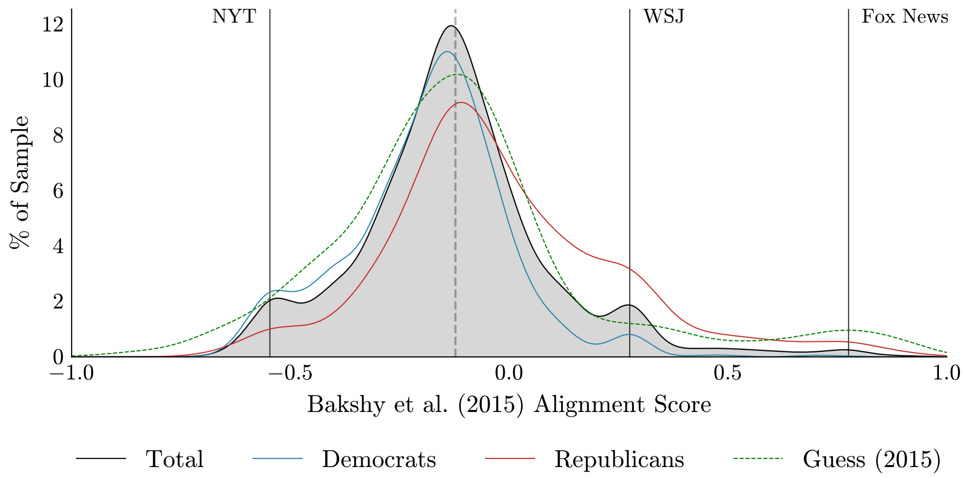
Note: This table presents summary statistics for the usage incentive experiment. “Age” is the respondent’s age in years. “Female”, “White”, and “At least some college” are binary indicators for gender, race, and highest level of education, respectively. Party affiliation is captured by the mutually exclusive binary indicators “Democrat”, “Independent”, and “Republican”. “News primarily online” is an indicator that equals one if the respondent consumes news primarily online; “Somewhat frequent news” equals one if the respondent consumes news at least somewhat frequently. “News apps” is the number of news apps on a respondent’s phone.

Table A10: Balance Table for the Usage Incentive Experiment

	(1) No customization	(2) Customization (political)	(3) Customization (Facts Only)	(4) p-val (2) vs. (1)	(5) p-val (3) vs. (1)
Demographics					
Age	37.456	37.901	38.262	0.331	0.819
	(12.072)	(12.460)	(12.031)		
Female	0.567	0.576	0.546	0.907	0.688
	(0.496)	(0.495)	(0.498)		
White	0.697	0.680	0.685	0.420	0.406
	(0.459)	(0.467)	(0.465)		
At least some college	0.880	0.901	0.850	0.167	0.524
	(0.325)	(0.299)	(0.358)		
Political leaning					
Democrat	0.443	0.411	0.404	0.076	0.519
	(0.497)	(0.492)	(0.491)		
Independent	0.328	0.336	0.340	0.541	0.620
	(0.470)	(0.473)	(0.474)		
Republican	0.229	0.253	0.256	0.176	0.201
	(0.420)	(0.435)	(0.437)		
News consumption					
News primarily online	0.907	0.900	0.903	0.452	0.581
	(0.290)	(0.301)	(0.297)		
Somewhat frequent news	0.587	0.597	0.545	0.596	0.056
	(0.493)	(0.491)	(0.498)		
News apps	1.842	1.814	1.572	0.582	0.015
	(1.620)	(1.648)	(1.532)		
Observations	2013	747	718		

Note: This table presents summary statistics for the usage incentive experiment by customization condition. “Age” is in years. “Female”, “White”, and “At least some college” are binary indicators for gender, race, and highest level of education, respectively. Party affiliation is captured by the mutually exclusive binary indicators “Democrat”, “Independent”, and “Republican”. “News primarily online” is an indicator that equals one if the respondent consumes news primarily online; “Somewhat frequent news” equals one if the respondent consumes news at least somewhat frequently. “News apps” is the number of news apps on a respondent’s phone. Columns (1), (2), and (3) show means with standard deviations in parentheses for the no-customization condition and the two customization conditions. Column (4) reports p-values for tests of equality between customization with political rewrites and no customization. Column (5) reports p-values for tests of equality between customization with *Facts Only* rewrites and no customization. All p-values are estimated using linear regressions that include fixed effects for the day on which the app download survey was completed.

Figure A7: Baseline News Consumption



Note: This figure overlays four kernel–density estimates of outlet-level ideology. For each respondent, we map every outlet they reported reading in the past twelve months onto the alignment scale of Bakshy et al. (2015), which ranges from –2 (very liberal) to +2 (very conservative), and weight the data so that each respondent contributes equally. The solid gray curve represents the full sample; the blue and red curves show Democrats and Republicans, respectively; and the green dotted curve reproduces the nationally weighted 2015 distribution reported by Guess (2021). Solid vertical reference lines mark the positions of three well-known outlets—The New York Times, The Wall Street Journal and Fox News—while the dashed gray line denotes the mean alignment score in our sample. The vertical axis expresses the resulting densities as the percentage of respondents.

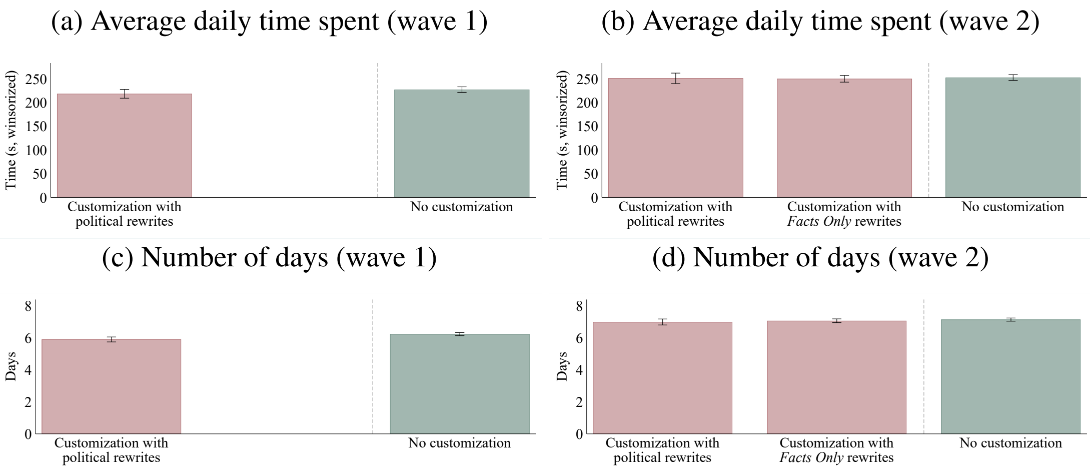
Table A11: Selection into Downloading the App

	Dependent variable: app download			
	(1)	(2)	(3)	(4)
Age	0.000216	0.000508**	-0.00171***	-0.00117**
	(0.000236)	(0.000259)	(0.000561)	(0.000580)
Female	-0.00486	-0.0119*	-0.0151	-0.0194
	(0.00615)	(0.00664)	(0.0131)	(0.0135)
White	0.0180***	0.0188***	0.0864***	0.0830***
	(0.00662)	(0.00708)	(0.0143)	(0.0148)
At least some college	-0.0312***	-0.0418***	-0.0424**	-0.0626***
	(0.00951)	(0.0109)	(0.0196)	(0.0207)
Democrat	-0.00438	-0.00605	0.0144	0.0137
	(0.00700)	(0.00756)	(0.0151)	(0.0157)
Republican	-0.0274***	-0.0294***	-0.108***	-0.107***
	(0.00817)	(0.00905)	(0.0171)	(0.0181)
News primarily online	0.0457***	0.0484***	0.128***	0.128***
	(0.00901)	(0.0100)	(0.0213)	(0.0227)
Somewhat frequent news	0.0248***	0.0212***	-0.00616	-0.00708
	(0.00619)	(0.00668)	(0.0136)	(0.0142)
News apps	0.0126***	0.0126***	-0.0117***	-0.0116***
	(0.00197)	(0.00211)	(0.00423)	(0.00439)
Trust in news	-0.0222***	-0.0283***	-0.0550***	-0.0601***
	(0.00382)	(0.00419)	(0.00823)	(0.00866)
Trust in AI	0.0629***	0.0631***	-0.0111	-0.0131*
	(0.00331)	(0.00356)	(0.00758)	(0.00793)
Alignment score		-0.0212		-0.0428
		(0.0130)		(0.0313)
Constant	0.0416**	0.0644***	0.871***	0.891***
	(0.0199)	(0.0221)	(0.0437)	(0.0455)
Wave FE	Yes	Yes	Yes	Yes
Adjusted R^2	0.030	0.029	0.054	0.057
N	17854	15904	5003	4639

Note: This table examines selection into downloading the app. Columns (1) and (2) estimate a linear probability model for the probability of downloading the app. Columns (3) and (4) estimate the same model but condition on having started the download survey. The dependent variable in all columns is a binary indicator for having downloaded the app. Demographics and news consumption variables are defined as in Table A9. "Trust in News" and "Trust in AI" is measured on a five-point Likert scale. Columns (2) and (4) include the additional variable "Alignment score", which refers to the ideological alignment scale from Bakshy et al. (2015). All columns include wave fixed effects. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A5.2 Measuring Customization Preferences

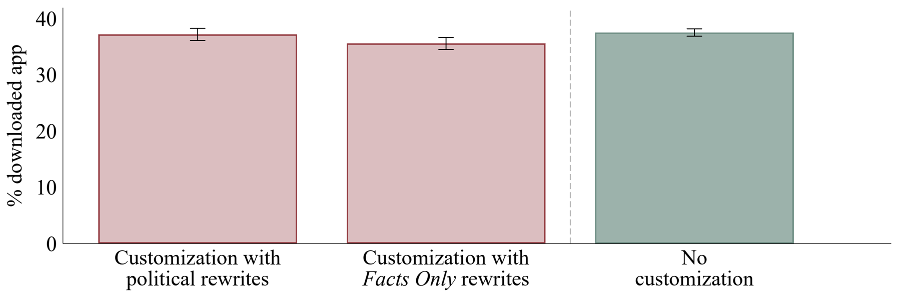
Figure A8: App Usage by Treatment Arm



(a) Average daily time spent (wave 1)
(b) Average daily time spent (wave 2)
(c) Number of days (wave 1)
(d) Number of days (wave 2)

Note: This figure shows app usage metrics during the incentive period across treatment arms and waves. Panels (a) and (b) show the average daily time spent using the app in seconds in Wave 1 and Wave 2, respectively. Panels (c) and (d) show the number of days with app usage in Wave 1 and Wave 2, respectively. The customization with *Facts Only* rewrites arm was only introduced in Wave 2. Error bars represent standard errors of the mean.

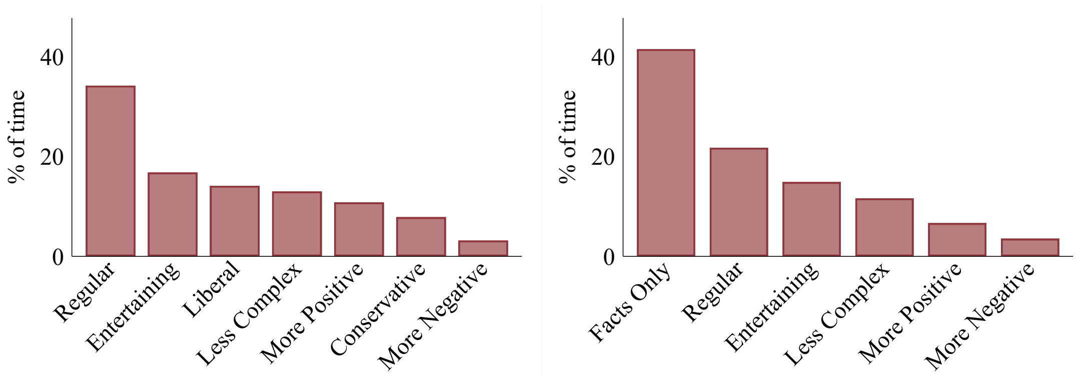
Figure A9: App Download Rate by Customization



Note: This figure shows app download rates across treatment arms. The bars indicate the share of participants who downloaded the app among those who expressed interest in early access in the recruitment survey. Error bars represent standard errors of the mean.

Figure A10: News Customization in the Incentive Period (Unweighted)

(a) Customization with Political Rewrites (b) Customization with *Facts Only* Rewrites

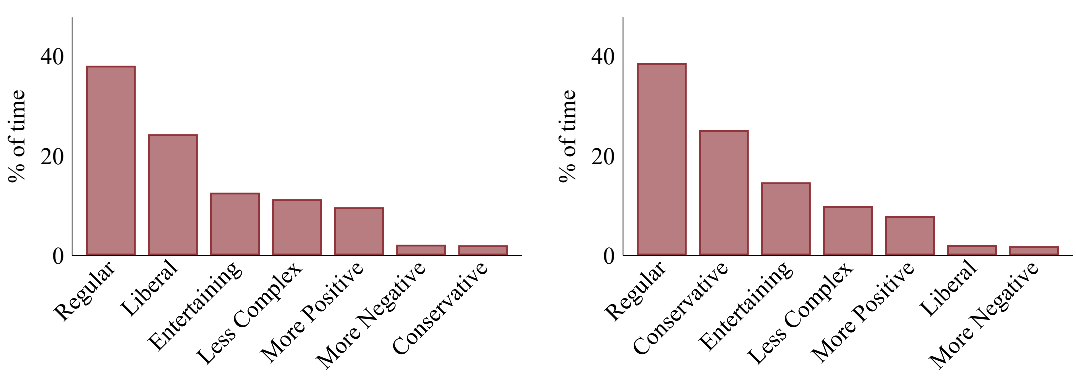


Note: This figure shows the share of time spent consuming news in different customization options during the incentive period, weighting the individual consumption shares equally. Panel (a) shows the shares for the respondents in the customization group with political rewrites ($N = 747$). Panel (b) shows the shares for the respondents in the customization group with *Facts Only* rewrites ($N = 718$).

A5.3 Heterogeneity in News Consumption With Customization

Figure A11: News Customization in the Incentive Period by Party Affiliation

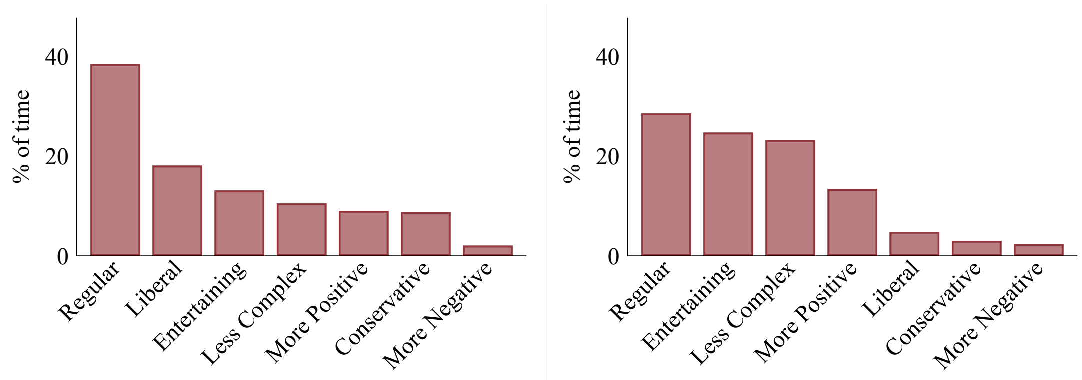
(a) Democrats (b) Republicans



Note: This figure shows the share of time spent consuming news in different customization options during the incentive period for the respondents in the customization group with political rewrites by party affiliation. Panel (a) shows time shares for Democrats. Panel (b) shows time shares for Republicans. Individual consumption shares are weighted by the amount of time spent consuming news in the app.

Figure A12: News Customization in the Incentive Period by Voting

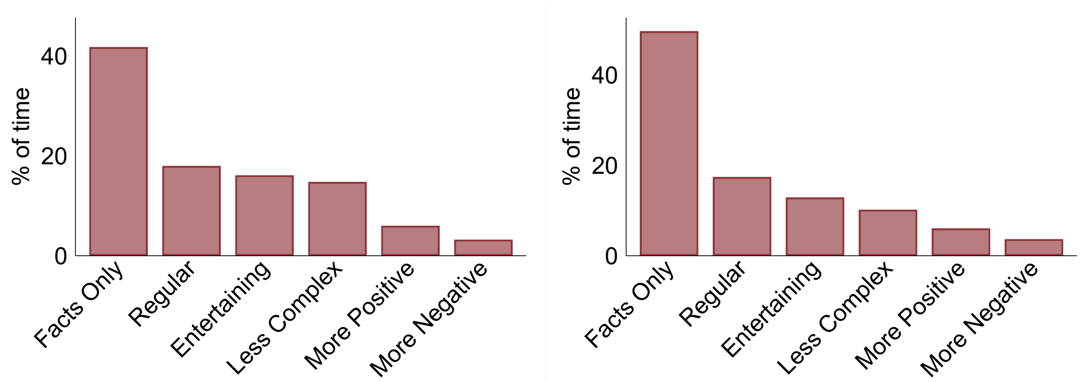
(a) Voters (b) Non-Voters



Note: This figure shows the share of time spent consuming news in different customization options during the incentive period for the respondents in the customization group with political rewrites by voting. Panel (a) shows time shares for voters (N=642). Panel (b) shows time shares for non-voters (N=105). Individual consumption shares are weighted by the amount of time spent consuming news in the app.

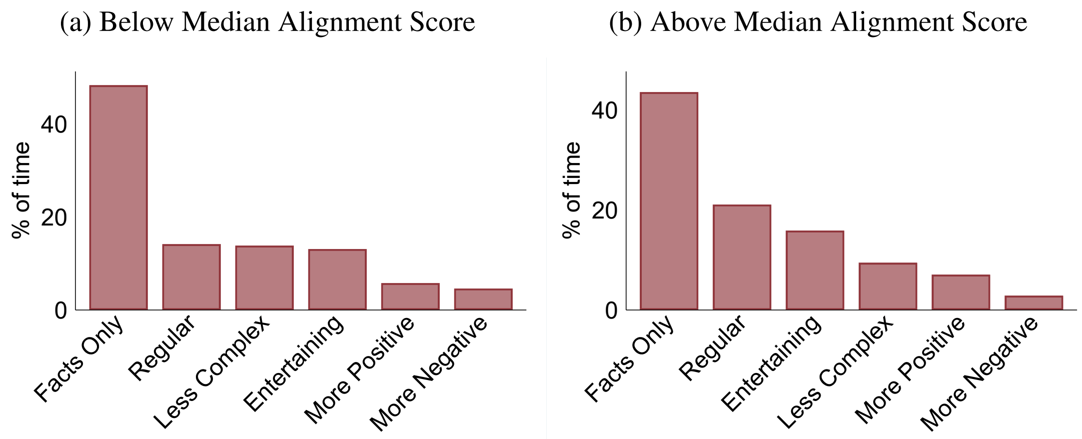
Figure A13: News Customization in the Incentive Period by Baseline News Consumption

(a) Less than somewhat frequent (b) At least somewhat frequent



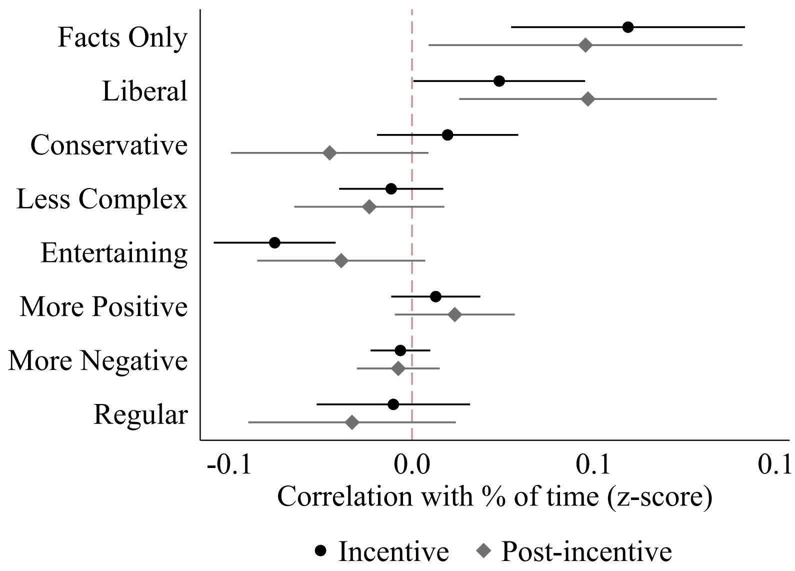
Note: This figure shows the share of time spent consuming news in different customization options during the incentive period for the respondents in the customization group with *Facts Only* rewrites by news consumption frequency at baseline. Panel (a) shows time shares for participants who consume news less than somewhat frequently (N=327). Panel (b) shows time shares for participants who consume news at least somewhat frequently (N=391). Individual consumption shares are weighted by the amount of time spent consuming news in the app.

Figure A14: News Customization in the Incentive Period by Baseline News Bias



Note: This figure shows the share of time spent consuming news in different customization options during the incentive period for the respondents in the customization group with *Facts Only* rewrites by news bias at baseline. We measure news bias using the Bakshy et al. (2015) alignment score. Panel (a) shows time shares for participants with an alignment score below the median (N=290). Panel (b) shows time shares for participants with an alignment score above the median (N=370). Individual consumption shares are weighted by the amount of time spent consuming news in the app.

Figure A15: News Dissatisfaction and Time Spent in Different Customizations



Note: This figure shows correlations between news dissatisfaction and the share of time respondents consume the news in different customization options during the incentive and post-incentive periods. News dissatisfaction is a z-score measure of dissatisfaction with the news at baseline. All specifications include fixed effects for the day on which the app download survey was completed. Error bars represent 95% confidence intervals.

Table A12: Effect of Customization Option Order on Usage

(a) Customization with Political Rewrites

	(1) Liberal	(2) Conservative	(3) Less Complex	(4) Entertaining	(5) More Positive	(6) More Negative	(7) Regular
Political	-0.0176	0.0237*	-0.00468	0.0195	-0.0164	0.00805	-0.0125
	(0.0211)	(0.0142)	(0.0187)	(0.0167)	(0.0156)	(0.00809)	(0.0263)
Less Complex	0.00102	0.00174	-0.0275	0.0112	0.00951	0.0128	-0.00867
	(0.0175)	(0.0138)	(0.0185)	(0.0169)	(0.0172)	(0.0120)	(0.0269)
Entertaining	-0.0155	0.0157	-0.00262	-0.0262	0.0226	0.00803	-0.00203
	(0.0176)	(0.0133)	(0.0190)	(0.0184)	(0.0173)	(0.00950)	(0.0265)
Download day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	390	390	390	390	390	390	390

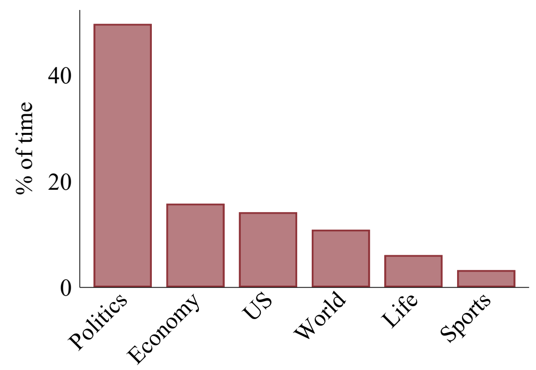
(b) Customization with *Facts Only* Rewrites

	(1) Facts Only	(2) Less Complex	(3) Entertaining	(4) More Positive	(5) More Negative	(6) Regular
Facts Only	-0.0149	-0.0211	0.0186	0.0114	0.00522	0.000864
	(0.0243)	(0.0157)	(0.0148)	(0.0114)	(0.00659)	(0.0207)
Less Complex	0.0144	-0.0545***	-0.00119	0.00796	0.0158*	0.0175
	(0.0243)	(0.0158)	(0.0163)	(0.0105)	(0.00852)	(0.0210)
Entertaining	0.0204	-0.00879	-0.00616	0.00114	0.0119*	-0.0185
	(0.0237)	(0.0146)	(0.0173)	(0.00993)	(0.00695)	(0.0194)
Download day FE	Yes	Yes	Yes	Yes	Yes	Yes
N	499	499	499	499	499	499

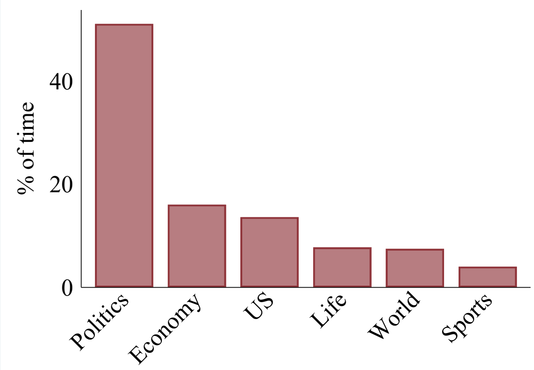
Note: This table shows how time spent in each rewrite version during the post-incentive period correlates with slider order. Panel (a) shows customization with political rewrites and Panel (b) shows customization with *Facts Only* rewrites. The sample includes respondents using the app in the post-incentive period. RHS variables (*Political* or *Facts Only*, *Less Complex*, and *Entertaining*) range from 1-4 indicating rank. The rank of *Sentiment* is omitted. All specifications include fixed effects for the day on which the app download survey was completed. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A16: News Sections by Treatment Arm During the Post-Incentive Period

(a) Customization with Political Rewrites



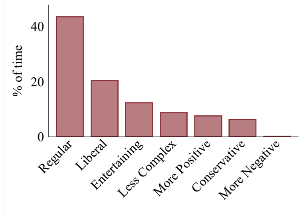
(b) Customization with *Facts Only* Rewrites



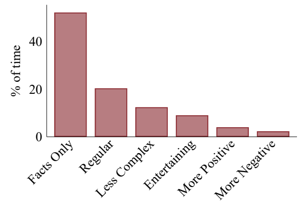
Note: This figure shows the time spent consuming news in different sections during the post-incentive period. Panel (a) shows the customization group with political rewrites. Panel (b) shows the customization group with *Facts Only* rewrites. Individual consumption shares are weighted by the amount of time spent consuming news in the app.

Figure A17: News Customization by Section During the Post-Incentive Period

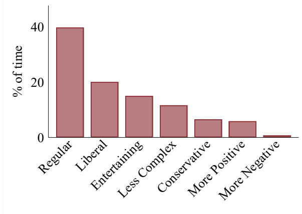
(a) Economy (Political)



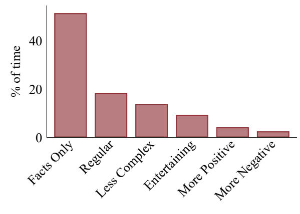
(b) Economy (Facts Only)



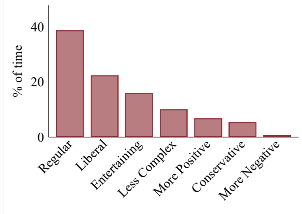
(c) Politics (Political)



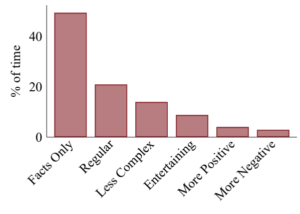
(d) Politics (Facts Only)



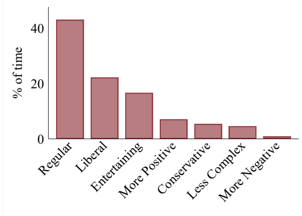
(e) US (Political)



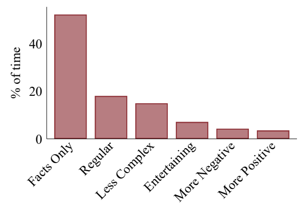
(f) US (Facts Only)



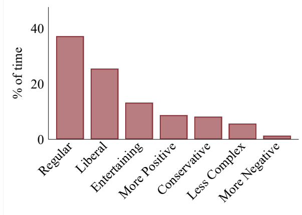
(g) World (Political)



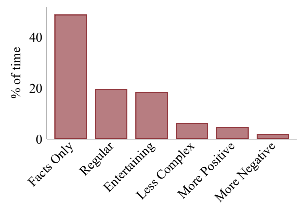
(h) World (Facts Only)



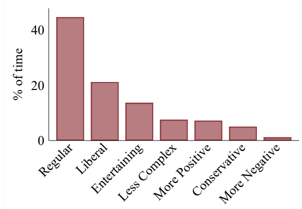
(i) Sports (Political)



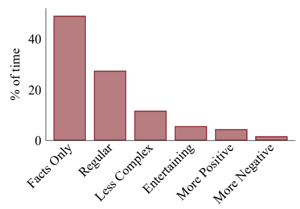
(j) Sports (Facts Only)



(k) Lifestyle (Political)



(l) Lifestyle (Facts Only)



Note: This figure shows the share of time spent consuming news in different customization options during the post-incentive period separately by section. The panels on the left show the shares for the respondents in the customization group with political rewrites. The panels on the right show the shares for the respondents in the customization group with *Facts Only* rewrites. Individual consumption shares are weighted by the amount of time spent consuming news in the app.

Table A13: News Customization at Different Times of the Day During the Post-Incentive Period

(a) Customization with Political Rewrites

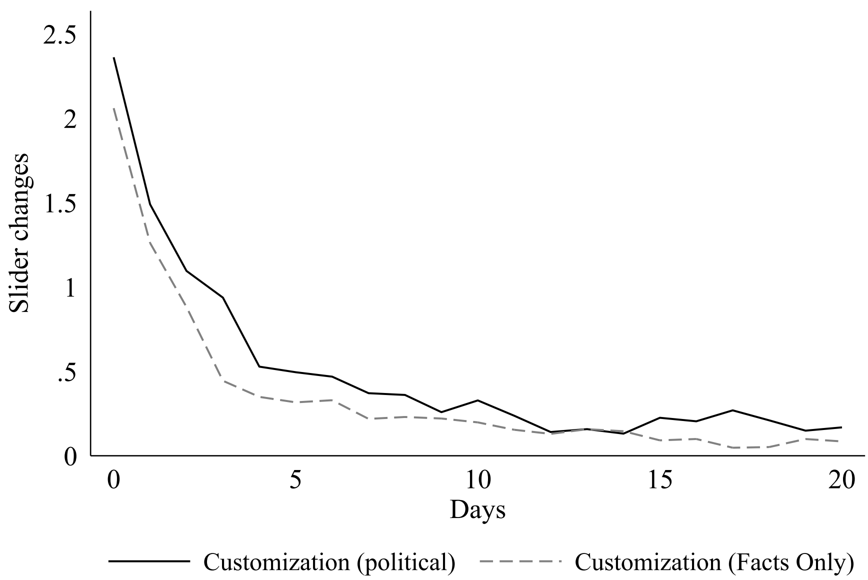
	(1) Regular	(2) Less Complex	(3) Entertaining	(4) More Positive	(5) More Negative	(6) Liberal	(7) Conservative
Evening	0.000499	0.00338	-0.00476	0.00221	-0.00287	-0.00322	0.00476
	(0.00480)	(0.00527)	(0.00526)	(0.00267)	(0.00280)	(0.00398)	(0.00343)
Constant	0.391***	0.133***	0.148***	0.0892***	0.0178***	0.145***	0.0766***
	(0.00130)	(0.00143)	(0.00143)	(0.000725)	(0.000761)	(0.00108)	(0.000932)
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	2161	2161	2161	2161	2161	2161	2161

(b) Customization with *Facts Only* Rewrites

	(1) Regular	(2) Less Complex	(3) Entertaining	(4) More Positive	(5) More Negative	(6) Facts Only
Evening	0.00410	0.00316	-0.00609	-0.00542	0.0107**	-0.00645
	(0.00456)	(0.00376)	(0.00492)	(0.00478)	(0.00505)	(0.00622)
Constant	0.176***	0.112***	0.109***	0.0477***	0.0277***	0.528***
	(0.00121)	(0.000998)	(0.00131)	(0.00127)	(0.00134)	(0.00165)
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
N	3259	3259	3259	3259	3259	3259

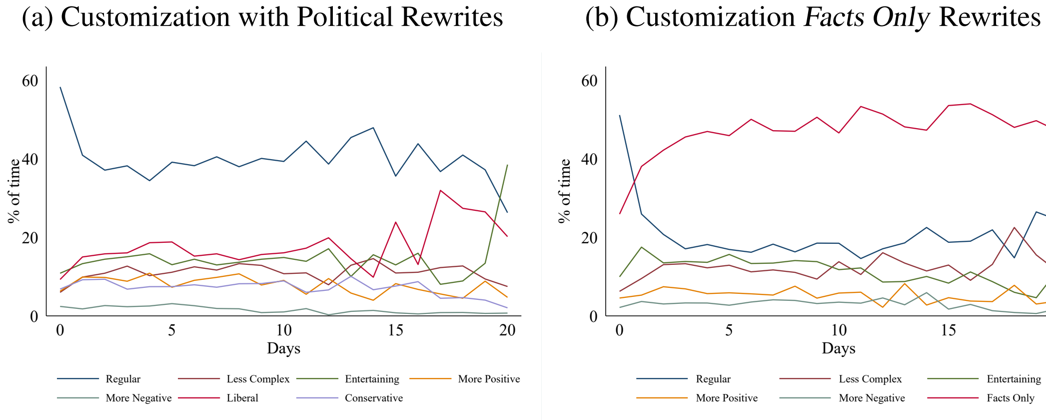
Note: This table reports how use of customization options varies between evening and daytime. Column show results of regressions of the customization option on an evening dummy (6 pm–midnight), excluding night (midnight–6 am). Panel (a) shows the customization arm with political rewrites and Panel (b) with *Facts Only* rewrites. The sample includes respondents using the app during the post-incentive period. All columns include individual fixed effects. Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A18: Number of Slider Changes over Time



Note: This figure shows the average number of slider changes for each participant on each day over time conditional on using the app on a given day. The solid line shows the customization group with the political rewrites and the dashed line shows the group with the *Facts Only* rewrites.

Figure A19: News Customization over Time



Note: This figure shows the use of the different customization settings over time. Panel (a) shows this for the customization arm with political rewrites and Panel (b) for the customization arm with *Facts Only* rewrites. Each line represents the average share of time participants spent with the respective customization setting on a given day.

Table A14: News Customization by Experimentation

(a) Customization with Political Rewrites

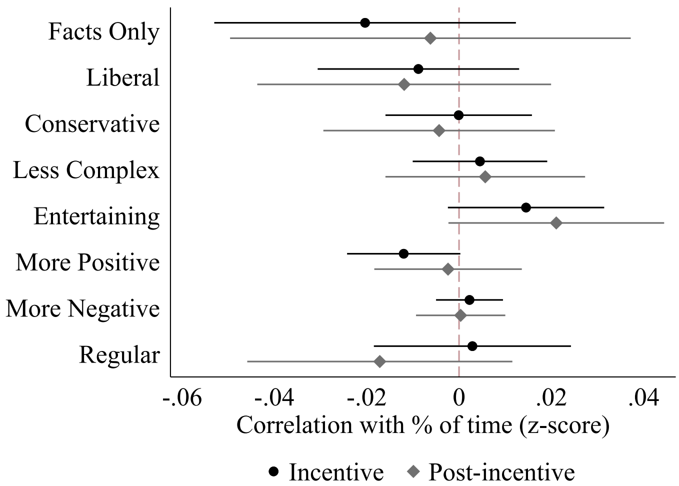
	(1) Liberal	(2) Conservative	(3) Less Complex	(4) Entertaining	(5) More Positive	(6) More Negative	(7) Regular
Slider changes $\geq$ median	0.00726	0.0228	0.0423	-0.0324	-0.0267	0.0230	-0.0362
	(0.0382)	(0.0256)	(0.0348)	(0.0369)	(0.0332)	(0.0151)	(0.0514)
Control mean	0.149	0.057	0.112	0.158	0.124	0.016	0.384
Download day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	390	390	390	390	390	390	390

(b) Customization with *Facts Only* Rewrites

	(1) Facts Only	(2) Less Complex	(3) Entertaining	(4) More Positive	(5) More Negative	(6) Regular
Slider changes $\geq$ median	0.0201	0.0990***	-0.0437	0.0147	0.0470***	-0.137***
	(0.0439)	(0.0272)	(0.0292)	(0.0197)	(0.0134)	(0.0358)
Control mean	0.443	0.067	0.152	0.049	0.004	0.284
Download day FE	Yes	Yes	Yes	Yes	Yes	Yes
N	499	499	499	499	499	499

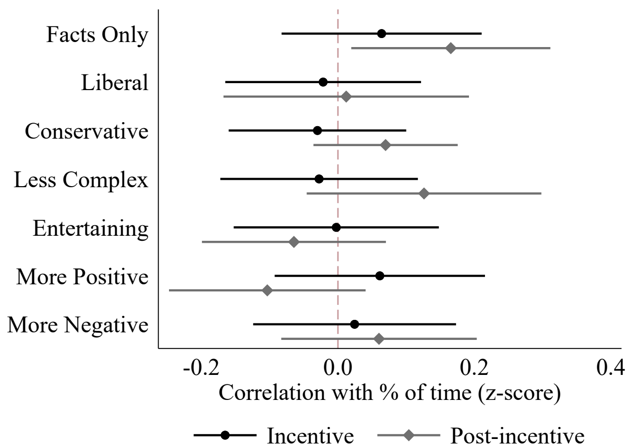
Note: This table shows the use of the different customization options during the post-incentive period by experimentation during the incentive period. Experimentation is measured by the number of slider changes made. Panel (a) shows this for the customization arm with political rewrites and Panel (b) for the customization arm with *Facts Only* rewrites. The sample is restricted to respondents who use the app during the post-incentive period. Coefficients are from a regression of the share of time spent with the respective customization option on a variable indicating above median number of slider changes. All specifications include fixed effects for the day on which the app download survey was completed. Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A20: Trust in AI and Time Spent in Different Customizations



Note: This figure shows correlations between trust in AI and the share of time respondents consume the news in different customization options during the incentive and post-incentive periods. Trust in AI is measured at baseline. Effects are shown in standardized z-scores. All specifications include fixed effects for the day on which the app download survey was completed. Error bars represent 95% confidence intervals.

Figure A21: Perceived News Mismatch and Time Spent by Control Arm

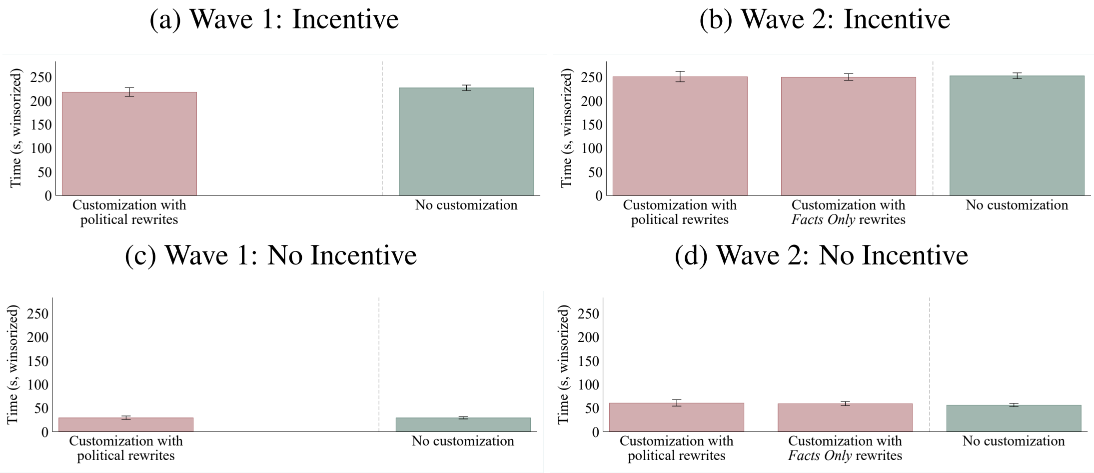


Note: This figure shows the correlations between perceived mismatch of preferences with the news and time spent across control arms. Coefficients are estimated from multivariate regressions of the time spent by respondents in a given control arm (e.g., *Facts Only*) on the full vector of mismatch dimensions as measured in the recruitment survey. For the mismatch dimensions, we elicit respondents' agreement with the statements "News articles are too [complex/negative/positive/liberal/conservative/opinionated]" and "News articles are not entertaining enough". We encode agreement such that larger values represent a stronger match with the customization option and standardize the values. Results are shown separately for the incentive (black circles) and post-incentive (gray diamonds) periods. All specifications include fixed effects for the day on which the app download survey was completed. Error bars represent 95% confidence intervals.

A5.4 Customization and News Consumption

We study whether the differences in news satisfaction and news perceptions also translate into consumption differences in the 10-day post-incentive period. First, we show that consumption is much lower during the post-incentive period than during the incentive period. But there is no difference across treatment arms (see Figure A22).

Figure A22: Time Spent During Incentive and No-Incentive Periods by Treatment Arm

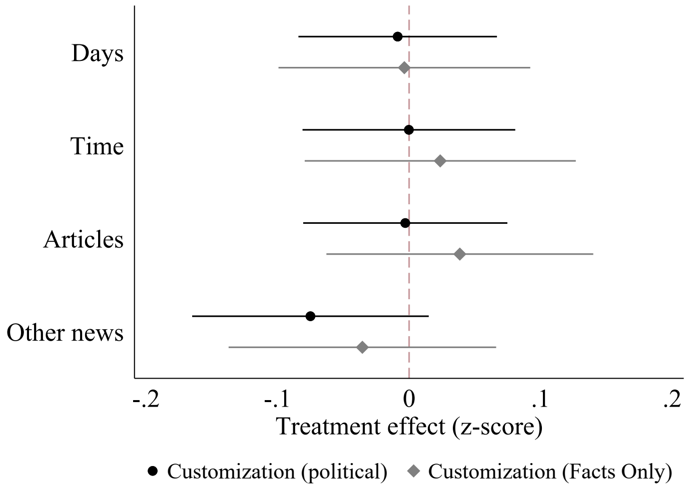


Note: This figure shows app usage time during the incentive and post-incentive periods across treatment arms. Panels (a) and (b) show average daily time in seconds for the incentive periods in Wave 1 and Wave 2, respectively. Panels (c) and (d) show average daily time for the post-incentive period in Wave 1 and Wave 2, respectively. The customization with *Facts Only* rewrites were only introduced in Wave 2. Error bars represent standard errors of the mean.

Next, Figure A23 compares consumption in the two customization conditions against all control conditions. We measure consumption using the winsorized total minutes spent in the app, winsorized number of articles read, and the number of days with any activity in the app. Overall consumption in the post-incentive period is not statistically distinguishable in any of the two customization conditions compared to the control conditions. Figure A23 also reveals that there are no spillovers to news consumption outside of the app (as measured with a self-reported question). The effect sizes we uncover are all relatively close to zero and are relatively precisely estimated.

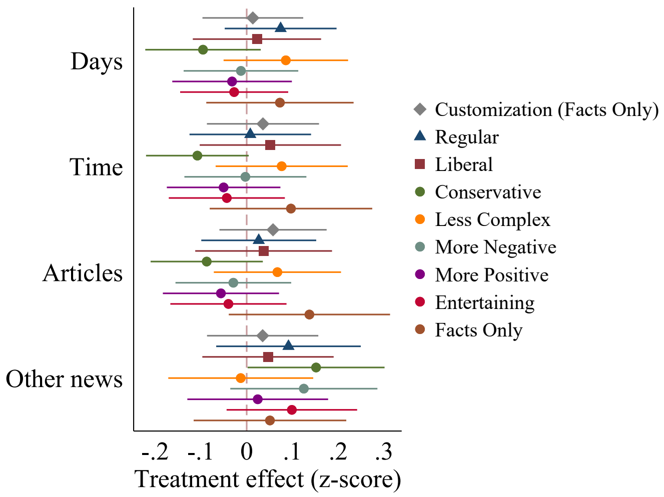
Decomposing effects Figure A24 provides estimates of post-incentive news consumption across all conditions. Despite pronounced differences in news satisfaction previously documented across conditions, we observe little systematic evidence of significant differences in subsequent consumption. Most treatment conditions—including factual customization—display effect sizes near zero, with broad confidence intervals consistently encompassing null effects. Overall, the evidence indicates relatively muted differences in news consumption behavior despite marked disparities in news satisfaction. The mismatch between our estimated effects on news satisfaction and news consumption in the post-incentive period is in line with the idea that engagement with news may not be a good measure of welfare (Bursztyn et al., 2023b; Kleinberg et al., 2024).

Figure A23: Effect of Customization on Consumption during the Post-Incentive Period



Note: This figure shows the effects of customization with political and *Facts Only* rewrites on news consumption in the app during the post-incentive period. Consumption is measured by days with app use, time spent, and articles seen, along with self-reported effects on news use from other sources. All specifications control for demographics (age, gender, race, income, some college, full-time employed), political leaning (party affiliation and leaning, past voting), news consumption patterns (primarily online, paid subscription, frequency, number of news apps, mobile as primary device), and trust in news and AI, and include fixed effects for the day on which the app download survey was completed. Error bars represent 95% confidence intervals.

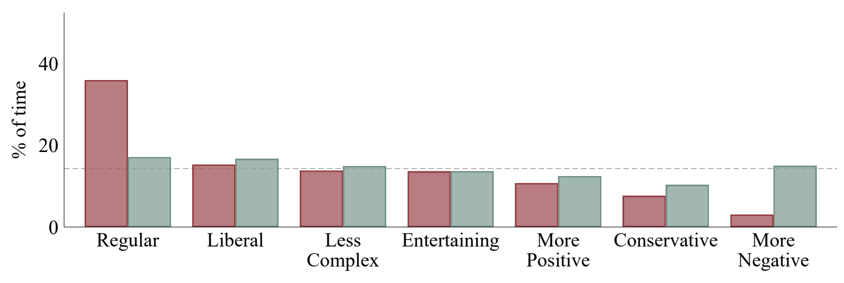
Figure A24: Disaggregated Effects on Consumption during the Post-Incentive Period



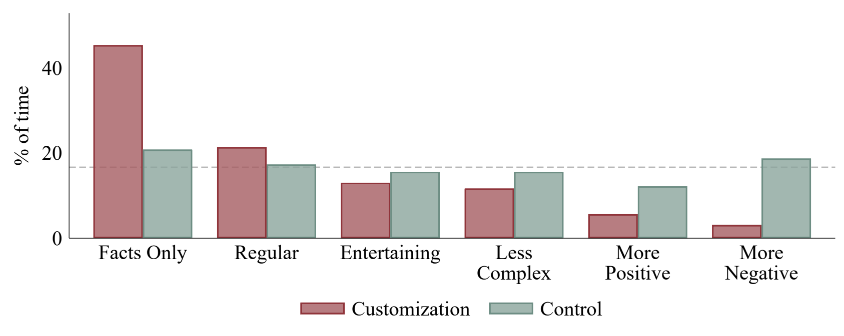
Note: This figure shows the effects of all control conditions and customization with *Facts Only* compared to the omitted category (customization with political rewrites) on news consumption in the app during the post-incentive period. Consumption is measured by days with app use, time spent, and articles seen, along with self-reported effects on news use from other sources. All specifications control for demographics (age, gender, race, income, some college, full-time employed), political leaning (party affiliation and leaning, past voting), news consumption patterns (primarily online, paid subscription, frequency, number of news apps, mobile as primary device), and trust in news and AI, and include fixed effects for the day on which the app download survey was completed. Error bars represent 95% confidence intervals.

Figure A25: News Customization in the Post-Incentive Period

(a) Customization with Political Rewrites



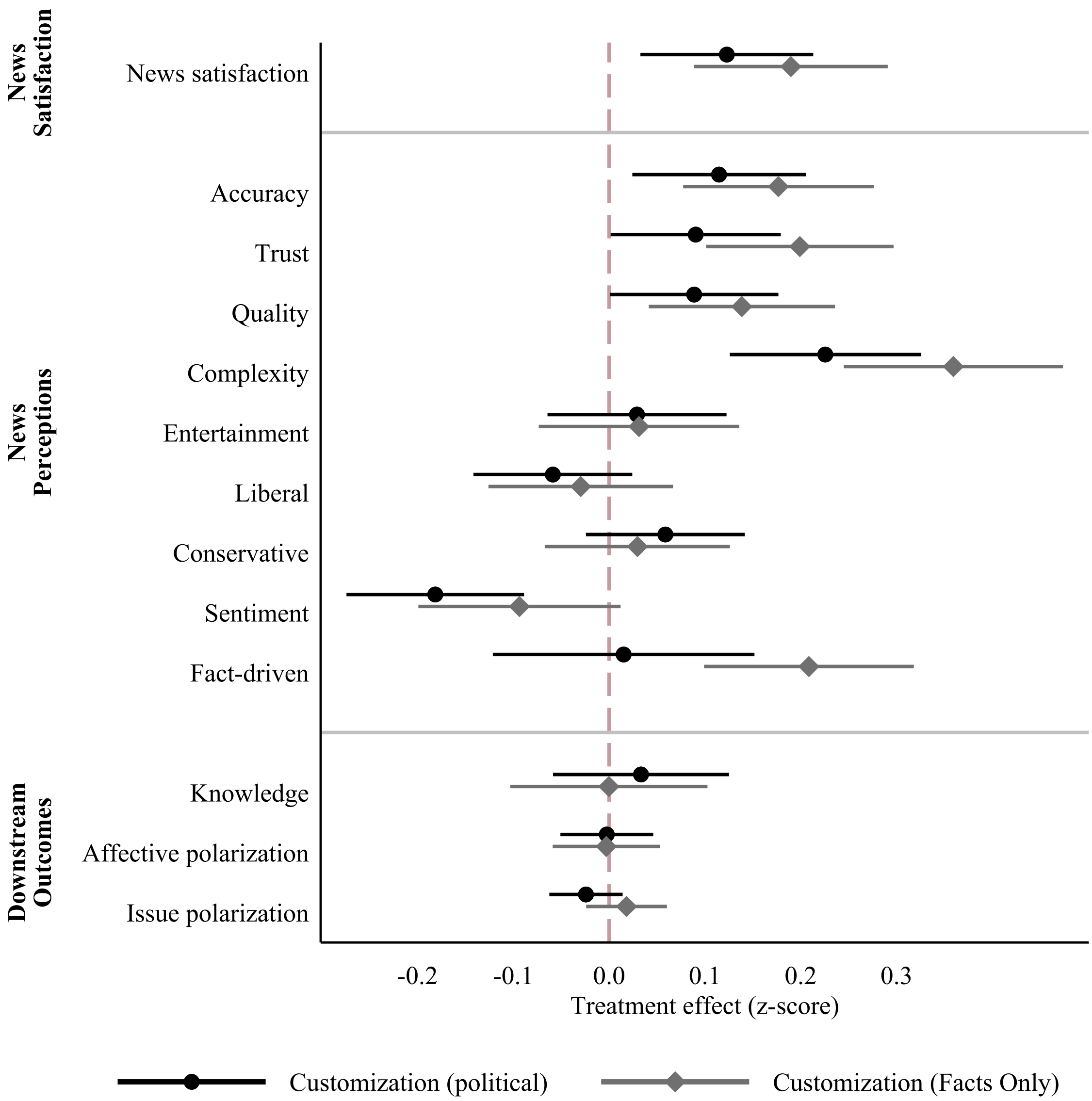
(b) Customization with *Facts Only* Rewrites



Note: This figure shows the share of time spent using the different customization options during the post-incentive period, weighting the individual consumption shares equally. Panel (a) shows the customization treatment with political rewrites and the corresponding control arms. Panel (b) shows the customization treatment with the *Facts Only* rewrites and the corresponding control arms. Panel (a) uses data from both waves, while Panel (b) uses only data from wave 2 given that we only introduced *Facts Only* rewrites in wave 2. The dashed horizontal lines show the shares that would be implied by equal time spent across settings.

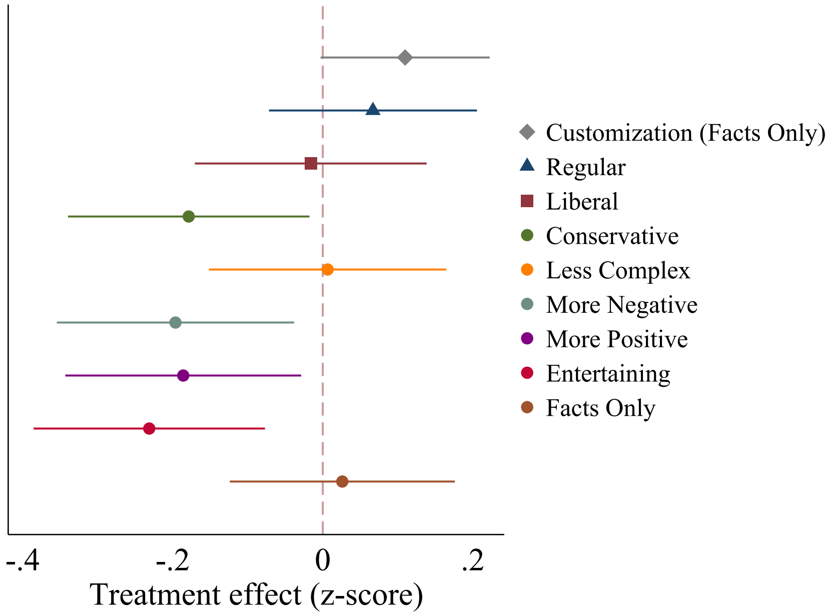
A5.5 Customization and News Satisfaction

Figure A26: Effects on News Satisfaction, Perceptions, and Downstream Outcomes at Endline



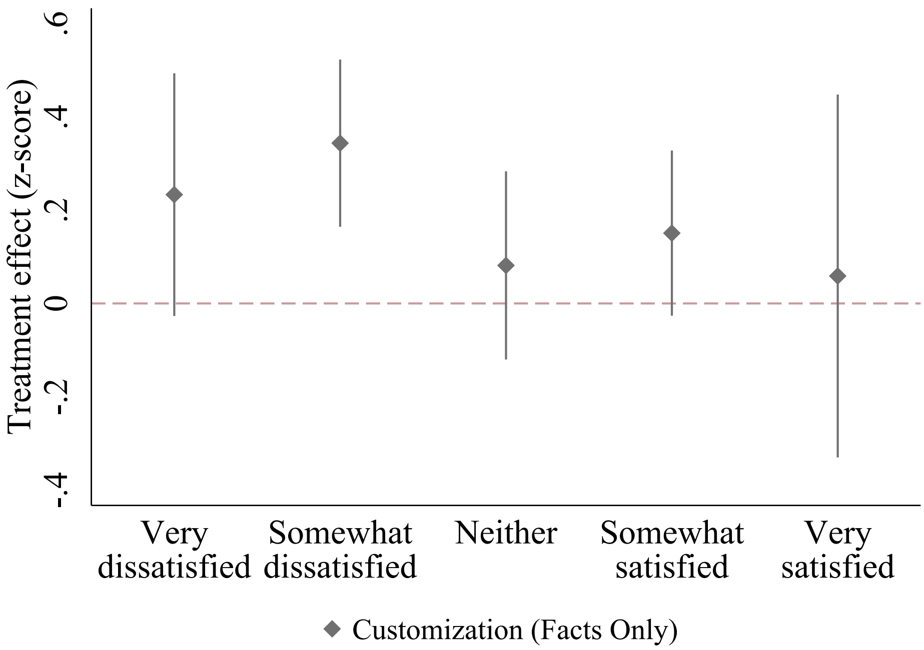
Note: This figure displays the effects of customization with political rewrites and with *Facts Only* rewrites on user satisfaction and perceptions of the app, as well as downstream outcomes, relative to all control conditions, at endline. The first block shows news satisfaction. The second block shows perceptions of accuracy, trustworthiness, quality, complexity, entertainment, political bias (liberal and conservative), sentiment, and whether the content is viewed as fact-driven. The third block shows effects on news knowledge, affective polarization, and issue polarization. All effects are shown in standardized z-scores. All specifications control for demographics (age, gender, race, income, some college, full-time employed), political leaning (party affiliation and leaning, past voting), news consumption patterns (primarily consuming news online, having a paid subscription to a news platform, consuming news at least somewhat frequently, number of news apps on the phone, primary device for news consumption is mobile), and trust in news and AI. All specifications include fixed effects for the day on which the app download survey was completed. Error bars represent 95% confidence intervals.

Figure A27: Disaggregated Effects on News Satisfaction



Note: This figure displays the effects of all different control conditions and customization with *Facts Only* compared to the omitted category (customization with political rewrites) on news satisfaction. All specifications include fixed effects for the day on which the app download survey was completed. Error bars represent 95% confidence intervals.

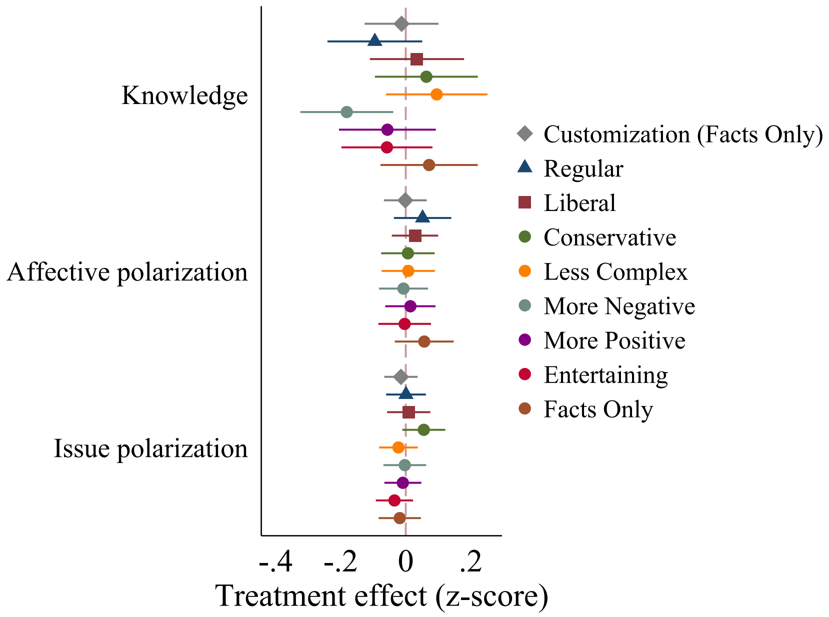
Figure A28: News Satisfaction Heterogeneity by Satisfaction at Baseline



Note: This figure displays the effects of customization with *Facts Only* rewrites on news satisfaction by dissatisfaction at baseline. Effects are shown in standardized z-scores. All specifications include fixed effects for the day on which the app download survey was completed. Error bars represent 95% confidence intervals.

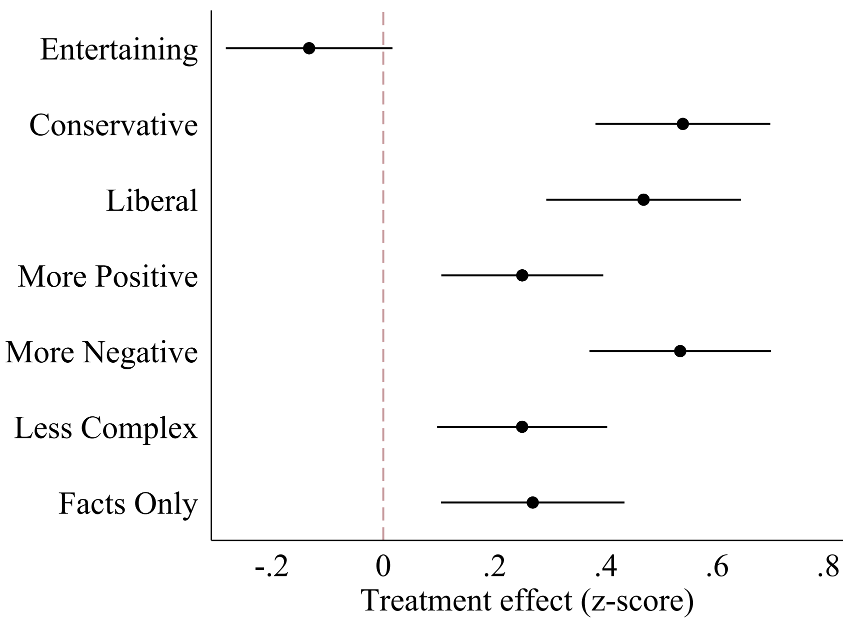
A5.6 Disaggregated Treatment Effects

Figure A29: Disaggregated Effects on Perceptions, Knowledge, and Polarization



Note: This figure displays the effects of all different control conditions and customization with *Facts Only* compared to the omitted category (customization with political rewrites) on knowledge, affective polarization, and issue polarization. All specifications include fixed effects for the day on which the app download survey was completed. Error bars represent 95% confidence intervals.

Figure A30: Perceptions by Control Arm



Note: This figure shows the effect of the control conditions on perceptions (entertaining, conservative, liberal, more positive, more negative, less complex, and whether the content is viewed as fact-driven) as standardized z-scores. Coefficients are from regressions of perceptions on the relevant control condition compared to all other control conditions. All specifications include fixed effects for the day on which the app download survey was completed. Error bars represent 95% confidence intervals.

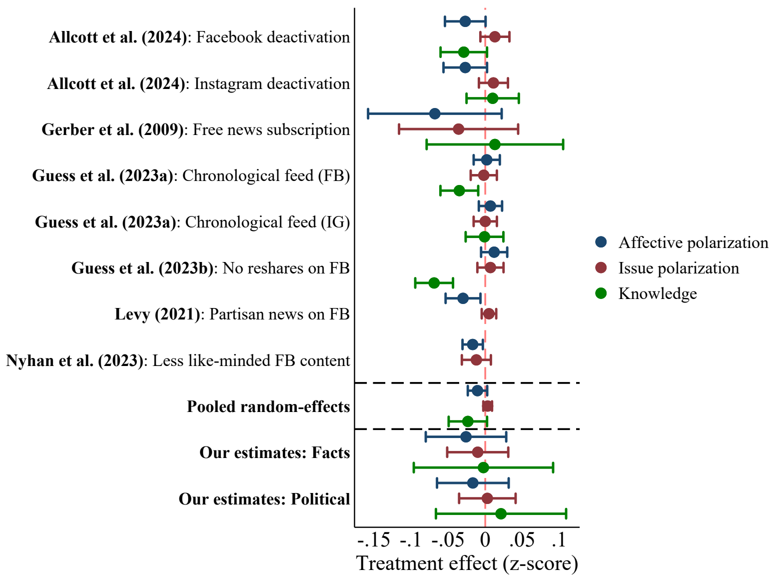
A5.7 Meta Study on Effects on Knowledge and Polarization

Table A15: Comparison to Other Experimental Estimates of Effects

Reference	Intervention	Outcomes	Included
Allcott et al. (2020)	Facebook deactivation	K, AP, IP	Yes
Allcott et al. (2024)	Facebook/Instagram deactivation	K, AP, IP	Yes
Bail et al. (2018)	Follow opposing-party Twitter bot	IP	No
Broockman and Kalla (2025)	Exposure to ideologically opposing news	K, AP, IP	No
Gerber et al. (2009)	Free subscription to newspapers (liberal vs. conservative)	K, AP, IP	Yes
Guess et al. (2021)	Exposure to left or right-leaning opposing news	K, AP, IP	No
Guess et al. (2023a)	Order Facebook/Instagram feeds reverse-chronologically	K, AP, IP	Yes
Guess et al. (2023b)	Feeds without reshares	K, AP, IP	Yes
Levy (2021)	Subscriptions to conservative or liberal news outlets on Facebook	AP, IP	Yes
Liu et al. (2025)	Manipulated YouTube recommendation algorithm	AP, IP	No
Nyhan et al. (2023)	Reduced exposure to content from like-minded sources on Facebook	AP, IP	Yes

Note: This table lists the studies considered for the comparison of effects on knowledge and polarization. "Intervention" describes the experimental intervention in the study. "Outcomes" lists the outcomes included in the study: knowledge ("K"), affective polarization ("AP"), and issue polarization ("IP"). "Included" indicates numerical values for standardized coefficients being available and being included in Figure A31.

Figure A31: Comparison to Other Experimental Estimates of Effects



Note: This figure shows treatment effects from experimental studies on knowledge and polarization. Affective polarization coefficients are in blue, issue polarization in red, and knowledge in green. All estimates are Intent-to-Treat effects shown as standardized z-scores. The y-axis shows the referenced studies (see Table A15). Pooled estimates for each outcome are computed via random-effects meta-analysis using restricted maximum likelihood. Our own estimates appear at the bottom of the figure.