Frictions in News Consumption: Evidence from Social Media*

Luca Braghieri, Ro'ee Levy, and Hannah Trachtman

February 2026

Abstract

We study the drivers of like-minded and low-reliability news-following on social media, as well as the effectiveness of interventions targeting them. In a five-week field experiment with more than 3,000 U.S. Facebook users, we document the importance of salience-based behavioral frictions in shaping users' news portfolios on the platform. Guided by a theoretical framework, the experiment varies: (i) whether participants are prompted to re-optimize the portfolio of news pages they follow on Facebook through a platform-integrated interface that increases the salience of a balanced set of news outlets, and (ii) whether they receive personalized information about outlet slant and reliability. We find that, consistent with our salience model and in contrast to canonical models of news demand, the re-optimization interface with a salient menu of news pages induces large portfolio changes even without the provision of information; conversely, the provision of information has no effect unless paired with the re-optimization interface. Our interventions produce two main implications for users' news portfolios. First, they move users' portfolios closer to their stated preferences, mitigating internalities. Second, they reduce the slant and increase the reliability of users' portfolios, thus potentially mitigating negative externalities for democracy. The induced portfolio changes persist for more than a month, translate into measurable changes in online news consumption, and are unlikely to be driven by experimentation motives or experimenter demand effects.

*Braghieri: Bocconi University and NHH, CEPR, CESifo, and IGIER. Email: luca.braghieri@unibocconi.it. Levy: Tel Aviv University, CESifo, and CEPR. Email: roeelevy@tauex.tau.ac.il. Trachtman: Hebrew University of Jerusalem. Email: hannah.Trachtman@mail.huji.ac.il. We thank Hunt Allcott, Valeria Burdea, Felix Chopra, Adrien D'Autheville, Sarah Eichmeyer, Ingar Haaland, Rafael Jiménez-Durán, Chris Roth, Lena Song, Peter Schwardmann, Dan Stone, Karine Van Der Straeten, Egon Tripodi, and participants at APSA, Bocconi University, Chicago Booth, Coller School of Management, CESifo, EEA Annual Congress, FAIR (NHH), Hebrew University, Technion, and the Toulouse Digital Economics Conference for helpful comments. We thank Giovanni Cecchi, Ittai Dayagi, Yuval Rubinstein, Heitor Sousa Borges, and Siying Wang for excellent research assistance. Funding for this project was provided by the European Research Council (ERC Grant 101116966), the Israel Science Foundation (791/23), the Tel Aviv University Center for AI & Data Science (TAD), the Pinchas Sapir Center for Development, and the Foerder Institute for Economic Studies. Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union, the European Research Council Executive Agency, or the other funding sources. Neither the European Union nor the granting authority can be held responsible for them. This research was approved by the Ethics Committees of Bocconi University (EA000764), Tel-Aviv University (0004305), and Hebrew University (IRB_2024_113). The experiment is registered at the AEA RCT Registry (Braghieri, Levy and Trachtman, 2025).

In 2025, for the first time, the share of U.S. adults who reported consuming news on social media (54%) overtook the share who reported watching news on television, marking a profound shift in the information ecosystem (Newman et al., 2025). This shift has intensified public concerns about the proliferation of like-minded (Braghieri et al., 2025) and low-reliability (Vosoughi, Roy and Aral, 2018) news on social media, as well as about potential negative externalities for democracy (Sunstein, 2017).

Recent research shows that users' own customization decisions on social media—especially which news pages to follow—are a central driver of like-minded and low-reliability news consumption for two reasons. First, a substantial share of the news content users see in their feeds comes from pages they follow.[1] Second, users have been found to encounter more polarized content and more misinformation from followed pages than from other sources (Levy, 2021; González-Bailón et al., 2023; Chouaki et al., 2024). This raises two crucial questions: why do users follow like-minded and low-reliability news pages on social media, and which interventions can reduce the slant and improve the reliability of the pages they follow?

These questions relate to a broader debate about why people consume slanted and low-reliability news.[2] Canonical explanations rely on preferences for like-minded outlets and incomplete information about reliability (Mullainathan and Shleifer, 2005; Gentzkow and Shapiro, 2006). We propose a complementary mechanism that we think is particularly relevant to social media: salience-driven behavioral frictions. Specifically, news customization decisions on social media occur in choice environments heavily shaped by friend and platform cues; if these cues disproportionately surface pro-attitudinal and low-reliability outlets, those outlets can become particularly salient in users' consideration sets and, in turn, overrepresented in the pages users choose to follow (Bordalo, Gennaioli and Shleifer, 2022).

In this paper, we study the drivers of news-following behavior on social media, as well as the effectiveness of interventions targeting them, in the context of a five-week field experiment on Facebook, the world's largest social media platform (Newman et al., 2025). Our theory-guided experiment, which couples changes to the choice architecture (Thaler and Sunstein, 2008) with the provision of tailored information (Haaland, Roth and Wohlfart, 2023), shows that news-following

[1] On Facebook, for instance, it is just under half (Levy, 2021).

[2] Slant captures a news outlet's political leaning on a left-right spectrum; reliability captures the degree to which a news outlet adheres to best-practice journalistic standards.

decisions on social media are substantially driven by salience-based behavioral frictions. By directly targeting those frictions, our intervention produces two key results: first, it increases alignment between users' stated preferences and their resulting news portfolios, thereby reducing internalities.[3] Second, it increases the reliability and reduces the slant of the news pages users follow, thereby mitigating potential negative externalities for democracy. Because the experiment is implemented on a leading platform with real users, real decisions, and a platform-integrated interface, our findings speak directly to the design of policies that could be deployed at scale.

We begin by establishing a few stylized facts about social-media news consumption using a baseline survey. Consistent with canonical models of news demand (Mullainathan and Shleifer, 2005; Gentzkow and Shapiro, 2006), participants interpret slant as a horizontal characteristic and reliability as a vertical one, they have relatively accurate perceptions of outlets' slant and much less accurate perceptions of outlets' reliability, and they hold pro-attitudinal news portfolios on Facebook. The main finding that is prima facie at odds with canonical models relying on the optimization of stable preferences under incomplete information is the way in which participants describe their news portfolio formation process. Specifically, respondents report that their news-following behavior on social media is generally passive—driven primarily by friends' suggestions and platform cues rather than deliberate search.

The tension above motivates a theoretical framework that contrasts two models of news-following behavior on social media. The first is a workhorse model featuring deliberate optimization under stable preferences and incomplete information that distills the canonical forces emphasized in the literature (Mullainathan and Shleifer, 2005; Gentzkow and Shapiro, 2006). In this model, users follow low-reliability outlets because they lack precise information about reliability, and they follow like-minded outlets both because they derive greater psychological utility from them and because they treat slant affinity as a proxy for high reliability. The second is a salience model in which friends' and platform cues substantially shape both when users make news-following decisions and which outlets enter their consideration sets. Because extreme pro-attitudinal and low-reliability content has been shown to be overrepresented on social media feeds (Braghieri et al., 2025; Vosoughi, Roy and Aral, 2018), this model can also generate pro-attitudinal and low-reliability news portfolios.

[3]We use the term "news portfolio" to indicate the set of news pages that an individual follows on Facebook.

To distinguish the two models, we turn to their comparative statics. In particular, the workhorse and salience models make sharply different predictions about how portfolios respond to information provision and to changes to the choice architecture. If portfolios reflect deliberate optimization under stable preferences and incomplete information, then providing information that users value but lack should induce them to modify their news portfolios. Conversely, offering users the opportunity to re-optimize their news portfolios using an interface that increases the salience of certain outlets without providing new information should have no effect. In contrast, if portfolios are substantially shaped by "bottom-up" attentional cues that determine which outlets enter users' consideration sets, a re-optimization interface that makes certain outlets more salient without providing new information should affect choices by shifting consideration sets. Conversely, information delivered outside a decision episode with a salient menu of news pages should be ineffective.

Guided by these predictions, we designed a five-week field experiment on Facebook that varies (i) whether users experience a prompted re-optimization episode with a salient menu of news pages, and (ii) whether they receive personalized information about the news pages they follow. The experiment proceeded as follows. We recruited more than 3,000 participants through Facebook ads and directed them to a baseline survey that collected demographics, elicited incentivized beliefs about outlet slant and reliability, and measured participants' bliss points in the slant–reliability space. All participants were required to grant us permission to observe the news pages they follow on Facebook via the platform's API. Participants using a Google Chrome browser were also incentivized to install a Chrome extension that tracked their browsing behavior on news websites.

Approximately one week after the baseline survey, participants were invited to a midline survey and were randomized into five main arms: a *Control* group; a ***R**e-optimization* treatment (**R**); a ***R**e-optimization with **R**eliability information* treatment (**RR**); a ***R**e-optimization with **S**lant information* treatment (**RS**); and a ***N**o-Reoptimization with **R**eliability information* treatment (**NR**). Participants assigned to the re-optimization treatments (R, RR, RS) were shown a user-friendly interface that allowed them, with a single click, to follow any news page from a curated list of the most popular news outlets on Facebook or to unfollow any news page they were already following on the platform. On top of the re-optimization interface, participants in RR received personalized information about the reliability of their Facebook news portfolios, whereas participants in RS received analogous information about slant. Finally, participants in NR received the same reliability

information as RR but were not shown the re-optimization interface. Around four weeks after midline, participants were invited to an endline survey, where they were given another opportunity to re-optimize their Facebook news portfolios using our interface and where we elicited survey-based outcomes.

We collected two behavioral datasets in addition to survey data. First, using the Facebook API, we recorded the news pages participants follow throughout the experiment. Second, for the subset of participants who installed our Chrome extension, we tracked visits to news websites. Our primary pre-specified outcome variable is the set of news pages that participants follow on Facebook after the midline survey. We focus on following behavior because it is the core customization decision faced by Facebook users, because it has been shown to be a key determinant of subsequent news exposure on the platform (Levy, 2021), and because the pages users follow on social media have been shown to affect their knowledge, beliefs, and political attitudes (Altay, Hoes and Wojcieszak, 2025; Chmel et al., 2025; Levy, 2021; Rathje et al., 2024).

A natural concern in the context of our experiment is that the results might be substantially affected by experimenter demand effects. To explore this possibility, our design features an additional treatment arm—the **N**o **E**xperimenter **D**emand treatment (**NED**)—which closely mirrors the RR treatment but explicitly removes potential experimenter oversight. Specifically, at the start of the midline survey, we asked participants in the NED group to revoke our permission to track the Facebook pages they follow. Later, at endline, we surprised participants by requesting permission to collect their following behavior retrospectively. By comparing the NED and RR treatments, which only differ in the extent to which participants thought their news following choices in the midline survey were observable by the experimenter, we can estimate the role of experimenter demand effects.

We organize our discussion of the experimental results in three steps. First, we relate the pattern of treatment effects to the theoretical framework and provide evidence in favor of the salience model, highlighting the importance of salience-driven behavioral frictions in shaping news-following decisions on social media. Second, we study the effects of targeting these frictions on outcomes related to both internalities and externalities. Third, we discuss the persistence of our effects, the impact of our intervention on downstream visits to news websites, and potential confounds.

The core empirical discriminator between the workhorse and salience models is participants' differential response to the **R** and **NR** treatments. First, the salience model predicts that **R**—which triggers a decision episode and presents a salient set of pages—should induce users to revise the portfolio of news pages they follow on Facebook, whereas the workhorse model predicts no effect from the re-optimization prompt alone. We find that, in line with the salience model, the vast majority (71%) of participants assigned to the **R** treatment actively modified their news portfolios by adding or removing at least one page using our re-optimization interface. This high degree of re-optimization—in the absence of new information, monetary incentives, or explicit encouragement—suggests that participants' portfolios were not already optimized and underscores the role of salience in shaping news customization decisions.[4]

Second, the workhorse model predicts that **NR**—which provides reliability information without prompting re-optimization—should shift following behavior, to the extent that users value this information and, as we established, suffer from misperceptions regarding reliability. The salience model makes the opposite prediction: absent a prompted decision episode and a salient menu of pages, information alone should not lead users to adjust their portfolios. Once again, the experimental results align with the salience model: the **NR** treatment has no detectable effect on the set of pages participants follow. Importantly, when the same information is delivered alongside the re-optimization interface in the **RR** treatment, participants make meaningful portfolio adjustments and the average reliability of the news pages they follow increases substantially. This suggests that participants value the information we provide, but that they act on it only if it is delivered in the context of an interface that surfaces a salient set of pages. Taken together, these results support the view that the salience-driven behavioral frictions are central to news-following decisions on social media.

The next set of results shows that our interventions targeting those frictions shift portfolios along dimensions that relate to both internalities and externalities. On the internality front, our interventions help close the gap between participants' portfolios and their stated preferences. Specifically, the three re-optimization treatment arms (R, RR, and RS) significantly reduce the distance between participants' actual portfolios and their self-reported bliss points in the slant-reliability

[4] An alternative explanation is that participants' re-optimization choices reflect experimentation motives. We address and rule out this explanation in Section 6.2.

space. Consistent with greater alignment between participants' actual portfolios and their bliss points, four weeks after the midline survey, treated participants reported higher trust in the news they encounter on Facebook and greater satisfaction with those news.

On the externality front, we study how our interventions address two major concerns about news consumption on social media: exposure to like-minded and low-reliability news. We find that, across all re-optimization arms, participants' portfolios become, on average, less slanted. Furthermore, the RR treatment, which pairs the re-optimization interface with reliability information, increases average portfolio reliability. Together, these shifts improve the composition of users' news portfolios along the very dimensions—slant and reliability—that are often implicated in concerns about potential negative externalities of social media news consumption for democracy.

Our last set of results explores persistence, effects on downstream online news consumption, and rules out potential confounds. Because following a page is a sticky customization decision, our one-time intervention can generate persistent changes in exposure and consumption. Consistent with this idea, changes in portfolios persist through the endline survey four weeks later. Moreover, for the subset of participants who installed our Chrome extension, we find that following a news page through our intervention more than doubles the probability of visiting that page in the subsequent month. This result links portfolio changes to meaningful changes in online news consumption, emphasizing the downstream consequences of shifting the set of pages that Facebook users follow on the platform. Lastly, we show that our results are unlikely to be driven by experimentation motives and, leveraging the comparison between the NED and RR treatments, by experimenter demand effects.

Our paper contributes to several strands of literature. First, we contribute to the literature studying news choices (Chan and Suen, 2008; Chopra, Haaland and Roth, 2022, 2023; Chopra et al., 2025; Gentzkow and Shapiro, 2006; Gentzkow, Shapiro and Stone, 2015; Mullainathan and Shleifer, 2005; Suen, 2004). Much of the theoretical literature has explored why individuals consume news aligned with their existing ideological beliefs, highlighting two main explanations: individuals might perceive pro-attitudinal news as more reliable (Gentzkow and Shapiro, 2006) and derive greater utility from consuming like-minded content (Mullainathan and Shleifer, 2005). We introduce and study a complementary driver that is likely to be especially important in platform-mediated environments: salience-driven behavioral frictions that shape *which* outlets enter users'

consideration sets and *when* users make customization decisions (Bordalo, Gennaioli and Shleifer, 2022). To the best of our knowledge, we provide the first systematic evidence that these behavioral frictions play a first-order role in decisions about which news pages users follow on social media.

Second, our paper adds to the growing body of work exploring interventions aimed at reducing the consumption and sharing of fake, low-reliability, or partisan news (Aslett et al., 2022; Henry, Zhuravskaya and Guriev, 2022; Guriev et al., 2025; Martel et al., 2024; Pennycook and Rand, 2021). The policy tools proposed so far—including content moderation, algorithmic re-engineering, fact-checking, and accuracy prompts—vary substantially in complexity, cost-effectiveness, scalability, and persistence. We introduce a simple, cheap, and scalable intervention, motivated by our salience model of news-following, that produces persistent shifts toward less slanted and more reliable news portfolios. The intervention operates as a form of soft commitment, in that a choice made in a relatively deliberative moment subsequently shapes day-to-day news exposure without requiring repeated attention or ongoing effort. Because analogous platform-mediated choices arise outside social media as well—for example, when users decide which news sites to bookmark, which mailing lists to subscribe to, or which apps to install—a similar approach may be valuable in other digital settings.

Third, our paper builds on the behavioral-economics literature on "bottom-up" attention processes and salience (Bordalo, Gennaioli and Shleifer, 2022), which shows, for instance, how salient dimensions of choice affect consumption (e.g., Bordalo, Gennaioli and Shleifer, 2013; Busse et al., 2013; Chetty, Looney and Kroft, 2009; Kroft et al., 2024; Taubinsky and Rees-Jones, 2018), how reminders can prompt decision episodes (e.g., Allcott and Rogers, 2014; Karlan et al., 2016), and how requiring active choice can affect behavior in highly consequential settings (e.g., Carroll et al., 2009; Keller et al., 2011). We extend these insights to news consumption in platform-mediated environments, highlighting how interventions that target salience-driven behavioral frictions can reduce the wedge between users' news portfolios and their stated preferences.

Finally, our paper contributes to the literature on experimenter demand effects (de Quidt, Haushofer and Roth, 2018; Dhar, Jain and Jayachandran, 2022; Mummolo and Peterson, 2019; Zizzo, 2010). We introduce a new methodology that leverages the possibility of collecting retrospective data to estimate experimenter demand effects.

1. Background

1.1 News Consumption on Social Media

Social media platforms are central players in the news ecosystem, with 54% of U.S. adults reporting they consumed news on social media within the past week (Newman et al., 2025). Among social media platforms, Facebook remains the dominant source of news in the United States. With approximately 190 million monthly active users domestically, Facebook's sheer scale positions it as a critical gateway to news consumption. Indeed, according to 2025 survey data from Reuters, 32% of U.S. adults regularly obtain news from Facebook—more than from any other social media platform and on par with the most popular U.S. news channel, Fox News (Newman et al., 2025). For comparison, only 23% and 12% of U.S. adults report regularly accessing news through Twitter (now "X") and TikTok, respectively (Newman et al., 2025).

The prevalence of partisan and low-reliability news on social media has generated widespread concerns (Sunstein, 2017). In line with these concerns, Braghieri et al. (2025) document a high degree of polarization in the news consumed on Facebook, showing that news accessed through Facebook is substantially more polarized than news accessed via other channels. For instance, it is 2.5 times more polarized than news accessed through search engines. In parallel, Vosoughi, Roy and Aral (2018) find that fake news spreads both more rapidly and more widely on social media platforms compared to truthful content.

Recent research shows that pro-attitudinal and low-reliability news consumption on social media is in large part driven by users' own feed customization decisions (Levy, 2021; González-Bailón et al., 2023; Chouaki et al., 2024). For instance, Levy (2021) finds that, within a person's feed and according to the isolation index from Gentzkow and Shapiro (2011), content from followed pages is more than twice as ideologically segregated as content from friends.[5] Similarly, Chouaki et al. (2024) shows that social media users encounter more misinformation from pages they follow than from incidental exposure. These findings matter both directly—because 41% of the news content that users observe on their feeds and around half of Facebook-referred visits to news sites come from followed pages (Levy, 2021)—and indirectly—because the pages that users

[5]Throughout, we use "following a news page" to refer to both "following" and "liking" a page. These actions have similar effects on the content users see in their feeds. The main difference is that following—introduced more recently—offers greater control over notifications and privacy settings.

follow provide signals to the content ranking algorithm, which might amplify subsequent exposure. Moreover, multiple studies show that the accounts individuals follow on social media affect their knowledge, beliefs, and attitudes (Altay, Hoes and Wojcieszak, 2025; Chmel et al., 2025; Levy, 2021; Rathje et al., 2024).

The evidence above motivates our focus on users' news customization decisions. To connect these decisions to concerns about slanted and low-reliability news content, we need empirical measures of both concepts. The next subsection describes how we operationalize slant and reliability in our setting.

1.2 Slant and Reliability

As far as slant is concerned, we adopt the measure from Bakshy, Messing and Adamic (2015), which assigns ideology scores to hundreds of English-language news outlets based on the average political leaning of Facebook users who shared content from those outlets. This measure satisfies three desiderata: first, it is one of the most comprehensive measures of outlet-level slant available in the literature; second, it correlates strongly with other established measures of slant, helping alleviate concerns about validity (Braghieri et al., 2025); third, it is relatively easy to describe to participants.

Following Aslett et al. (2022), we operationalize the notion of reliability using NewsGuard ratings, which score outlets along nine criteria of journalistic quality, including passing fact-checks, appropriately sourcing material, clearly flagging paid content, distinguishing news from opinion, and having a transparent ownership structure. NewsGuard ratings correlate strongly with several other measures of reliability used in the literature (Lin et al., 2023) and also have high coverage. Appendix B provides additional details on the slant and reliability measures employed in this project.

Our final dataset of outlet slant and reliability comprises 262 news outlets for which we have both a slant score and a reliability score, and that we successfully matched to a corresponding Facebook page. These outlets account for the lion's share of online news consumption: according to Comscore data (Comscore, 2022), they capture 93% of visits to online news websites.[6] Appendix

[6]To compute the total number of visits to news websites in the 2022 Comscore Web Behavior Database Panel, we sum visits to all domains that NewsGuard classifies as U.S. English news sites focusing on "Local News," "General

Figure A.1 presents a scatter plot of these 262 outlets in the slant–reliability space. As noted by various media watchdogs (e.g., AdFontes Media), the relationship between slant and reliability is inverse-U-shaped: outlets with extreme slant—particularly those on the right-end of the political spectrum—tend to have lower reliability scores.

2. Motivating Evidence

We document six stylized facts about news consumption on Facebook using a baseline survey of 3,356 Facebook users that we conducted as part of our field experiment.[7] These stylized facts, summarized for convenience in Appendix Table A.1, motivate our theoretical framework. To our knowledge, several of these facts—specifically those about misperceptions and the wedge between actual and ideal portfolios—have not been empirically documented before and contribute to the literature by providing new descriptive evidence on people's news perceptions, preferences, and portfolios.

2.1 Self-Reported Bliss Points

We begin by documenting participants' self-reported bliss points in the slant–reliability space, elicited as shown in Appendix F.4. Consistent with workhorse models of news consumption (Gentzkow and Shapiro, 2006; Mullainathan and Shleifer, 2005), participants' self-reported bliss points support interpreting slant as a horizontal attribute and reliability as a vertical one. Specifically, Appendix Figure A.2a shows that slant bliss points exhibit the hallmark pattern of a horizontal characteristic: participants who identify as conservative report desiring right-leaning portfolios, and vice-versa for liberals. In contrast, Appendix Figure A.2b shows that reliability bliss points have the footprint of a vertical characteristic: the vast majority of both liberals and conservatives report preferring high-reliability portfolios.[8]

News," and/or "Political news or commentary." We manually exclude domains that clearly do not fit these categories (e.g., substack.com).

[7] Additional details about recruitment, the sample, and data collection are described in Section 4.4.

[8] Some participants express a preference for reliability below 100. This could reflect demand for entertainment (Chopra et al., 2025)—if respondents perceive a trade-off between reliability and entertainment—misunderstanding, or noise.

Fact 1. *Stated bliss points are largely consistent with slant being a horizontal characteristic and reliability being a vertical characteristic.*

2.2 Perceptions of Slant and Reliability

The next set of stylized facts documents participants' (mis)perceptions of the slant and reliability of their Facebook news portfolios and of various news outlets. In the baseline survey, we carefully described our measures of slant and reliability, reminded participants which Facebook news pages they follow, and elicited participants' perceptions of the average slant and reliability of those pages. To mitigate expressive responses, the questions were incentivized as described in Appendix Section F.4. We find that, in line with workhorse models of news consumption (Gentzkow and Shapiro, 2006; Mullainathan and Shleifer, 2005), participants perceive slant relatively accurately and perceive reliability much less accurately.

The left panel of Figure 1 plots perceived portfolio slant (y-axis) against actual portfolio slant (x-axis). The figure shows that participants' slant perceptions are reasonably on target: observations cluster around the 45-degree line, and the correlation between perceived and actual average portfolio slant is 0.61. To assess accuracy categorically, we classify portfolios as left-leaning (slant < -0.2), centrist ($-0.2 \leq$ slant ≤ 0.2), or right-leaning (slant > 0.2). Appendix Table A.2 shows that only 7%-13% of participants substantially misclassify their portfolio's ideological orientation (e.g., perceiving it as right-leaning when it is left-leaning).

Reliability perceptions are markedly noisier and much less accurate than slant perceptions. The right panel of Figure 1 plots perceived portfolio reliability (y-axis) against actual portfolio reliability (x-axis). The correlation is only 0.19, indicating substantial misperceptions.[9] To characterize reliability errors categorically, we group portfolios into low reliability (score < 60), moderate reliability ($60 \leq$ score ≤ 80), and high reliability (score > 80). Appendix Table A.3 shows that a substantial fraction (31%) of participants with low-reliability portfolios incorrectly place them in the high-reliability category, thus suggesting that the provision of targeted information about relia-

[9] A potential concern is that, despite our explicit instructions and comprehension questions, respondents' interpretation of "reliability" is different from NewsGuard's. Appendix Figure A.3 shows robustness to alternative measures of slant and reliability: we compare perceived reliability to Media Bias Fact Check's factual reporting rating and perceived slant to the slant measure from Braghieri et al. (2025). The same qualitative pattern holds: slant perceptions are substantially more accurate than reliability perceptions.

bility might be helpful to some participants. Appendix E shows a similar pattern for major outlets: participants are more accurate about slant rankings than about reliability rankings.

Fact 2. *Slant perceptions are relatively accurate; reliability perceptions are relatively inaccurate.*

Finally, in line with Gentzkow and Shapiro (2006), perceptions of slant and reliability are correlated: participants tend to predict higher reliability for outlets that align with their own ideology. In particular, Appendix Table A.4 shows that Democrats perceive left-leaning outlets as more reliable, whereas Republicans perceive right-leaning outlets as more reliable.

Fact 3. *Perceptions of slant and reliability are correlated: pro-attitudinal outlets are generally considered higher reliability than counter-attitudinal ones.*

2.3 Facebook News Portfolios

We now turn to participants' actual news portfolios on Facebook. As shown in Appendix Table A.5, Democrats tend to hold more left-leaning portfolios and Republicans more right-leaning ones, consistent with existing evidence on pro-attitudinal news consumption (Gentzkow and Shapiro, 2010).

Fact 4. *Participants hold news portfolios that, on average, are pro-attitudinal.*

Next, we compare participants' actual portfolios to their stated bliss points and document a systematic wedge. Figure 2 plots actual portfolio slant (reliability) against desired slant (reliability). The left panel shows that, although actual and desired slant are positively correlated, participants typically report wanting less extreme portfolios than they currently have: the fitted line is flatter than the 45-degree line. The average gap between actual and desired absolute average slant (the absolute value of the average slant of participants' portfolio) is 0.39, roughly the ideological distance between MSNBC and *The New York Times*.

The right panel shows little relationship between actual and desired reliability: most participants want high reliability regardless of their baseline portfolio, implying that desired reliability typically exceeds actual reliability. The average reliability gap is 15.1 points on the 0–100 scale, roughly the difference between *The Huffington Post* and *The New York Times*.

Fact 5. *Participants hold news portfolios that, on average, are more partisan and less reliable than their stated bliss points.*

A natural interpretation of the reliability wedge is informational: when most users want high reliability but perceive it noisily as in Gentzkow and Shapiro (2006), realized portfolios will, on average, have lower reliability than participants' stated bliss points. The slant wedge is more puzzling from the perspective of workhorse models of news consumption: despite perceiving slant relatively accurately, participants still hold portfolios that are more extreme than their stated bliss points. This wedge could be explained by salience-driven behavioral frictions as we discuss in the theoretical framework.

The last stylized fact is that participants describe their portfolio-formation process as passive and unstructured. Figure 3 shows that 53% of participants report encountering news online incidentally rather than intentionally—that is, while doing something else rather than actively seeking it out. The same figure shows that, when asked how they typically come to follow a news page on Facebook, only 28% said they had searched for it intentionally; 72% reported following a page because someone shared content from it or because Facebook recommended it. Finally, few participants appear to think of the pages they follow as a coherent portfolio. When asked whether they consider how a page complements the ones they already follow, only 16% of them answered affirmatively.

Fact 6. *Participants describe their news-portfolio formation process as largely passive and driven by salient suggestions from the platform or from friends.*

The next section interprets these motivating facts through the lens of two theoretical models: a workhorse model of news consumption and a salience-based model. The comparison between the two models guides our experimental design.

3. Theoretical Framework

We develop our theoretical framework in three steps. First, Section 3.1 presents a workhorse model that captures canonical demand forces behind the consumption of like-minded news and assesses the extent to which this model can account for the motivating evidence presented in the previous section. Second, Section 3.2 adds a minimal salience mechanism that restricts when choices are made and which outlets enter agents' consideration sets, aiming to match the self-reported passive,

platform-driven nature of news following on social media. Finally, Section 3.3 derives distinct predictions about how each model responds to changes in the choice architecture and to the provision of reliability information. These predictions inform the experimental design and structure the empirical analysis.

3.1 Workhorse model of news consumption

We begin by laying out a simple framework that captures the main forces in two canonical models of news consumption—Mullainathan and Shleifer (2005) and Gentzkow and Shapiro (2006). In Mullainathan and Shleifer (2005), individuals consume ideologically congenial news because they derive greater psychological utility from reading it. In Gentzkow and Shapiro (2006), individuals care about how accurately news conveys information about the state of the world, which, in this paper, we proxy by reliability. In the model of Gentzkow and Shapiro (2006), individuals have heterogeneous priors about the state of the world and, rationally, treat the proximity of an outlet's slant to their prior as a signal of higher reliability. Both the mechanism emphasized in Mullainathan and Shleifer (2005) and the one emphasized in Gentzkow and Shapiro (2006) produce pro-attitudinal patterns of news demand.[10]

Setup We capture the core mechanisms in Mullainathan and Shleifer (2005) and Gentzkow and Shapiro (2006) through a unified framework, which we refer to as the *workhorse* model. The economy consists of a finite set of users I and a countably infinite set of news pages J.[11] Each user i has an ideological position $s_i \in \mathbb{R}$, and each page j has a slant $s_j \in \mathbb{R}$ (observed by users). Each page also has a true reliability $r_j \in \mathbb{R}$ that is not directly observed by users.

We assume a locally dense supply of outlets in slant: for every s_i and every $\varepsilon > 0$, there exists an outlet j with $|s_j - s_i| < \varepsilon$. Furthermore, we assume that the distribution of r_j in any neighborhood of s_i has nondegenerate support.[12] Finally, the conditional mean reliability is unimodal and peaks at moderate slant (normalized to $s = 0$).

[10] Another important mechanism behind pro-attitudinal news consumption is delegation, as in Suen (2004). We abstract from delegation here to keep the model tractable.

[11] The assumption of countably many pages simplifies notation and the statement of propositions; a large but finite set would yield the same qualitative results.

[12] These assumptions are stronger than necessary but they simplify exposition; qualitatively we only need a sufficiently high number of both left- and right-leaning outlets.

The assumptions above provide a clean baseline that aligns with the relevant stylized facts from the previous section. Treating slant as known and reliability as uncertain reflects evidence that people accurately perceive an outlet's slant but have noisier beliefs about its reliability. Assuming a locally dense supply of outlets in slant and a unimodal conditional mean for reliability captures key features of the actual news market on social media: users face a rich set of options differing in both ideology and reliability, and more moderate outlets have, on average, higher reliability scores than those at the extremes (see Appendix Figure A.1).

We next describe users' beliefs about reliability. We denote user i's prior about outlet j's reliability by r_{ij} and we let $r_{ij} \sim N(\mu_{ij}, 1)$. We assume that $\mu_{ij} = r_j - \pi|s_i - s_j|$ with $\pi > 0$. Thus, prior means are anchored at the truth r_j but decrease with ideological distance $|s_i - s_j|$. This captures, in a reduced form way, the Gentzkow and Shapiro (2006) intuition that slant serves as a proxy for reliability: an agent with ideological position s_i expects outlets with slant close to her position to have higher reliability.

With these beliefs in place, the user chooses which news pages to follow on the platform. To capture the insights of Mullainathan and Shleifer (2005) and Gentzkow and Shapiro (2006), we assume users receive negative utility γ from ideological distance and positive utility α from high reliability. Consequently, individual i's benefit of following outlet j is given by $b_{ij} = \alpha\hat{r}_{ij} - \gamma|s_i - s_j|$, where $\hat{r}_{ij}$ is user i's expectation of the reliability of outlet j and where $\alpha, \gamma > 0$.

On the cost side, following m pages entails a convex displacement cost $C_i(m)$ with $C_i(0) = 0$, $C_i(m+1) > C_i(m)$, and $C_i(m+1) - C_i(m) > C_i(m) - C_i(m-1)$ for $m \geq 1, m \in \mathbb{N}$. The fact that adding news pages is costly reflects the idea that news pages displace valued non-news content on a user's feed; the fact that those costs are convex aims to capture the idea that the marginal cost of crowding out non-news content increases in the amount of news content already present in one's feed. In other words, we assume that social media users prefer to have some balance between news and non-news content in their feeds.

User's problem and qualitative features of the solution. User i chooses a set of pages $A_i \subseteq J$ to maximize

$$\max_{A_i \subseteq J} \sum_{j \in A_i} b_{ij} - C_i(|A_i|).$$

The solution set $A_i^\star$ can be found by simply applying a greedy algorithm: the user ranks all outlets by their net benefit b_{ij} and adds pages sequentially until the marginal cost of adding an additional page exceeds its marginal benefit. At the optimum, the user's portfolio includes all outlets with b_{ij} above this cutoff. We denote the average slant of a user's portfolio by $\bar{s}(A_i^\star)$ and the average reliability of that portfolio by $\bar{r}(A_i^\star)$.

Two qualitative implications follow directly. First, because the menu of outlets is locally dense, the average slant of the user's portfolio aligns closely with her ideology, $\bar{s}(A_i^\star) \approx s_i$. This alignment arises through two reinforcing channels: the user gains psychological utility from like-minded outlets, and she perceives such outlets as more reliable. Second, since reliability is only imperfectly perceived, the average reliability of the chosen portfolio, $\bar{r}(A_i^\star)$, is lower than what she could achieve under full information.

The workhorse model—built on the core assumptions of Mullainathan and Shleifer (2005) and Gentzkow and Shapiro (2006)—is consistent with Facts 1 through 4: preferences for like-minded and high-reliability outlets, relatively accurate slant perceptions and relatively inaccurate reliability perceptions, a correlation between slant and reliability in users' beliefs, and pro-attitudinal news portfolios. Conversely, Facts 5 and 6, namely that participants' realized portfolios are systematically more slanted than what they report wanting and their self-reported passivity in portfolio formation, do not easily follow from the workhorse model and motivate the salience model of news consumption described below.

3.2 A salience model of news consumption

On social media, attention is often captured "bottom up" by the content that appears in one's feed rather than directed "top down" through deliberate search (e.g., typing a query into a search bar) (Bordalo, Gennaioli and Shleifer, 2022). This is consistent with the self-reports in Fact 6: users primarily follow news pages after encountering them through friend activity or algorithmic recommendations, not after intentionally searching for them. We therefore model news-page following as a choice shaped by platform cues and salience-driven behavioral frictions.

Setup. The primitives (users, outlets, beliefs, preferences, and costs) are as in Section 3.1. We add a minimal salience layer that affects both *when* choices occur and *which* outlets enter users'

consideration sets.

First, time is partitioned into decision episodes. In each episode, a binary prompt $P \in \{0,1\}$ indicates whether the platform surfaces a menu of news pages. If $P = 1$, the platform shows user i a salient set of pages $S_i \subseteq J$ and an active choice occurs. If $P = 0$, no menu is surfaced and the user does not revise her followed pages in that episode. In words, users adjust which pages they follow only when the platform prompts the decision.

Second, each user i is endowed with a salience vector $\sigma_i = (\sigma_{i1}, \ldots, \sigma_{iJ})$, where σ_{ij} measures how salient outlet j is to user i in that episode.[13] Salience reflects many factors we do not model explicitly—recent conversations, recent consumption, familiarity, and past experience—but, crucially, it is also shaped by prominence within one's social media feed, which makes some outlets more likely to be encountered and top-of-mind than others (Bordalo, Gennaioli and Shleifer, 2022).

When $P = 1$, only a subset of outlets enters the user's consideration set. For simplicity, we assume S_i consists of the k outlets with the highest salience parameter σ_{ij} (ties broken arbitrarily), with k small so selection is sharp.[14]

User's problem and qualitative features of the solution. In an episode with $P = 1$, user i chooses $A_i \subseteq S_i$ to maximize

$$\max_{A_i \subseteq S_i} \sum_{j \in A_i} b_{ij} \; - \; C_i(|A_i|). \tag{1}$$

When $P = 0$, no choice occurs and the portfolio is unchanged.

The solution mirrors the workhorse case but with the feasible set restricted to S_i. The user ranks outlets in S_i by b_{ij} and adds them sequentially until the marginal cost $C_i(|A_i|+1) - C_i(|A_i|)$ exceeds the marginal benefit of the next outlet, or until she exhausts the salient set. Equivalently, the chosen portfolio contains exactly those $j \in S_i$ with b_{ij} above a cutoff.

If, as suggested by recent evidence, misinformation spreads more quickly on social media than accurate news and extreme pro-attitudinal content is over-represented in users' feeds (Vosoughi, Roy and Aral, 2018; Braghieri et al., 2025), then platform prominence is likely to skew

[13] Since our model is static, we suppress episode indices. In a dynamic model, we would include them because salience and consideration sets may vary across episodes.

[14] All results are robust to a probabilistic consideration-set formulation (e.g., softmax) in which outlets enter S_i with probability increasing in σ_{ij}. The top-k rule we adopt corresponds the limit of the softmax function as the temperature parameter goes to zero.

most users' salience vector σ_i toward extreme, like-minded outlets.

Three patterns follow from the considerations above. First, chosen portfolios remain pro-attitudinal on average because S_i contains many like-minded outlets. Second, when k is sufficiently small, average portfolio slant shifts further in the direction of s_i relative to the full-menu benchmark, reflecting selection from a skewed consideration set. Third, average reliability remains below the full-information benchmark because users still have imperfect information about reliability.

Overall, the salience model reproduces the stylized facts explained by the workhorse model and, thanks to its additional salience assumptions, also accounts for Facts 5 and 6. This success is unsurprising, as we explicitly introduced the salience model to capture these additional patterns. The more substantive distinction between the two models lies in how they respond to interventions, as discussed below.

3.3 Models' Responses to Interventions

The workhorse and the salience models respond very differently to two natural interventions: changes to the choice architecture and the provision of targeted information about reliability.

Changes to the choice architecture. The first intervention where the two models yield sharply different predictions is a re-optimization intervention that triggers a decision episode by prompting users to re-optimize the set of news pages they follow on social media from a salient set of pages.

In the workhorse model, users are already optimizing over the full set of outlets. Since they are assumed to have stable preferences, a re-optimization prompt with an interface that surfaces a salient set of news pages but that does not provide additional information should not meaningfully affect their news portfolios.

In the salience model, in contrast, choice architecture is first-order. Users revise their portfolios only when a decision episode is triggered and only among outlets that enter their consideration sets. As a result, interventions that (i) trigger a decision episode by prompting re-optimization and (ii) make a balanced menu of outlets salient at the moment of choice are predicted to induce meaningful changes even in the absence of information provision. In terms of the model, a re-optimization interface that increases the salience of a set of news pages corresponds to triggering a decision episode $(P = 1)$ and to increasing σ_{ij} for the outlets surfaced in the re-optimization

interface, thereby changing the composition of the user's consideration set S_i.

Targeted information about reliability. The second intervention where the two models differ sharply is the provision of information about reliability.

In the workhorse model, the provision of reliability information directly shifts perceived outlet reliability and thus induces agents to update their news portfolios even if the platform does not explicitly prompt a re-optimization episode.

In the salience model, information by itself has little bite. If reliability information is delivered when users are not in a decision episode, users are predicted not to act on that information. Information influences news-following behavior only when it is paired with an active-choice prompt and a salient menu: in that case, reliability information affects the ranking of options within the consideration set, shifting portfolios toward higher-reliability sources. In terms of the model, providing information may increase σ_{ij} for the highlighted outlets, but—absent a salient re-optimization prompt—it does not trigger a decision episode ($P = 0$).

Experimental implications. The differential responses of the two models to changes in the choice architecture and to the provision of reliability information motivate our experimental design. First, to isolate the role of choice architecture, we implement a re-optimization intervention that triggers a decision episode and presents a salient menu of news outlets without providing any new information. Second, to isolate the role of information absent a triggered decision episode, we provide reliability information without a re-optimization prompt.

These interventions allow us to test the key differential predictions of the two models. Changes to the choice architecture without the provision of information are effective under the salience model but not under the workhorse model; conversely, the provision of reliability information is effective under the workhorse model but not under the salience model.

4. Experimental Design and Implementation

4.1 Experiment Overview

Figure 4 presents a schematic overview of the experimental design, which we summarize below. Screenshots of the experimental interface are presented in Appendix F.

We recruited participants for a five-week experiment via Facebook ads. To avoid attracting a sample with specific views about news consumption on social media, the ads did not disclose the nature of the study. Upon clicking an ad, users were shown a consent form explaining that participation required logging into the study via Facebook and granting the research team permission to access the list of pages they follow on the platform. Participants who accessed the baseline survey using the Google Chrome browser were offered a bonus to install a Chrome extension that tracked their visits to a large set of news websites, allowing us to observe off-platform online news consumption.

Participants who consented and logged in via Facebook were directed to a baseline survey. The survey collected demographic data, information about news consumption habits, and pre-treatment measures of various outcome variables.[15] The baseline survey also included detailed explanations of our slant and reliability measures, comprehension questions ensuring participants understood these measures, and incentivized questions to elicit participants' perceptions. The questions asked about the slant and reliability of the portfolio of news pages participants follow on the platform, as well as of major news outlets.[16] Finally, participants reported their "bliss points" regarding the ideal slant and reliability of their Facebook news portfolios.

Around one week after the baseline survey, participants were invited to complete a midline survey. During that survey, they were randomized into one of six study arms: a *Control* group, three *Re-optimization* treatment arms, a *Reliability Information* treatment (NR), and a *No Experi-*

[15] We excluded participants who met any of the following criteria: were not U.S. residents, accessed the survey via a VPN, did not follow any news pages on Facebook, reported never obtaining news from Facebook, did not grant us permission to observe the pages they follow on the platform or whose Facebook data we could not access (e.g., because the participant had a professional profile), failed comprehension checks twice regarding our definitions of slant and reliability, were flagged as potential scammers or duplicates, and requested data deletion. Appendix Figure A.4 provides more details on participants who were screened out before randomization.

[16] The measures were incentivized using a binarized scoring rule, where the probability of winning a fixed prize, independently drawn for each participant, increases in the accuracy of the participant's forecasts (Hossain and Okui, 2013).

menter Demand treatment (NED). All study arms other than the NED treatment are described in this section; the NED treatment—employed to check for potential experimenter demand effects—is described in Section 6. Randomization was stratified based on the average slant and reliability of participants' baseline news portfolios and based on whether they installed the Chrome extension.[17]

Participants in the control group received no intervention. We implemented three re-optimization treatments: Re-optimization (R), Re-optimization with Reliability Information (RR), and Re-optimization with Slant Information (RS). Participants assigned to any of the re-optimization treatment arms were shown a simple, user-friendly interface that allowed them to follow or unfollow news pages on Facebook with a single click. Specifically, participants in the re-optimization treatments were shown a list of news outlets and were informed that they could update their Facebook news portfolio by clicking the "follow" or "unfollow" button next to each outlet. The interface was integrated with the Facebook API, so that any changes made through the survey were immediately reflected in the participants' Facebook accounts.

The re-optimization interface was divided into two tables. The first table showed each participant 12 outlets they did not already follow and asked them whether they wanted to follow any of those outlets going forward. The table always included a mix of outlets that spanned the slant spectrum and had either medium or high reliability.[18] The second table showed each participant the news outlets they were already following and provided an option to unfollow any of them going forward.[19] We did not provide an incentive to follow or unfollow pages, nor did we encourage users to make changes to their portfolios. Appendix C provides more information about how we selected the news pages displayed in the re-optimization interface and Appendix F.2 shows screenshots of the interface.

We randomized the provision of additional information across the treatment arms. Participants assigned to the Re-optimization (R) treatment received no additional information. In the

[17] As specified in the pre-analysis plan, participants who installed the Chrome extension were randomized only into the control group or one of the three re-optimization treatments. The reason is that these participants were more expensive to recruit; thus, to maximize power, we assigned them only to the treatment arms that we expected to have an effect.

[18] We omitted low-reliability pages from the table because we did not want to spread misinformation. One might worry that doing so mechanically boosts reliability—for instance, because "random" choices from the re-optimization interface would then be drawn from a more reliable pool. Appendix Figure A.6 shows this is not the case; if anything, random selection from the interface reduces portfolio reliability on average.

[19] Participants were shown up to 12 news pages they already followed. If a participant's Facebook news portfolio contained more than 12 pages, we randomly selected 12 pages to display.

Re-optimization Reliability Info (RR) treatment, participants received information about the reliability of each news outlet they followed on Facebook, the average reliability of their current news portfolio on the platform, and the reliability of each outlet shown in the re-optimization interface. Participants in the Re-optimization Slant Info (RS) treatment received analogous information regarding the slant of news outlets.

Finally, guided by our model, we deployed the Reliability Info (NR) treatment in order to study the consequences of mere information provision. Participants in the NR treatment were not shown the re-optimization interface, but were given the same reliability information as participants in the RR treatment—both for the outlets they followed and for those they would have seen in the re-optimization interface had they gone through it.

Approximately four weeks after the midline survey, participants were invited to complete an endline survey. This final survey elicited secondary outcomes, including participants' satisfaction with their Facebook news feed and trust in the news they see on Facebook. At the end of the endline survey, participants were given another opportunity to re-optimize their news portfolios using the same interface and the same set of news pages. This allowed us to assess whether participants remained satisfied with the portfolio choices they made earlier in the experiment.

4.2 Pre-analysis Plan and Procedures

Pre-analysis Plan. We pre-registered the study by submitting a pre-analysis plan (Braghieri, Levy and Trachtman, 2025) in which we specified the sample size, experimental design, outcome variables—including a distinction between primary and secondary outcomes, as well as the construction of each outcome variable—and our main regression specifications. The only notable deviation from the pre-analysis plan is that we did not manage to recruit as large a sample as we had anticipated, due to higher-than-expected recruitment costs. Thus, the sample size in the experiment is approximately 70% of the one specified in the pre-analysis plan.

Procedures. We implemented the experiment in three separate but overlapping waves, beginning on January 14, 2025, and concluding in April 2025. Participants received $5 for completing the baseline survey and $8 for each of the midline and endline surveys. Those who installed the Chrome extension and kept it active for the full duration of the experiment received an additional $20.

Incentive payments for accurately answering the perception questions were disbursed at the end of the study. We also provided additional incentives for specific survey invitations to decrease attrition. Appendix D contains additional information about the implementation of the experiment.

4.3 Outcome Variables

Following our pre-analysis plan, we distinguish between primary and secondary outcome variables.

Primary Outcome Variables. We pre-specified the following as our primary outcomes:

- **Perceptions of slant and reliability.** Participants provided incentivized assessments of the average slant and reliability of the portfolio of news pages they follow on Facebook, as well as of various news outlets.
- **News-following behavior on Facebook.** Using data obtained directly from the Facebook API, we track the set of news pages each participant follows on Facebook at various points throughout the experiment. From this list, we compute the average reliability and the absolute average slant of each participant's news portfolio.
- **Distance to bliss points in slant–reliability space.** Participants reported their ideal combination of slant and reliability in a two-dimensional space. We calculate the distance between each participant's bliss point and the actual slant and reliability of her news portfolio. We construct both dimension-specific distances and a composite Euclidean distance after standardizing our measures of slant and reliability.
- **Off-platform news consumption.** For participants who installed the Chrome extension, we record the number of visits to a large set of news websites.
- **Self-reported news exposure.** We record participants' reports of the average slant and reliability of the news content they encountered on Facebook between the midline and endline surveys.

Secondary Outcome Variables. Our pre-specified secondary outcomes consist of survey-based measures of trust in news, satisfaction with one's news diet, and overall news consumption habits on Facebook.

We did not pre-specify downstream outcomes related to political attitudes or knowledge for two reasons. First, a core objective of this paper is to understand the drivers of news customization decisions in platform-mediated environments, where "bottom-up" attention and salience-driven behavioral frictions might distort choice. Doing so requires a multi-arm design that varies both choice architecture and information provision; as a result, the study was neither designed nor powered to detect changes in downstream political outcomes. Second, a large literature has already documented that media exposure (including social media) can affect downstream political outcomes (e.g. Adena et al., 2015; Allcott et al., 2020; Altay, Hoes and Wojcieszak, 2025; Broockman and Kalla, 2025; Bursztyn et al., 2023; DellaVigna and Kaplan, 2007; Chen and Yang, 2019; Djourelova, 2023; Enikolopov, Petrova and Zhuravskaya, 2011; Fujiwara, Müller and Schwarz, 2024; Levy, 2021; Martin and Yurukoglu, 2017) and, as we show in Section 5.5, our intervention affects media exposure.

4.4 Summary Statistics

Table 1 summarizes the number of participants at each stage of the experiment. We also report the number of participants, at each stage, for whom we successfully collect the set of pages they follow on Facebook from the platform's API.[20]

Appendix Table A.6 compares the demographic characteristics of our sample to those of the broader Facebook and U.S. populations. Relative to the average Facebook user and the average U.S. adult, our sample is more likely to be female, white, older than fifty, and college-educated. Our sample is also more likely to identify as Democrat compared to the overall U.S. adult population.[21] Appendix Table A.7 shows randomization achieved balance along the expected number of baseline observable characteristics.

[20] We do not observe the set of followed pages on Facebook for all participants at all points in the experiment because participants can revoke these permissions and because, in a small fraction of cases, our calls to the Facebook API fail. We provide more details in Appendix D.

[21] In Appendix Tables A.9-A.10 we show that our results are robust to weighting the sample to mimic the US population on age, gender, education, race, ethnicity, and party affiliation. There could, of course, remain unobservable differences between our experimental sample and the overall Facebook population.

Appendix Table A.8 reports attrition rates across treatment arms. Attrition between midline, where randomization occurs, and endline was minor—approximately 15%. Only one out of our four main treatment arms, namely the R treatment, exhibits evidence of differential attrition at the endline survey. Such differential attrition is not a concern for our primary outcome, namely the set of pages participants follow on Facebook right after the midline survey, because that outcome is not subject to survey attrition.[22]

4.5 Main Empirical Strategy

To evaluate the effects of our interventions, we estimate the impact of our treatments on each outcome variable from Section 4.3. We estimate the following OLS regression:

$$Y_i = \alpha + \sum_{j \in J} \beta_j T_i^j + \rho \cdot Y_i^b + \eta \cdot \mathbf{X_i} + \mu_s + \varepsilon_i \tag{2}$$

where Y_i is an outcome variable, Y_i^b is the baseline value of Y_i (when available), $T_i^j \in \{0,1\}$ is an indicator for assignment to treatment arm $j \in J = \{R, RS, RR, NR\}$ (with the Control group as the reference group), μ_s denote recruitment and randomization strata fixed effects,[23] ε_i is an idiosyncratic error term, and $\mathbf{X_i}$ is a vector of pre-specified baseline covariates: political ideology, reliance on Facebook for news, interest in politics, recruitment wave fixed effects, and device fixed effects (desktop vs. mobile). When estimating Equation 2, we use robust standard errors.

5. Results

We present our experimental results in five subsections. First, we show that the evidence is more consistent with the salience model of news following on Facebook than with the workhorse model. Second, we show that interventions reduce the gap between participants' Facebook news portfo-

[22] In Section 5.3 we show that the effects on our primary outcomes collected in the midline survey, where no attrition occurs, are remarkably similar to the primary outcomes collected in the endline survey, further allaying concerns about attrition at endline. Moreover, Section 5.3 also shows that all the re-optimization treatments had a similar effect on slant, and Appendix Table A.8 does not find evidence for differential attrition when pooling these treatments. Finally, Appendix Table A.11 shows that balance on observable characteristics is largely preserved at endline.

[23] As described in Section 4.1, these include average slant and reliability of baseline news portfolios and an indicator for installing the browser extension.

lios and their stated bliss points, thereby mitigating internalities. Third, we examine whether the intervention reduces portfolio slant and increases average reliability—two dimensions central to concerns about the potential negative externalities of social-media news consumption for democracy. Fourth, we study effects on secondary outcomes—trust in and satisfaction with the news in users' feeds. Fifth, leveraging participants who installed our Chrome extension, we document that changes in users' news portfolios affected downstream news consumption.

5.1 Evidence in Favor of the Salience Model

Re-optimization (R) Treatment The workhorse model predicts that our re-optimization treatment—which simply offers participants a user-friendly interface to update the set of news pages they follow on Facebook—should have no effect. The intuition is straightforward: if the transaction costs of following or unfollowing pages are negligible, then participants should already be close to their optimal portfolios.[24] The salience model yields the opposite prediction. Because choices are made only when a decision episode is triggered and only from outlets that are salient at that moment, a re-optimization prompt should be effective: it both triggers a decision episode and presents a curated set of pages that differs from the algorithmic and friend-based suggestions users typically encounter.

The results align with the salience model. As shown in the first panel of Appendix Figure A.5, the vast majority (71%) of participants assigned to the R treatment adjusted their portfolios by following or unfollowing at least one news page, despite receiving no new information.[25]

Beyond predicting that the R treatment induces participants to make changes to their Facebook news portfolios, the salience model also disciplines how those changes occur. In particular, it makes two additional predictions. First, because the interface makes a particular menu of pages salient, participants should primarily re-optimize by selecting outlets from within the salient menu.

[24]Following a page typically requires typing its name into the Facebook search bar and clicking "Follow." Unfollowing is even easier: users can click "Unfollow" directly from any post that appears in their feed.

[25]Strictly speaking, our experiment identifies the joint effect of behavioral frictions and the minimal transaction cost of following or unfollowing a news page. We interpret the results as evidence of behavioral frictions because the implied transaction costs are extremely small. We estimate that following a page takes around 3 seconds. Participants were willing to take a survey for a compensation equivalent to a $30/hour wage. Thus, an implied upper bound on their time cost of following a news page outside of the re-optimization interface is about $0.025. As shown in the next few sections, following a news page affects both what users see on Facebook and the news sites they visit over time. In order for transaction costs alone to explain our re-optimization results, the total value of these effects to participants would have to be less than 2.5 cents.

Second, within that menu, participants should not choose outlets at random; instead, they should select outlets in ways that reflect their preferences.

Both predictions hold in the data. First, in the four weeks after the midline survey, only 6.3% of participants in the R treatment add to their portfolios news outlets that were not included in the re-optimization interface. In other words, nearly all of participants in the R group who make changes to their portfolios do so from within the set of pages made salient by the interface.

Second, participants' choices within the interface are systematic rather than random. To test this, we simulate counterfactual portfolios under random choice.[26] Appendix Figure A.6 panel (a) reports simulated effects on absolute average slant and panel (b) reports simulated effects on reliability. The gray bars show the distribution of average treatment effects obtained under random choice. In both panels, the actual average treatment effect of the R treatment lies far in the tails of this distribution, leading us to reject the hypothesis that participants simply choose randomly from the re-optimization interface. Panel (a) also reports a distribution generated by simulating random choice from among pro-attitudinal outlets; we reject this version of random choice as well.[27] Consistent with this interpretation, we show in Section 5.2 that re-optimization treatments reduce the distance between participants' portfolios and their stated bliss points, indicating that participants use the interface to move portfolios in directions consistent with their stated preferences.

No-Reoptimization Reliability Information (NR) Treatment The workhorse model also predicts that providing participants with outlet- and portfolio-level information they value—such as reliability information—should induce re-optimization even without an explicit prompt. The salience model again reverses this prediction: information delivered outside a decision episode should have little effect, because users will not act on it unless they are simultaneously prompted to adjust their portfolios and given a salient set of options.

The evidence again favors the salience model. By the end of the experiment, participants assigned to the NR treatment added, on average, only 0.09 news pages to their Facebook news portfolios, compared to 1.90 in the R treatment. Put differently, a prompt that triggers re-optimization

[26] For each participant in the R treatment, we randomly draw from the offered choice set the same number of outlets the participant actually followed via the interface. We then compute the resulting portfolio slant and reliability, estimate Equation 2, and repeat the procedure one thousand times.

[27] The fact that the R treatment differs from the pro-attitudinal random-choice benchmark also illustrates that the balanced menu in the re-optimization interface does not mechanically force moderation.

without providing any new information induces portfolio changes roughly twenty times larger than an intervention that provides targeted reliability information without a re-optimization interface. Consistent with this, and as shown in Section 5.3, the NR treatment does not increase the average reliability of participants' Facebook news portfolios relative to the control group.

A natural question is whether NR is ineffective because participants do not value the information, or because they value it but fail to act on it in the absence of a prompted decision episode and a salient menu. Only the latter interpretation is consistent with the salience model. We can rule out the former interpretation for two reasons. First, as documented in Section 2—and as assumed in canonical models of news demand (Gentzkow and Shapiro, 2006)—participants substantially misperceive outlet reliability, so information about reliability is plausibly useful. Second, and more importantly, when the same reliability information is paired with the re-optimization interface in the RR treatment, it meaningfully raises average portfolio reliability relative to the control group as shown in Section 5.3. Together, these findings indicate that information alone is not sufficient to change behavior: it becomes behaviorally relevant only when the same set of pages is made salient in the re-optimization interface.

Overall, the pattern of results from the R and NR treatments aligns more closely with the salience model than with the workhorse model, highlighting the role of salience-driven behavioral frictions in news-following decisions on social media.

5.2 Internalities and Distance to Stated Bliss Points

Salience-driven behavioral frictions might drive a wedge between users' preferences and their actual portfolios, generating internalities. On top of that, limited information about outlet reliability may lead users to hold portfolios with lower-than-desired reliability. Our interventions target both the internality generated by salience-driven behavioral frictions and the lack of information about reliability; therefore, they have the potential to increase alignment between participants' realized portfolios and their stated bliss points.

To quantify alignment, we construct distance measures between each participant's portfolio and her bliss point in the slant–reliability space.[28] Along each dimension, we compute the absolute gap between the average slant (reliability) of the news pages the participant follows on Facebook

[28] We elicited bliss-points in the baseline survey before the intervention.

and her corresponding slant (reliability) bliss point. We also construct a composite measure: the Euclidean distance between the portfolio and the bliss point in two dimensions, after standardizing slant and reliability relative to the control-group distribution.

We find evidence that, indeed, our interventions reduce the gap between participants' Facebook news portfolios and their stated bliss points. Table 2 reports treatment effects using our main specification (Equation 2), with the slant distance, reliability distance, and Euclidean distance as outcomes. All three re-optimization treatments significantly reduce distance to the slant bliss point. In addition, the R and RR treatments significantly reduce distance on the reliability dimension. Together, these results suggest that our intervention—by addressing behavioral frictions and, when applicable, providing targeted reliability information—helps participants bring their Facebook news portfolios closer to the portfolios they themselves report wanting.

5.3 Externalities

We next ask whether our interventions move participants' Facebook news portfolios along the two dimensions that are central to concerns about news consumption on social media producing negative externalities on democracy (Sunstein, 2017): slant and reliability.

Slant Table 3 reports treatment effects on the absolute average slant of participants' Facebook news portfolios.[29] Relative to the control group, all three re-optimization treatments significantly reduce portfolio slant. The effects appear immediately after the midline survey—when participants complete the re-optimization task—and persist through the endline survey roughly four weeks later (columns (1) and (2), respectively). In contrast, the NR treatment has no detectable effect. Because the post-midline effect sizes are very similar across the three re-optimization arms, we streamline the exposition by pooling them in what follows; the pooled regression appears in column (1) of Appendix Table A.12.

The magnitude of the slant effects is substantial. The re-optimization treatments decreased absolute average slant by 0.24 standard deviation units—equivalent to 51% of the baseline gap between self-identified independents and partisans in our sample. As an alternative benchmark, the

[29]The treatment effect estimates in this section are obtained from Equation 2 and weight each observation equally. Appendix Tables A.9-A.10 show similar results when estimating population average treatment effects using weights that align the sample with the U.S. population on age, gender, race, education, ethnicity, and party affiliation.

estimate corresponds to more than half the distance between the slant of The Washington Post (a liberal outlet) and that of Bloomberg (a centrist outlet). Appendix Table A.13 shows that the effect applies to both Democrats and Republicans, as both groups reduce portfolio slant in response to our intervention.

A reduction in absolute average slant could arise through two channels: (i) greater exposure to centrist outlets (low slant in absolute value), or (ii) greater exposure to counter-attitudinal outlets (outlets whose ideological orientation differs from the participant's own). Appendix Table A.12 points to the second channel. Column (2) shows no statistically significant decline in average absolute slant (as opposed to absolute average slant), suggesting that increased emphasis on centrist outlets is not the primary driver. Column (3), in contrast, shows that the share of counter-attitudinal pages in a participant's portfolio more than triples under the pooled re-optimization treatments. Therefore, treated participants do not simply shift toward more neutral content; they diversify their news portfolios by adding outlets from the other side of the ideological spectrum.[30] Columns (3)–(6) of Appendix Table A.13 show that the diversification effects apply to both Democrats and Republicans.

Figure 5 provides a complementary view by focusing on participants in the pooled re-optimization treatments and splitting them by baseline portfolio type: left-leaning (average slant < -0.2), moderate ($-0.2 \leq$ slant ≤ 0.2), and right-leaning (average slant > 0.2). For each group, we plot desired portfolio slant, perceived portfolio slant, actual baseline slant, and post-treatment slant. Three patterns stand out. First, participants with partisan portfolios perceive them as more moderate than they are, indicating some mild slant misperceptions. Second, these participants nonetheless report wanting portfolios that are even more moderate than what they perceive, consistent with the presence of salience-driven behavioral frictions. Third, the re-optimization interface moves partisan portfolios toward these more moderate bliss points, in line with the results shown in Section 5.2.

Reliability Table 4 reports treatment effects on the average reliability of participants' Facebook news portfolios. The Re-optimization with Reliability Information (RR) treatment produces a large and statistically significant increase in portfolio reliability. As with slant, the effect emerges im-

[30] To the best of our knowledge, a taste for counter-attitudinal exposure is not typically incorporated into canonical models of news consumption. It is straightforward, however, to enrich the primitives in Section 3 to allow for a small demand for ideological diversification alongside a primary preference for like-minded news.

mediately at midline—when re-optimization occurs—and persists through endline. In contrast, the Re-optimization (R) and Re-optimization with Slant Information (RS) treatments generate much smaller, statistically insignificant changes in reliability.

The comparison between RR and NR further underscores the importance of pairing information with an active-choice moment. Participants in the NR arm receive the same reliability information as those in RR, but without the re-optimization interface. By the end of the experiment, RR participants added, on average, 2.00 pages to their Facebook news portfolios, whereas NR participants added only 0.09. Consistent with this limited re-optimization, the last row of Table 4 shows that the NR treatment does not significantly increase portfolio reliability relative to the control group.[31]

The magnitude of the RR effect is sizable. Relative to the control group, RR increases average portfolio reliability by 0.21 standard deviation units. This corresponds to closing 85% of the baseline reliability gap between college-educated and non-college-educated participants. As an additional benchmark, the average improvement is roughly comparable to adding The New York Times to a portfolio consisting of The Huffington Post, Jezebel, and Slate.

Relation between Treatment Effects and Misperceptions The pattern of treatment effects mirrors the asymmetry in participants' baseline beliefs. As shown in Section 2, participants perceive outlet slant relatively accurately but substantially misperceive reliability. Consistent with this, re-optimization alone is sufficient to reduce portfolio slant, whereas improving reliability requires coupling re-optimization with targeted reliability information.

5.4 Secondary Outcomes

Appendix Table A.14 examines two pre-specified survey outcomes using our main specification: satisfaction with the news participants see on Facebook (column 1) and trust in that news (column 2). We find that the RR treatment significantly increases satisfaction with the news participants encounter on Facebook and so do the pooled re-optimization treatments, albeit with a p-value

[31] These findings echo Aslett et al. (2022), who study a browser-extension intervention that displayed NewsGuard ratings and find no discernible effect on news consumption. Our results reinforce this conclusion—information alone has limited power to shift behavior—and also suggest a solution: coupling the provision of information with a decision prompt.

that straddles significance at the 5% level (p-value=0.052). Column (2) shows that the pooled re-optimization treatments significantly increase trust in the news participants see on Facebook.

Appendix Table A.15 reports analogous regressions for participants' perceptions of their Facebook feeds: perceived slant (column 1) and perceived reliability (column 2). The RR treatment increases perceived feed reliability, consistent both with belief updating among participants who initially underestimate reliability and with the observed shift toward higher-reliability portfolios under RR. In contrast, we find no significant effects on perceived slant. This may reflect limited sensitivity of our survey measure: the slant question is relatively coarse and may not capture changes in absolute average slant even when portfolio composition shifts in that direction.

5.5 Downstream News Consumption

We have shown that our intervention induces meaningful changes in the set of news pages participants follow on Facebook. We now ask whether these portfolio changes translate into changes in actual news consumption. To study whether following an outlet's Facebook page increases consumption of content from that outlet, we follow our pre-analysis plan and implement an instrumental variables (IV) strategy, using treatment assignment as an instrument for following behavior.

IV Specification We consider the following regression equation, where each observation is a participant-page:

$$Y_{i,k} = \phi + \tau \cdot F_{i,k} + \psi \cdot Y_{i,k}^{b} + \chi \cdot \mathbf{X_i} + \mu_s + \upsilon_{i,k} \tag{3}$$

where $Y_{i,k}$ is an outcome that relates to participant i and outlet k, $Y_{i,k}^{b}$ is the baseline value of $Y_{i,k}$ (when available), and $F_{i,k}$ is an indicator for whether participant i follows outlet k on Facebook. We instrument for $F_{i,k}$ using an indicator that takes value one if the participant was assigned to any of the three re-optimization treatments (R, RR, RS), and zero otherwise. We include the same set of controls $\mathbf{X_i}$, and μ_s, as in our main specification and cluster standard errors at the participant level.

To define the set of outlet–participant pairs (i,k) used in the IV specification, we construct, for each participant who completed the midline survey—including those in the control group—the set of outlets they would have been shown had they been assigned to one of the three re-optimization

treatments. We refer to these as "potentially offered outlets." Importantly, these outlets exclude outlets that the participant already followed at baseline.[32]

IV Results As a first check, Appendix Table A.16 shows that participants who follow a news page as a result of treatment report seeing more posts from that page in their Facebook feeds at endline. This shows that following behavior has noticeable consequences for on-platform exposure.

To examine off-platform news consumption, we use browsing data from the subset of participants who installed our Chrome extension. The extension tracks browsing only on the device where it is installed (a desktop or laptop computer). Because much online news consumption occurs on mobile devices, the estimates in this section should therefore be interpreted as lower bounds on total changes in consumption.

Table 5 shows that following an outlet's Facebook page as a result of assignment to one of our re-optimization treatments increases visits to that outlet's website in the month after the midline survey. We estimate the causal effect of following on visits using the instrumental-variables specification in Equation 3, pooling participants assigned to the three re-optimization treatments. Column (2) shows that following a page increases the probability of a Facebook-originating visit by 4.02 percentage points, and column (4) shows that it more than doubles the overall probability of a visit (an increase of 10.55 percentage points). The effects operate on both the extensive margin and the intensive margin: total visits also increase, particularly visits originating from Facebook (columns (3) and (5)).[33]

Table A.17 reports reduced form effects of the pooled re-optimization treatments on visits. Unlike the IV specification, these estimates rely only on random assignment and do not require the monotonicity and exclusion-restriction assumptions. Taken together, the results from this section provide direct evidence that changes in Facebook following behavior generate meaningful shifts in off-platform news consumption.

[32] One might be worried about possible violations of monotonicity and of the exclusion restriction. To assuage those concerns, we also present reduced-form effects.

[33] As pre-specified, we also examine the slant and reliability of the websites participants visited in the month following the intervention. Since the subset of participants for whom we observe browsing data is small, we have limited statistical power. Table A.18 presents these results and shows suggestive evidence that the reoptimization treatments increased visits to counter-attitudinal outlets; in light of the limited power, however, we refrain from drawing strong conclusions.

6. Mechanisms and Robustness

6.1 Experimenter Demand

Because our interventions—especially RR—are relatively transparent, a natural concern is experimenter demand: participants may adjust their behavior to align with what they believe the research team wants to find. Our downstream consumption results make this explanation prima facie less compelling. In particular, it seems unlikely that participants would incur meaningful time costs over the course of a month to visit specific news sites simply to conform to perceived experimental aims.

To directly test for experimenter demand effects, we implemented an additional arm that closely mirrors the Re-optimization with Reliability Information (RR) treatment but removes perceived and actual experimenter oversight. In the No Experimenter Demand (NED) treatment, participants were explicitly told during the midline survey that we would no longer observe the pages they followed on Facebook. To reinforce this, we revoked the Facebook permissions participants had granted us at baseline. Participants could remove the permissions themselves using instructions we provided or authorize us to revoke them automatically. Participants' knowledge that we no longer had access to their page-following data from that point forward substantially reduces the scope for demand-driven behavior during the midline survey.

At endline, we surprised participants by re-requesting permission to observe their followed pages. Nearly all participants (99%) who took the endline survey consented. This design allows us to compare endline portfolios between the NED and RR groups and thereby gauge how much of the RR effect could plausibly be attributed to experimenter demand.

Appendix Table A.19 reports the results. Like RR, the NED treatment significantly increases the average reliability of participants' Facebook news portfolios. The point estimate under NED is somewhat smaller than under RR, but the difference is not statistically significant at conventional levels. Thus, we can be confident that our results are not primarily driven by experimenter demand effects.

6.2 Experimentation Motives

Throughout the paper, we interpret the high rate of re-optimization in the R treatment as evidence that salience-driven behavioral frictions shape participants' news-following decisions on Facebook. A natural alternative explanation is experimentation: participants may have temporarily followed outlets out of curiosity, rather than to correct misalignment between their portfolios and their preferences.

Several pieces of evidence weigh against the experimentation hypothesis. First, a substantial share of participants (20%) unfollowed at least one news page during the midline survey. This behavior is difficult to reconcile with experimentation, which typically entails sampling new options rather than disengaging from familiar ones. Second, among participants who followed at least one new outlet at midline, 86% reported being familiar with at least one of the outlets they chose to follow, suggesting that many additions reflect deliberate choices among familiar options rather than exploratory clicks. Third, experimentation could in principle improve portfolio alignment with stated bliss points in the long run if participants "try out" pages and then keep only the ones they like. However, this mechanism does not predict the sharp, immediate reduction in distance to bliss points we observe right after midline, when portfolio changes occur (Section 5.2). Fourth, if midline changes were primarily trial runs, we would expect substantial churn by endline. In fact, we make it easy for participants in the endline survey to unfollow pages they started following at midline by offering them another opportunity to revise their Facebook news portfolios using the same interface and the same set of pages. Instead of substantial churn, we find that only 5% of pages followed at midline are unfollowed by endline. This low reversal rate is hard to square with a transient experimentation story.

Taken together, these patterns suggest that experimentation is unlikely to be the main driver of re-optimization in the R treatment and support our interpretation that the intervention operates primarily by reducing salience-driven behavioral frictions.

7. Conclusion

In this paper, we showed that salience-driven behavioral frictions strongly shape which news pages users follow on social media. We furthermore showed that interventions that target these frictions

and reveal information move users' Facebook news portfolios closer to their stated bliss points in the slant–reliability space, reduce portfolio slant, and increase portfolio reliability.

These results speak directly to questions of external validity and policy relevance. The intervention is both cheap and scalable: it does not rely on algorithmic overhauls, regulatory enforcement, or heavy-handed content moderation. Instead, it mirrors the kind of nudges already deployed on platforms (e.g., privacy check-ups or account security prompts) and could be implemented widely with minimal cost. The fact that our intervention produces measurable behavior change in a real-world setting offers strong grounds for optimism about external validity.

More broadly, our results highlight a general lesson about platform-mediated environments: users' choices are shaped not only by their preferences and the options available, but also by when platforms trigger decision episodes and which options they make salient in those moments. Hence, salience-driven behavioral frictions can drive a systematic wedge between users' choices and their preferences. Consistent with this mechanism, information that users value may have little effect when delivered outside a decision episode, whereas small changes to the choice architecture can produce large and persistent shifts in behavior. This suggests that, in platform-mediated environments, users and stakeholders should pay close attention to when platforms prompt actions and to the menu of options they surface in those moments.

References

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya. 2015. "Radio and the Rise of the Nazis in Prewar Germany." *The Quarterly Journal of Economics*, 130(4): 1885–1939.

Allcott, Hunt, and Todd Rogers. 2014. "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." *American Economic Review*, 104(10): 3003–37.

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

Allen, Jennifer, Antonio A Arechar, Gordon Pennycook, and David G Rand. 2021. "Scaling Up Fact-Checking Using the Wisdom of Crowds." *Science Advances*, 7(36): eabf4393.

Altay, Sacha, Emma Hoes, and Magdalena Wojcieszak. 2025. "Following news on social media boosts knowledge, belief accuracy and trust." *Nature Human Behaviour*.

American National Election Study. 2024. "American National Election Study."

Aslett, Kevin, Andrew M Guess, Richard Bonneau, Jonathan Nagler, and Joshua A Tucker. 2022. "News credibility labels have limited average effects on news diet quality and fail to reduce misperceptions." *Science Advances*, 8(18).

Bakshy, Eytan, Solomon Messing, and Lada A Adamic. 2015. "Exposure to Ideologically Diverse News and Opinion on Facebook." *Science*, 348(6239): 1130–1132.

Bhadani, Saumya, Shun Yamaya, Alessandro Flammini, Filippo Menczer, Giovanni Luca Ciampaglia, and Brendan Nyhan. 2022. "Political Audience Diversity and News Reliability in Algorithmic Ranking." *Nature Human Behaviour*, 6(4): 495–505.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2013. "Salience and Consumer Choice." *Journal of Political Economy*, 121(5): 803–843.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2022. "Salience." *Annual Review of Economics*, 14(1): 521–544.

Braghieri, Luca, Ro'ee Levy, and Hannah Trachtman. 2025. "Demand for Online News, Inertia, and Misperceptions." AEA RCT Registry. https://doi.org/10.1257/rct.14392.

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

Broockman, David E, and Joshua L Kalla. 2025. "Consuming cross-cutting media causes learning and moderates attitudes: A field experiment with Fox News viewers." *The Journal of Politics*, 87(1): 246–261.

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

Busse, Meghan R., Nicola Lacetera, Devin G. Pope, Jorge Silva-Risso, and Justin R. Sydnor. 2013. "Estimating the Effect of Salience in Wholesale and Retail Car Markets." *American Economic Review*, 103(3).

Carroll, Gabriel D, James J Choi, David Laibson, Brigitte C Madrian, and Andrew Metrick. 2009. "Optimal Defaults and Active Decisions." *The Quarterly Journal of Economics*, 124(4): 1639–1674.

Chan, Jimmy, and Wing Suen. 2008. "A Spatial Theory of News Consumption and Electoral Competition." *The Review of Economic Studies*, 75(3): 699–728.

Chen, Yuyu, and David Y Yang. 2019. "The impact of media censorship: 1984 or brave new world?" *American economic review*, 109(6): 2294–2332.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review*, 99(4): 1145–77.

Chmel, Kirill, Eunji Kim, John Marshall, Tiffany Fisher-Love, and Nathaniel Lubin. 2025. "How social media creators shape mass politics: A field experiment during the 2024 US elections." *arXiv preprint arXiv:2512.15401*.

Chopra, Felix, Ingar Haaland, and Christopher Roth. 2022. "Do People Demand Fact-Checked News? Evidence from US Democrats." *Journal of Public Economics*, 205: 104549.

Chopra, Felix, Ingar Haaland, and Christopher Roth. 2023. "The Demand for News: Accuracy Concerns versus Belief Confirmation Motives." *CEPR Discussion Paper No. DP17169*.

Chopra, Felix, Ingar Haaland, Fabian Roeben, Christopher Roth, and Vanessa Sticher. 2025. "News Customization with AI." *Working Paper*.

Chouaki, Salim, Abhijnan Chakraborty, Oana Goga, and Savvas Zannettou. 2024. "What News Do People Get on Social Media? Analyzing Exposure and Consumption of News through Data Donations." *WWW '24*, 2371–2382. New York, NY, USA:Association for Computing Machinery.

Comscore. 2022. *Web Behavior Database Panel 2022*. Wharton Research Data Services, University of Pennsylvania.

DellaVigna, Stefano, and Ethan Kaplan. 2007. "The Fox News effect: Media bias and voting." *The Quarterly Journal of Economics*, 122(3): 1187–1234.

de Quidt, Jonathan, Johannes Haushofer, and Christopher Roth. 2018. "Measuring and Bounding Experimenter Demand." *American Economic Review*, 108(11).

Dhar, Diva, Tarun Jain, and Seema Jayachandran. 2022. "Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India." *American Economic Review*, 112(3): 899–927.

Djourelova, Milena. 2023. "Persuasion through slanted language: Evidence from the media coverage of immigration." *American economic review*, 113(3): 800–835.

Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya. 2011. "Media and political persuasion: Evidence from Russia." *American economic review*, 101(7): 3253–3285.

Fujiwara, Thomas, Karsten Müller, and Carlo Schwarz. 2024. "The effect of social media on elections: Evidence from the united states." *Journal of the European Economic Association*, 22(3): 1495–1539.

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

Gentzkow, Matthew, and Jesse M Shapiro. 2010. "What Drives Media Slant? Evidence from U.S. Daily Newspapers." *Econometrica*, 78(1): 35–71.

Gentzkow, Matthew, and Jesse M Shapiro. 2011. "Ideological Segregation Online and Offline." *The Quarterly Journal of Economics*, 126(4): 1799–1839.

Gentzkow, Matthew, Jesse M Shapiro, and Daniel F Stone. 2015. "Media Bias in the Marketplace: Theory." In *Handbook of Media Economics*. Vol. 1, 623–645. Elsevier.

González-Bailón, Sandra, David Lazer, Pablo Barberá, Meiqing Zhang, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Deen Freelon, Matthew Gentzkow, Andrew M Guess,

Shanto Iyengar, Young Mie Kim, Neil Malhotra, Devra Moehler, Brendan Nyhan, Jennifer Pan, Carlos Velasco Rivera, Jaime Settle, Emily Thorson, Rebekah Tromble, Arjun Wilkins, Magdalena Wojcieszak, Chad Kiewiet de Jonge, Annie Franco, Winter Mason, Natalie Jomini Stroud, and Joshua A Tucker. 2023. "Asymmetric ideological segregation in exposure to political news on Facebook." *Science*, 381(6656): 392–398.

Guess, Andrew M. 2021. "(Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets." *American Journal of Political Science*, 65(4): 1007–1022.

Guriev, Sergei, Emeric Henry, Théo Marquis, and Ekaterina Zhuravskaya. 2025. "Curtailing False News, Amplifying Truth." *Working Paper*.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart. 2023. "Designing Information Provision Experiments." *Journal of Economic Literature*, 61(1).

Hainmueller, Jens. 2012. "Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies." *Political analysis*, 20(1): 25–46.

Henry, Emeric, Ekaterina Zhuravskaya, and Sergei Guriev. 2022. "Checking and Sharing Alt-Facts." *American Economic Journal: Economic Policy*, 14(3).

Hossain, Tanjim, and Ryo Okui. 2013. "The Binarized Scoring Rule." *Review of Economic Studies*, 80(3).

Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman. 2016. "Getting to the Top of Mind: How Reminders Increase Saving." *Management Science*, 62(12): 3393–3411.

Keller, Punam Anand, Bari Harlam, George Loewenstein, and Kevin G Volpp. 2011. "Enhanced active choice: A new method to motivate behavior change." *Journal of Consumer Psychology*, 21(4): 376–383.

Kroft, Kory, Jean-William Laliberté, René Leal-Vizcaíno, and Matthew J. Notowidigdo. 2024. "Salience and Taxation with Imperfect Competition." *Review of Economic Studies*, 91(1).

Levy, Ro'ee. 2021. "Social Media, News Consumption, and Polarization: Evidence from a Field Experiment." *American Economic Review*, 111(3): 831–870.

Lin, Hause, Jana Lasser, Stephan Lewandowsky, Rocky Cole, Andrew Gully, David G Rand, and Gordon Pennycook. 2023. "High level of correspondence across different news domain quality rating sets." *PNAS Nexus*, 2(9): pgad286.

Martel, Cameron, Jennifer Allen, Gordon Pennycook, and David G Rand. 2024. "Crowds can effectively identify misinformation at scale." *Perspectives on Psychological Science*, 19(2): 477–488.

Martin, Gregory J, and Ali Yurukoglu. 2017. "Bias in cable news: Persuasion and polarization." *American Economic Review*, 107(9): 2565–2599.

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

Mummolo, Jonathan, and Erik Peterson. 2019. "Demand effects in survey experiments: An empirical assessment." *American Political Science Review*, 113(2): 517–529.

Newman, Nic, Amy Ross Arguedas, Craig T. Robertson, Rasmus Klein Nielson, and Richard Fletcher. 2025. "Reuters Institute Digital News Report 2025."

Pennycook, Gordon, and David G Rand. 2021. "The Psychology of Fake News." *Trends in Cognitive Sciences*, 25(5): 388–402.

Peterson, Erik, Sharad Goel, and Shanto Iyengar. 2021. "Partisan Selective Exposure in Online News Consumption: Evidence from the 2016 Presidential Campaign." *Political Science Research and Methods*, 9(2): 242–258.

Pew Research Center. 2024. "Social Media and News Factsheet." *Pew Research Center*.

Rathje, Steve, Clara Pretus, James He, Trisha Harjani, Jon Roozenbeek, Kurt Gray, Sander van der Linden, and Jay Van Bavel. 2024. "Unfollowing hyperpartisan social media influencers durably reduces out-party animosity."

Suen, Wing. 2004. "The Self-Perpetuation of Biased Beliefs." *The Economic Journal*, 114(495): 377–396.

Sunstein, Cass R. 2017. *#Republic: Divided Democracy in the Age of Social Media.* Princeton University Press.

Taubinsky, Dmitry, and Alex Rees-Jones. 2018. "Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment." *Review of Economic Studies*, 85(4).

Thaler, Richard H, and Cass R Sunstein. 2008. "Nudge: Improving decisions about health, wealth, and happiness."

U.S. Census Bureau. 2024. "American Community Survey."

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. "The spread of true and false news online." *Science*, 359(6380): 1146–1151.

Zizzo, Daniel John. 2010. "Experimenter demand effects in economic experiments." *Experimental Economics*, 13: 75–98.

Tables and Figures

Table 1: Sample Sizes

Phase	Sample Size Description
Baseline	N = 3,356 consented to participate and completed the baseline survey.
Midline	N = 2,184 participated in the midline survey and were randomized, of which: N = 1,874 were assigned to the control group, the re-optimization treatments (R, RS, and RR), or the Reliability Info treatment (NR). N = 1,866 were in the **midline impact evaluation sample**, which means we successfully collected the set of pages they followed on Facebook at midline. N = 310 were assigned to the No-Experimenter Demand treatment (NED).
Endline	N = 1,862 participated in the endline survey, of which: N = 1,610 were in the control group, the re-optimization treatments (R, RS, and RR), or the Reliability Info treatment (NR). N = 1,582 were in the **endline impact evaluation sample**, which means we successfully collected the set of pages they followed on Facebook at the beginning of the endline survey. N=252 were in the No-Experimenter Demand treatment (NED)

Notes: This table shows the size of our sample at different stages of the experiment. The main results on the effect of our treatments on the set of news pages that participants follow on Facebook rely on the midline impact evaluation sample. The endline impact evaluation sample is primarily used to show persistence.

Table 2: Distance to Bliss Points

	Slant (1)	Reliability (2)	Euclidean (3)
Re-optimization (R)	-0.120***	-0.050***	-0.119***
	(0.029)	(0.019)	(0.022)
Re-optimization + Slant Info (RS)	-0.152***	-0.021	-0.118***
	(0.029)	(0.017)	(0.022)
Re-optimization + Reliability Info (RR)	-0.092***	-0.103***	-0.121***
	(0.027)	(0.017)	(0.022)
Reliability Info (NR)	0.000	0.002	-0.000
	(0.014)	(0.009)	(0.011)
p-value, R=RS	0.42	0.25	0.96
p-value, R=RR	0.47	0.03	0.97
p-value, RR=NR	0.00	0.00	0.00
Control Mean	0.00	-0.00	-0.00
Control SD	1.00	1.00	1.00
Observations	1866	1866	1866
R-squared	0.78	0.91	0.86

Notes: This table explores the effect of our intervention on the distance between the slant and reliability of participants' news portfolios on Facebook and participants' stated bliss points in the slant-reliability space. Specifically, it presents estimates of coefficients β_j from Equation 2 with distance to bliss point as the outcome variable. Column (1) estimates Equation 2 on the standardized absolute difference between a participant's stated bliss point on the slant dimension and the average slant of the portfolio of news pages the participant follows on Facebook. Column (2) estimates Equation 2 on the standardized absolute difference between a participant's stated bliss point on the reliability dimension and the average reliability of the portfolio of news pages the participant follows on Facebook. Column (3) estimates Equation 2 on the Euclidean distance between a participant's stated bliss point in the two-dimensional slant-reliability space and the average position in that space of the portfolio of news pages the participant follows on Facebook. The standardization was performed as follows: for each participant, we subtracted from the outcome the average outcome among participants in the control group, and divided the resulting number by the standard deviation of the outcome among participants in the control group. The regressions are run on the midline impact evaluation sample. All regressions include pre-specified controls. Our controls consist of: baseline value of the outcome variable (when available), baseline political ideology, baseline degree of reliance on Facebook for news, baseline interest in politics, wave fixed effects, device fixed effects (desktop or mobile), recruitment and randomization strata fixed effects. Standard errors in parentheses are robust.

Table 3: ATE on Absolute Average Slant

	(1) Absolute Average Slant (post-midline)	(2) Absolute Average Slant (pre-endline)
Re-optimization (R)	-0.058***	-0.049***
	(0.008)	(0.009)
Re-optimization + Slant Info (RS)	-0.052***	-0.050***
	(0.009)	(0.010)
Re-optimization + Reliability Info (RR)	-0.074***	-0.069***
	(0.008)	(0.010)
Reliability Info (NR)	-0.001	-0.006
	(0.004)	(0.006)
p-value, R=RS	0.64	0.92
p-value, R=RR	0.15	0.09
p-value, RR=NR	0.00	0.00
Control Mean	0.45	0.46
Control SD	0.24	0.24
Observations	1866	1582
R-squared	0.70	0.67

Notes: This table explores the effect of our intervention on the absolute average slant of participants' news portfolios on Facebook. Specifically, it presents estimates of coefficients β_j from Equation 2 with the absolute average slant of a participant's news portfolio on Facebook as the outcome variable. Column (1) estimates Equation 2 using data collected right after the midline survey; thus, it employs the midline impact evaluation sample. Column (2) estimates Equation 2 using data collected at the beginning of the endline survey; thus, it employs the endline impact evaluation sample. All regressions include pre-specified controls. Our controls consist of: baseline value of the outcome variable (when available), baseline political ideology, baseline degree of reliance on Facebook for news, baseline interest in politics, wave fixed effects, device fixed effects (desktop or mobile), recruitment and randomization strata fixed effects. Standard errors in parentheses are robust.

Table 4: ATE on Reliability

	(1) Avg. Reliability (post-midline)	(2) Avg. Reliability (pre-endline)
Re-optimization (R)	0.504	0.394
	(0.321)	(0.365)
Re-optimization + Slant Info (RS)	0.092	0.036
	(0.300)	(0.350)
Re-optimization + Reliability Info (RR)	2.505***	2.314***
	(0.301)	(0.377)
Reliability Info (NR)	-0.089	-0.047
	(0.220)	(0.282)
p-value, R=RS	0.31	0.41
p-value, R=RR	0.00	0.00
p-value, RR=NR	0.00	0.00
Control Mean	86.53	86.50
Control SD	12.17	12.13
Observations	1866	1582
R-squared	0.82	0.79

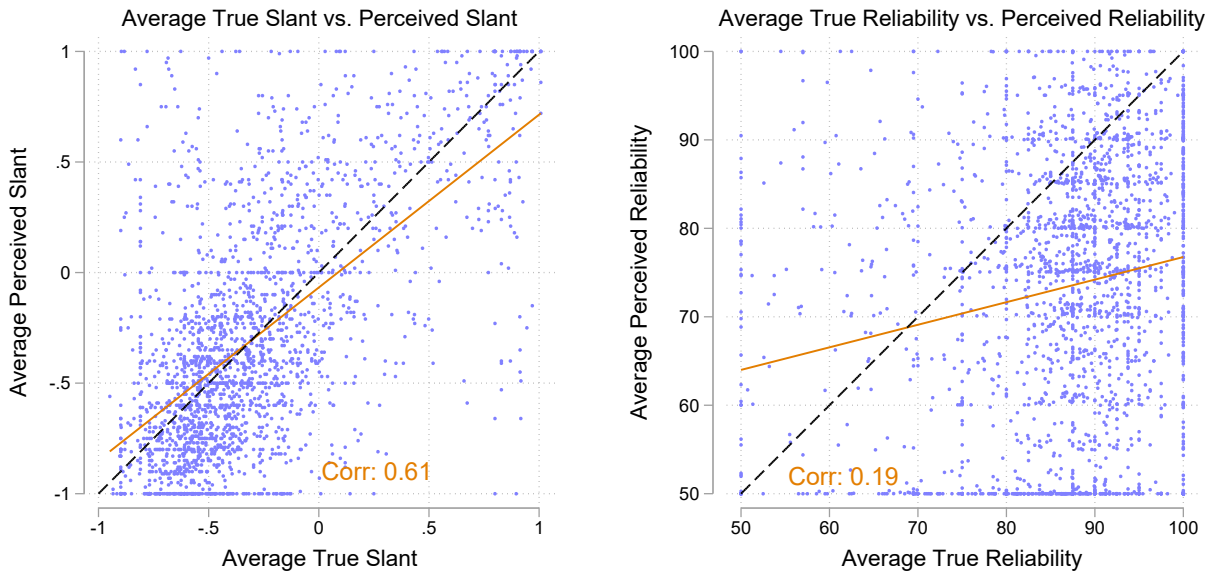
Notes: This table explores the effect of our intervention on the average reliability of participants' news portfolios on Facebook. Specifically, it presents estimates of coefficients β_j from Equation 2 with reliability of a participant's news portfolio on Facebook as the outcome variable. Column (1) estimates Equation 2 using data collected right after the midline survey; thus, it employs the midline impact evaluation sample. Column (2) estimates Equation 2 using data collected at the beginning of the endline survey; thus, it employs the endline impact evaluation sample. All regressions include pre-specified controls. Our controls consist of: baseline value of the outcome variable (when available), baseline political ideology, baseline degree of reliance on Facebook for news, baseline interest in politics, wave fixed effects, device fixed effects (desktop or mobile), recruitment and randomization strata fixed effects. Standard errors in parentheses are robust.

Table 5: Effects on Page Visits

	(1) First Stage Page Followed	(2) Extensive margin (Facebook referred)	(3) IHS(Number of Visits) (Facebook referred)	(4) Extensive margin	(5) IHS(Number of Visits)
Pooled Reoptimization Treatments	0.195***				
	(0.010)				
Page Liked		0.040**	0.047**	0.105**	0.118
		(0.019)	(0.021)	(0.050)	(0.075)
First-stage F-stat	404.95				
Control Mean, Dep Var	0.00	0.01	0.01	0.09	0.16
Control SD, Dep Var	0.00	0.10	0.11	0.29	0.65
Observations	6156	6156	6156	6156	6156
R-squared	0.07	0.03	0.04	0.19	0.55

Notes: This table explores the effects of our intervention on online news consumption for the subset of participants who installed the Chrome browser extension that tracks their behavior on news websites. The analysis is run on a dataset where an observation is a user-page. Pooled treatments refer to the three main re-optimization treatments, namely R, RS, and RR. Belonging to any of those treatments is used as an instrument for following a page in the re-optimization interface. Column (1) displays the results of the first-stage regression on the instrument and the same set of controls as in Equation 3. Columns (2) through (5) display the results of the second-stage regression; i.e., estimates of parameter τ in Equation 3. Columns (2) and (3) consider visits to news websites originating from Facebook. Columns (4) and (5) consider all visits to news websites, independently of whether the visit originates from Facebook. Columns (2) and (4) present extensive margin results: the outcome variable is an indicator that takes value one if a user visited a news page at least once in the 31 days after the midline survey and zero otherwise. Columns (3) and (5) capture both the extensive and intensive margins: the outcome variable is the inverse hyperbolic sign of the total number of visits by a user to a page in the 31 days after the midline survey. Standard errors are clustered at the participant level.

Figure 1: Perceptions vs. Reality



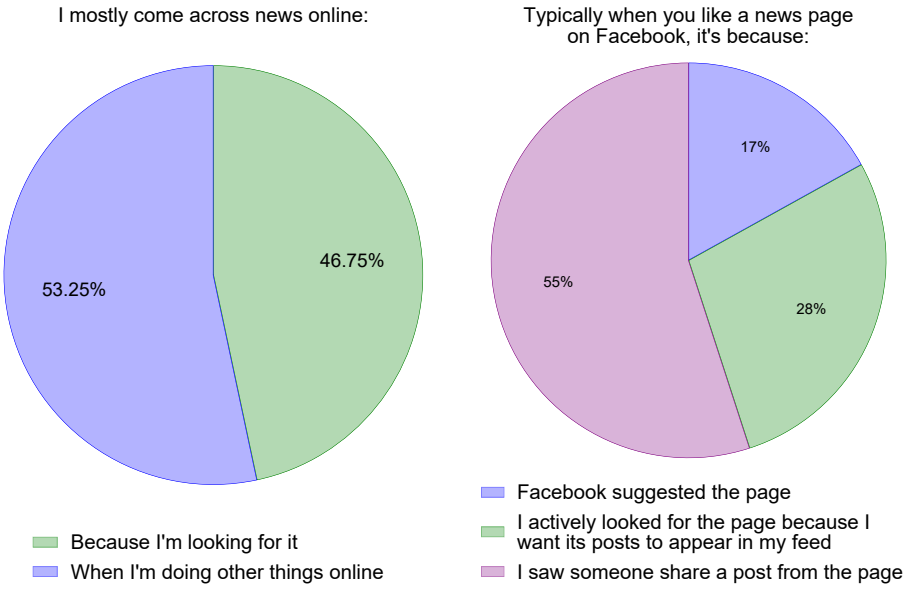
Notes: This figure explores the relationship between participants' perceptions of the average slant and reliability of their news portfolios on Facebook and the actual average slant and reliability of those portfolios. The left panel presents a scatter plot where the x-axis indicates the actual average slant of a participant's news portfolio and the y-axis indicates a participant's perception of the average slant of their news portfolio. Each dot in the figure represents an individual. The right panel presents a scatter plot where the x-axis indicates the actual average reliability of a participant's news portfolio and the y-axis indicates the participant's perceptions of the average reliability of their news portfolio. Each dot in the figure represents an individual. The solid orange lines are lines of best fit. The dashed black lines are 45-degree lines. The figure also reports Pearson correlations. To improve the clarity of the figure, we winsorize reliability values below 50 to 50. The Pearson correlation is still calculated using the non-winsorized data

Figure 2: Bliss Points vs. Reality



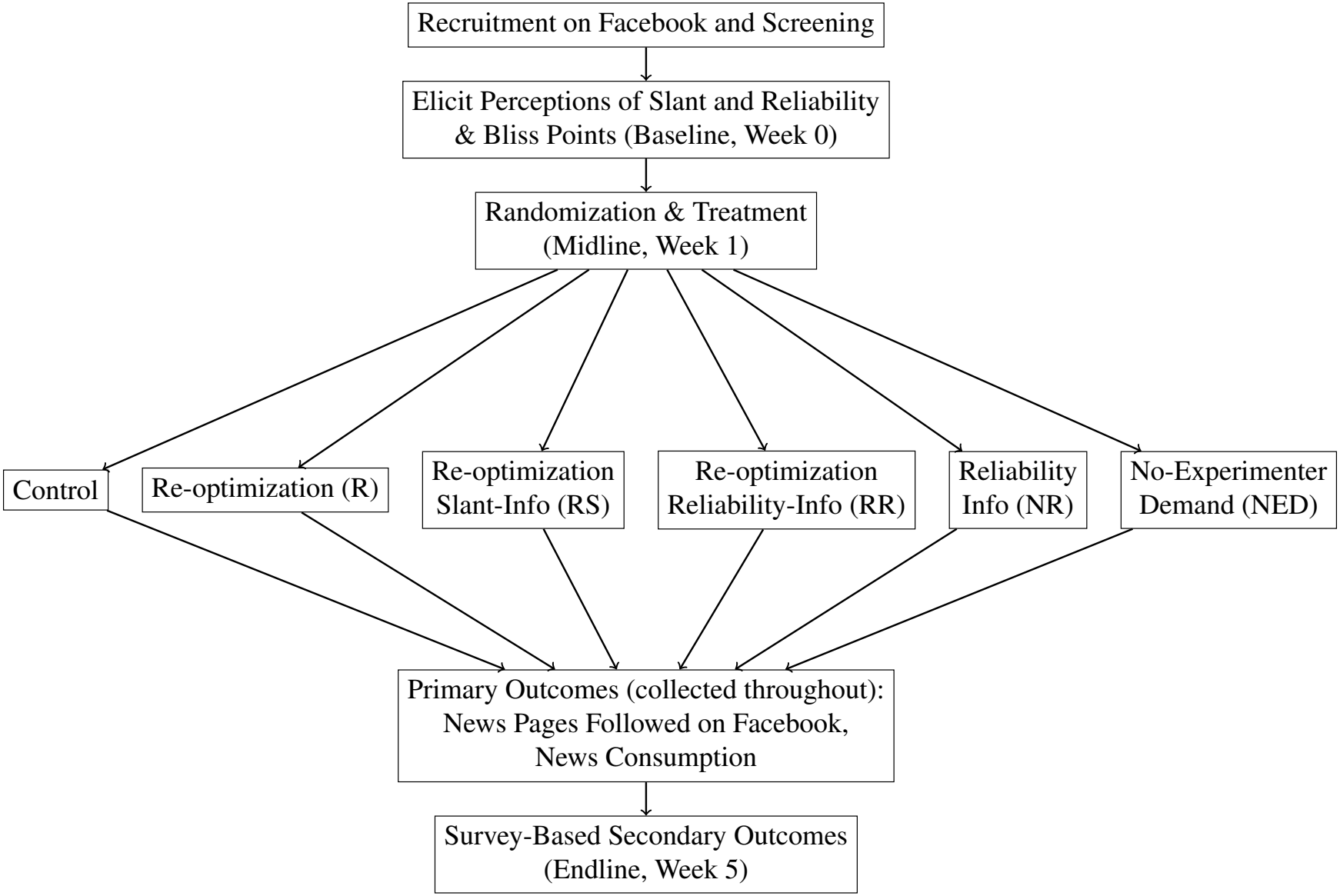
Notes: This figure explores the relationship between participants' stated bliss points in the slant-reliability space and the actual average slant and reliability of their Facebook news portfolios. The left panel presents a scatter plot where the x-axis indicates the actual average slant of a participant's news portfolio and the y-axis indicates a participant's stated bliss point in the slant dimension. Each dot in the figure represents an individual. The right panel presents a scatter plot where the x-axis indicates the actual average reliability of a participant's news portfolio and the y-axis indicates a participant's stated bliss point in the reliability dimension. Each dot in the figure represents an individual. The solid orange lines are lines of best fit. The dashed black lines are 45-degree lines. The figure also reports Pearson correlations. To improve the clarity of the figure, we winsorize reliability values below 50 to 50. The Pearson correlation is still calculated using the non-winsorized data.

Figure 3: Evidence of Passive News Consumption Online



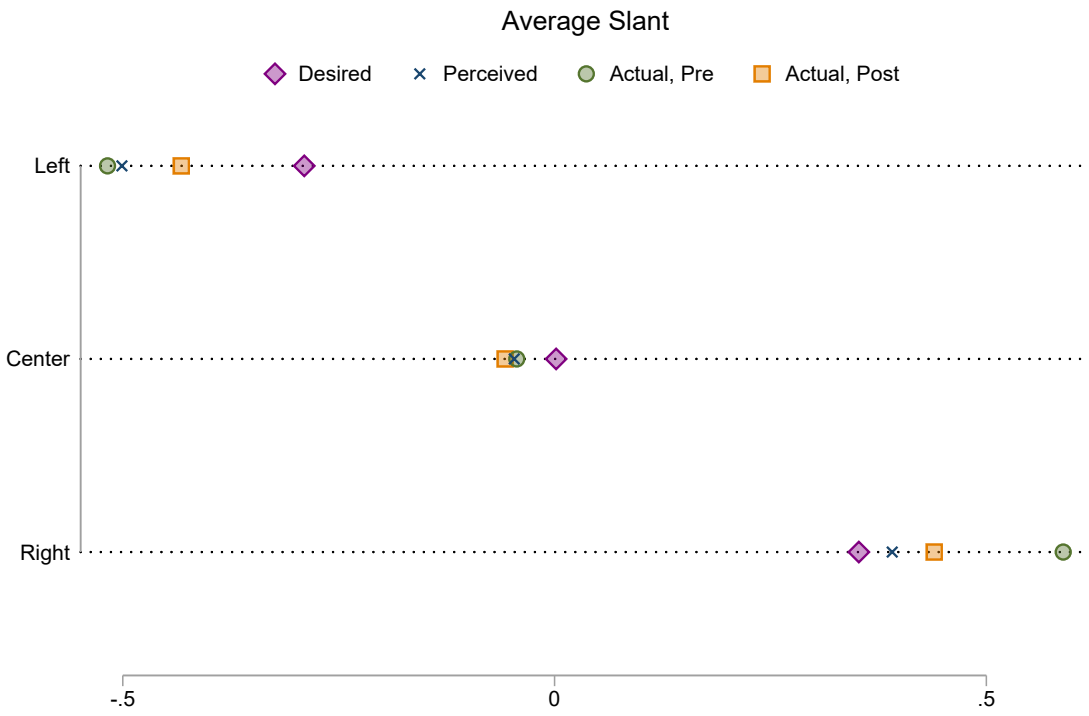
Notes: This figure presents the distribution of responses to two questions about participants' interactions with the online news ecosystem. The first question asked participants: "Which statement best describes how you get news online, whether on a computer, phone, or tablet, even if neither is exactly right?" The answer options were: "I mostly come across news online because I'm looking for it; I mostly come across news online when I'm doing other things online." The second question asked participants: "Typically when you like a news page on Facebook, does this happen because:" The answer options were: "I saw someone share a post from the page; I actively looked for the page because I want its posts to appear in my feed; Facebook suggested the page; I do not like Facebook pages; None of the above." We dropped from this analysis participants who said 'none of the above'. We also dropped 4% of the remaining participants who answered that they did not like any Facebook pages, because we knew those participants were mistaken. Specifically, had they really not liked any pages on Facebook, they would have automatically been screened out prior to this question.

Figure 4: Experimental Design



Notes: This figure illustrates graphically the key components of our experimental design.

Figure 5: Average Desired, Perceived, Pre-Treatment, and Post-Period Slant by Baseline Portfolio



Notes: This figure shows the average portfolio desired slant at baseline (purple diamond), the average perceived portfolio slant at baseline (navy X), the average actual portfolio slant at baseline (green circle), and the average actual portfolio slant at midline, post-treatment (orange square). The averages are shown for participants in the re-optimization treatments (R, RS, RR) and computed separately for participants whose portfolios were left-leaning (average slant of less than -0.2), participants whose portfolios were moderate (average slant between -0.2 and 0.2), and participants whose portfolios were right-leaning (average slant of greater than 0.2).

Appendix For Online Publication

Table of Contents

A Appendix Figures and Tables **A.2**

B Measuring Slant and Reliability of News **A.24**

C Selection of News Pages **A.25**

D Implementation: Additional Details **A.27**

E Perceptions of Slant and Reliability of Mainstream Outlets **A.28**

F Survey Screenshots **A.36**

F.1 Information Interventions A.36

F.2 Like and Unlike Tables A.38

F.3 Experimenter Demand A.45

F.4 Explanation of Slant and Reliability Measures and Belief Elicitation A.46

A. Appendix Figures and Tables

Table A.1: Stylized Facts on News Consumption on Facebook

	Stylized Fact
1	Stated bliss points are largely consistent with slant being a horizontal characteristic and reliability being a vertical characteristic.
2	Slant perceptions are relatively accurate; reliability perceptions are relatively inaccurate.
3	Perceptions of slant and reliability are correlated: pro-attitudinal outlets are generally considered higher reliability than counter-attitudinal ones.
4	Participants hold news portfolios that, on average, are pro-attitudinal.
5	Participants hold news portfolios that, on average, are more partisan and less reliable than their stated bliss points.
6	Participants describe their news-portfolio formation process as largely passive and driven by salient suggestions from the platform or from friends.

Notes: This table summarizes the motivating facts discussed in Section 2. Most stylized facts are consistent with canonical models of news demand. Fact 6 and the first part of Fact 5 are prima facie at odds with those models.

Table A.2: Confusion Matrix: Slant

		Perceived		
		Left-leaning	Centrist	Right-leaning
Actual	Left-leaning	0.81	0.12	0.07
	Centrist	0.43	0.28	0.29
	Right-leaning	0.13	0.13	0.74

Notes: This table explores the relationship between participants' perceptions of the average slant of their Facebook news portfolios and the actual average slant of those portfolios. The table is row-stochastic, meaning the entries of each row sum to one. The entries of the first row show the conditional probabilities that a participant whose actual Facebook news portfolio is, on average, left-leaning perceives the average slant of her portfolio as left-leaning, centrist, and right-leaning, respectively. The other rows show similar probabilities, but conditional on the actual portfolio being, on average, centrist (row 2) or right-leaning (row 3). A portfolio is considered left-leaning if its average slant < -0.2, centrist if $-0.2 \leq$ average slant ≤ 0.2, and right-leaning if its average slant > 0.2.

Table A.3: Confusion Matrix: Reliability

		Perceived		
		Low Reliability	Medium Reliability	High Reliability
Actual	Low Reliability	0.35	0.34	0.31
	Medium Reliability	0.34	0.39	0.27
	High Reliability	0.19	0.40	0.41

Notes: This table explores the relationship between participants' perceptions of the average reliability of their Facebook news portfolios and the actual average reliability of those portfolios. The table is row-stochastic, meaning the entries of each row sum to one. The entries of the first row show the conditional probabilities that a participant whose actual Facebook news portfolio is, on average, low-reliability perceives the average reliability of her portfolio as low-reliability, medium-reliability, or high reliability, respectively. The other rows show similar probabilities, but conditional on the actual portfolio being, on average, medium-reliability (row 2) or high-reliability (row 3). A portfolio is considered low-reliability if its average reliability < 60, medium-reliability if $60 \leq$ average reliability ≤ 80, and high-reliability if its average reliability > 80.

Table A.4: Correlations between Slant and Reliability Perceptions

	(1) Democrats Perceived Reliability	(2) Republicans Perceived Reliability
Perceived Slant	-20.450***	17.110***
	(0.591)	(1.254)
Observations	7056	1952
R-squared	0.23	0.24
Clusters	1189	330

Notes: This table shows the results from a regression at the individual-outlet level of the perception of reliability on the perception of slant. We run this regression separately for Democrats and for Republicans and cluster standard errors at the participant level.

Table A.5: Baseline Portfolio Statistics

	(1) Democrats	(2) Republicans	(3) All
Avg. Number News Pages Followed	6.52	4.86	5.77
Avg. Portfolio Slant [-1, +1]	-0.46	0.19	-0.30
Avg. Portfolio Reliability [0, 100]	88.99	76.70	86.15
Observations	1029	281	1874

Notes: This table presents descriptive statistics about the baseline portfolios of news pages that participants in our experiment follow on Facebook. The table shows these statistics for our midline impact evaluation sample: participants who were assigned to the control group or to one of the main treatments, and for whom we successfully collected the set of pages they followed on Facebook during the midline survey (see Section 4.4 for details). Column (1) shows descriptive statistics for self-identified Democrats. Column (2) does the same for self-identified Republicans. Column (3) includes all participants. The number of observations in column (3) is not simply the sum of the observations in columns 1 and 2, because the former columns exclude independents.

Table A.6: Sample Demographics

	(1) Midline impact evaluation sample	(2) US Facebook users	(3) US population
Female	0.73	0.57	0.51
White	0.94	0.73	0.72
Age under 50	0.44	0.57	0.54
College	0.62	0.35	0.34
Democrat	0.55		0.20
Republican	0.15		0.15

Notes: This table compares the demographic characteristics of our midline impact evaluation sample to those of U.S. adult Facebook users and of the overall U.S. adult population. Column (1) presents the average demographics for participants in our sample (N=1,866): participants who were assigned to the control group or to one of the main treatments, and for whom we successfully collected the set of pages they followed on Facebook during the midline survey (see Section 4.4 for details). Column (2) presents our estimate of the average demographics of U.S. adults with a Facebook account. The numbers in column (2) are inferred from a Pew Research Center survey of social media use by demographic group (Pew Research Center, 2024). Column (3) presents the average demographics of American adults. The top four numbers are from the 2023 American Community Survey (U.S. Census Bureau, 2024), and the Republican and Democrat shares are from the 2024 American National Election Study (American National Election Study, 2024).

Table A.7: Balance at Midline

	Group Means						p-value treatment-control difference				
	Control	R	RS	RR	NR	NED	p(R)	p(RS)	p(RR)	p(NR)	p(NED)
Female	0.75	0.72	0.70	0.75	0.77	0.72	0.35	0.17	0.92	0.55	0.48
White	0.94	0.94	0.94	0.94	0.96	0.93	0.94	0.88	0.80	0.27	0.54
Age	51.86	53.43	53.60	52.73	51.49	53.25	0.10	0.07	0.37	0.71	0.17
College Grad	0.62	0.65	0.65	0.62	0.57	0.65	0.36	0.41	0.95	0.24	0.30
Democrat	0.57	0.54	0.56	0.56	0.50	0.52	0.43	0.77	0.67	0.06	0.14
Republican	0.11	0.17	0.16	0.15	0.16	0.16	0.02	0.06	0.13	0.05	0.05
News SM Often	0.82	0.83	0.85	0.83	0.84	0.82	0.60	0.18	0.58	0.37	1.00

Notes: This table presents balance checks on the following baseline characteristics: gender (indicator for identifying as female), race (indicator for identifying as white), age, education (indicator for being a college graduate), political ideology (indicators for identifying as Democrat or Republican), and baseline consumption of news on social media (indicator for consuming news on social media often or sometimes as rarely or never). The first six columns present means across our six study arms. The last five columns present p-values related to the null hypothesis of no difference in means between each treatment arm and the control arm. The sample includes all individuals who were randomized.

Table A.8: Differential Attrition

	(1)	(2)
	Pre-endline likes collected	Completed Endline + post-endline likes collected
Pooled Reoptimization Treatments	0.018	0.017
	(0.021)	(0.021)
Re-optimization (R)	0.052**	0.049**
	(0.024)	(0.025)
Re-optimization + Slant Info (RS)	0.016	0.013
	(0.026)	(0.026)
Re-optimization + Reliability Info (RR)	-0.013	-0.011
	(0.027)	(0.027)
Reliability Info (NR)	0.018	0.014
	(0.027)	(0.028)
No Experimenter Demand (NED)	-0.043	-0.053*
	(0.030)	(0.030)
Control Mean	0.84	0.84
Observations	2184	2184
R-squared	0.01	0.01

Notes: This table presents the results of a battery of tests for differential attrition in our experiment. The independent variables are each treatment. We also run a separate regression where the independent variable is the pooled reoptimization treatments (R, RS, and RR). In column (1), the dependent variable is an indicator that takes the value one if and only if we were able to collect the set of news pages followed by a participant at the beginning of the endline survey. In column (2), the outcome variable is an indicator for completing the endline survey and the collection of post-endline likes. The sample includes all participants who reached the randomization stage. Standard errors in parentheses are robust.

Table A.9: PATE on Absolute Average Slant

	(1) Absolute Average Slant (post-midline)	(2) Absolute Average Slant (pre-endline)
Re-optimization (R)	-0.082***	-0.074***
	(0.015)	(0.016)
Re-optimization + Slant Info (RS)	-0.080***	-0.086***
	(0.017)	(0.019)
Re-optimization + Reliability Info (RR)	-0.073***	-0.074***
	(0.010)	(0.011)
Reliability Info (NR)	-0.014*	-0.021**
	(0.008)	(0.010)
p-value, R=RS	0.95	0.58
p-value, R=RR	0.54	0.97
p-value, RR=NR	0.00	0.00
Control Mean	0.45	0.46
Control SD	0.24	0.24
Observations	1866	1582
R-squared	0.75	0.74

Notes: This table explores the population average treatment effects of our intervention on the absolute average slant of participants' news portfolios on Facebook. To create these estimates, we run the same regression we present in Table 3 but add weights to make our sample similar to the US population. We create the weights using entropy balancing (Hainmueller, 2012) to match the following characteristics to the US population: gender (male or female), age (bins for 20-39, 40-59, and 60+), race (White or not), Hispanic (binary), party affiliation (Democrats, Independents, Republicans) and education (Bachelor's degree and above or not). The US population shares for party affiliation are based on the 2024 American National Election Survey; all other variables are based on 2024 American Community Survey 1-year estimates. We included the one participant who was 19-year-old-participant in the 20-year-old bin in order to not exclude any observations.

Table A.10: PATE on Reliability

	(1) Average Reliability (post-midline)	(2) Average Reliability (pre-endline)
Re-optimization (R)	0.865	0.627
	(0.540)	(0.592)
Re-optimization + Slant Info (RS)	0.293	0.276
	(0.546)	(0.594)
Re-optimization + Reliability Info (RR)	2.795***	2.370***
	(0.517)	(0.516)
Reliability Info (NR)	0.564	0.597
	(0.441)	(0.461)
p-value, R=RS	0.40	0.64
p-value, R=RR	0.00	0.01
p-value, RR=NR	0.00	0.00
Control Mean	86.53	86.50
Control SD	12.17	12.13
Observations	1866	1582
R-squared	0.83	0.82

Notes: This table explores the population average treatment effects of our intervention on the average reliability of participants' news portfolios on Facebook. To create these estimates, we run the same regression we present in Table 4 but add weights to make our sample similar to the US population. We create the weights using entropy balancing (Hainmueller, 2012) to match the following characteristics to the US population: gender (male or female), age (bins for 20-39, 40-59, and 60+), race (White or not), Hispanic (binary), party affiliation (Democrats, Independents, Republicans) and education (Bachelor's degree and above or not). The US population shares for party affiliation are based on the 2024 American National Election Survey; all other variables are based on 2024 American Community Survey 1-year estimates.

Table A.11: Balance at Endline

	Group Means						p-value treatment-control difference				
	Control	R	RS	RR	NR	NED	p(R)	p(RS)	p(RR)	p(NR)	p(NED)
Female	0.75	0.71	0.69	0.75	0.76	0.70	0.29	0.13	0.98	0.63	0.22
White	0.93	0.94	0.94	0.93	0.95	0.92	0.55	0.70	0.87	0.30	0.55
Age	51.20	53.19	53.35	52.01	50.56	52.85	0.05	0.04	0.44	0.55	0.14
College Grad	0.62	0.66	0.65	0.62	0.58	0.69	0.23	0.51	0.91	0.35	0.09
Democrat	0.56	0.55	0.54	0.57	0.50	0.50	0.73	0.62	0.90	0.12	0.16
Republican	0.11	0.16	0.17	0.14	0.15	0.16	0.06	0.02	0.28	0.15	0.10
News SM Often	0.85	0.84	0.85	0.85	0.84	0.81	0.55	0.92	0.99	0.63	0.14

Notes: This table presents balance checks on the following baseline characteristics: gender (indicator for identifying as female), race (indicator for identifying as white), age, education (indicator for being a college graduate), political ideology (indicators for identifying as Democrat or Republican), and baseline consumption of news on social media (indicator for consuming news on social media often or sometimes as rarely or never). The first six columns present means across our six study arms. The last five columns present p-values related to the null hypothesis of no difference in means between each treatment arm and the control arm. The sample includes all individuals for whom we were able to collect the set of news pages they followed at the beginning of the endline survey.

Table A.12: Effect on Slant, Channels

	(1) Absolute Average Slant (post-midline)	(2) Average Absolute Slant (post-midline)	(3) Fraction Counter-attitudinal Pages Followed (post-midline)
Pooled Reoptimization Treatments	-0.061***	-0.005	0.058***
	(0.005)	(0.004)	(0.004)
Control Mean	0.45	0.51	0.02
Control SD	0.24	0.21	0.06
Observations	1552	1552	1279
R-squared	0.64	0.72	0.30

Notes: This table presents the effect of our intervention on the absolute average slant of participants' news portfolios on Facebook (our primary outcome), the average absolute slant of those portfolios, and the fraction of counter-attitudinal pages in those portfolios. Specifically, it presents the estimate of the β coefficient for a version of Equation 2 in which all three main re-optimization treatment arms are pooled together. A page is deemed counter-attitudinal if its slant and the average slant of the participant's baseline news portfolio on Facebook have opposite signs. The regressions whose outcomes are reported in columns 1 and 2 use data collected right after the midline survey; thus, they employ the midline impact evaluation sample. The regression whose outcome is reported in column (3) uses the subset of the midline impact evaluation sample that does not have an initial portfolio with a centrist slant (i.e., their baseline average slant does not fall between -0.2 and 0.2). The regression includes pre-specified controls. Our controls consist of: baseline value of the outcome variable (when available), baseline political ideology, baseline degree of reliance on Facebook for news, baseline interest in politics, wave fixed effects, device fixed effects (desktop or mobile), recruitment and randomization strata fixed effects. Standard errors in parentheses are robust.

Table A.13: Effect on Slant, Channels by Party

	(1) Democrats Absolute Average Slant	(2) Republican Absolute Average Slant	(3) Democrats Average Absolute Slant	(4) Republican Average Absolute Slant	(5) Democrats Fraction Counter-attitudinal Pages Followed	(6) Republican Fraction Counter-attitudinal Pages Followed
Pooled Reoptimization Treatments	-0.055***	-0.051**	-0.005	0.007	0.050***	0.096***
	(0.007)	(0.023)	(0.005)	(0.015)	(0.005)	(0.022)
Control Mean	0.47	0.39	0.50	0.55	0.02	0.04
Control SD	0.21	0.31	0.20	0.26	0.06	0.09
Observations	868	229	868	229	775	152
R-squared	0.63	0.64	0.71	0.76	0.29	0.32

Notes: This table presents the effect of our intervention on the absolute average slant of participants' news portfolios on Facebook, the average absolute slant of those portfolios, and the fraction of counter-attitudinal pages in those portfolios. The table presents the effects separately for Democrats and Republicans. Specifically, it presents the estimate of the β coefficient for a version of Equation 2 in which all three main re-optimization treatment arms are pooled together. A page is deemed counter-attitudinal if its slant and the average slant of the participant's baseline news portfolio on Facebook have opposite signs. The regressions whose outcomes are reported in columns 1 and 2 use data collected right after the midline survey; thus, they employ the midline impact evaluation sample. The regression whose outcome is reported in column (3) uses the subset of the midline impact evaluation sample that does not have an initial portfolio with a centrist slant. The regression includes pre-specified controls. Our controls consist of: baseline value of the outcome variable (when available), baseline political ideology, baseline degree of reliance on Facebook for news, baseline interest in politics, wave fixed effects, device fixed effects (desktop or mobile), recruitment and randomization strata fixed effects. Standard errors in parentheses are robust.

Table A.14: ATE on Secondary Outcomes: News Satisfaction and Trust

	News Satisfaction (endline)	News Trust (endline)
Pooled Reoptimization Treatments	0.104*	0.082**
	(0.053)	(0.039)
Re-optimization (R)	0.081	0.089*
	(0.065)	(0.047)
Re-optimization + Slant Info (RS)	0.101	0.081*
	(0.065)	(0.047)
Re-optimization + Reliability Info (RR)	0.130**	0.081*
	(0.064)	(0.048)
p-value, R=RS	0.77	0.87
p-value, R=RR	0.45	0.87
Control Mean	2.98	2.53
Control SD	0.88	0.65
Observations	1610	1610
R-squared	0.16	0.21

Notes: This table explores the effect of our intervention on the secondary outcomes of news satisfaction and trust described in Section 4.3. Specifically, it presents estimates of coefficients β_j from Equation 2 for each secondary outcome variable, as well as the coefficient on a version of the regression that pools our main re-optimization treatments. The sample is the set of individuals who were randomized and who completed endline. Column (1) estimates Equation 2 using a measure of news satisfaction on Facebook as the outcome variable. The measure is constructed from the following question: "How satisfied are you with the news appearing in your Facebook feed?" The answer options were: "Very dissatisfied (coded as 0); Dissatisfied (coded as 1); Neutral (coded as 2); Satisfied (coded as 3); Very satisfied (coded as 4)." Column (2) estimates Equation 2 using a measure of trust in news on Facebook as the outcome variable. The measure is constructed from the following question: "How much, if at all, do you trust the news you get from your Facebook feed?" The answer options were: "Not at all (coded as 0); Not too much (coded as 1); Some (coded as 2); A lot (coded as 3)." All regressions include pre-specified controls. Our controls consist of: baseline value of the outcome variable (when available), baseline political ideology, baseline degree of reliance on Facebook for news, baseline interest in politics, wave fixed effects, device fixed effects (desktop or mobile), recruitment and randomization strata fixed effects. Standard errors in parentheses are robust.

Table A.15: ATE on Secondary Outcomes: Perceived Slant and Reliability

	Perceived Slant (endline)	Perceived Reliability (endline)
Pooled Reoptimization Treatments	-0.015	0.066
	(0.060)	(0.046)
Re-optimization (R)	-0.033	0.002
	(0.073)	(0.059)
Re-optimization + Slant Info (RS)	0.059	0.077
	(0.072)	(0.056)
Re-optimization + Reliability Info (RR)	-0.065	0.126**
	(0.072)	(0.056)
p-value, R=RS	0.19	0.20
p-value, R=RR	0.65	0.03
Control Mean	-0.11	2.82
Control SD	0.87	0.72
Observations	1419	1610
R-squared	0.02	0.02

Notes: This table explores the effect of our intervention on the secondary outcomes of perceived slant and reliability described in Section 5.4. Specifically, it presents estimates of coefficients β_j from Equation 2 for each secondary outcome variable, as well as the coefficient on a version of the regression that pools our main re-optimization treatments. The sample is the set of individuals who were randomized and who completed endline. Column (1) estimates Equation 2 using a measure of the slant of the news in one's Facebook feed. The measure is constructed from the following question: "In the last four weeks, relative to your typical Facebook feed, was the news you saw in your news feed...?" The answer options were: "Much more liberal than usual; Somewhat more liberal than usual; Had a similar slant compared to what I am used to; Somewhat more conservative than usual; Much more conservative than usual." If, at baseline, a participant reported primarily consuming conservative (liberal) news on Facebook, the answer options above were ranked in ascending (descending) order. This way, a negative treatment effect always indicates exposure to more counter-attitudinal news. Column (1) has fewer observations than the other columns, because we exclude individuals who, at baseline, reported primarily consuming moderate news. Column (2) estimates Equation 2 using a measure of reliability of the news in one's Facebook feed. The measure is constructed from the following question: "In the last four weeks, relative to your typical Facebook feed, was the news you saw in your news feed...?" The answer options were: "Much less reliable than usual (coded as a 0), Somewhat less reliable than usual (coded as a 1), Had a similar reliability compared to what I am used to (coded as a 2), Somewhat more reliable than usual (coded as a 3), Much more reliable than usual (coded as a 4)." All regressions include pre-specified controls. Our controls consist of: baseline value of the outcome variable (when available), baseline political ideology, baseline degree of reliance on Facebook for news, baseline interest in politics, wave fixed effects, device fixed effects (desktop or mobile), recruitment and randomization strata fixed effects. Standard errors in parentheses are robust.

Table A.16: Effects on (Self-reported) Changes to Facebook Feed

	(1) First Stage Page Followed	(2) Number of Times Seen in Feed
Pooled Reoptimization Treatments	0.165***	
	(0.005)	
Page Followed (instrumented)		0.534**
		(0.252)
Control Mean, Dep Var	0.00	0.76
Control SD, Dep Var	0.00	1.46
Observations	18660	18660
R-squared	0.05	0.23

Notes: This table explores the effects of our interventions on the content participants reported seeing on their Facebook feeds. The analysis is run on a dataset where an observation is a user-page. The users are the individuals who completed the endline survey and were assigned either to the control group or to one of our three main re-optimization treatments, namely R, RS, and RR. The pages are the pages users were shown or the potential pages they would have been shown had they been assigned to one of the three re-optimization treatments. Pooled treatments refer to the three main re-optimization treatments. Belonging to any of those treatments is used as an instrument for following a page in the re-optimization interface. Column (1) displays the results of the first-stage regression of following a news page on Facebook on the instrument and the same set of controls as in Equation 3. Column (2) displays the results of the second-stage regression. The outcome variable in column (2) is a participant's self-report of the number of times she encountered a post from the page in her Facebook feed. Standard errors are clustered at the participant level.

Table A.17: Reduced Form Effects on Page Visits

	(1) Extensive margin (Facebook referred)	(2) IHS(Number of Visits) (Facebook referred)	(3) Extensive margin	(4) IHS(Number of Visits)
Pooled Reoptimization Treatments	0.008**	0.009**	0.020**	0.023
	(0.004)	(0.004)	(0.010)	(0.014)
Control Mean, Dep Var	0.01	0.01	0.09	0.16
Control SD, Dep Var	0.10	0.11	0.29	0.65
Observations	6156	6156	6156	6156
R-squared	0.01	0.03	0.17	0.54

Notes: This table explores the reduced form effects of our intervention on online news consumption for the subset of participants who installed the Chrome browser extension that tracks their behavior on news websites. The analysis is run on a dataset where an observation is a user-page. The users are the individuals who installed the extension. The pages are the pages users were shown or the pages they would have been shown had they been assigned to one of the three re-optimization treatments. Pooled treatments refer to the three main re-optimization treatments, namely R, RS, and RR. The table presents estimates of the reduced-form of the IV specification detailed in Equation 3; in other words, we regress $V_{i,k}$ on an indicator for assignment to the pooled re-optimization treatments and include the same set of controls. Columns (1) and (2) consider visits to news websites originating from Facebook. Columns (3) and (4) consider all visits to news websites, independently of whether the visit originates from Facebook. Columns (1) and (3) present extensive margin results: the outcome variable is an indicator that takes value one if a user visited a news page at least once in the 31 days after the midline survey and zero otherwise. Columns (2) and (4) capture both the extensive and intensive margins: the outcome variable is the inverse hyperbolic sign of the total number of visits by a user to a page in the 31 days after the midline survey. Standard errors are clustered at the participant level.

Table A.18: ATE on Slant and Reliability of Visited Pages

	Extensive Margin High-reliability	IHS(visits) High-reliability	Extensive Margin Counter-Attitudinal	IHS(visits) Counter-Attitudinal
Pooled Reoptimization Treatments	0.058	0.142	0.034	0.065*
	(0.039)	(0.121)	(0.021)	(0.034)
Re-optimization (R)	0.009	0.132	0.030	0.041
	(0.047)	(0.142)	(0.028)	(0.040)
Re-optimization + Slant Info (RS)	0.077*	0.217	0.074**	0.122***
	(0.047)	(0.145)	(0.032)	(0.046)
Re-optimization + Reliability Info (RR)	0.086*	0.069	-0.008	0.026
	(0.048)	(0.150)	(0.026)	(0.053)
Reliability Info (NR)	0.059	0.006	0.004	0.035
	(0.074)	(0.205)	(0.041)	(0.055)
Control Mean	0.74	2.51	0.04	0.04
Control SD	0.44	2.10	0.19	0.23
Observations	538	538	451	451
R-squared	0.26	0.71	0.28	0.65

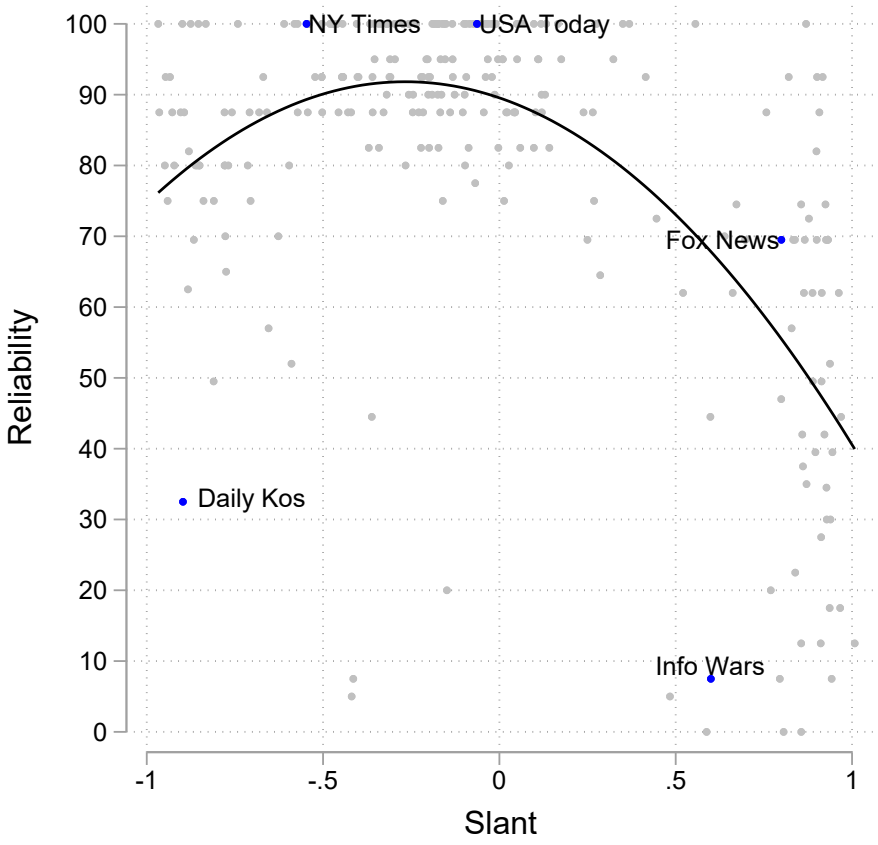
Notes: This table examines the effect of our intervention on the slant and reliability of news websites visited by participants who installed our Chrome extension. Because treatment assignment affects the likelihood of visiting any news website at all, directly regressing the average slant or reliability of visited pages on treatment indicators could yield biased estimates, as the data would be missing not at random. We therefore construct outcome variables defined for all Chrome extension users and estimate treatment effects using Equation 2. The table reports estimates of the coefficients β_j for each outcome, as well as the coefficient from a specification pooling the main re-optimization treatments. The outcome in column (1) is an indicator for visiting at least one high-reliability news website (defined as the website of an outlet with a reliability score above 80) in the 31 days after the midline survey. Column (2) uses the inverse hyperbolic sine (IHS) transformation of the number of visits to high-reliability news websites. Column (3) is an indicator for visiting at least one counter-attitudinal news website during the same period, where a website is considered counter-attitudinal if its slant has the opposite sign of the average slant of the participant's baseline Facebook news portfolio. Column (4) reports the IHS of the number of visits to counter-attitudinal news websites using the same definition. Columns (3) and (4) exclude participants with a baseline slant between -0.2 and +0.2, as counter-attitudinal news sites do not exist for such participants. All regressions include the pre-specified controls: baseline political ideology, baseline reliance on Facebook for news, baseline political interest, wave fixed effects, device fixed effects (desktop or mobile), and recruitment and randomization fixed effects. Standard errors in parentheses are robust.

Table A.19: ATE of the No-Experimenter Demand (NED) Treatment

	Absolute Average Slant (pre-endline) (1)	Average Reliability (pre-endline) (2)
Re-optimization (R)	-0.049***	0.418
	(0.010)	(0.365)
Re-optimization + Slant Info (RS)	-0.050***	0.054
	(0.010)	(0.351)
Re-optimization + Reliability Info (RR)	-0.069***	2.327***
	(0.010)	(0.375)
Reliability Info (NR)	-0.006	-0.030
	(0.007)	(0.282)
No Experimenter Demand (NED)	-0.078***	1.785***
	(0.012)	(0.387)
p-value, R=RS	0.90	0.41
p-value, R=RR	0.09	0.00
p-value, RR=NED	0.54	0.24
Control Mean	0.46	86.50
Control SD	0.24	12.13
Observations	1824	1824
R-squared	0.64	0.79

Notes: This table explores the effect of our intervention on the absolute average slant and average reliability of participants' news portfolios on Facebook. Specifically, it presents estimates of coefficients β_j from Equation 2, including not only the three re-optimization treatment arms, but also the No-Experimenter Demand (NED) treatment. Column 1 estimates Equation 2 with the absolute average slant of a participant's news portfolio on Facebook as the outcome variable. Column 3 estimates Equation 2 with the average reliability of a participant's news portfolio on Facebook as the outcome variable. The sample is the endline impact evaluation sample + the sample of participants assigned to the NED treatment for whom we could collect the pages they follow on Facebook during the endline survey. We need to rely on the endline impact evaluation sample rather than the midline impact evaluation sample, because, by design, permissions to observe the set of news pages followed by participants in the NED treatment were revoked at midline and re-obtained at endline. All regressions include pre-specified controls. Our controls consist of: baseline political ideology, baseline degree of reliance on Facebook for news, baseline interest in politics, wave fixed effects, device fixed effects (desktop or mobile), recruitment and randomization fixed effects. Standard errors in parentheses are robust.

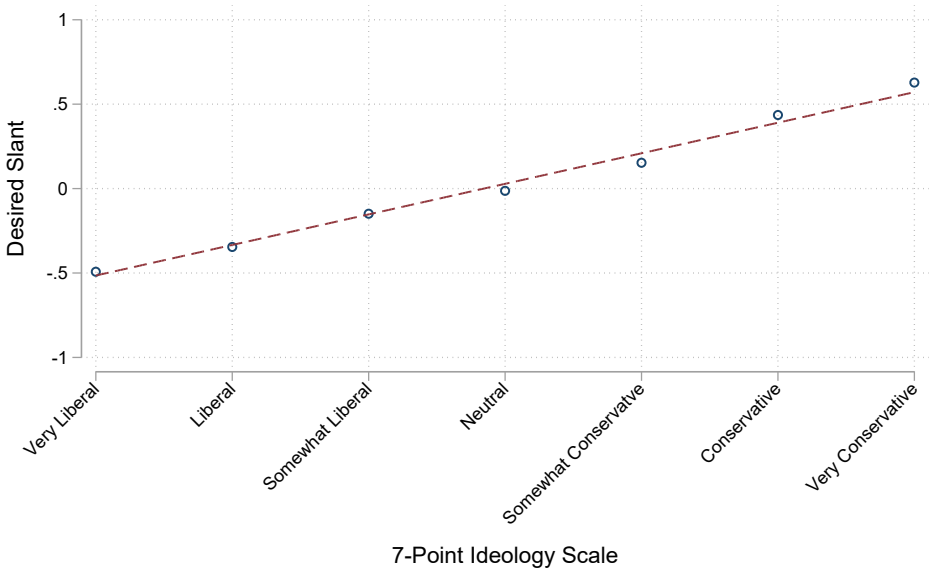
Figure A.1: Joint Distribution of Slant and Reliability for News Outlets



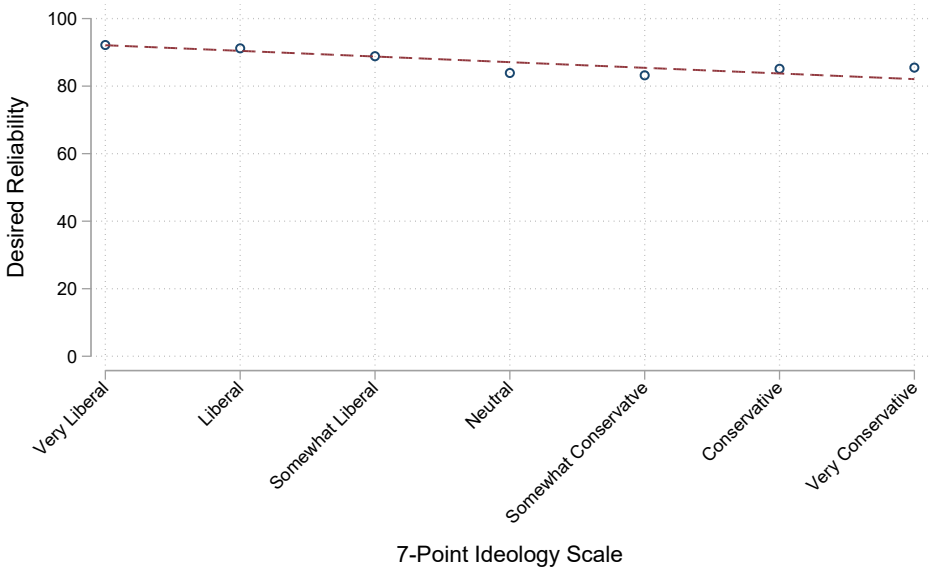
Notes: This figure shows the joint distribution of slant and reliability of the 262 news outlets for which we have both a slant rating from Bakshy, Messing and Adamic (2015) and a reliability measure from NewsGuard.

Figure A.2: bliss point and Political Ideology

(a) Slant

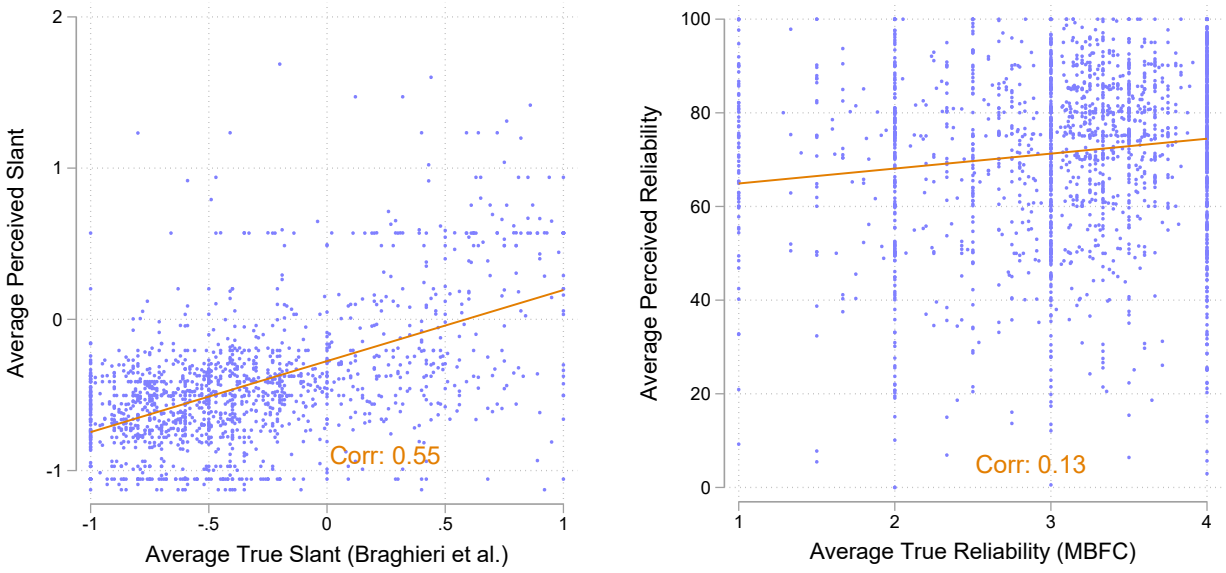


(b) Reliability



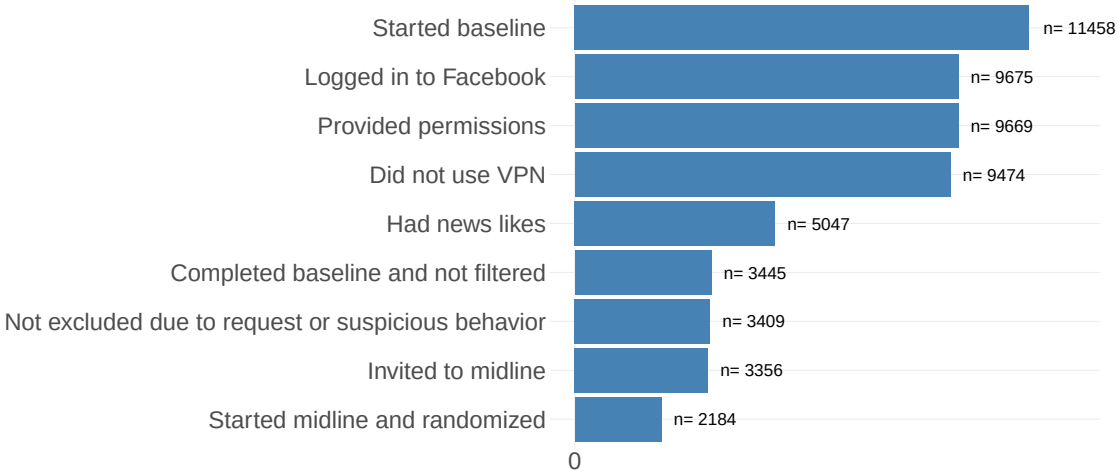
Notes: These figures explore the relationship between participants' stated bliss points on the slant and reliability dimensions and their self-reported political ideology. Specifically, the x-axis shows seven bins of self-reported political ideology, and the y-axis shows the average slant and reliability bliss points for participants whose self-reported ideology falls in the bin on the x-axis.

Figure A.3: Comparing Perceptions to Alternative Definitions of Slant and Reliability



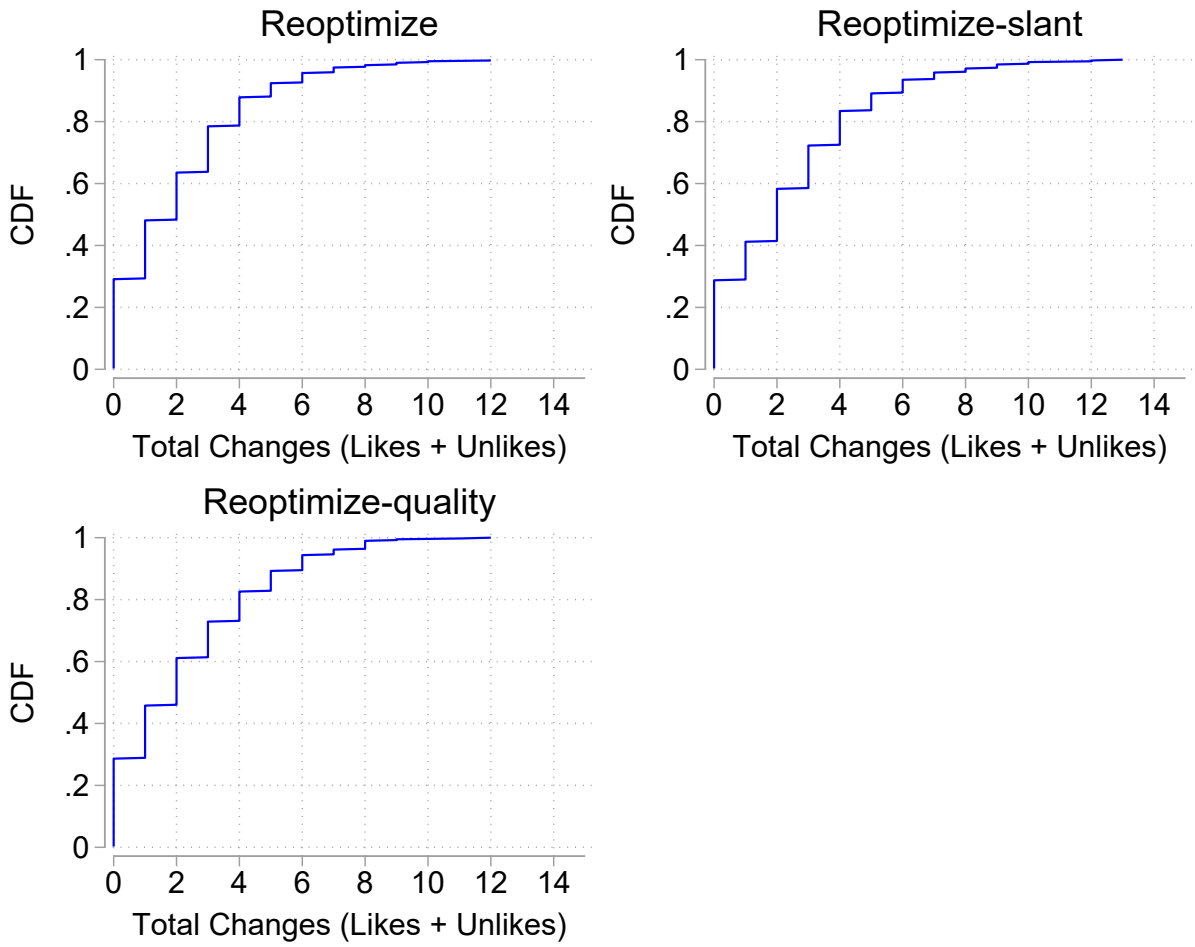
Notes: This figure explores the relationship between participants' perceptions of the average slant and reliability of their news portfolios on Facebook and the actual average slant and reliability of those portfolios, based on alternative measures of slant and reliability. The left panel focuses on slant and the right panel focuses on reliability. Each dot in each figure represents an individual. The y-axis indicates the participants' perceptions of the average reliability and slant of their news portfolio. Slant and reliability perceptions are based on our baseline survey. When these perceptions were elicited, reliability and slant were defined in the same way as in Figure 1, with slant measured based on Bakshy, Messing and Adamic (2015) and reliability based on NewsGuard. The x-axis presents the actual slant and reliability based on alternative measures. On the left panel, actual slant is measured according to Braghieri et al. (2025), where the slant of an outlet is the average slant of all articles published by the outlets. On the right panel, actual reliability is based on Media Bias Fact Check (MBFC), with outlets with low and very low reliability according to MBFC receiving a reliability score of 1, outlets with mixed reliability receiving a score of 2, mostly factual outlets receiving a score of 3, and high or very high reliability outlets receiving a score of 4. The figure also reports Pearson correlations.

Figure A.4: Participants Screened Out Before Randomization



Notes: This figure presents the number of participants in different stages of recruitment. "Completed baseline and not filtered" refers to respondents who live in the USA, self-report getting at least some news on Facebook, passed an attention check, submitted an email address, and did not take the survey before. "Not excluded due to request or suspicious behavior" refers to respondents who did not ask to have their Facebook data deleted and were suspected of being a fraudulent or duplicate response.

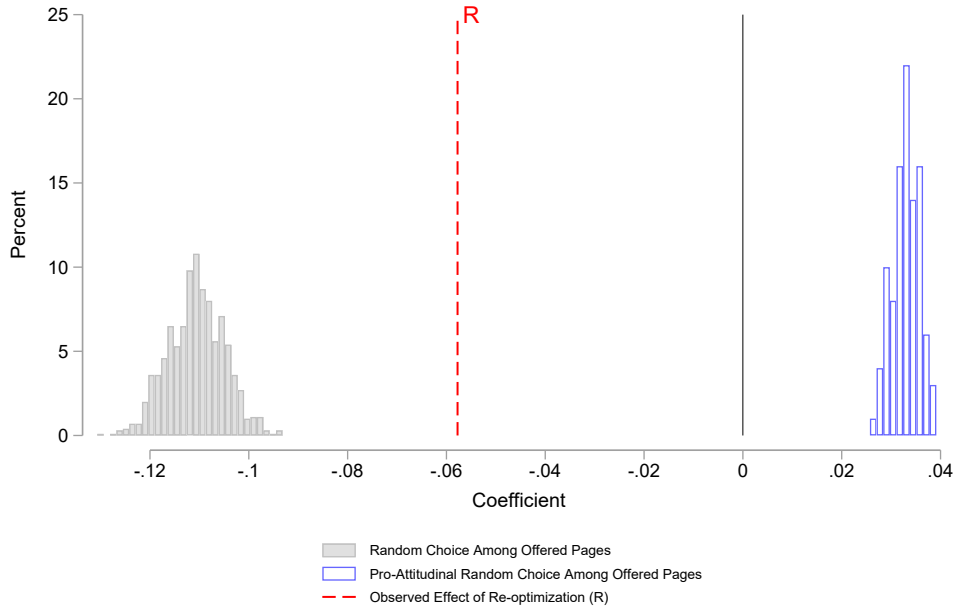
Figure A.5: CDF of portfolio changes by participants at midline



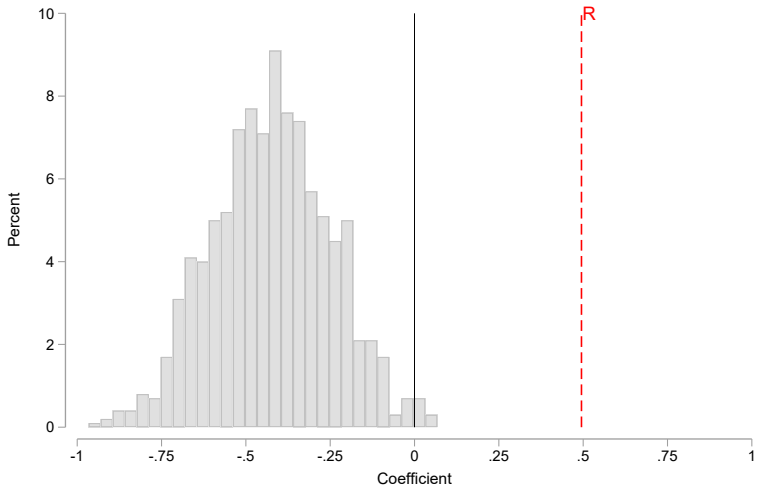
Notes: This figure shows the cumulative distribution function (CDF) of the number of total portfolio changes by participants in the re-optimization treatments during the midline survey. Following F new pages to one's portfolio and unfollowing U pages from one's original portfolio is counted as $F+U$ changes.

Figure A.6: Simulated Estimates of ATE on Absolute Average Slant and Reliability

(a) Absolute Average Slant



(b) Average Reliability



Notes: The gray bars show the distribution of simulated treatment effects on absolute average portfolio slant (Panel a) and average portfolio reliability (Panel b) when pages in our re-optimization interface are followed and unfollowed at random. Specifically, for each participant in the Re-optimization (R) treatment, we first determine how many pages they followed (F) and unfollowed (U) at midline. Next, we randomly select F pages from the set of potentially offered pages as "followed" and randomly select U pages from the participant's original set as "unfollowed." Using these simulated choices, we calculate the resulting absolute average slant and average reliability for each participant's portfolio and re-estimate Equation 2. Each gray bar corresponds to an estimate of Equation 2. The purple bars in Panel a show the distribution of simulated treatment effects on actual absolute average portfolio slant when pro-attitudinal pages are followed. Specifically, for each participant in the Re-optimization (R) treatment, we randomly select F pages from the set of pro-attitudinal potentially offered pages. This means that participants who at baseline had portfolios with average slant < -0.2 (> 0.2) randomly follow F pages from the set of potentially offered pages with slant < -0.2 (> 0.2). Participants who at baseline had portfolios with an average slant between -0.2 and 0.2 randomly follow F pages from the full set of potentially offered pages. Unfollows are not pro-attitudinal: as before, we randomly select U pages from the participant's original set as "unfollowed." Using these simulated choices, we calculate the resulting absolute average slant for each participant's portfolio and re-estimate Equation 2. Each purple bar corresponds to an estimate of Equation 2. We perform 1000 simulations of each type. In both panels, the red dotted vertical line indicates the actual estimated treatment effects on absolute average portfolio slant (Panel a) and average portfolio reliability (Panel b) for the Re-optimization treatment.

B. Measuring Slant and Reliability of News

In this section, we provide additional detail about our measures of slant and reliability.

Slant We measure news slant using the methodology introduced by Bakshy, Messing and Adamic (2015), which determines slant based on the average ideology of Facebook users who share content from a particular news domain. This measure has been widely validated and correlates well with alternative slant metrics (Gentzkow and Shapiro, 2010; Braghieri et al., 2025). It has also been extensively adopted in prior literature studying news slant (Guess, 2021; Peterson, Goel and Iyengar, 2021). The primary advantage of the Bakshy, Messing and Adamic (2015) measure is its comprehensiveness, covering 500 news outlets.

Reliability We measure news reliability using NewsGuard's Source Reliability Ratings from January 2024.[34] NewsGuard employs a team of analysts who assess outlets based on nine criteria of journalistic quality. Each criterion carries a specific number of points:

1. Does not repeatedly publish false content (22 pts)
2. Gathers and presents information responsibly (18 pts)
3. Regularly corrects or clarifies errors (12.5 pts)
4. Handles the difference between news and opinion responsibly (12.5 pts)
5. Avoids deceptive headlines (10 pts)
6. Website discloses ownership and financing (7.5 pts)
7. Clearly labels advertising (7.5 pts)
8. Reveals who is in charge, including possible conflicts of interest (5 pts)
9. Provides the names of content creators, along with either contact or biographical information (5 pts)

If an outlet meets a given criterion, it receives the full points allocated; otherwise, it receives zero points. Analysts prepare a detailed "Nutrition Report" documenting and justifying each rating. NewsGuard also reaches out to outlets for comment, and each assessment is reviewed by an experienced editor. Ultimately, each outlet receives a reliability rating ranging from 0 to 100. NewsGuard ratings are also highly comprehensive, covering thousands of outlets, and have been increasingly utilized in recent studies on news reliability (Pennycook and Rand, 2021; Allen et al., 2021; Bhadani et al., 2022).

We merge the datasets containing slant and reliability measures, manually matching them to active Facebook pages. The resulting dataset used in this study consists of 262 unique domains.[35]

[34] www.newsguardtech.com/solutions/news-reliability-ratings

[35] We combine outlets with similar domains, such as huffingtonpost.ca and huffingtonpost.com, into a single domain.

C. Selection of News Pages

We provided each participant in our re-optimization treatments with up to 12 news pages that they could follow on Facebook. The selection of which pages to show each participant is described below.

Outlets were classified into six distinct categories based on their slant and reliability, and within each category, outlets were ranked according to their popularity on Facebook. We only included U.S.-focused outlets dedicated to hard news. For each category, we chose as defaults two prominent outlets among the top four most popular outlets. If a participant already followed one of these outlets, we offered the next most popular outlet in that category instead. We detail these categories and their rankings below. Table A.20 lists the top 3-4 outlets potentially offered in each category. In rare cases where a user already liked all or most of these outlets, they were offered the next most popular outlet in that category (not shown in the table) if such an outlet was available.

The six categories used for selecting outlets were as follows: higher-reliability conservative, lower-reliability conservative, higher-reliability moderate, lower-reliability moderate, higher-reliability liberal, and lower-reliability liberal.

As discussed, we measured slant according to the methodology of Bakshy, Messing and Adamic (2015). Outlets with a slant score greater than 0.25 were classified as conservative, those with a slant between -0.25 and 0.25 as moderate, and those with a slant below -0.25 as liberal.

We assessed reliability using NewsGuard's Source Reliability Ratings. Higher-reliability outlets were defined as those scoring 92.5 or above, meaning they failed at most one low-weight criterion as detailed in Appendix B. Lower-reliability outlets were defined as those scoring between 60 and 75. We did not include outlets with scores below 60 in the re-optimization interface to avoid promoting misinformation.

To determine popularity, we manually collected the number of Facebook follows (or likes) for each page as of January 2022. We also manually verified whether each outlet focused specifically on hard news (news covering recent events of general significance) and whether the outlet primarily targeted U.S. audiences.

Table A.20: Most common outlets offered

Name	Group	Rank	Reliability	Slant
The Wall Street Journal	Higher reliability conservative	1	100	0.28
National Review	Higher reliability conservative	2	92.5	0.9
The Daily Caller	Higher reliability conservative	3	100	0.87
Washington Examiner	Higher reliability conservative	4	92.5	0.82
Fox News	Lower reliability conservative	1	69.5	0.8
The Political Insider	Lower reliability conservative	2	69.5	0.9
Opposing Views	Lower reliability conservative	3	75	0.27
MRCTV	Lower reliability conservative	4	62	0.86
Business Insider	Higher reliability moderate	1	100	-0.06
USA TODAY	Higher reliability moderate	2	100	-0.06
Yahoo News	Higher reliability moderate	3	100	0.05
Today Show	Higher reliability moderate	4	95	-0.17
ABC News	Lower reliability moderate	1	75	-0.16
AOL	Lower reliability moderate	2	75	0.01
New York Post	Lower reliability moderate	3	69.5	0.25
The New York Times	Higher reliability liberal	1	100	-0.55
NPR	Higher reliability liberal	2	100	-0.61
NBC News	Higher reliability liberal	3	100	-0.27
TIME	Higher reliability liberal	4	100	-0.33
Mother Jones	Lower reliability liberal	1	69.5	-0.87
VICE	Lower reliability liberal	2	70	-0.63
Upworthy	Lower reliability liberal	3	75	-0.81
Atlanta Black Star	Lower reliability liberal	4	75	-0.71

Notes: This table presents the two outlets we offered participants, along with the top two alternatives. The top two outlets in each category are the ones we offered participants by default (these are always two of the four most popular outlets in each category). The rest of the outlets are ranked by popularity and were offered in case previous outlets were already followed. Only three outlets are shown for lower reliability moderate outlets because only three outlets fit our inclusion criteria. For more details, see Appendix C.

D. Implementation: Additional Details

Participants were invited to take the surveys in three sequential batches. Recruitment for the baseline survey took place in 2025 through targeted Facebook ads. Ads for the first batch ran from January 14th to January 24th, ads for the second batch from January 25th to February 13th, and ads for the final batch from February 21st to March 2nd.

On average, participants received invitations for the midline survey approximately one week after completing the baseline survey, and for the endline survey approximately one month after baseline. Invitations and multiple reminders for both midline and endline surveys were distributed via email.

Upon starting the baseline survey, participants were asked to grant us permission to collect information about the pages they follow on Facebook. We collected the data using three complementary methods:

1. Survey-based follows: Real-time collection of the full set of pages participants follow. The data was collected at the beginning and the end of each survey.

2. Webhooks-based follows: Notifications from Facebook's webhooks whenever a participant followed or unfollowed a page.

3. Snapshot-based follows: Weekly snapshots capturing all pages each participant followed and the corresponding dates they began following them.

Utilizing these three methods allowed us to maintain high-quality data and promptly detect and rectify any potential inconsistencies or errors.

We nonetheless encountered two issues during data collection:

1. We temporarily lost access to participants' follow data on Feb 13^{th} 2025 due to technical issues with our Facebook app. Consequently, we paused data collection during this period. Fortunately, most participants re-granted us access to their data upon completing the subsequent survey. This interruption means that for the first batch, data between the midline and endline surveys is partially missing, while for the second batch, data between baseline and midline surveys is partially missing (no webhooks-based follows or snapshots were collected during this period). Importantly, data on pages followed or unfollowed during the survey itself remained unaffected. We describe below how we addressed these gaps.

2. In certain cases, data obtained through survey-based and snapshot-based methods was incomplete. Whenever we suspected partial data, we cross-referenced with the webhooks data, which was not affected by this issue, to ensure accuracy.

We used these collected data to construct the following variables:

- Baseline follows: Determined using survey-based follow data. Participants for whom we could not collect survey-based data at baseline were not invited to participate in the midline survey. We cross-checked this data with later snapshots. In rare cases, we retrospectively identified partial baseline data, resulting in participants having been offered pages they already followed. These 53 participants were excluded from our final analysis dataset.

- Post-midline follows: We primarily relied on webhooks data to capture changes in follows between baseline and midline surveys, as well as pages followed or unfollowed during the midline survey. For the minority of participants affected by the temporary Facebook permissions issue (and therefore lacking data between baseline and midline), we assumed no changes occurred during this period. This assumption is reasonable, as data from unaffected users indicated minimal activity between baseline and midline surveys.

- Pre-endline and post-endline follows: To identify the list of follows before and after the endline survey, we first located the earliest complete snapshot taken after endline completion. We then combined this snapshot with webhooks data to account for any changes occurring between the beginning of the endline survey and the snapshot.[36] This approach ensured accuracy despite the temporary loss of Facebook access prior to the endline survey for some users. In the very few cases where this procedure did not yield complete data, we used the survey-based follow records as a fallback.

E. Perceptions of Slant and Reliability of Mainstream Outlets

In this appendix, we examine participants' perceptions of the slant and reliability of several mainstream news outlets. The findings echo those in Section 2.2, which focuses on the perceptions of the slant and reliability of the outlets participants actually follow. As before, individuals appear to have a better grasp of ideological slant than of reliability. We also find that Democrats and Republicans hold similar perceptions of outlets' slant, but differ markedly in their perceptions of reliability.

In the baseline survey, in addition to eliciting perceptions of the average slant and reliability of participants' own news portfolios, we also asked about six major outlets: MSNBC, The New York Times, CNN, USA Today, The Wall Street Journal, and Fox News.

One potential concern is that participants may correctly understand which outlets are more conservative or more reliable relative to each other, but misperceive the overall scales; for instance, they might believe that all outlets are less reliable than they actually are. To address this, we also analyze the implied rankings of outlets. For each participant, we construct a rank order of the six outlets based on perceived slant or reliability, then compute the average rank for each outlet across participants.

Figures A.7, A.8, A.9, and A.10 display average perceptions among baseline participants. In each figure, green dots indicate the overall average perception; black diamonds show the true value; blue plus signs mark averages among self-identified Democrats; and red plus signs among self-identified Republicans.

As in the left panel of Figure 1, Figures A.7 and A.8 show that participants' perceptions of slant are broadly accurate. They correctly recognize, for instance, that Fox News is more conservative than The Wall Street Journal, which in turn is to the right of USA Today, while USA Today is more conservative than CNN, The New York Times, and MSNBC. Strikingly, Democrats and Republicans tend to agree on the relative slant of these outlets, even for those that are counter-attitudinal.

[36] To detect incomplete snapshots, we first determined the maximum number of pages followed in any snapshot. We considered any snapshot incomplete if it recorded at least 50 fewer pages than the maximum snapshot for that user.

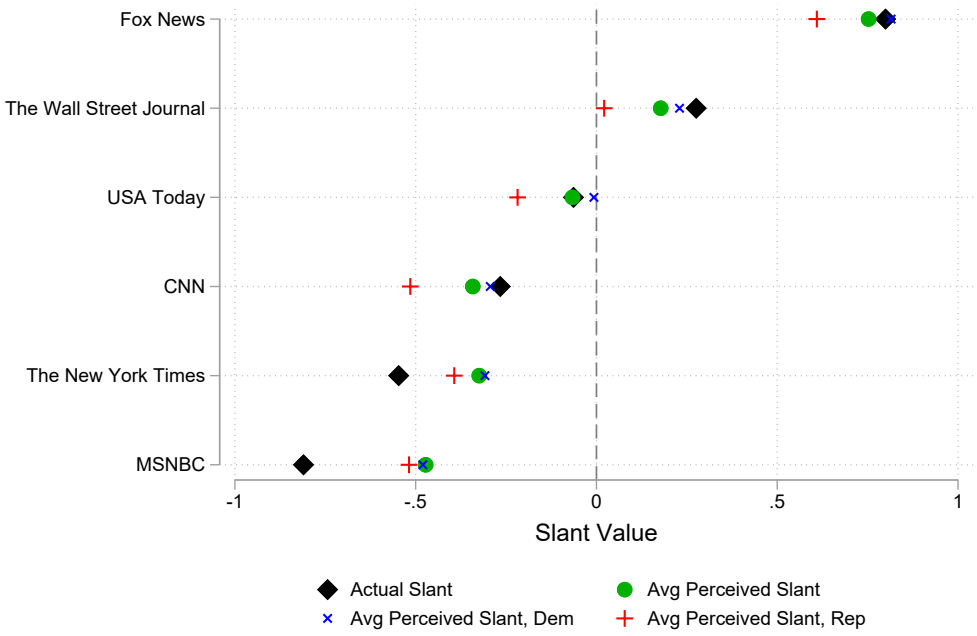
In contrast, as in the right panel of Figure 1, Figures A.9 and A.10 show that participants consistently underestimate the reliability of most outlets and misrank them relative to the truth. For example, many believe that MSNBC and CNN are at least as reliable as The Wall Street Journal, contrary to NewsGuard ratings. Moreover, Republicans and Democrats diverge sharply in their assessments of the reliability of The New York Times, CNN, Fox News, and MSNBC, with each group rating ideologically aligned outlets as more reliable (consistent with (Gentzkow and Shapiro, 2006))

We quantify the relative accuracy of slant and reliability perceptions by comparing participants' perceived rankings to the actual rankings using Kendall's tau-b, computed individually for each baseline respondent. Figure A.11 confirms that perceived rankings of slant are substantially more accurate than those for reliability, consistent with the analysis in Section 2.2.

Finally, we test whether our information treatments affected participants' perceptions. At midline, we re-elicited perceptions for five mainstream outlets—two already included at baseline (MSNBC, Fox News), and three new ones (The Atlantic, The Washington Post, The National Review). For each participant, we compute the absolute distance between perceived and actual slant and reliability, averaged across outlets, and estimate Equation 2. Appendix Table A.21 presents the results. Column (1) reports treatment effects on slant misperceptions; column (2) on reliability misperceptions.

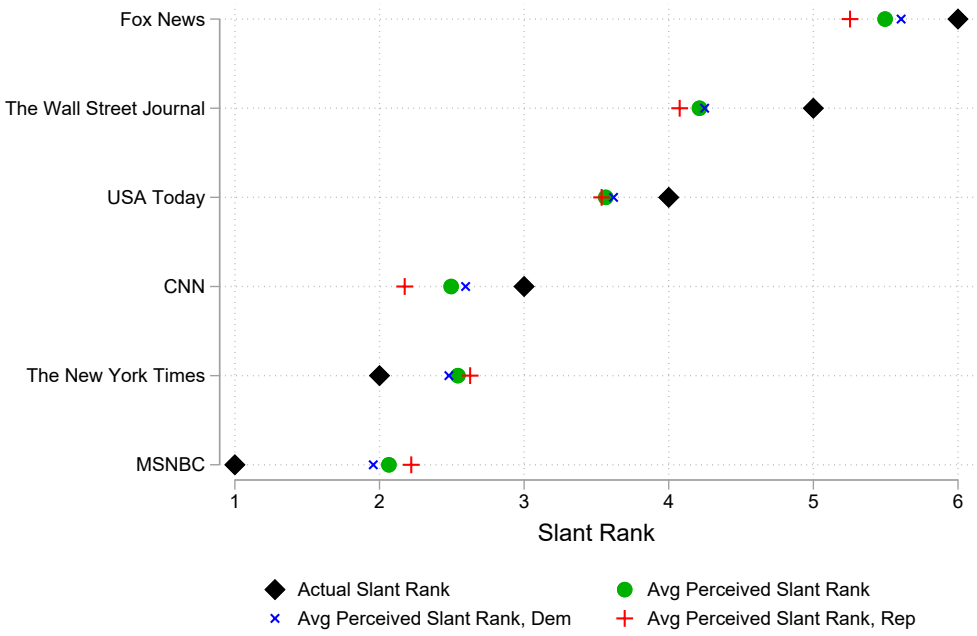
We find no detectable effect of the RS treatments on slant misperceptions. However, the RR treatment significantly reduces reliability misperceptions by 5.6 points. The NR treatment, which provides the same information without the re-optimization interface, also reduced reliability misperceptions. These findings reinforce the baseline pattern: participants already have reasonably accurate perceptions of slant, but substantial misperceptions about reliability. This helps explain why information about reliability—but not about slant—was a valuable addition to the re-optimization interface. Interestingly, the RR treatment increased slant misperceptions slightly, possibly reflecting a perceived correlation between slant and reliability.

Figure A.7: Average Perceived Slant of Mainstream Outlets



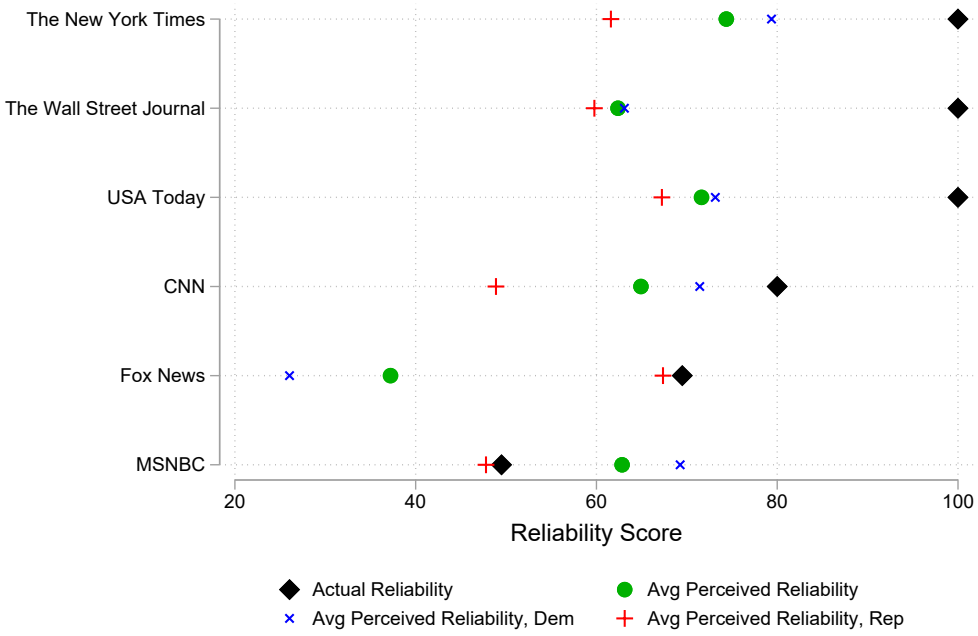
Notes: This figure plots the average perceived slant of six mainstream outlets (green dots) compared to the actual slant of the same outlets (black diamonds). The blue plus signs indicate the average perceived slant among self-reported Democrats, and the red plus signs indicate the average perceived slant among self-reported Republicans.

Figure A.8: Average Perceived Slant Rank of Mainstream Outlets



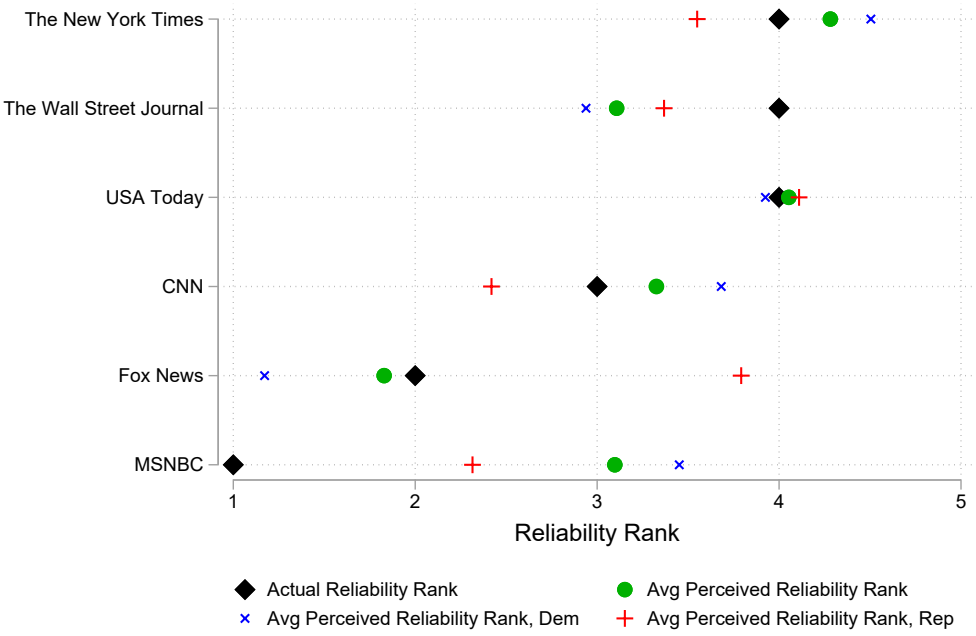
Notes: This figure plots the average perceived slant rank of six mainstream outlets (green dots) compared to the actual slant rank of the same outlets (black diamonds). In other words, for each individual, we create a ranking over the six outlets based on perceived slant (ties are given the same rank, and a gap is left in the ranking). Then, for each outlet, we average the rankings across participants. The blue plus signs indicate the average perceived slant rank among self-reported Democrats, and the red plus signs indicate the average perceived slant rank among self-reported Republicans.

Figure A.9: Average Perceived Reliability of Mainstream Outlets



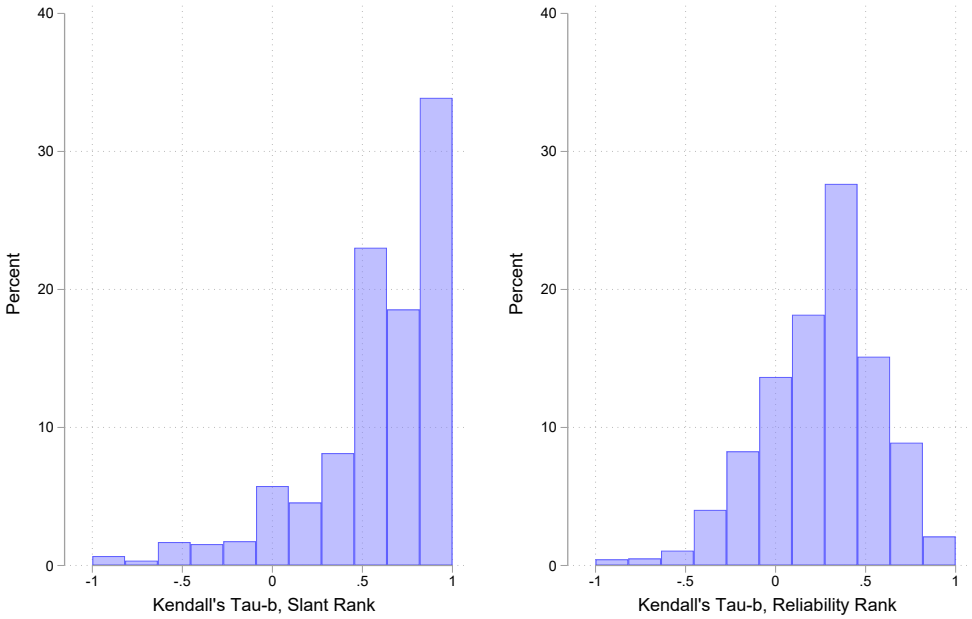
Notes: This figure plots the average perceived reliability of six mainstream outlets (green dots) compared to the actual reliability of the same outlets (black diamonds). The blue plus signs indicate the average perceived reliability among self-reported Democrats, and the red plus signs indicate the average perceived reliability among self-reported Republicans.

Figure A.10: Average Perceived Reliability Rank of Mainstream Outlets



Notes: This figure plots the average perceived reliability rank of six mainstream outlets (green dots) compared to the actual reliability rank of the same outlets (black diamonds). In other words, for each individual, we create a ranking over the six outlets based on perceived reliability (ties are given the same rank, and a gap is left in the ranking). Then, for each outlet, we average the rankings across participants. The blue plus signs indicate the average perceived reliability rank among self-reported Democrats, and the red plus signs indicate the average perceived reliability rank among self-reported Republicans.

Figure A.11: Ranking Accuracy: Kendall's Tau-b



Notes: For each participant in the baseline survey, we compute Kendall's tau-b to measure the association between the participant's perceived ranking of the outlet and the outlet's actual ranking, over six outlets, for both slant and reliability. Kendall's tau-b is a rank correlation coefficient that measures the ordinal association between two rankings, adjusting for ties in both. It is based on the difference between the number of concordant and discordant item pairs, normalized by a factor that accounts for tied values. A value of 1 indicates perfect agreement, 0 indicates no association, and –1 indicates perfect disagreement. In the above figures, we plot the distribution of Kendall's tau-b, for rankings of slant (left) and reliability (right).

Table A.21: ATE on Misperceptions

	(1) Average Slant Misperceptions	(2) Average Reliability Misperceptions
Re-optimization (R)	0.022	-0.068
	(0.016)	(0.946)
Re-optimization + Slant Info (RS)	-0.013	-0.384
	(0.016)	(0.933)
Re-optimization + Reliability Info (RR)	0.048***	-5.829***
	(0.016)	(0.975)
Reliability Info (NR)	0.043**	-3.818***
	(0.017)	(0.978)
p-value, R=RS	0.04	0.74
p-value, R=RR	0.12	0.00
p-value, RR=NR	0.77	0.05
Control Mean	0.41	34.84
Control SD	0.21	12.71
Observations	1813	1824
R-squared	0.011	0.032

Notes: This table explores the effect of our intervention on misperceptions of slant and reliability. Specifically, it presents estimates of coefficients β_j from Equation 2, where the outcome variable is the distance between the perception and the true value, averaged over the five mainstream outlets that were asked about at the end of the midline survey. We report the results for slant misperceptions in column (1), and for reliability misperceptions in column (2). All regressions include pre-specified controls. Our controls consist of: baseline political ideology, baseline degree of reliance on Facebook for news, baseline interest in politics, wave fixed effects, device fixed effects (desktop or mobile), recruitment and randomization fixed effects. Standard errors in parentheses are robust.

F. Survey Screenshots

This section shows screenshots from our baseline and midline surveys. Note that the surveys were personalized and these screenshots are just examples. Different participants received different information and offers.

F.1 Information Interventions

Figure A.12: Slant Information for RS



Notes: This is the information about slant that was provided in the RS treatment, early in the midline survey, prior to the re-optimization interface with slant labels.

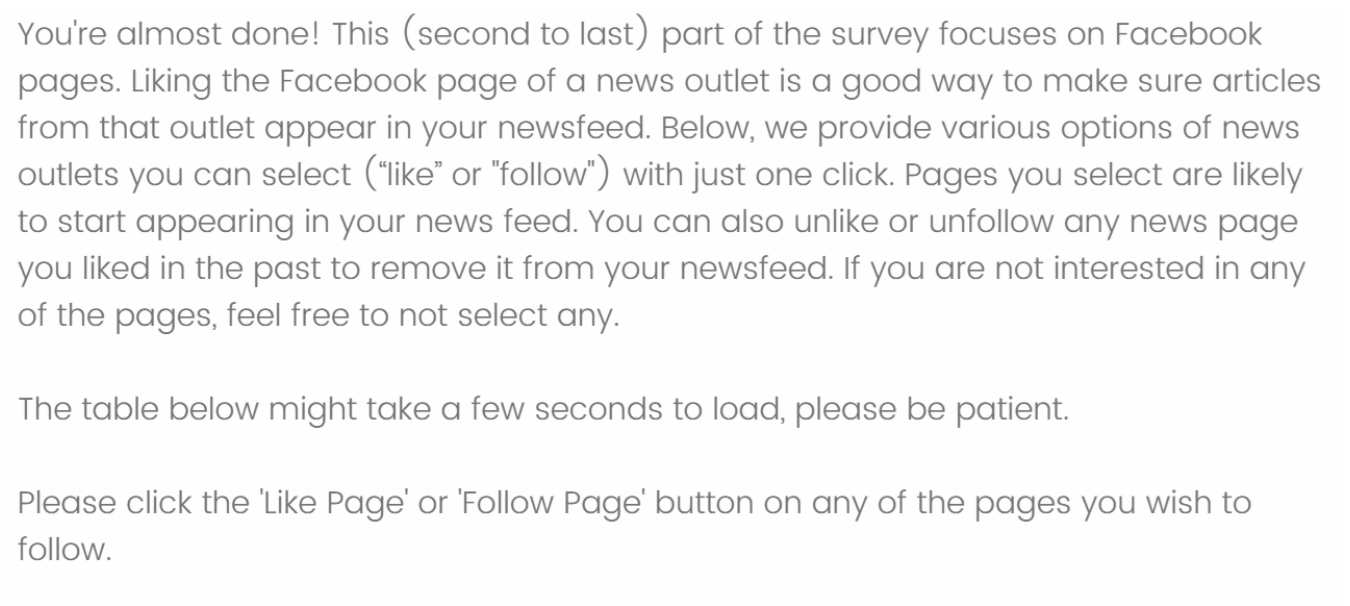
Figure A.13: Reliability Information for RR and NR



Notes: This is the information about reliability that was provided in the RR and NR treatments, early in the midline survey. Participants in the RR treatment were shown this page prior to the re-optimization interface with reliability labels. Participants in the NR were not shown the re-optimization interface, but they were shown a list of outlets that would have been offered in the re-optimization interface had they been assigned to RR, along with their reliability labels.

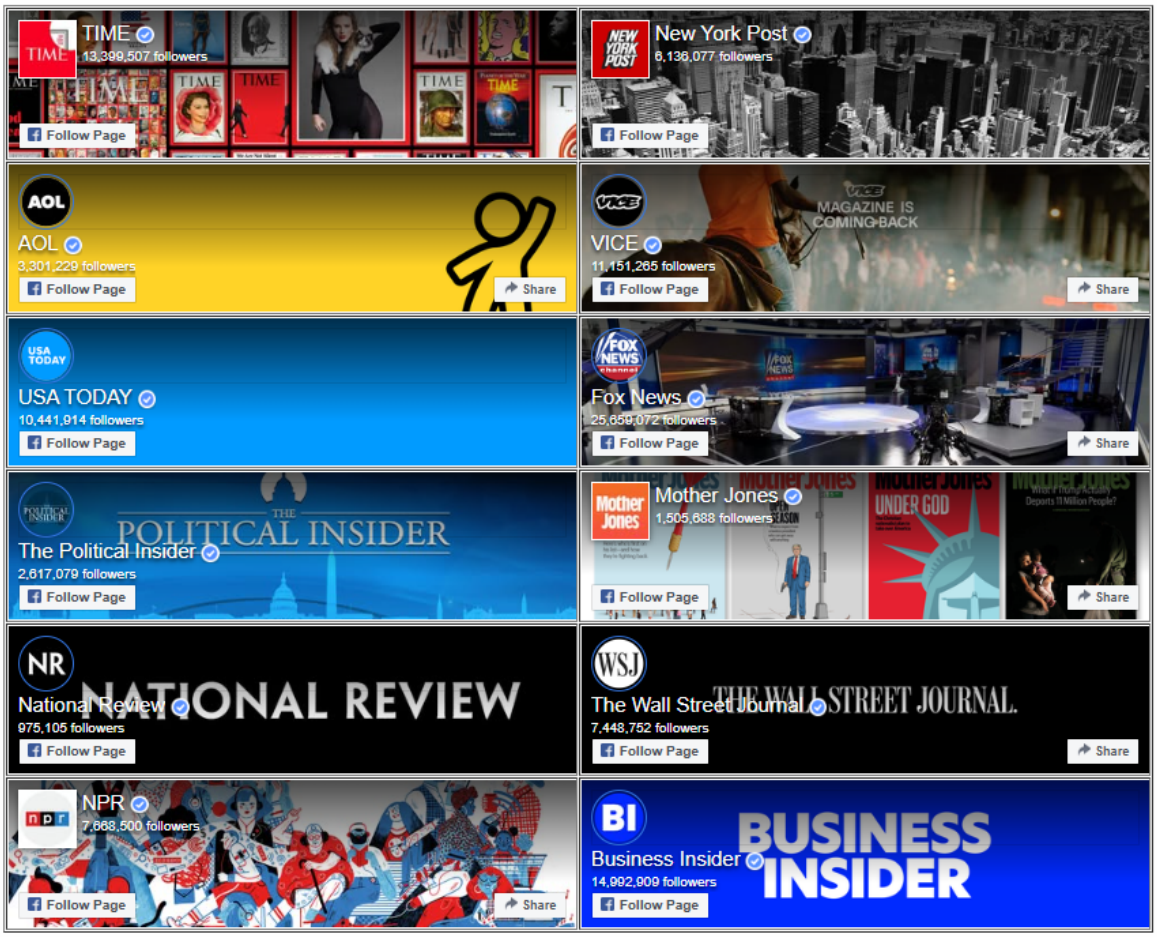
F.2 Like and Unlike Tables

Figure A.14: Instructions for Liking/Unliking Pages



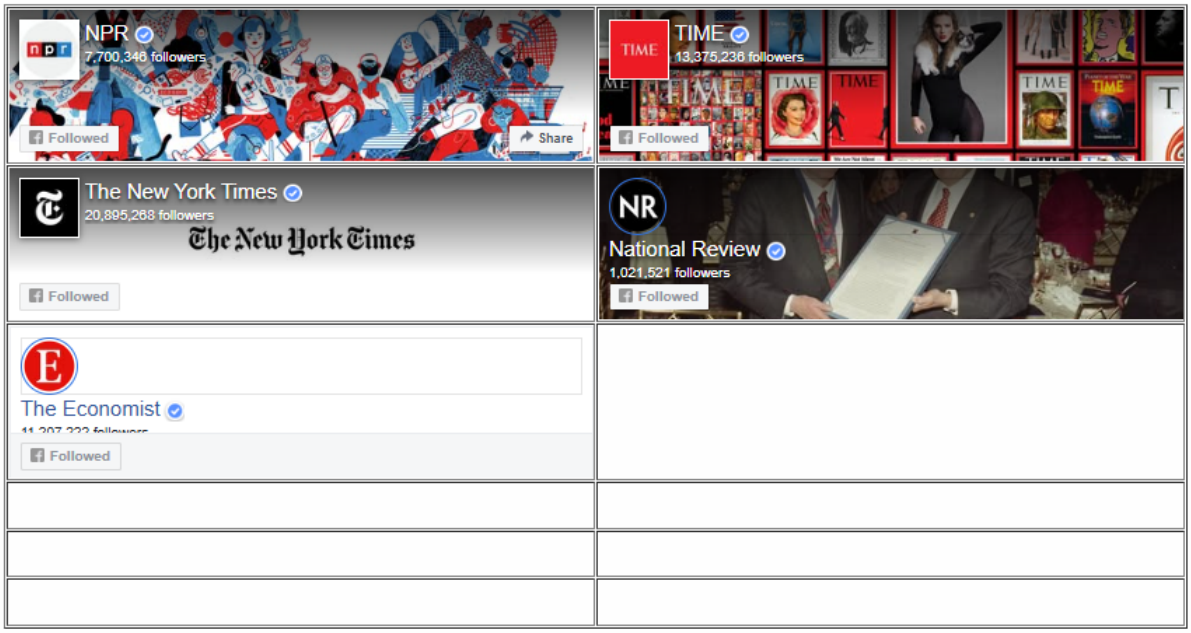
Notes: These are the instructions provided to participants regarding the re-optimization interface toward the end of the midline survey.

Figure A.15: Likes Table, Re-optimization (R) Treatment



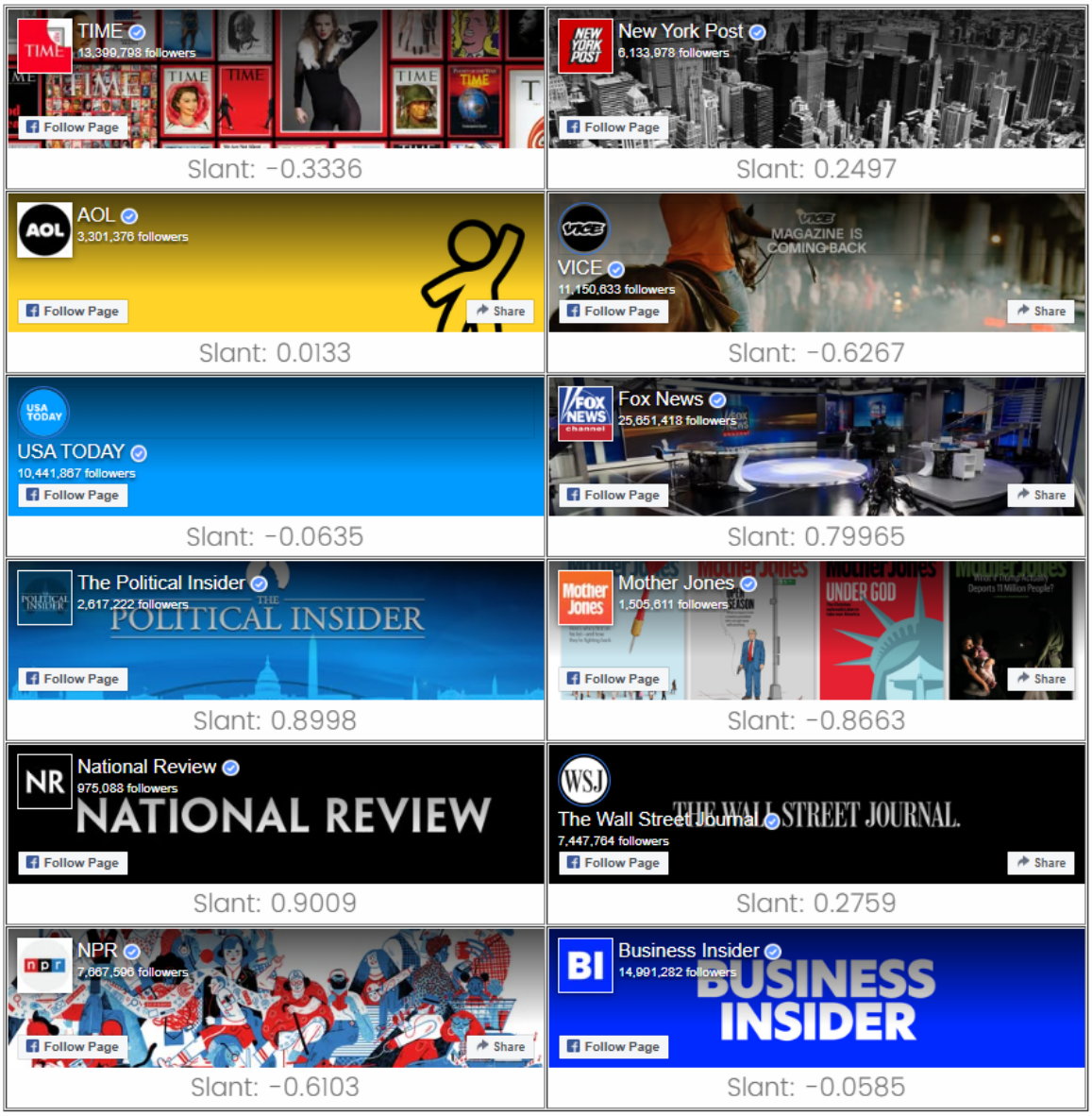
Notes: This is the first table within the re-optimization interface at the end of the midline survey. This version of the table has no labels and was shown to participants in the R treatment. Participants were able to directly click the "Follow Page" button to follow any of the pages in the table, in which case the button would turn gray and would read "Followed". The table included 12 outlets that the participant did not already follow, spanning the slant spectrum and with either medium or high reliability. Appendix C provides more information about how we selected the news pages for this table.

Figure A.16: Unlikes Table, Re-optimization (R) Treatment



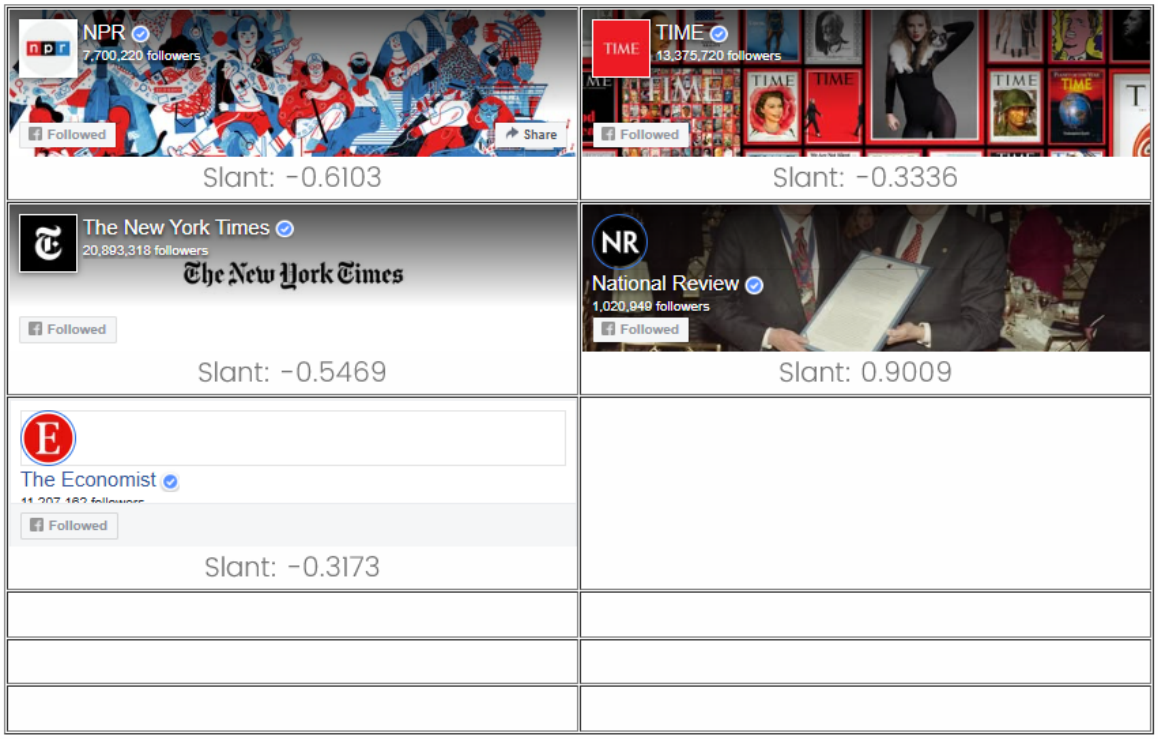
Notes: This is the second table within the re-optimization interface at the end of the midline survey. This version of the table has no labels and was shown to participants in the R treatment. Participants were able to directly click the "Followed" button to unfollow any of the pages in the table, in which case the button would turn blue and would read "Follow Page". Participants were shown up to 12 news pages they already followed. If a participant's Facebook news portfolio contained more than 12 pages, we randomly selected 12 pages to display.

Figure A.17: Likes Table, Re-optimization Slant-Info (RS) Treatment



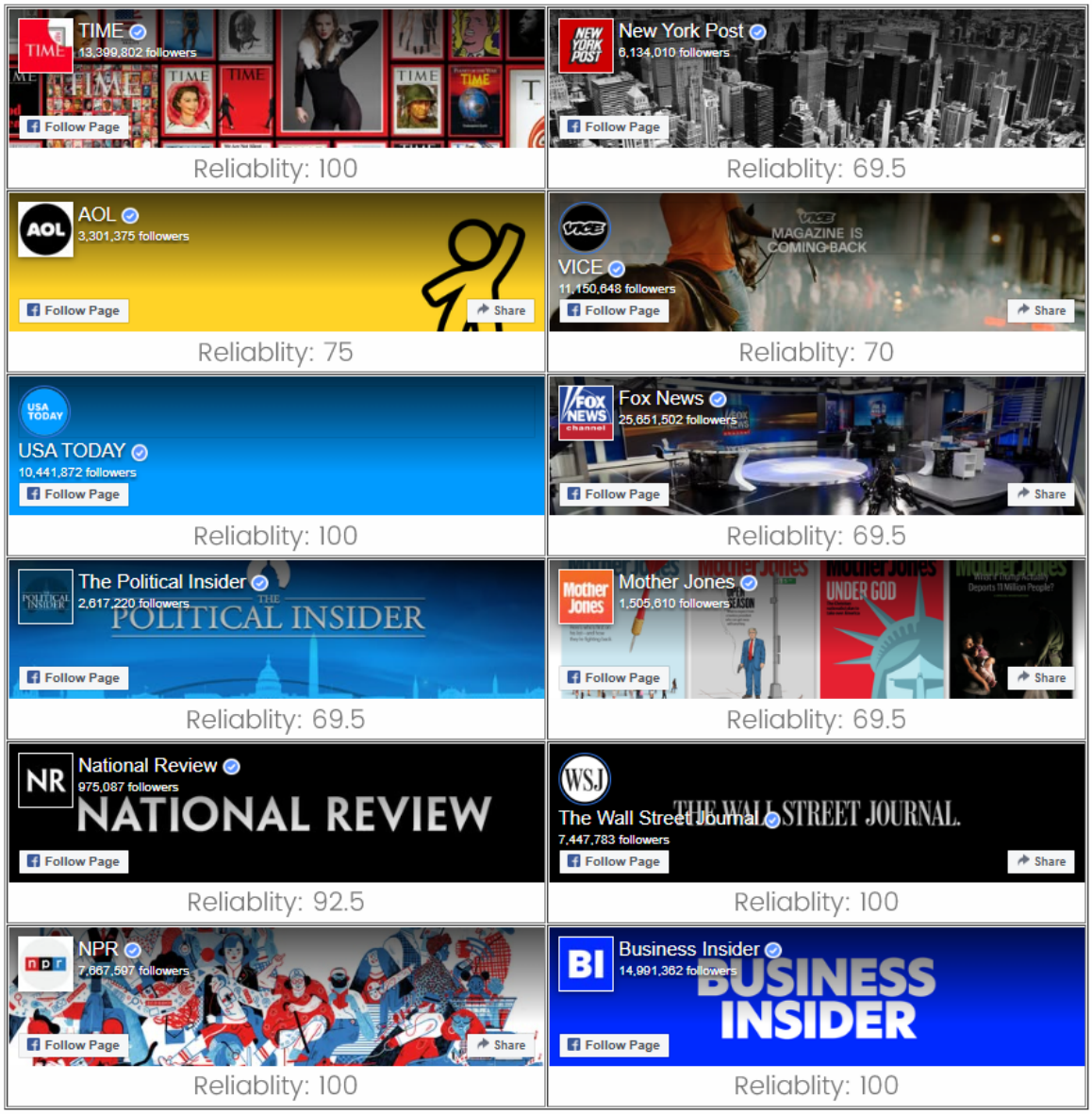
Notes: This is the first table within the re-optimization interface at the end of the midline survey. This version of the table has slant labels and was shown to participants in the RS treatment. Participants were able to directly click the "Follow Page" button to follow any of the pages in the table, in which case the button would turn gray and would read "Followed". The table included 12 outlets that the participant did not already follow, spanning the slant spectrum and with either medium or high reliability. Appendix C provides more information about how we selected the news pages for this table.

Figure A.18: Unlikes Table, Re-optimization Slant-Info (RS) Treatment



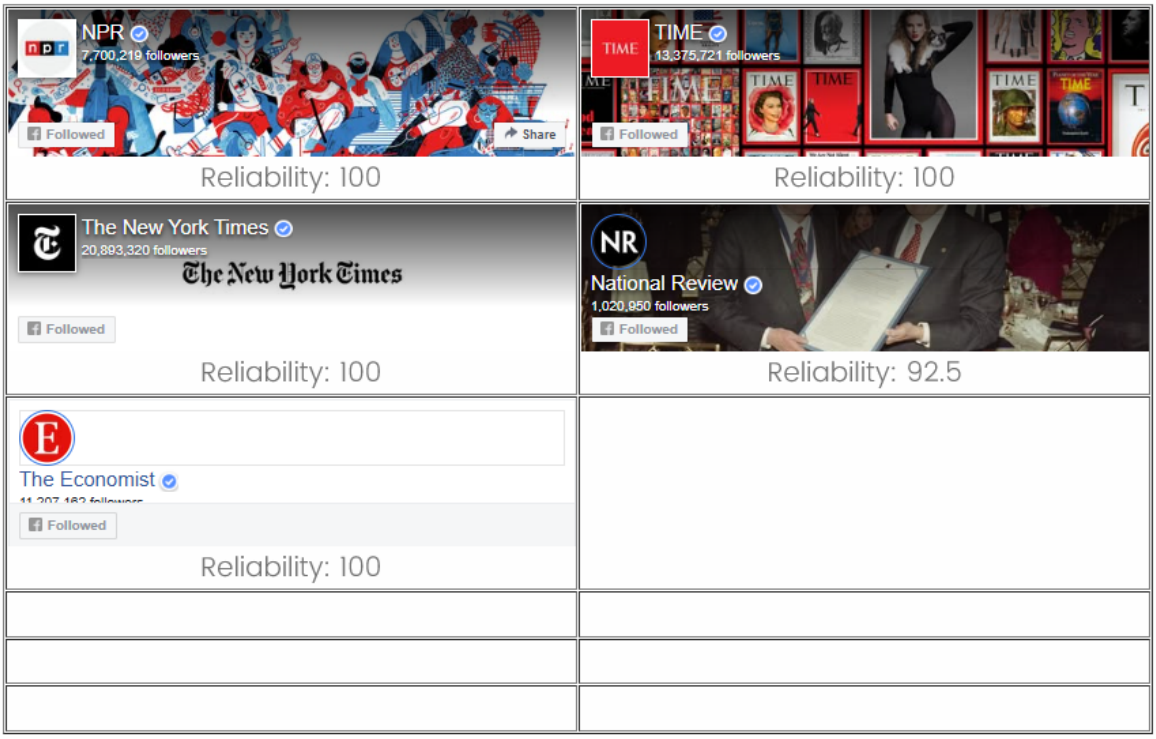
Notes: This is the second table within the re-optimization interface at the end of the midline survey. This version of the table has slant labels and was shown to participants in the RS treatment. Participants were able to directly click the "Followed" button to unfollow any of the pages in the table, in which case the button would turn blue and would read "Follow Page". Participants were shown up to 12 news pages they already followed. If a participant's Facebook news portfolio contained more than 12 pages, we randomly selected 12 pages to display.

Figure A.19: Likes Table, Re-optimization Reliability-Info (RR) Treatment



Notes: This is the first table within the re-optimization interface at the end of the midline survey. This version of the table has reliability labels and was shown to participants in the RR treatment. Participants were able to directly click the "Follow Page" button to follow any of the pages in the table, in which case the button would turn gray and would read "Followed". The table included 12 outlets that the participant did not already follow, spanning the slant spectrum and with either medium or high reliability. Appendix C provides more information about how we selected the news pages for this table. Participants in the NR treatment were not shown this interface, but they were shown a list of the outlets they would have been offered had they been assigned RR, along with their reliability labels.

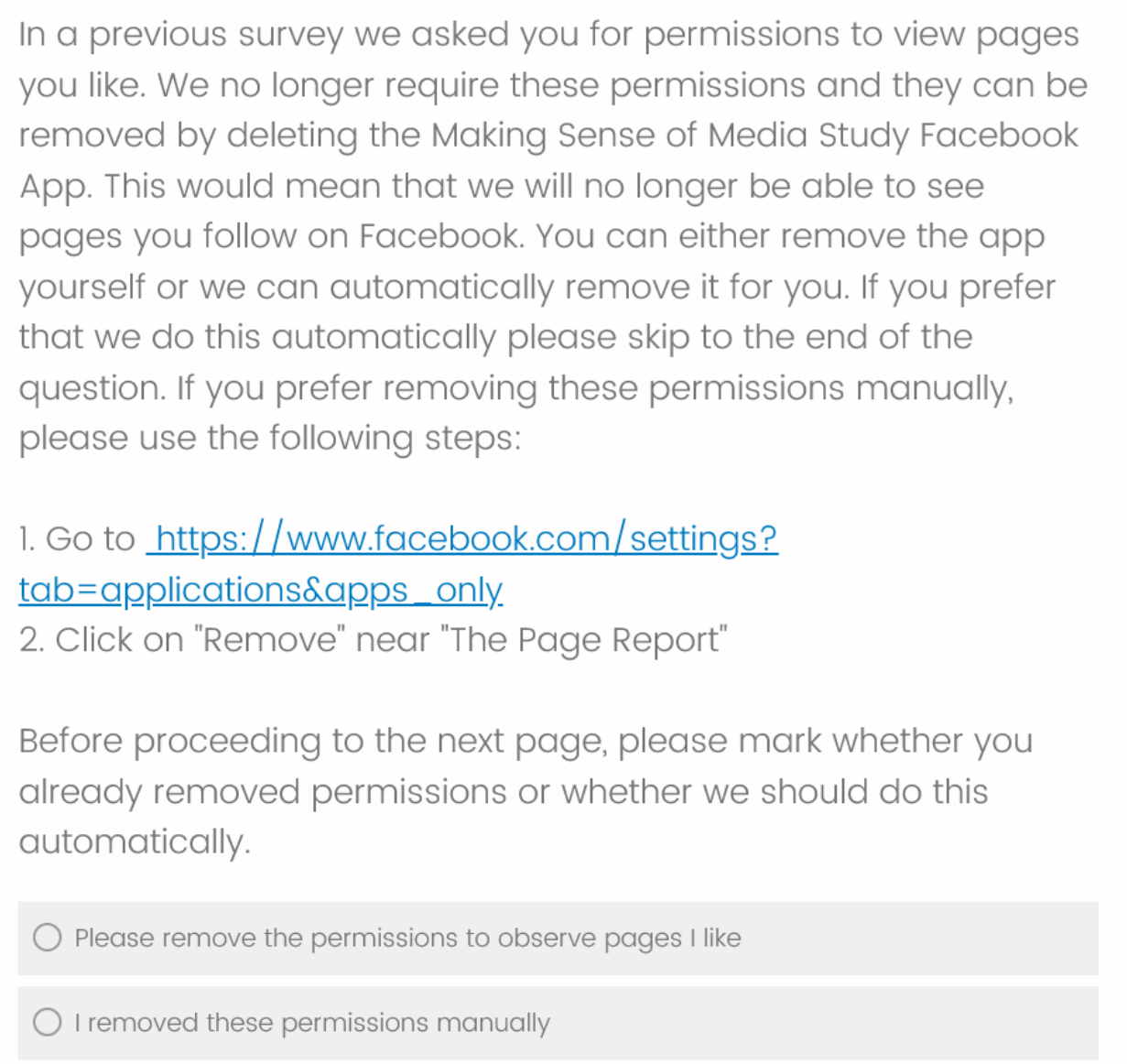
Figure A.20: Unlikes Table, Re-optimization Reliability-Info (RR) Treatment



Notes: This is the second table within the re-optimization interface at the end of the midline survey. This version of the table has reliability labels and was shown to participants in the RR treatment. Participants were able to directly click the "Followed" button to unfollow any of the pages in the table, in which case the button would turn blue and would read "Follow Page". Participants were shown up to 12 news pages they already followed. If a participant's Facebook news portfolio contained more than 12 pages, we randomly selected 12 pages to display.

F.3 Experimenter Demand

Figure A.21: Instructions for NED (No Experimenter Demand) Group



Notes: These are the instructions provided to participants in the No Experimenter Demand (NED) group. They were shown near the beginning of the midline survey, immediately after the randomization (and before information about reliability was provided). If participants clicked "Please remove the permissions to observe pages I like," we removed the permissions for them. If participants clicked "I removed these permissions manually" then we assumed they did so themselves.

F.4 Explanation of Slant and Reliability Measures and Belief Elicitation

Figure A.22: Explanation of Bakshy Measure for Slant

The slant of an outlet is its political leaning. Some outlets may have a liberal slant in that their content is more aligned with the views of the Democratic party, while others may have a conservative slant in that their content is more aligned with the views of the Republican party. It is also possible that an outlet is not biased and does not lean toward either party.

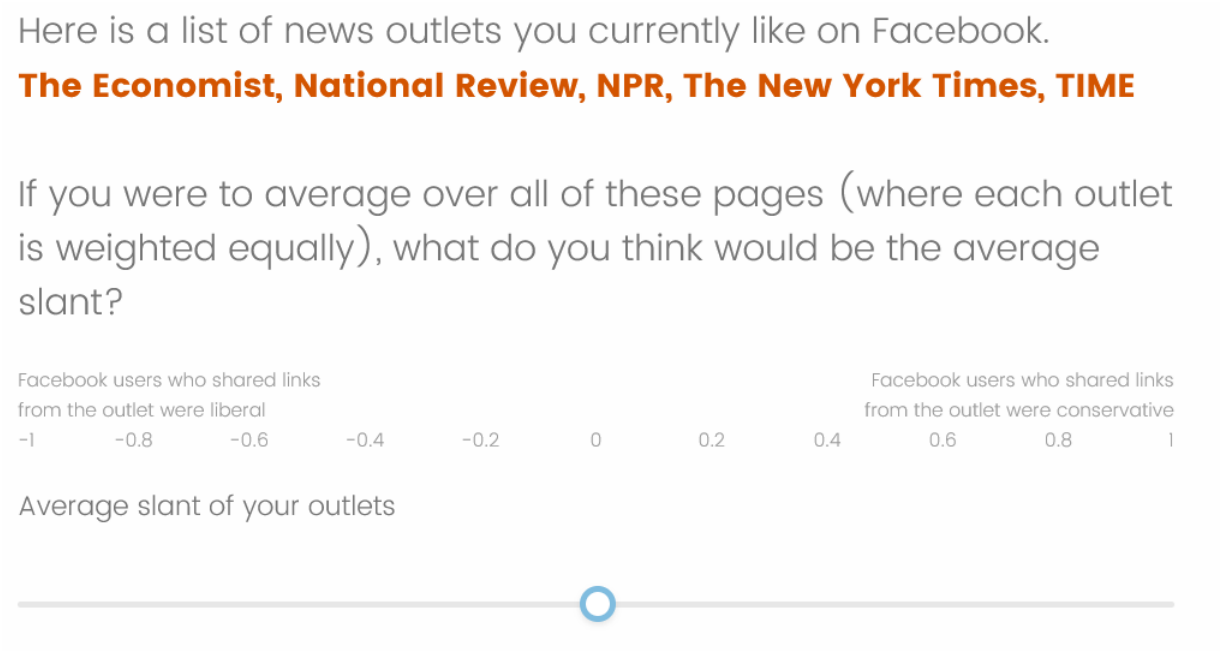
We use a measure of slant based on the ideology of people who share an outlet's content on social media. Here's how the measure works. First, it takes all of the hard news that an outlet posted on Facebook. ("Hard news" is typically used to refer to topics like politics, international affairs and business news, as opposed to "soft news" which refers to topics like entertainment, celebrity, and lifestyle news.) Second, it calculates the ideology of individuals who shared hard news on a five-point scale: -2 for individuals who are very liberal, -1 for liberals, 0 for moderate, +1 for conservative, and +2 for individuals who are very conservative. Finally, slant is defined as the average ideology of all the individuals who shared links from the outlet.

In practice the slant almost always ranges between -1 (very liberal) and +1 (very conservative).If everyone who shared links from the outlet is a liberal, the outlet gets a score of -1; if everyone who shared links is moderate or if an equal share of liberals and conservatives shared links, the article gets a score of 0; if everyone who shared links from the outlet is conservative, the outlet gets a score of +1.

The Bottom Line:
From now on, whenever we talk about the "slant" of an outlet, we are referring to the ideology of the people who share its content on social media.

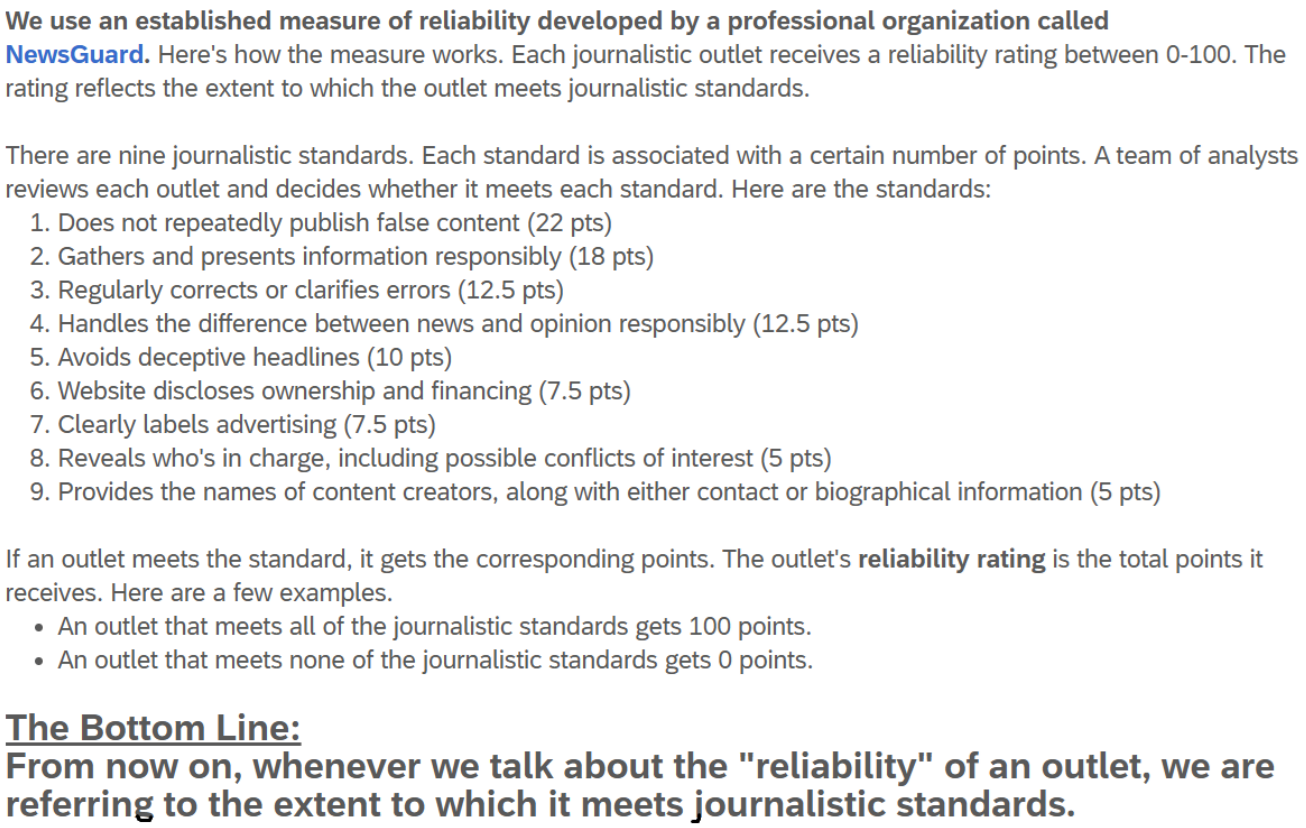
Notes: This is the explanation provided to participants of the slant measure from Bakshy, Messing and Adamic (2015). It was shown to participants near the beginning of the baseline survey, before we elicited perceptions about slant. It was followed by a comprehension check, which participants had to answer correctly before moving on.

Figure A.23: Elicitation of Slant Perceptions



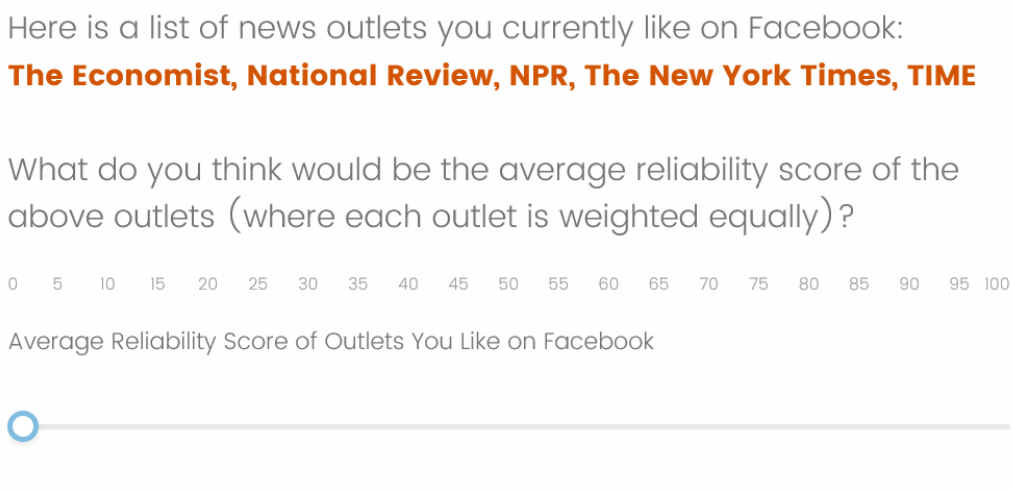
Notes: This is the question that we used to elicit perceptions of the average slant of each participant's portfolio.

Figure A.24: Explanation of NewsGuard Measure for Reliability



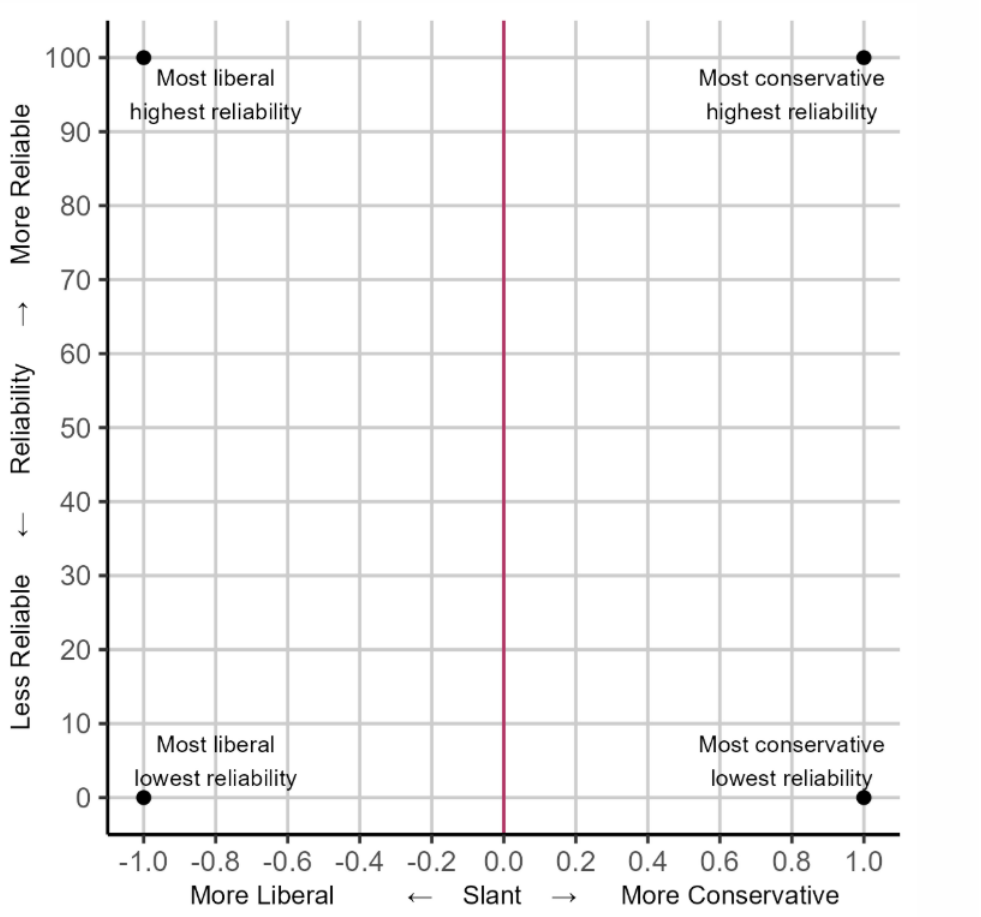
Notes: This is the explanation provided to participants of the NewsGuard reliability measure. It was shown to participants near the beginning of the baseline survey, before we elicited perceptions about reliability. It was followed by an understanding check, which participants had to answer correctly before moving on.

Figure A.25: Elicitation of Reliability Perceptions



Notes: This is the question that we used to elicit perceptions of the average reliability of each participant's portfolio.

Figure A.26: Elicitation of Bliss Point



Notes: This is the interface that we used to elicit the participant's "bliss point". The question accompanying the figure explained: "The figure below combines the reliability and the slant ratings. In particular, the left side of the figure represents more liberal outlets, and the right side represents more conservative outlets. The higher up on the figure, the more reliable the outlet is, and the lower down, the less reliable it is. For example, a point in the top-right corner of the figure represents an outlet that is highly conservative and very reliable." Participants were then asked: "If you could choose, what would you want the average slant and reliability of the pages you like/follow on Facebook to be?"