Frictions in News Consumption: Evidence from Social Media*

Luca Braghieri, Ro'ee Levy, and Hannah Trachtman

August 2025

Abstract

We document pervasive frictions in news consumption on social media, and test interventions designed to address them through a large field experiment on Facebook. We find that a simple, platform-integrated interface prompting active choice significantly and persistently reduces the partisan slant of users' Facebook news portfolios, while coupling this interface with information substantially improves portfolio reliability. Our interventions also increase alignment between users' stated preferences and their Facebook portfolios and influence off-platform news consumption. The results underscore the promise of simple, behaviorally-informed, and scalable interventions to improve news consumption on social media platforms.

*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, the Tel Aviv University Center for AI & Data Science (TAD), the Pinchas Sapir Center for Development, and the Foerder Institute for Economic Studies. 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).

The increased reliance on social media for news access has generated widespread concern for two main hypothesized reasons (Allcott and Gentzkow, 2017; Pariser, 2011; Sunstein, 2017). First, the structure of the social networks and the algorithms governing users' feeds might primarily expose them to content that matches their political ideology, thus generating so-called "echo-chambers" and "filter-bubbles" that promote the consumption of partisan news (Pariser, 2011; Sunstein, 2017; Gauthier et al., 2025). Second, the lack of editorial oversight on social media, together with users' tendency to share sensationalist content, can lead to the proliferation of low-quality and fake news (Allcott and Gentzkow, 2017; Grinberg et al., 2019). These concerns are echoed in the academic literature, which documents a high degree of pro-attitudinal news consumption on social media (Braghieri et al., 2025), as well as the diffusion of fake news (Vosoughi, Roy and Aral, 2018). In light of these findings, there is a pressing need to gain a deeper understanding the drivers of low-quality and partisan news consumption on social media and develop cost-effective remedies.

In this paper, we present the results of a five-week-long field experiment on Facebook that: i) identifies hitherto largely neglected behavioral and information frictions in news consumption on social media, and ii) shows how cheap and scalable interventions targeting those frictions lead to an increase in the reliability and a decrease in the partisanship of the news that users follow on Facebook. Our interventions couple changes to the choice architecture (Thaler and Sunstein, 2008) with the provision of information. This redesign serves three objectives: (i) reducing the "hassle costs" of following or unfollowing news pages on Facebook; (ii) helping users make an active, informed choice about which news pages to follow; and (iii) offering a form of soft commitment, whereby a decision made in a cold state of mind has persistent effects on day-to-day news exposure on the platform. Because our experiment is deployed on the most widely used social media platform for news consumption, with real users, real behavior, and a scalable platform-integrated intervention, the findings speak directly to potential real-world policies.

We conceptualize news consumption as having two important dimensions: a horizontal dimension (henceforth *slant*) that captures a news outlet's political leaning on a left-right spectrum, and a vertical dimension (henceforth *reliability*) that captures the degree to which a news outlet adheres to best-practice journalistic standards.[1] Since slanted and low-reliability news consumption on social media has been argued to exert negative externalities in the political domain (Sunstein, 2017; Aral, 2021), we focus on whether our intervention reduces the slant and increases the reliability of the news followed by social media users.

The design of our five-week-long experiment follows a three-phase structure. We first recruited more than 3,000 participants through Facebook ads and directed them to a baseline survey. This survey includes a consent form, a Facebook login button that asks participants to grant the research team permission to observe the pages they follow on the platform, and a set of questions eliciting demographics, incentivized perceptions about the slant and reliability of various news outlets, and participants' news portfolio bliss-points in the slant-reliability space. 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 complete a midline survey, where they were randomized. Our study arms include a *Control* group, a ***R**e-optimization* treatment (**R**), a ***R**e-optimization with **S**lant Information* treatment (**RS**), and a ***R**e-optimization with **R**eliability Information* treatment (**RR**).[2] Participants in the control group were not subjected to any intervention. Participants assigned to the three re-optimization treatment arms were shown a user-friendly interface that allowed them, with a single click, to follow any outlet from a curated list of the most popular news outlets on Facebook—spanning the slant-reliability spectrum—or to unfollow any they were already following. Participants in the RS treatment, on top of going through the re-optimization interface, received personalized information about the **S**lant of news outlets. Participants in the RR treatment received the same information, but about **R**eliability rather than slant. Around four weeks after the midline survey, participants were invited to an endline survey, where we elicited secondary outcome measures and asked participants in the treatment arms to go through the re-optimization interface once again.

In addition to self-reported survey data, we collected two complementary datasets. First,

[1] As discussed in detail in Section 1.2, we operationalize the notion of slant by employing the methodology developed in Bakshy, Messing and Adamic (2015), which assigns a measure of slant to a news page based on the ideology of the Facebook users that share content from that page. We operationalize the notion of reliability by means of expert ratings produced by media watchdog NewsGuard, as in Aslett et al. (2022).

[2] Our experiment features two additional, smaller treatment arms aimed at exploring mechanisms and probing the robustness of our findings. One arm explores the effectiveness of information provision alone; the other addresses potential experimenter demand effects. We discuss the design and results of these additional treatment arms in Section 6.

using the Facebook API, we recorded the news pages participants follow on the platform throughout the experiment. Second, for the subset of participants who installed our Chrome extension, we tracked their visits to a large set of news websites. Our primary pre-specified outcome variable, obtained directly from the Facebook API, is the set of news pages that participants follow after the midline survey. We focus on following behavior because prior research shows that following news pages on Facebook influences significant downstream outcomes such as political polarization and news knowledge (Levy, 2021; Altay, Hoes and Wojcieszak, 2025).

Our first finding highlights the role of behavioral frictions—such as inertia and sensitivity to minor "hassle factors"—in shaping users' news customization decisions on Facebook. We find that the vast majority (73%) of participants assigned to the R treatment actively modified their news portfolios by adding or removing at least one page.[3] This high degree of re-optimization in the absence of new information, monetary incentives, or explicit encouragement suggests that participants' news portfolios were not optimized to begin with, which in turn points to the presence of behavioral frictions.[4] The passive nature of participants' behavior on social media is underscored by their survey responses: most participants reported that their Facebook news portfolio is not the result of deliberate optimization over the full set of available news pages, but instead reflects almost exclusive reliance on algorithmic recommendations and friend suggestions.

Our second finding highlights nuanced information frictions: participants have reasonably accurate perceptions of slant, but significantly misperceive reliability. The Pearson correlation between participants' perceptions of the average slant of their Facebook news portfolio and the actual average slant of their portfolio is 0.61. In contrast, the correlation between perceived and actual portfolio reliability is only 0.19.

The influence of behavioral and informational frictions is reflected in the portfolios of news pages participants follow on Facebook. For most participants, we observe a significant wedge between the position of their actual news portfolios in the slant-reliability space and their stated "bliss-point" in that space. In particular, participants' actual Facebook portfolios tend to be ideo-

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

[4]As in virtually all active choice experiments, our intervention identifies the joint effect of behavioral frictions and the minimal transaction cost involved in following or unfollowing a news page. We interpret the substantial degree of re-optimization as evidence of behavioral frictions, given that the transaction costs involved—essentially a single click—are negligible. Section 5.1.1 provides more details. An alternative explanation is that participants' re-optimization choices may reflect experimentation. We address this concern in Section 6.3.

logically more extreme and to contain less reliable sources than what participants indicate as ideal.

Our third finding is that our simple re-optimization interface successfully moderates the portfolios of news pages followed on Facebook. All re-optimization treatments lead to a decrease in the absolute average slant of participants' news portfolios. Across the three re-optimization treatments, the mean reduction in absolute average slant is 0.24 standard deviation units, which corresponds to 51% of the difference between the absolute average slant of the news portfolios of self-identified partisans and independents in our study. The effectiveness of re-optimization alone—evident in the Re-optimization (R) treatment that provides no explicit slant information—is consistent with our result showing that participants exhibit only minor misperceptions about slant.

Our fourth finding is that combining information on reliability with the re-optimization interface (the RR treatment) significantly improves the average reliability of the news pages followed by participants, increasing it by 0.21 standard deviation units. This increase is approximately the same size as the gap between the average reliability of the Facebook news portfolios of individuals with and without a college degree. In contrast to slant, re-optimization without information does not have a detectable effect on reliability. This difference aligns with our earlier results showing substantial baseline misperceptions regarding reliability.

Our fifth finding is that the intervention not only shifts users' portfolios toward outlets that are arguably more desirable from a societal perspective—less slanted and more reliable—it also helps close the gap between participants' portfolios and their stated preferences. Specifically, the three treatment arms significantly reduce the distance between participants' actual portfolios and their self-reported bliss-points in the slant-reliability space.

The changes in the portfolio of news pages participants follow on Facebook persist beyond the endline survey four weeks later and translate into changes in online news consumption. In order to study downstream effects on news consumption, we analyze browsing data from the subset of participants who installed our Chrome extension and implement an instrumental variables strategy. Specifically, we use assignment to one of the re-optimization treatments as an instrument for whether a participant follows a given news page. We find that following a news page through our intervention more than doubles the probability of visiting that page in the subsequent month.

Our final set of results offers important caveats and sheds light on underlying mechanisms. Since we find that the combination of information and re-optimization improves reliability, it is natural to wonder whether information provision alone is sufficient to do the same. To study this question, our experiment features an additional treatment arm—the Reliability Information treatment—which provides the same information as the RR treatment but omits the re-optimization interface (we refer to the Reliability Info treatment as **NR**, where **N** stands for **N**o re-optimization and **R** stands for **R**eliability information). This treatment produces no detectable effects, highlighting the necessity of combining information with the re-optimization interface in order to increase the reliabilty of participants' Facebook news portfolios.

We also explore the possibility that our results are primarily driven by experimenter demand effects. To examine this, 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. Specifcally, at the start of the midline survey, we informed participants in the NED group that we had revoked our permission to track the Facebook pages they follow, and we provided instructions for them to verify this. Later, at endline, we requested permission once again to collect their following behavior retrospectively. The NED treatment significantly increases portfolio reliability, ruling out experimenter demand as the main driver of our results.

Lastly, we disentangle the roles played by our offering a balanced set of news pages in the re-optimization treatments and by participants' actual choices from within that set. We find that both factors matter. Specifically, participants in the re-optimization treatments rarely add news pages from outside the provided set, emphasizing the importance of the menu of pages we offer. However, when choosing among the offered pages, participants gravitate towards outlets that reduce the gap between their actual and desired portfolios, underscoring the additional importance of active user choice within the provided set.

The results of this study carry important policy implications. In principle, Facebook (and other social media platforms) could implement a version of our intervention by prompting users to actively manage the news pages they follow—just as it currently encourages them, on a regular basis, to review their privacy settings through a simple "privacy check-up" interface. If social media platforms are unwilling to implement such an intervention, governments could, in principle, mandate it; alternatively, third-party apps integrated with the platforms could offer it as a service.

Our paper contributes to several strands of literature. First, it 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; Pennycook and Rand, 2021).[5] The set of policy tools proposed so far—which include content moderation, algorithmic re-engineering, fact-checking, and accuracy prompts—varies substantially in terms of complexity, cost-effectiveness, scalability, and persistence. We introduce a novel policy tool, grounded in choice architecture and information provision, that addresses low-quality and partisan news consumption on social media in a manner that is simple, cost-effective, scalable, and persistent in the medium term.

Second, our paper contributes to the literature studying news preferences. Much of the theoretical literature has explored why individuals consume news aligned with their existing ideological beliefs, highlighting three main explanations: individuals might perceive pro-attitudinal news as more credible, derive greater utility from consuming like-minded content, or view ideologically aligned outlets as ”delegates” that filter content to provide more relevant information (Chan and Suen, 2008; Chopra, Haaland and Roth, 2022, 2023; Gentzkow and Shapiro, 2006; Gentzkow, Shapiro and Stone, 2015; Mullainathan, Schwartzstein and Shleifer, 2008; Suen, 2004). Our paper introduces and highlights a complementary set of mechanisms—behavioral and informational frictions—that are likely to be especially important for news consumption on social media. Specifically, we provide, to the best of our knowledge, the first systematic evidence that (i) individuals hold substantial misperceptions about the reliability of news outlets, and (ii) inertia and minor ”hassle costs” significantly shape users' decisions about which news sources to consume online.[6] Finally, our finding that the re-optimization interface reduces the slant of participants' news portfolios dovetails with recent research showing that when people are offered tools to customize the language of their news, they opt for neutral rather than slanted framing (Chopra et al., 2025).

Lastly, 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). Specifically, we introduce a new methodology that leverages the possibility of collecting retrospective data to estimate the role of experimenter demand effects.

[5] A related literature documents the prevalence of partisan and low-quality content in digital media environments (Allcott and Gentzkow, 2017; Aridor et al., 2025; Bakshy, Messing and Adamic, 2015; Braghieri et al., 2025; Flaxman, Goel and Rao, 2016; Gauthier et al., 2025; Gentzkow and Shapiro, 2011; Vosoughi, Roy and Aral, 2018).

[6] A related literature finds that ”hassle costs” strongly influence the types of information individuals choose to share on social media (Henry, Zhuravskaya and Guriev, 2022; Guriev et al., 2025; Pennycook and Rand, 2021).

1. Background

1.1 News Consumption on Social Media

As of 2025, social media platforms have become 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, 26% of U.S. adults regularly obtain news from Facebook—more than from any other social media platform. In comparison, only 11% and 10 % report regularly accessing news through Twitter (now ”X”) and TikTok, respectively (Newman et al., 2025).

Two primary concerns have emerged regarding news consumption on social media: the prevalence of partisan news and the proliferation of low-quality content. 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.

Several features of social media platforms are likely to systematically encourage the consumption of partisan and low-reliability content. First, ideologically congenial sharing behavior, together with the political homophily of social media network structures, can give rise to "echo chambers" in which individuals are disproportionately exposed to like-minded news (Sunstein, 2017). Second, personalized ranking algorithms can reinforce these patterns by creating "filter bubbles" that prioritize ideologically aligned posts (Pariser, 2011). Third, digital platforms give users substantial autonomy in shaping their own news feeds—by following or unfollowing pages, joining groups, or hiding certain types of content (Negroponte, 1995).

This paper focuses on users' customization decisions, which have been shown to contribute to partisan news exposure and consumption. As Gauthier et al. (2025) document and as we elaborate upon in Section 7.1, these choices are often influenced by the platforms' recommendation algorithms, which tend to suggest ideologically skewed pages to follow. Instead, we examine a setting in which users are presented with a balanced set of news pages—covering a broad range of slant and reliability—and are invited to actively manage their news portfolios.

1.2 Slant and Reliability

For this paper, we operationalize the notions of slant and reliability as follows. 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 to 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 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, 2018), they capture 93% of all visits to online news websites.[7]

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.

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

2. Experimental Design and Implementation

2.1 Experiment Overview

Figure 1 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.[8] 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 proceeded to a baseline survey. The survey collected demographic data, information about news consumption habits, and pre-treatment measures of various outcome variables.[9] The survey also included personalized incentivized questions to elicit participants' perceptions of the slant and reliability of the portfolio of news pages they currently followed on the platform as well as of major news outlets.[10] 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, the three main *Re-optimization* treatment arms, a *Reliability Information* treatment (NR), and a *No Experimenter Demand* treatment (NED). The control group and the three main re-optimization

[8]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.

[9]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, participated in an earlier pilot survey, were flagged as potential scammers, and requested data deletion.

[10]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).

treatment arms are described in this section. The NR and NED treatments—employed to study mechanisms and probe robustness—are 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.[11]

Participants in the control group received no intervention. Those in the three 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.[12] 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.[13] The second table showed each participant the news outlets they were already following and provided an option to unfollow any of them going forward.[14]

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

Approximately four weeks after the midline survey, participants were invited to complete an

[11]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 main re-optimization treatments. The reason is that these participants were more expensive to recruit; thus, to maximize power, we assigned them only to our main study arms.

[12]Appendix F.2 show screenshots of the re-optimization interface.

[13]We omitted low-reliability pages from the re-optimization interfacec because we did not want to spread misinformation. Appendix C provides more information about how we selected the news pages that participants were shown in the re-optimization interface.

[14]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.

endline survey. This final survey elicited secondary outcomes, including participants' satisfaction with their Facebook news feed and trust in news outlets. 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.

2.2 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.

2.3 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.

- **Bliss-points in slant–reliability space.** Participants reported their ideal combination of slant and reliability in a two-dimensional space. To assess the optimality of a participant's news portfolio on Facebook, 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.** 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 secondary outcomes include survey-based measures of trust in news, satisfaction with one's news diet, and overall news consumption habits on Facebook.

3. Summary Statistics

3.1 Sample Size, Balance, and Attrition

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.[15]

Table A.1 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,

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

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.

Table A.2 shows randomization achieved balance along most baseline observable characteristics.

Table A.3 reports attrition rates across treatment arms. Attrition between midline, where randomization occurs, and endline was minor—approximately 15%. Only one out of our three main treatment arms, namely the R treatment, exhibits evidence of differential attrition at the endline survey. As discussed in section Section 6.2, the set of pages followed at endline by participants in the No Experimenter Demand (NED) treatment also exhibits some evidence of differential attrition.

Importantly, our primary outcome, namely the set of pages participants follow on Facebook right after the midline survey is not subject to survey attrition. Reassuringly for our secondary outcomes and for our robustness test aimed at assuaging concerns about experimenter demand effects, Table A.4 shows that balance on observable characteristics is largely preserved at endline. Where relevant, we also present results employing Lee (2009) bounds.

3.2 Baseline Patterns of News Consumption

To provide context about our sample, we report baseline statistics about participants' news consumption habits.

Figure A.2 shows participants' bliss-point in the slant dimension as a function of their self-reported political ideology. The figure is consistent with our interpretation of slant as a horizontal characteristic: the more individuals identify with the conservative end of the political spectrum, the more right-wing their desired news portfolios on Facebook.

Figure A.3 shows participants' bliss-points in terms of reliability. The figure is consistent with our interpretation of reliability as a vertical characteristic, in that the vast majority of participants reports wanting highly reliable news. The fact that most participants report preferring high-reliability news portfolios stands in contrast to recent concerns about the emergence of a post-truth environment in which individuals no longer care about journalistic standards.

Table A.5 presents the average number of news pages followed on Facebook, as well as the average slant and reliability of participants' news portfolios at baseline. Overall, participants in our study follow, on average, 5.84 news pages. In line with previous research, we find that Democrats tend to have portfolios that are more left-leaning and Republicans portfolios that are more right-leaning (Braghieri et al., 2025). We also find that Democrats, on average, have higher reliability portfolios than Republicans.

4. Empirical Strategy

To evaluate the effects of our interventions, we estimate the impact of our treatments on each outcome variable from Section 2.3.

Let Y_i denote an outcome variable, and let Y_i^b represent the baseline value of Y_i (when available). Let $T_i^j \in \{0,1\}$ be an indicator for assignment to treatment arm $j \in J = \{R, RS, RR\}$, let the reference group be the Control group, and let μ_s denote recruitment and randomization strata fixed effects.[16] We also include a vector of pre-specified baseline covariates, $\mathbf{X_i}$, which includes: political ideology, reliance on Facebook for news, interest in politics, recruitment wave fixed effects, and device fixed effects (desktop vs. mobile).

We estimate the following OLS regression:

$$Y_i = \alpha + \sum_{j \in J} \beta_j T_i^j + \rho \cdot Y_i^b + \eta \cdot \mathbf{X_i} + \mu_s + \varepsilon_i \tag{1}$$

where ε_i is an idiosyncratic error term. When estimating Equation 1, we use robust standard errors.

To study whether following an outlet's Facebook page increases consumption of content from that outlet, we implement an instrumental variables (IV) strategy, using treatment assignment as an instrument for following behavior. Specifically, we consider the following regression equation, where each observation is a participant-page:

$$V_{i,k} = \phi + \tau \cdot F_{i,k} + \psi \cdot Y_i^b + \chi \cdot \mathbf{X_i} + \mu_s + \upsilon_{i,k} \tag{2}$$

where $V_{i,k}$ is the number of visits by participant i to outlet k 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 re-optimization treatments, and zero otherwise. We

[16] As described in Section 2.1, these include average slant and reliability of baseline news portfolios and an indicator for installing the browser extension.

include the same set of controls, Y_i^b, $\mathbf{X_i}$, 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 re-optimization treatments. We refer to these as "potentially offered outlets." Importantly, these outlets exclude outlets that the participant already followed at baseline, helping preserve the monotonicity of the instrument.[17]

5. Results

We articulate the presentation of our results in five subsections. First, we document the role of behavioral and information frictions in news consumption on Facebook. Second, we study whether our treatment is successful at mitigating the slant and improving the reliability of the news pages that participants follow on the platform. Third, we analyze whether the treatments induce greater alignment between participants' news portfolios on Facebook and their stated bliss-points. Fourth, we study treatment effects on off-platform news consumption. Fifth, we present treatment effects on our secondary outcomes, such as trust in the news participants encounter on Facebook, and satisfaction with one's news diet on the platform.

5.1 Behavioral and Information Frictions

Despite the widespread use of Facebook as a news source in our sample, only 22% of participants report being "satisfied" or "very satisfied" with the news content in their feeds, and only 52% say they trust the news they see there "some" or "a lot."

A possible explanation for this dissatisfaction is the relatively passive and unstructured way in which participants report engaging with news on Facebook. As shown in Figure 2, 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

[17] That is, assignment to treatment should only increase the probability of following an outlet, never decrease it.

it intentionally; 72% reported following a page because someone shared content from it or because Facebook recommended it.[18] Lastly, 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. The behavior above seems at odds with workhorse economic models of news consumption, which typically characterize individuals' news choices as highly deliberate.

These patterns point to a possible wedge between participants' actual news portfolios on Facebook and the portfolios they would ideally like to have. Indeed, when we compare participants' self-reported bliss-points in the slant–reliability space to the observed slant and reliability of their pre-treatment portfolios, we find a substantial degree of divergence. Figure 3 presents a scatter plot of the average slant (reliability) of participants' actual news portfolios on Facebook and their stated bliss-point in the slant (reliability) dimension. On average, participants seem to desire Facebook news portfolios that are less slanted and more reliable than the one they currently have. In terms of magnitudes, the gap between one's actual and ideal average portfolio slant is 0.38 points on the -1 to 1 scale on which slant is measured, corresponding to the ideological distance between MSNBC and the New York Times. Similarly, the reliability gap is, on average, 15.1 points on the 0-100 scale on which reliability is measured, corresponding roughly to the difference between the Huffington Post and the New York Times.

Taken together, the findings above offer suggestive evidence of a passive, unstructured approach to news consumption on Facebook. This passivity may help explain the notable wedges between users' actual and ideal news portfolios. In the next two subsections, we turn to more direct evidence of the behavioral and information frictions that support such passivity. These frictions not only help account for the divergence between participants' observed and desired news diets, but also provide leverage for our intervention relying on changes to the choice architecture and information provision.

[18] This aligns with Gauthier et al. (2025), who show that social media algorithms play a major role in shaping which political pages users follow.

5.1.1 Behavioral Frictions

Do users always make deliberate and informed choices about which pages to follow on Facebook? Our evidence suggests they do not, and that behavioral frictions are pervasive.

Our main finding pointing to the presence of behavioral frictions relies on the following logic. Absent behavioral frictions, participants' news portfolios on Facebook should be subjectively optimal, because the transaction costs of following or unfollowing a Facebook page are arguably negligible.[19] In contrast, as shown in Figure A.4, we find that the vast majority (73%) of participants in the R treatment adjusted their portfolios by following or unfollowing news pages, despite receiving no new information. A substantial degree of re-optimization in the absence of new information and in an environment with negligible transaction costs suggests the presence of behavioral frictions.[20,21]

The findings above dovetail with participants' own reports about the passivity of their online news consumption. For instance, the widespread habit of only following pages that are suggested by either the Facebook algorithm or by one's friends has direct implications for optimality: by restricting their attention to a narrow, externally curated set of options, users are effectively optimizing over a constrained choice set. As a result, their portfolios are, at best, weakly optimal and often systematically misaligned with their stated preferences.

Taken together, these results provide evidence that behavioral frictions—such as inertia in decision-making and passivity in the face of small "hassle costs"—prominently shape news-following behavior on Facebook. As such, they are likely to be partly responsible for the wedge between participants' actual and ideal news portfolios.

[19] 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.

[20] Strictly speaking, our experiment identifies the joint effect of behavioral frictions and the minimal transaction cost involved in following or unfollowing a news page on Facebook. We interpret the results as evidence of behavioral frictions, because we consider the transaction costs negligible. To support this claim, we estimate the time cost of re-optimization. Participants in the R treatment follow, an average, 1.96 additional pages from our re-optimization interface, which we estimate takes at most 8 seconds. Given that participants were compensated with gift cards equivalent to a $30/hour wage, the implied cost of the 8 seconds needed to re-optimize their protfolio is about $0.07, which we view as negligible. Since following news pages affects both the posts participants see on Facebook and the external news sites they visit over time (as shown in Section 5.3), the total value of these downstream differences would have to be less than seven cents for transaction costs alone to explain participants' prior inaction. In any case, attributing the observed behavior entirely to transaction costs would not alter the policy implications.

[21] One might wonder whether the observed re-optimization simply reflects experimentation motives—participants trying out unfamiliar outlets out of curiosity. In Section 6.3, we present evidence suggesting that experimentation is not the primary explanation for our results.

5.1.2 Information Frictions

In addition to behavioral frictions, the discrepancy between participants' actual news portfolios on Facebook and their stated preferences may also be driven by information frictions. Specifically, participants might hold misperceptions about the slant or reliability of news outlets. Indeed, we find that while participants have a reasonably accurate understanding of slant, they exhibit pronounced misperceptions about reliability.

The left panel of Figure 4 presents a scatterplot comparing participants' perceptions of the average slant of their Facebook news portfolios (on the y-axis) with the actual average slant (on the x-axis). Although the data exhibit some noise, the scatterplot shows that most participants have a reasonably accurate understanding of their portfolio's ideological slant, with observations clustering around the 45-degree line: participants following more conservative outlets generally perceive their news consumption as more conservative, and vice versa. The correlation between perceived and actual average slant is 0.61.

To evaluate accuracy categorically, we classify participants' news portfolios into three groups based on their slant scores: left-leaning (slant < -0.2), centrist ($-0.2 \leq$ slant ≤ 0.2), and right-leaning (slant > 0.2). Table A.6 presents the resulting confusion matrix, comparing participants' perceptions to the actual ideological lean of their portfolios. Overall, only between 7% and 13% of participants substantially misclassify their portfolio's ideological orientation—for example, perceiving it as right-leaning when it is, in fact, left-leaning.

In contrast to slant, participants' perceptions of outlet reliability are markedly less accurate. The right panel of Figure 4 presents a scatterplot where the x-axis denotes the actual average reliability of participants' news portfolios on Facebook and the y-axis represents participants' perceptions the reliability of their portfolios. The correlation is only 0.13, indicating substantial misperceptions. Consistent with the secular decline in trust in the media, many participants appear overly pessimistic about the quality of the outlets they follow, underestimating the reliability of their own news portfolios.

At the same time, participants with low-reliability portfolios tend to make the opposite error, overestimating how trustworthy their news sources are. To classify these misperceptions, we group portfolios into three categories: low-reliability (score < 60), moderate-reliability ($60 \leq$ score ≤ 80), and high-reliability (score > 80). Table A.7 presents the resulting confusion matrix, showing that a substantial fraction (31%) of participants with low-reliability portfolios mistakenly place them in the high-reliability category.

Granting the caveat that both our belief elicitation and our ground truth measures of slant and reliability are noisy, our findings suggest that participants have relatively accurate perceptions of slant and relatively inaccurate ones of reliability. In Appendix E, we analyze participants' perceptions of the slant and reliability of major outlets and arrive at a similar conclusion—participants have more accurate perceptions of slant than of reliability. The treatment effect estimates we present in the next section are highly consistent with this interpretation.

The asymmetry between the accuracy of participants' perceptions of slant and reliability could be due to slant being more observable than reliability: it can often be inferred from language, tone, and partisan cues. In contrast, assessing reliability requires insight into factors that are less readily visible—such as an outlet's track record on fact-checking, transparency around ownership and funding, and the labeling of sponsored content.

5.2 Treatment Effects

Having documented the presence of behavioral and information frictions in the formation of participants' Facebook news portfolios, we turn to examining whether our interventions successfully moderated the slant and improved the reliability of those portfolios. In this section, we focus on the three re-optimization treatments. We defer discussion of the NR treatment—hypothesized in the pre-analysis plan to have no detectable effects—and the NED treatment to Section 6.1 and Section 6.2, respectively. The treatment effect estimates described in this section are obtained from Equation 1.

Slant. Table 2 reports the impact of the re-optimization treatments on the absolute average slant of participants' news portfolios on Facebook.[22] Relative to the control group, all three re-optimization treatments significantly reduce portfolio slant. These effects materialize immediately after the midline survey—where participants complete the re-optimization task—and persist at the endline survey, approximately four weeks later—columns (1) and (2), respectively. Since, after the midline

[22]In terms of absolute average slant, portfolios range between 0 and 1, where 0 indicates a moderate portfolio and 1 indicates an extremely slanted one.

survey, the effect sizes are very similar, we streamline exposition by pooling them in our discussion. The pooled regression is shown in column (1) of Table A.8.

The magnitude of the effects is substantial. The pooled treatments reduce absolute average slant by 0.24 standard deviation units—equivalent to 51% of the gap between the average slant of self-identified independents and partisans in our sample. As an alternative benchmark, the effect corresponds to more than half the distance between the slant of the Washington Post and that of Bloomberg.

Importantly, the slant-reducing effect of the intervention holds across partisan subgroups. As shown in Table A.9, our treatments induce both Democrats and Republicans to reduce the slant of their news portfolios, though the relatively smaller number of Republicans in our sample leads to some comparisons being underpowered.

There are two conceptually distinct channels through which our interventions might have reduced absolute average slant. The first is through greater exposure to centrist outlets—those with low slant in absolute terms. The second is through increased exposure to counter-attitudinal outlets—i.e., outlets with an ideological orientation opposite to that of a participant. To distinguish between these channels, we examine changes in average absolute slant. If moderation occurs primarily through greater exposure to centrist outlets, we would expect a decline in the average absolute slant of participants' news portfolios. If, instead, moderation arises from following counter-attitudinal outlets, average absolute slant would not necessarily decrease.

As shown in column (2) of Table A.8, we find no statistically significant reduction in average absolute slant as a result of our pooled re-optimization treatments, suggesting that ideological moderation in our setting is primarily driven by an increase in following counter-attitudinal rather than centrist attitudes. In other words, participants in the re-optimization treatments are not simply opting for more neutral content; rather, they are willing to diversify their portfolios by incorporating viewpoints from the opposite side of the political spectrum.[23] Column (3) of Table A.8 corroborates the argument above by providing direct evidence that the fraction of counter-attitudinal news pages in a participant's portfolio almost triples as a result of being assigned to any of our re-optimization

[23]As shown in Table A.10, the RR treatment does lead to a small reduction in the average absolute slant of participants' news portfolios. This effect likely arises because participants in this treatment tend to select higher-reliability outlets, and—as shown in Figure A.1—there is an inverse-U-shaped relationship between slant and reliability: outlets with higher reliability typically exhibit lower absolute slant.

treatments.

Reliability Table 3 reports the effects of our re-optimization treatments on the average reliability of participants' Facebook news portfolios. While the effects on slant are consistent across all treatment arms, the results for reliability are more nuanced. The Re-optimization Reliability Info (RR) treatment leads to a substantial and statistically significant improvement in portfolio reliability. As with slant, the effect materializes immediately—at the midline survey, when re-optimization occurs—and persists through the endline survey, four weeks later. In contrast, the Re-optimization (R) and Re-optimization Slant Info (RS) treatments produce much smaller and insignificant effects.

The magnitude of the effect of coupling the re-optimization interface with reliability information is sizable. Relative to the control group, the RR treatment increases average portfolio reliability by 0.21 standard deviation units. This effect is equivalent to closing 91% of the baseline gap in portfolio reliability between college-educated and non-college-educated participants. As an additional benchmark, the average improvement is roughly equivalent to adding the New York Times to a portfolio consisting of The Huffington Post, Jezebel, and Slate.

Unlike slant, where treatment effects are broadly similar across partisan subgroups, the reliability results show some divergence between Democrats and Republicans. On average, the news portfolios of Republicans in our sample have significantly lower reliability than those of Democrats—a gap of 1.00 standard deviation units at baseline. Table A.11 presents suggestive evidence that the estimated treatment effects on reliability are slightly larger among Republicans, though the heterogeneity margin is not significant at conventional statistical levels due to the relatively small number of Republicans in our sample (p-value 0.199).

Since participants appear to conceptualize reliability as a vertical attribute (see Section 3.2), one might reasonably worry about experimenter demand effects. We address this concern directly Section 6.2, where we describe and analyze the NED treatment, specifically designed to test for such effects.

Distance to bliss-points We have shown that our intervention reduces the slant and, in the RR condition, increases the reliability of the news pages participants follow on Facebook. These portfolio changes can help mitigate the negative political externalities of partisan and low-reliability news consumption on the platform. We next ask whether these changes also align participants' news portfolios with their stated preferences, or whether they move users away from their ideal news diets.

We find that our intervention brings participants' news portfolios closer to their self-reported "bliss-points"—the combinations of slant and reliability they consider ideal. To formally assess this, we construct measures of the distance between a participant's news portfolio and her bliss-point along the two dimensions of slant and reliability. For each participant, we compute the absolute difference between the average slant (reliability) of the news pages they follow on Facebook and their corresponding slant (reliability) bliss-point. We also calculate a composite measure: the Euclidean distance between a participant's portfolio and her bliss-point in the two-dimensional slant–reliability space, with both variables standardized relative to the control group means.

Table 4 presents the results, using our main regression specification (Equation 1), with the slant distance, reliability distance, and Euclidean distance as outcome variables. All three re-optimization treatments significantly reduce the distance between participants' portfolios and their slant bliss-points. Additionally, both the R and RR treatments significantly reduce the distance on the reliability dimension. These findings reinforce the interpretation that our intervention, by alleviating behavioral and information frictions, helps participants better align their news consumption on social media with their own stated preferences.

Treatment Effects Takeaways The contrasting patterns of treatment effects for slant and reliability mirror the asymmetries in participants' baseline beliefs. As shown in Section 5.1.2, participants held relatively accurate perceptions of the ideological slant of news outlets but substantially misperceived their reliability. Consistent with this, we find that re-optimization alone is sufficient to reduce portfolio slant, but not to improve reliability. Improvements in reliability only emerge when re-optimization is paired with targeted information.

Importantly, our intervention promotes alignment on two fronts. It brings participants' news portfolios closer to what is arguably desirable from a societal perspective—less slanted, more reliable content—and closer to participants' own stated preferences. Thus, in this setting, there is no apparent trade-off between individual and societal objectives. Instead, the intervention facilitates a mutually reinforcing alignment: what participants say they want tends to resembles what society tends to hope they consume.

5.3 News Consumption

We have shown that our intervention leads to meaningful changes in the set of news pages participants follow on Facebook. We now ask whether these changes affect actual news consumption behavior. The evidence suggests that they do.

As a starting point, Table A.14 shows that following a news page as a result of one of our main treatments leads participants to report seeing more posts from that page in their Facebook feeds. This serves as a sanity check: following a page on Facebook does appear to influence the content shown in users' feeds, at least according to self-reports.

To examine off-platform news consumption behavior in a way that does not rely on self-reports, we leverage data from the subset of participants who agreed to install our Chrome extension that tracks browsing activity on news websites. Thus, for each participant who installed the browser extension, we are able to track the number of times they visited the website of each of the potentially offered outlets—outlets that a participant did not already follow at baseline and that the participant would have encountered in the re-optimization interface had she been assigned to one of the re-optimization treatments. We note that our extension captures behavior only on the specific device where it was installed, namely a desktop or laptop computer. Most news consumption today occurs on mobile devices. Thus, most of the results in this section should be seen as a lower bound, because we are not able to track the bulk of participants' news consumption.

We estimate the causal effect of following a Facebook news page on visits to that outlet's website in the month after midline using the instrumental variables specification in Equation 2, pooling all participants assigned to the re-optimization treatments.[24]

Table 5 presents the results. Column (1) reports the first-stage results, columns (2) and (3) show the effect of following a news page on visits to that outlet's website when the visit originates from Facebook, and columns (4) and (5) report the effect on total visits to the outlet's website, regardless of the referral source. Columns (2) and (4) capture extensive-margin effects (i.e., whether

[24] To increase statistical power, we also pool the Reliability Information (NR) treatment with the control group. This approach was anticipated in footnote 9 of the pre-analysis plan, which noted that these two study arms would be pooled together if the NR treatment showed no effect on page follows—an assumption supported by the results in Section 6.1.

the participant visits the website at all in the 31 days after midline), while columns (3) and (5) incorporate both the extensive and intensive margins.[25]

The results indicate that following an outlet's Facebook page as a result of being assigned to one of our re-optimization treatments increases the likelihood of visiting that outlet's website in the month after the midline survey. Specifically, it raises the probability of a Facebook-originating visit by 4.01 percentage points and it more than doubles the overall probability of a visit (raising it by 10.74 percentage points). This finding echoes previous work (Levy, 2021) and provides direct evidence that changes in Facebook following behavior lead to meaningful shifts in off-platform news consumption. The effects operate both on the extensive margin (discussed above) and for the total number of visits, especially ones originating from Facebook—columns (3) and (5) of Table 5.[26]

5.4 Secondary Outcomes

In our pre-analysis plan, we pre-specified the following survey-based measures collected at endline as secondary outcomes: self-reported satisfaction and trust with the news seen on Facebook, as well as slant and reliability of the news seen on the platform.

Although the point estimates have the expected sign, we are a bit underpowered to consistently detect significant effects on these secondary outcomes. As shown in Table A.13, however, we find significant evidence that the RR treatment increases both the extent to which participants perceived the news they encountered on Facebook as reliable and satisfaction with those news. We also find suggestive evidence that the R treatment increased trust with the news seen on the platform.

[25]The number of visits to potentially offered outlets in the month after midline is right-skewed and frequently takes a value of zero. We apply the inverse hyperbolic sine transformation to the outcome variables in columns (3) and (5), acknowledging the caveats discussed in Chen and Roth (2024).

[26]As pre-specified, we also examine, for participants who installed our Chrome extension, the slant and reliability of the websites they visited in the month following the intervention. Unlike the analysis of visits to potentially offered pages described in the rest of this section, this analysis cannot pool the re-optimization treatments, because different treatments have distinct effects on the slant and reliability of participants' Facebook news portfolios. The infeasibility of pooling treatments, combined with the fact that only a fraction of participants in our experiment installed the Chrome extension, substantially reduces statistical power and limits our ability to detect systematic effects. Table A.12 presents the results and shows some evidence that one of our treatments (RS) may have induced participants to visit more counter-attitudinal news pages. However, given the overall limited power of the analysis, we refrain from drawing strong conclusions from this result.

6. Mechanisms and Robustness

6.1 The Ineffectiveness of Mere Information Provision

Our interventions are built around two key components: a re-design of the choice architecture and the provision of information. As shown in Section 5.2, modifying the choice architecture alone—through a streamlined re-optimization interface—induces substantial changes in the set of news pages participants follow on Facebook, but needs to be combined with information in order to improve reliability. In this section we show that information provision alone does not produce detectable effects.

In order to study the consequences of mere information provision, we deployed the Reliability Info (NR) treatment. 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 currently followed and for those they would have seen in the re-optimization interface had they gone through it.

Despite receiving the same information as participants in the RR treatment, participants in the NR treatment do not meaningfully re-optimize their news portfolios. By the end of the experiment, participants in the RR group added, on average, 2.02 pages to their Facebook news portfolio, whereas participants in the NR group added, on average, only 0.09. As a result, the last row of Table A.15 shows that the portfolios of participants in the NR treatment do not differ significantly from those of participants in the control group along the slant and reliability dimensions. In short, information alone did not change the composition of participants' news portfolios.

These findings are consistent with those in Aslett et al. (2022), who evaluate an intervention that provided participants with NewsGuard reliability ratings via a browser extension. Although the ratings were prominently displayed, the intervention had no discernible effect on news consumption. Our results reinforce this conclusion: without a mechanism that reduces behavioral friction and prompts active engagement, information alone is likely insufficient to shift behavior in the online news setting.

6.2 Experimenter Demand

Since the intervention was quite transparent, a natural concern is experimenter demand effects—the possibility that subjects may have adjusted their behavior to align with their expectations of the research team's objectives. Our results on downstream news consumption make this explanation less likely. Specifically, it seems improbable that participants would consistently recall which sites were being tracked, or that they would incur meaningful time costs to visit specific news pages over the course of a month in order to conform with perceived experimental aims.

Still, experimenter demand might drive our main effects on following news pages on Facebook. To measure the extent of such effects, we introduced a dedicated treatment arm that closely mirrored the Re-optimization Reliability Info (RR) condition, with one key difference: in the No Experimenter Demand (NED) treatment, participants were explicitly informed during the midline survey that we would no longer collect information about the pages they followed on Facebook. To reinforce this message, we revoked the Facebook permissions they had previously granted us at baseline. Participants could either manually remove these permissions using instructions we provided or opt to have us revoke them automatically. In both cases, we told participants how to verify that we did not have those permissions anymore. As a result of revoking the permissions, we lost access to their page-following data from that point forward, thereby substantially reducing the scope for experimenter demand effects.

At endline, we re-requested access to the Facebook data of participants in the NED treatment. This request was unanticipated, and the vast majority (approximately 99%) of participants consented. The new permissions allow us to compare the portfolio of participants in the NED and RR treatment at endline, and thus isolate and measure experimenter demand effects.

Table A.16 shows the results of the NED treatment. Like the RR treatment, the NED treatment significantly increases the average reliability of the portfolio of news pages that participants follow on Facebook. The point estimate of being assigned to the NED treatment on average reliability is a bit smaller than that of being assigned to the RR treatment, but not significantly so at conventional level.[27]

[27] As shown in Table A.3, the NED treatment exhibits suggestive evidence—significant only at the 10% level—of differential attrition. Table A.4 confirms that, despite this modest differential attrition, the NED treatment and control groups remain balanced on most observable characteristics, thus strenghtening our interpretation that experimenter demand effects are not the main drivers of our results. For thoroughness, however, Table A.17 also provides Lee

We end this section by noting that both the NED and RR estimates are potentially useful from a policy perspective. If a platform implemented a policy similar to the re-optimization and reliability information treatment, demand effects may still encourage people to improve the reliability of the portfolio, and then visit these outlets when they appear in the feed. Thus, even though some of the effect of the RR treatment effects might be due to experimenter demand, it may still have more external validity than the NED treatment. However, if one is interested in determining whether individuals are actually interested in improving the reliability of their outlets (in contrast to impressing an experimenter), the NED estimate is more relevant.

6.3 Experimentation Motives

Throughout our analysis, we have interpreted the high degree of re-optimization in the R treatment as evidence of behavioral frictions in participants' news-following decisions on Facebook. A natural alternative explanation, however, is that participants were simply driven by experimentation motives—temporarily following outlets out of curiosity rather than to address misalignment in their news portfolios.

We believe several pieces of evidence cast doubt on the experimentation hypothesis. First, a substantial share of participants (20%) chose to unfollow at least one news page during the midline survey. This behavior is difficult to reconcile with experimentation, which typically entails exploration of unfamiliar options rather than disengagement with familiar ones. Second, among participants who followed at least one new outlet at midline, 86% chose to follow a page they reported being familiar with. Third, if experimentation were the primary driver of portfolio changes, we would expect to observe a substantial degree of unfollowing behavior by the endline survey. However, only 5% of the pages followed at midline are unfollowed by the end of the endline survey, even though participants are offered the same re-optimization interface, which allows them to unfollow pages at virtually no cost. This behavior is largely inconsistent with experimentation motives: if portfolio additions in the midline survey were just "trial runs," we would have likely observed a higher post-trial churn than the 5% percent recorded over the four-week window.

Together, these patterns point away from experimentation as the main mechanism driving

(2009) bounds on the average absolute slant and average reliability of participants' portfolios at the start of the endline survey.

re-optimization in the R treatment, and bolster our interpretation that re-optimization reflects a reduction in behavioral frictions.

6.4 Role of the Choice Set

The re-optimization interface offered in our main treatments (R, RS, and RR) shape participants' Facebook news portfolios by influencing two key dimensions of user behavior. First, it helps participants make deliberate, informed decisions about the set of news pages they follow on the platform. Second, it provides participants with an explicit and easily actionable set of news pages that they can follow with a single click.

In this section, we examine these two dimensions separately to better understand their respective contributions to the effectiveness of our intervention. We structure this exploration around three questions. First, does our treatment motivate users to actively curate their news portfolios *outside* the set of pages explicitly offered in our re-optimization interface? Second, what would be the impact of participants randomly adding news pages to their portfolios from the balanced set provided in the re-optimization interface? Third, how does deliberate and informed user choice *within* our explicitly provided choice set further shape users' portfolios?

The reason for asking the first question stems from the possibility that encouraging deliberate re-optimization might lead users to actively select pages beyond those explicitly offered in our interface. Our findings clearly indicate otherwise. On average, the share of pages participants followed from the 12 explicitly listed options in our re-optimization interface is 17%. In contrast, the share of pages participants followed between the midline and endline surveys from our full dataset of Facebook news pages beyond our suggested set is 0.1%. Thus, the vast majority of new page additions occurred strictly within the confines of the options explicitly presented in our re-optimization interface. This observation highlights the crucial role that the composition of the suggested choice set plays in guiding user behavior.[28]

Given that users seldom add pages outside those we offered, our second question asks: what role does the specific composition of pages in the re-optimization interface play? To investigate this, we simulated participants randomly selecting news pages from the interface.[29] As illustrated

[28]These findings are consistent with the results of the NR treatment, which showed that merely providing reliability information without offering the re-optimization interface had virtually no effect on participants' news portfolios.

[29]Specifically, for each participant in the R, RS, and RR treatments, we randomly selected the same number of

in Figure A.5, even random selection from the choice set we provided in the re-optimization interface significantly reduces the gap between the slant of participants' actual news portfolios and their stated preferences.[30] This result stems from the interaction of two forces: first, most participants' Facebook news portfolios are more extreme than their stated preferences. Second, our re-optimization interface offers a balanced set of pages that uniformly spans the slant spectrum. Thus, for a person with a pro-attitudinal news portfolio, it is more likely to randomly click on relatively more moderate or counter-attitudinal outlets than on relatively more pro-attitudinal ones.

The finding that random selection from our re-optimization interface decreases the wedge between participants' Facebook news portfolios and their stated bliss points prompts our third question: what is the additional value of (informed) user choice within the offered choice set? We find that when users make deliberate and informed choices within the set of pages offered in the re-optimization interface, their portfolios align significantly more closely with their stated bliss-points compared to random selection. Specifically, random choice from the balanced set closes 9% of the slant gap between actual and desired portfolios and 6% of the reliability gap. Conversely, the RS treatment closes 11% of the slant gap and the RR treatment 11% of the reliability gap. Thus, while offering a balanced choice set alone generates greater alignment between participants' news portfolios and their stated preferences, enabling users to make informed decisions provides further substantial improvements.

7. Discussion and Conclusion

7.1 Discussion

Collectively, our results suggest a two-stage model of news-page selection on social media. First, a subset of news pages is made salient to users, either via our experimental interface or through social sharing and algorithmic recommendation. Second, users select specific pages to follow almost exclusively within this salient subset. This two-stage process provides a coherent framework that rationalizes many of our findings and offers broader insights into news consumption behavior on

outlets they actually chose to follow in the re-optimization interface and computed the resulting changes in portfolio slant and reliability. We repeated this process one thousand times to generate a distribution of simulated effects under random selection.

[30]For reliability, the effects of random selection are substantially more muted and only significant at the 10% level.

social media.

In typical social media environments, users have minimal control over which news pages become salient: these pages are determined largely by social interactions and algorithmic recommendations. Indeed, as shown in Section 5.1, participants overwhelmingly report selecting news pages because either friends shared content from them or Facebook's algorithm explicitly recommended them.

Importantly, these algorithmically and socially derived choice sets likely differ substantially from the balanced menu provided in our experiment. Whereas our set was explicitly designed to span the broad ideological spectrum and include medium-to-high reliability sources, typical algorithmic and social suggestions are likely to be systematically more ideologically polarized and less reliable.[31] Indeed, our findings indirectly support this interpretation. Given that users predominantly optimize within sets defined by algorithms or their social network, and given that users' observed portfolios tend to be both more ideologically polarized and less reliable than their stated preferences, it is reasonable to infer that standard, algorithmically and socially determined choice sets skew ideologically extreme and toward lower reliability.

Our experiment shows how thoughtful restructuring of these salient choice sets, together with information provision, can be a promising approach for improving users' online news diets.

7.2 Conclusion

The growing prevalence of partisan and low-reliability news on social media has raised urgent concerns about the effects of this content on polarization, democratic functioning, and public trust in information.

In this paper, we showed: i) that news consumption on social media is significantly shaped by behavioral and information frictions, and ii) that a simple intervention targeting these frictions meaningfully reduces the slant and improves the reliability of the news pages followed by Facebook users. Our results thus offer a promising approach to mitigating the broader societal threats posed by polarized and low-quality news on social media.

Importantly, our intervention also helps participants align their observed behavior with their

[31] Braghieri et al. (2025) provide evidence that individuals' Facebook friends disproportionately share ideologically extreme rather than moderate news content.

self-reported bliss-points in the slant–reliability space. This finding suggests that many users prefer a more balanced and reliable news environment when given simple tools to achieve it.

These results also 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 scalability.

More broadly, our findings point to a class of interventions that improve online information diets by reducing frictions, enhancing user agency, and respecting individual autonomy.

References

Allcott, Hunt, and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives*, 31(2): 211–236.

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."

Aral, Sinan. 2021. *The hype machine: How social media disrupts our elections, our economy, and our health–and how we must adapt.* Crown Currency.

Aridor, Guy, Tevel Dekel, Rafael Jiménez Durán, Ro'ee Levy, and Lena Song. 2025. "Digital News Consumption: Evidence from Smartphone Content in the 2024 US Elections."

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.

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*.

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, Jiafeng, and Jonathan Roth. 2024. "Logs with Zeros? Some Problems and Solutions." *The Quarterly Journal of Economics*, 139(2).

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*.

Comscore. 2018. *Web Behavior Database Panel 2022.* Wharton Research Data Services, University of Pennsylvania.

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.

Flaxman, Seth, Sharad Goel, and Justin M Rao. 2016. "Filter Bubbles, Echo Chambers, and Online News Consumption." *Public Opinion Quarterly*, 80(S1): 298–320.

Gauthier, Germain, Roland Hodler, Philine Widmer, and Ekaterina Zhuravskaya. 2025. "The Political Effects of X's Recommender Algorithm." *Working Paper*.

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.

Grinberg, Nir, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. "Fake news on Twitter during the 2016 U.S. presidential election." *Science*, 363(6425): 374–378.

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*.

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).

Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies*, 76(3).

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.

Mullainathan, Sendhil, Joshua Schwartzstein, and Andrei Shleifer. 2008. "Coarse Thinking and Persuasion." *The Quarterly Journal of Economics*, 123(2): 577–619.

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

Negroponte, Nicholas. 1995. *Being Digital.* Alfred A. Knopf.

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

Pariser, Eli. 2011. *The Filter Bubble: What the Internet is Hiding from You.* Penguin UK.

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*.

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.

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 = 3,411 consented to participate and completed the baseline survey.
Midline	N = 2,186 participated in the midline survey and were randomized, of which:
	N = 1,560 were assigned to the control group or one of our three main re-optimization treatments (R, RS, and RR).
	N = 1,554 were in the **midline impact evaluation sample**, which means we successfully collected the set of pages they followed on Facebook at midline.
	N = 626 were assigned to the robustness & mechanism treatments (NR and NED).
Endline	N = 1,863 participated in the endline survey, of which:
	N = 1,337 were in the control group or one of our three main re-optimization treatments (R, RS, and RR).
	N = 1,314 were in the **endline impact evaluation sample**, which means we successfully collected the set of pages they followed on Facebook at endline.
	N=526 were in the robustness & mechanism treatments (NR and 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: ATE on Absolute Average Slant

	(1) Absolute Avgerage Slant (post-midline)	(2) Absolute Avgerage Slant (pre-endline)
Re-optimization (R)	-0.055***	-0.047***
	(0.008)	(0.009)
Re-optimization + Slant Info (RS)	-0.048***	-0.046***
	(0.008)	(0.010)
Re-optimization + Reliability Info (RR)	-0.068***	-0.060***
	(0.008)	(0.010)
p-value, R=RS	0.53	0.94
p-value, R=RR	0.21	0.28
Control Mean	0.44	0.45
Control SD	0.24	0.24
Observations	1554	1314
R-squared	0.66	0.64

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 1 with the absolute average slant of a participant's news portfolio on Facebook as the outcome variable. Column (1) estimates Equation 1 using data collected right after the midline survey; thus, it employs the midline impact evaluation sample. Column (2) estimates Equation 1 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 3: ATE on Reliability

	(1) Avg. Reliability (post-midline)	(2) Avg. Reliability (pre-endline)
Re-optimization (R)	0.502	0.368
	(0.324)	(0.372)
Re-optimization + Slant Info (RS)	0.087	0.060
	(0.311)	(0.365)
Re-optimization + Reliability Info (RR)	2.461***	2.198***
	(0.312)	(0.406)
p-value, R=RS	0.28	0.47
p-value, R=RR	0.00	0.00
Control Mean	86.53	86.50
Control SD	12.17	12.13
Observations	1554	1314
R-squared	0.78	0.76

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 1 with reliability of a participant's news portfolio on Facebook as the outcome variable. Column (1) estimates Equation 1 using data collected right after the midline survey; thus, it employs the midline impact evaluation sample. Column (2) estimates Equation 1 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 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.

Table 4: Distance to Bliss Points

	Slant (1)	Reliability (2)	Euclidean (3)
Re-optimization (R)	-0.109***	-0.050***	-0.113***
	(0.029)	(0.019)	(0.023)
Re-optimization + Reliability Info (RR)	-0.092***	-0.104***	-0.120***
	(0.027)	(0.018)	(0.022)
Re-optimization + Slant Info (RS)	-0.143***	-0.022	-0.116***
	(0.029)	(0.017)	(0.022)
Control Mean	0.00	-0.00	0.00
Control SD	1.00	1.00	1.00
Observations	1554	1554	1554
R-squared	0.75	0.88	0.83

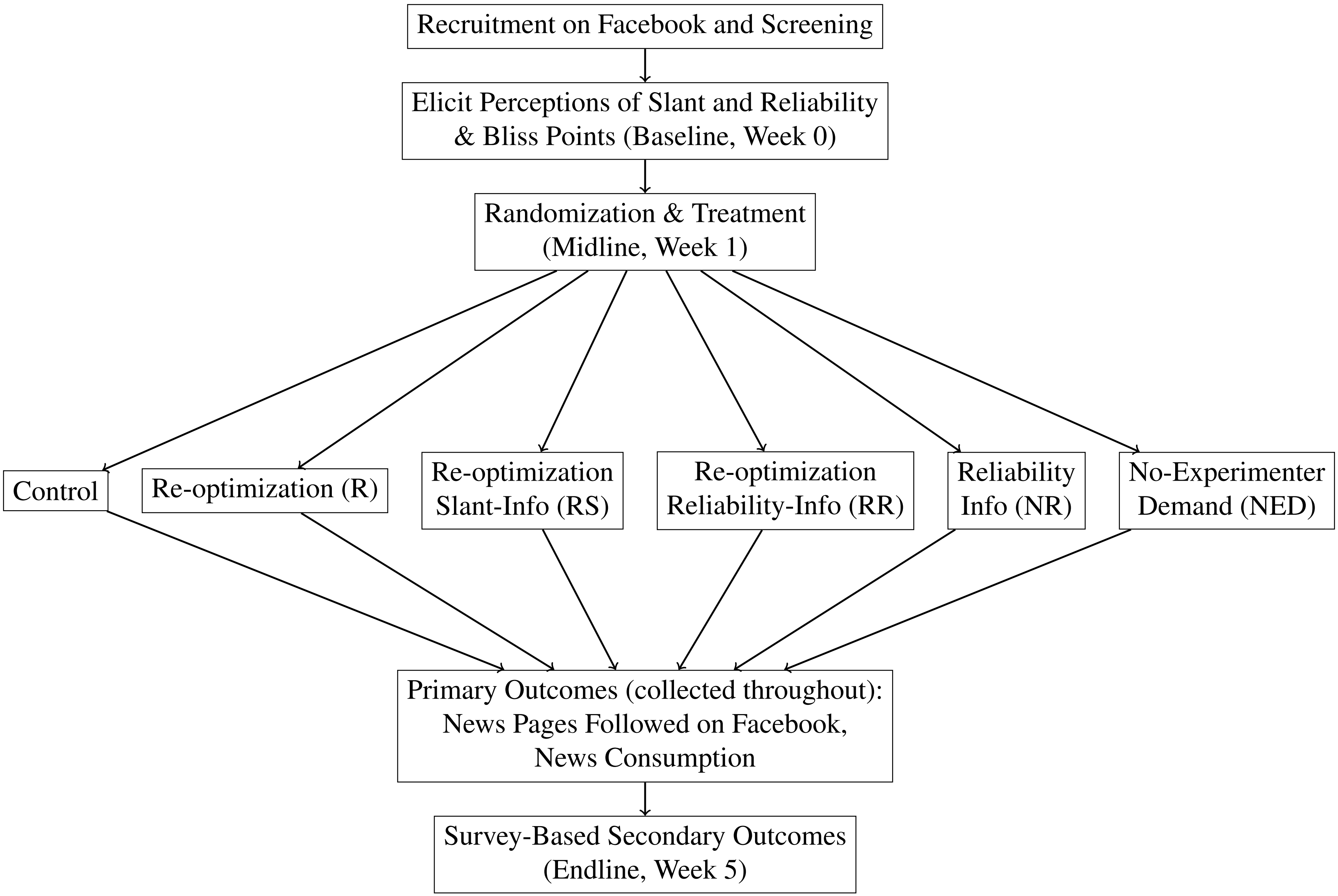
Notes: This table explores the effect of our intervention on the distance between the slant and reliability of participants' news portfolios on Facebook and participants' stated bliss points in the slant-reliability space. Specifically, it presents estimates of coefficients β_j from Equation 1 with distance to bliss point as the outcome variable. Column (1) estimates Equation 1 on the standardized absolute difference between a participants' stated bliss point on the slant dimension and the average slant of the portfolio of news pages the participant follows on Facebook. The standardization was performed as follows: for each participant, taking the distance between their bliss point on the slant dimension and their average slant, we subtracted the average of such distance among participants in the control group, and divided the resulting number by the standard deviation of this distance among participants in the control group. Column (2) estimates Equation 1 on the standardized absolute difference between a participants' stated bliss point on the reliability dimension and the average reliability of the portfolio of news pages the participant follows on Facebook. The standardization was performed as follows: for each participant, taking the distance between their bliss point on the reliability dimension and their average reliability, we subtracted the average of such distance among participants in the control group, and divided the resulting number by the standard deviation of this distance among participants in the control group. Column (3) estimates Equation 1 on the Euclidean distance between a participants' 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, taking the Euclidean distance between their bliss point and their average position, we subtracted the average Euclidean distance among participants in the control group, and divided the resulting number by the standard deviation of the Euclidean distance 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 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.

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 Treatments	0.193^{***}				
	(0.009)				
Page Liked		0.040^{**}	0.048^{**}	0.107^{**}	0.128
		(0.020)	(0.021)	(0.052)	(0.080)
First-stage F-stat	428.52				
Control Mean, Dep Var	0.00	0.01	0.01	0.09	0.15
Control SD, Dep Var	0.00	0.09	0.11	0.28	0.64
Observations	6455	6455	6455	6455	6455
R-squared	0.07	0.03	0.04	0.19	0.54

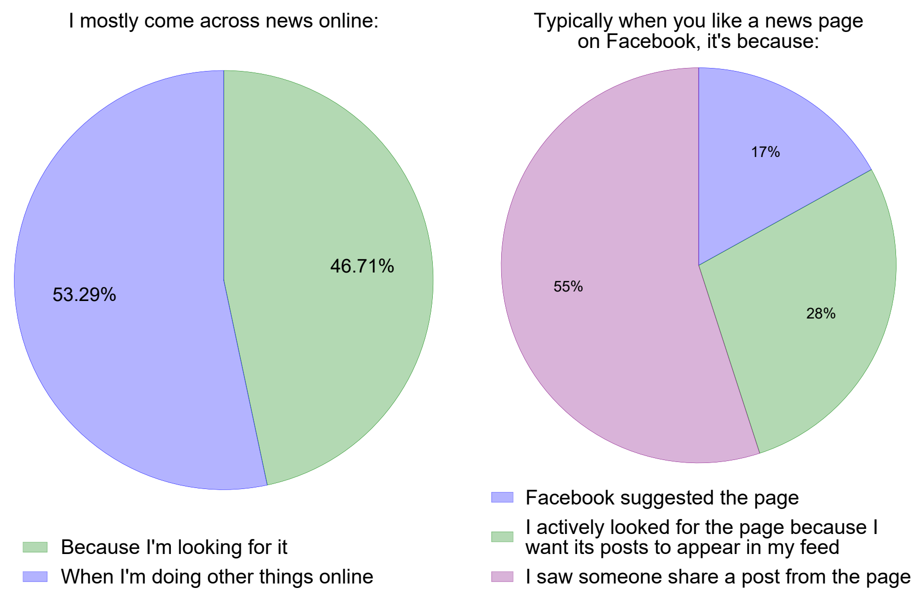
Notes: This table explores the effects of our intervention on online news consumption for the subset of partcipants who installed the Chrome browser extension that tracks their behavior on news websites. The analysis is run on a dataset where an observation is a user-page. Pooled treatments refer to 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. As anticipated in footnote 9 of our pre-analysis plan, we pool the NR treatment and the control group in order to increase power. Column (1) displays the results of the first-stage regression of $F_{i,k}$ on the instrument and the same set of controls as in Equation 2. Columns (2) through (5) display the results of the second-stage regression; i.e., estimates of parameter τ in Equation 2. 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: Experimental Design



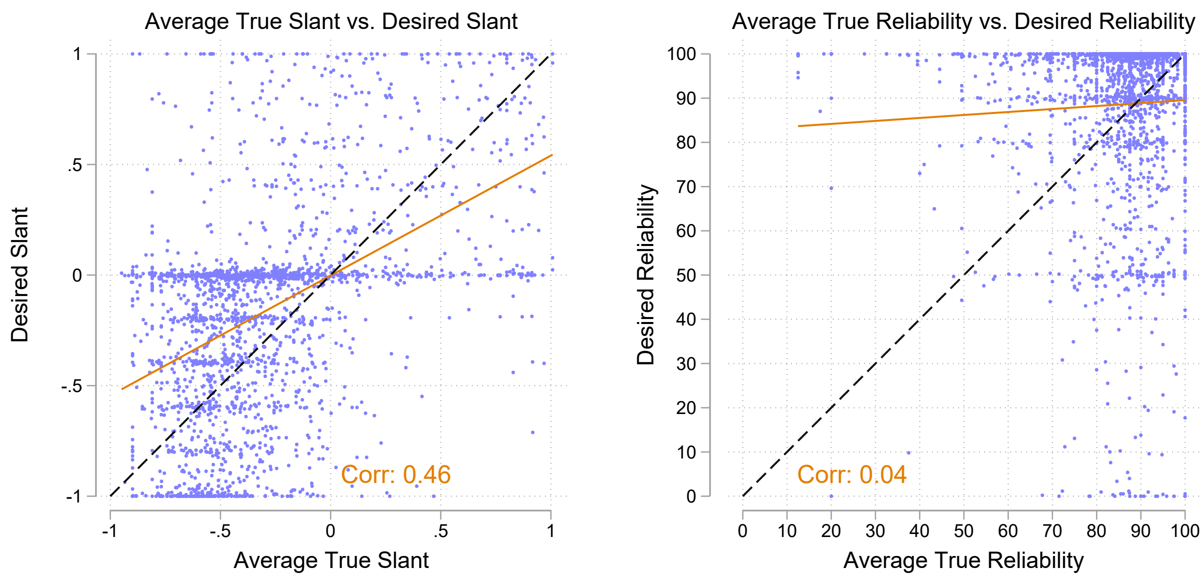
Notes: This figure illustrates graphically the key components of our experimental design.

Figure 2: Evidence of Passive News Consumption Online



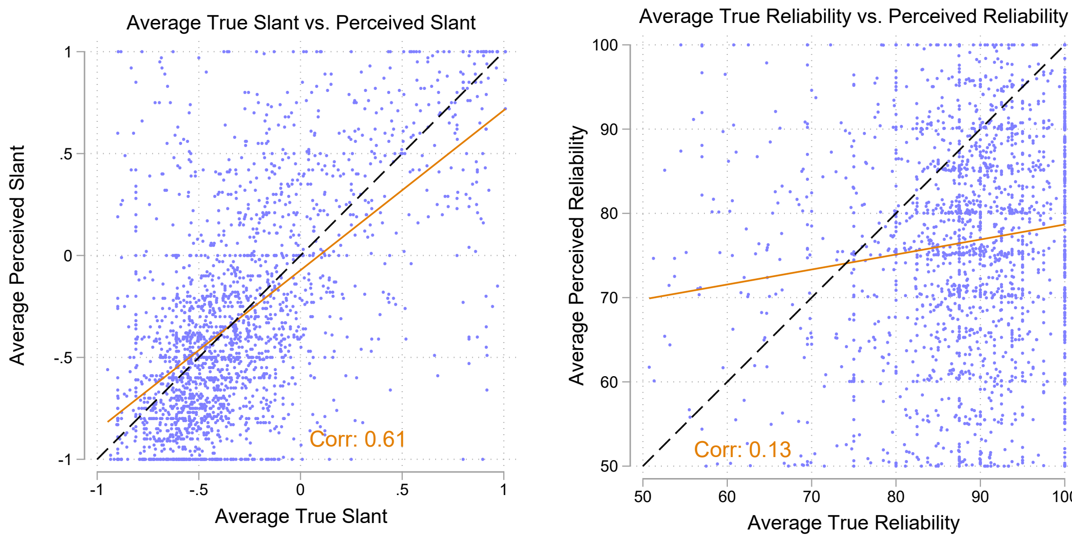
Notes: This figure presents the distribution of responses to two questions about participants' interactions with the online news ecosystem. The first question asked participants: ”Which statement best describes how you get news online, whether on a computer, phone, or tablet, even if neither is exactly 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 4% of participants who answered that they did not like any Facebook pages, because we knew those participant were mistaken. Specifically, had they really not liked any pages on Facebook, they would have automatically been screened out prior to this question.

Figure 3: 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.

Figure 4: 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.

Appendix For Online Publication

Table of Contents

A. Appendix Figures and Tables

Table A.1: 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	0.43	0.57	0.54
College	0.63	0.35	0.34
Democrat	0.56		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 the midline impact evaluation sample (N=1,558): participants who were assigned to the control group or to one of the three main re-optimization treatments, and for whom we successfully collected the set of pages they followed on Facebook during the midline survey (see Section 3 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 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.2: 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	0.77	0.72	0.75	0.70	0.48	0.55	0.36	0.92	0.18
White	0.94	0.93	0.96	0.94	0.94	0.94	0.54	0.27	0.93	0.80	0.87
Age	51.86	53.25	51.49	53.42	52.73	53.56	0.17	0.71	0.10	0.37	0.08
College Grad	0.62	0.65	0.57	0.65	0.62	0.64	0.30	0.24	0.39	0.95	0.44
Democrat	0.57	0.52	0.50	0.54	0.56	0.56	0.14	0.06	0.41	0.67	0.74
Republican	0.11	0.16	0.16	0.17	0.15	0.16	0.05	0.05	0.01	0.13	0.06
News SM Often	0.82	0.82	0.84	0.83	0.83	0.85	1.00	0.37	0.59	0.58	0.17

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.3: Differential Attrition

	(1) Completed Endline	(2) Endline ”Follows” Collected
Re-optimization (R)	0.052**	0.050**
	(0.024)	(0.025)
Re-optimization + Slant Info (RS)	0.011	0.011
	(0.026)	(0.026)
Re-optimization + Reliability Info (RR)	-0.013	-0.011
	(0.026)	(0.027)
Reliability Info (NR)	0.023	0.014
	(0.027)	(0.028)
No Experimenter Demand (NED)	-0.032	-0.053*
	(0.029)	(0.030)
Control Mean	0.84	0.84
Observations	2186	2186
R-squared	0.01	0.01

Notes: This table presents the results of a battery of tests for differential attrition in our experiment. Column (1) shows the coefficients in a regression of an indicator for completing the endline survey on a constant and indicators for being assigned to the various treatment arms. The regression in Column (2) is the same, but the outcome variable is an indicator that takes 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. The sample includes all participants who reached the randomization stage. Standard errors in parentheses are robust.

Table A.4: 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	0.70	0.76	0.71	0.75	0.69	0.22	0.63	0.30	0.98	0.13
White	0.93	0.92	0.95	0.94	0.93	0.94	0.55	0.30	0.55	0.87	0.70
Age	51.20	52.85	50.56	53.17	52.01	53.35	0.14	0.55	0.06	0.44	0.04
College Grad	0.62	0.69	0.58	0.66	0.62	0.65	0.09	0.35	0.25	0.91	0.51
Democrat	0.56	0.50	0.50	0.55	0.57	0.54	0.16	0.12	0.70	0.90	0.62
Republican	0.11	0.16	0.15	0.16	0.14	0.17	0.10	0.15	0.05	0.28	0.02
News SM Often	0.85	0.81	0.84	0.84	0.85	0.85	0.14	0.63	0.56	0.99	0.92

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.5: Baseline Portfolio Statistics

	Democrats	Republicans	All
Avg. Number News Pages Followed	6.58	5.01	5.84
Avg. Portfolio Slant	-0.45	0.18	-0.30
Avg. Portfolio Reliability	89.14	76.85	86.32
Observations	871	231	1560

Notes: This table presents descriptive statistics about the baseline portfolios of news pages that participants in our experiment follow on Facebook. Column (1) shows descriptive statistics for self-identified Democrats. Column (2) does the same for self-identified Republicans. Column (3) includes all participants who were randomized into the control group or into one of our three main re-optimization treatments. 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: Confusion Matrix: Slant

		Perceived		
		Left-leaning	Centrist	Right-leaning
Actual	Left-leaning	0.81	0.12	0.07
	Centrist	0.44	0.27	0.28
	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.7: Confusion Matrix: Reliability

		Perceived		
		Low Reliability	Medium Reliability	High Reliability
	Low Reliability	0.35	0.34	0.31
Actual	Medium Reliability	0.33	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.8: ATE on Absolute Average Slant and Average Absolute Slant (pooled)

	(1) Absolute Average Slant (post-midline)	(2) Average Absolute Slant (post-midline)	(3) Fraction Counter-attitudinal pages followed (post-midline)
Pooled Treatments	-0.057***	-0.003	0.059***
	(0.005)	(0.004)	(0.004)
Control Mean	0.44	0.51	0.02
Control SD	0.24	0.21	0.06
Observations	1554	1554	1237
R-squared	0.66	0.73	0.30

Notes: This table explores 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-attittuinal pages in those portfolios. Specifically, it presents the estimate of coefficient β for a version of Equation 1 in which all three main re-optimization treatment arms are pooled together. The outcome variable in column (1) is the absolute average slant of a participant's news portfolio on Facebook. The outcome variable in column (2) is the average absolute slant of a participant's news portfolio. The outcome variable in column (3) is the fraction of counter-attitudinal pages in a participant's news portfolio. 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 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.

Table A.9: ATE on Absolute Average Slant by Political Ideology

	(1) Democrats Absolute Average Slant	(2) Republicans Absolute Average Slant
Re-optimization (R)	-0.050***	-0.056*
	(0.010)	(0.029)
Re-optimization + Slant Info (RS)	-0.044***	-0.068**
	(0.011)	(0.031)
Re-optimization + Reliability Info (RR)	-0.059***	-0.026
	(0.009)	(0.024)
Control Mean	0.46	0.39
Control SD	0.21	0.32
Observations	868	230
R-squared	0.65	0.67

Notes: This table explores, separately for self-identified Democrats and Republicans, the effect of our intervention on the absolute average slant of their news portfolios on Facebook. Specifically, it presents estimates of coefficients β_j from Equation 1 with the absolute average slant of a participant's news portfolio on Facebook as the outcome variable. Column (1) estimates Equation 1 using only data of self-identified Democrats. Column (2) estimates Equation 1 using only data of self-identified Republicans. The regression uses data collected right after the midline survey and employs the subset of the midline impact evaluation sample that self-identifies as Democrat or Republican. 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.

Table A.10: ATE on Average Absolute Slant

	(1) Average Absolute Slant (post-midline)	(2) Average Absolute Slant (pre-endline)
Re-optimization (R)	0.007	0.005
	(0.006)	(0.007)
Re-optimization + Slant Info (RS)	-0.001	-0.001
	(0.006)	(0.007)
Re-optimization + Reliability Info (RR)	-0.014**	-0.014*
	(0.006)	(0.007)
p-value, R=RS	0.29	0.48
p-value, R=RR	0.00	0.02
Control Mean	0.51	0.51
Control SD	0.21	0.21
Observations	1554	1314
R-squared	0.73	0.71

Notes: This table explores the effect of our intervention on the average absolute slant of participants' news portfolios on Facebook. Specifically, it presents estimates of coefficients β_j from Equation 1 with the average absolutea slant of a participant's news portfolio on Facebook as the outcome variable. Column (1) estimates Equation 1 using data collected right after the midline survey; thus, it employs the midline impact evaluation sample. Column (2) estimates Equation 1 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 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.

Table A.11: ATE on Average Reliability by Political Ideology

	(1) Democrats Average Reliability	(2) Republicans Average Reliability
Re-optimization (R)	-0.209	1.604
	(0.346)	(1.229)
Re-optimization + Slant Info (RS)	-0.391	1.919
	(0.301)	(1.437)
Re-optimization + Reliability Info (RR)	1.693***	3.229***
	(0.335)	(1.191)
Control Mean	89.21	76.08
Control SD	8.28	18.00
Observations	868	230
R-squared	0.68	0.80

Notes: This table explores, separately for self-identified Democrats and Republicans, the effect of our intervention on the average reliability of their news portfolios on Facebook. Specifically, it presents estimates of coefficients β_j from Equation 1 with the average reliability of a participant's news portfolio on Facebook as the outcome variable. Column (1) estimates Equation 1 using only data of self-identified Democrats. Column (2) estimates Equation 1 using only data of self-identified Republicans. The regression uses data collected right after the midline survey and employs the subset of the midline impact evaluation sample that self-identifies as Democrat or Republican. 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.

Table A.12: ATE on Slant and Reliability of Visited Pages

	(1) Extensive Margin High-reliability News Sites	(2) IHS(number of visits) High-reliability News Sites	(3) Extensive Margin Counter-Attitudinal News Sites	(4) IHS(number of visits) Counter-Attitudinal News Sites
Re-optimization (R)	0.005	0.140	0.026	0.031
	(0.045)	(0.132)	(0.023)	(0.032)
Re-optimization + Slant Info (RS)	0.058	0.190	0.057**	0.088**
	(0.044)	(0.136)	(0.027)	(0.037)
Re-optimization + Reliability Info (RR)	0.069	0.051	-0.010	0.014
	(0.045)	(0.141)	(0.023)	(0.044)
Control Mean	0.74	2.50	0.03	0.04
Control SD	0.44	2.09	0.18	0.21
Observations	539	539	539	539
R-squared	0.27	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. To address this, we construct outcome variables that are defined for all Chrome extension users and regress these outcomes on our treatment indicators using Equation 1. 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. 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.13: ATE on Secondary Outcomes

	News satisfaction (endline)	News Trust (endline)	Relative Slant (endline)	Relative Reliability (endline)
Re-optimization (R)	0.087	0.091*	-0.030	-0.000
	(0.065)	(0.047)	(0.073)	(0.059)
Re-optimization + Slant Info (RS)	0.100	0.075	0.065	0.077
	(0.065)	(0.047)	(0.072)	(0.056)
Re-optimization + Reliability Info (RR)	0.130**	0.074	-0.062	0.125**
	(0.065)	(0.048)	(0.072)	(0.056)
Pooled Treatments	0.105*	0.080**	-0.010	0.065
	(0.054)	(0.039)	(0.060)	(0.047)
Observations	1337	1337	1178	1337

Notes: This table explores the effect of our intervention on the secondary outcomes described in Section 5.4. Specifically, it presents estimates of coefficients β_j from Equation 1 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 1 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 1 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)." Column (3) estimates Equation 1 using a measure of 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 (3) has fewer observations than the other columns, because we exclude individuals who, at baseline, reported primarily consuming moderate news. Column (4) estimates Equation 1 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.14: Effects on (Self-reported) Changes to Facebook Feed

	(1) First Stage Page Followed	(2) Number of Times Seen in Feed
Pooled Treatments	0.165***	
	(0.005)	
Page Followed (instrumented)		0.540**
		(0.253)
Control Mean, Dep Var	0.00	0.76
Control SD, Dep Var	0.00	1.46
Observations	18681	18681
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 includes all 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 analysis is run on a dataset where an observation is a user-page. 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 2. 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.15: ATE of the Reliability Info (NR) Treatment

	Absolute Average Slant (pre-endline) (1)	Average Reliability (pre-endline) (2)
Re-optimization (R)	-0.047***	0.351
	(0.009)	(0.367)
Re-optimization + Slant Info (RS)	-0.046***	0.039
	(0.009)	(0.353)
Re-optimization + Reliability Info (RR)	-0.060***	2.221***
	(0.009)	(0.395)
Reliability Info (NR)	-0.004	-0.064
	(0.006)	(0.285)
p-value, RR=NR	0.00	0.00
Control Mean	0.45	86.50
Control SD	0.24	12.13
Observations	1583	1583
R-squared	0.69	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 1, including not only the three re-optimization treatment arms, but also the Reliability Info (NR) treatment. Column (1) estimates Equation 1 with the absolute average slant of a participant's news portfolio on Facebook as the outcome variable. Column (2) estimates Equation 1 with the average reliability of a participant's news portfolio on Facebook as the outcome variable. The sample is the midline impact evaluation sample + the set of participants assigned to the NR treatment for whom we could collect the pages they follow on Facebook. 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.

Table A.16: ATE of the No-Experimenter Demand (NED) Treatment

	Abs Avg Slant (pre-endline) (1)	Abs Avg Slant (post-endline) (2)	Avg Reliability (pre-endline) (3)	Avg Reliability (post-endline) (4)
Re-optimization (R)	-0.047***	-0.064***	0.403	0.784**
	(0.010)	(0.010)	(0.371)	(0.393)
Re-optimization + Slant Info (RS)	-0.046***	-0.057***	0.081	0.328
	(0.010)	(0.010)	(0.365)	(0.388)
Re-optimization + Reliability Info (RR)	-0.059***	-0.092***	2.208***	3.163***
	(0.010)	(0.010)	(0.403)	(0.426)
No Experimenter Demand (NED)	-0.075***	-0.109***	1.925***	2.382***
	(0.012)	(0.013)	(0.435)	(0.460)
p-value, R=RS	0.94	0.54	0.45	0.29
p-value, R=RR	0.29	0.02	0.00	0.00
p-value, RR=NED	0.25	0.23	0.57	0.13
Control Mean	0.45	0.45	86.50	86.47
Control SD	0.24	0.24	12.13	12.14
Observations	1556	1552	1556	1552
R-squared	0.61	0.57	0.75	0.71

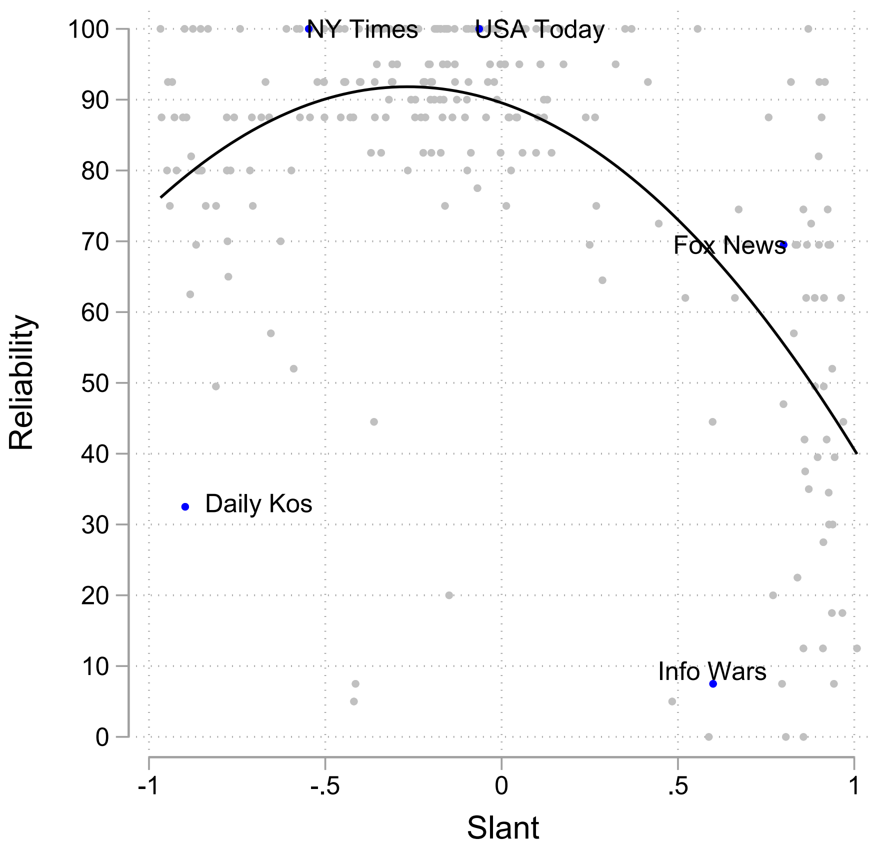
Notes: This table explores the effect of our intervention on the absolute average slant and average reliability of participants' news portfolios on Facebook. Specifically, it presents estimates of coefficients β_j from Equation 1, including not only the three re-optimization treatment arms, but also the No-Experimenter Demand (NED) treatment. Columns 1 and 2 estimate Equation 1 with the absolute average slant of a participant's news portfolio on Facebook as the outcome variable. Columns 3 and 4 estimate Equation 1 with the average reliability of a participant's news portfolio on Facebook as the outcome variable. The difference between columns 1 and 2, and 3 and 4 is that the odd columns estimate Equation 1 using the set of pages that participants follow at the *beginning* of the endline survey, whereas the even columns estimate Equation 1 using the set of pages that participants follow at the *end* of the endline survey. The two sets of pages are different, because the endline survey offered participants in the R, RS, RR, and NED treatment arms another opportunity to update the set of pages they follow on Facebook using our re-optimization interface. 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.

Table A.17: No-Experimenter Demand (NED) treatment: Lee (2009) Bounds

Outcome	Lower Bound	Upper Bound
Average Reliability (pre-endline)	0.752	3.704***
	(1.279)	(1.122)

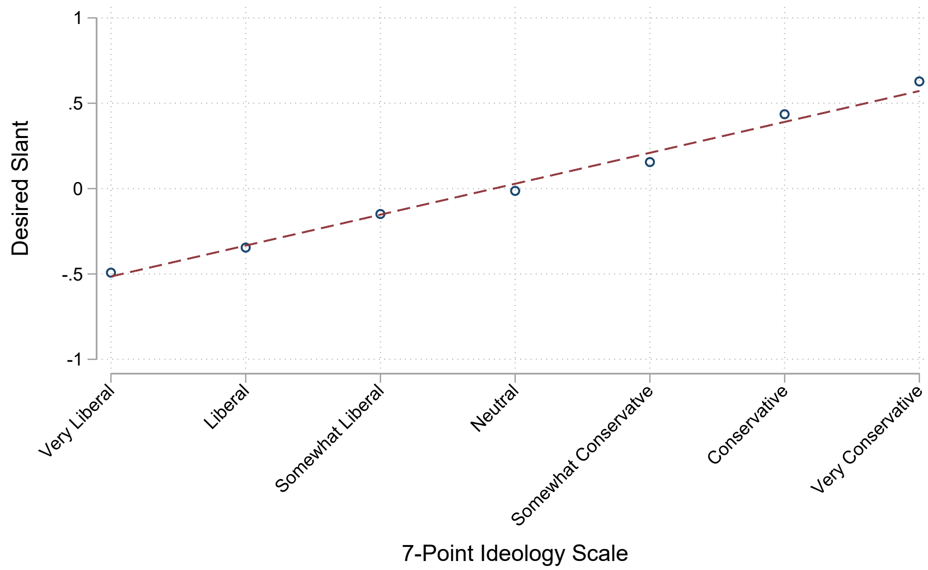
Notes: This table reports bounds on the estimated treatment effects for the NED treatment, following the procedure outlined in Lee (2009). We compute these bounds because, as shown in Table A.3, the NED treatment experiences mild differential attrition due to the fact that a small fraction of participants declined to grant us permission to access their Facebook data during the endline survey. The outcome variable is the average reliability of participants' Facebook news portfolios at the beginning of the endline survey. Standard errors are provided in parentheses.

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



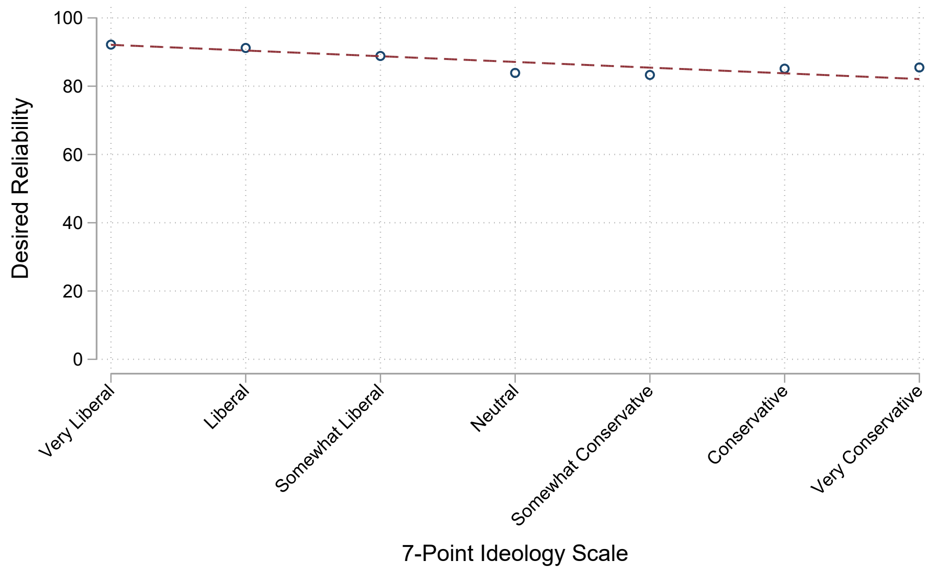
Notes: This figure shows the joint distribution of slant and reliability of the 262 news outlets for which we have both a slant rating from Bakshy, Messing and Adamic (2015) and a reliability measure from NewsGuard.

Figure A.2: Bliss-point and Political Ideology: Slant



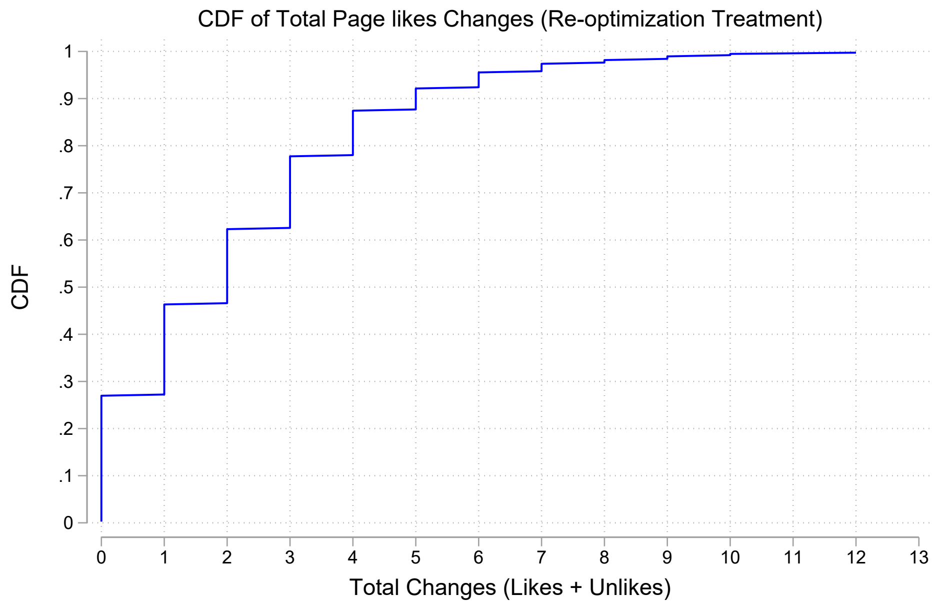
Notes: This figure explores the relationship between participants' stated bliss points on the slant dimension and their self-reported political ideology. Specifically, the x-axis shows seven bins of self-reported political ideology, and the y-axis shown the average slant bliss-point for participants whose self-reported ideology falls in the bin on the x-axis.

Figure A.3: Bliss-point and Political Ideology: Reliability



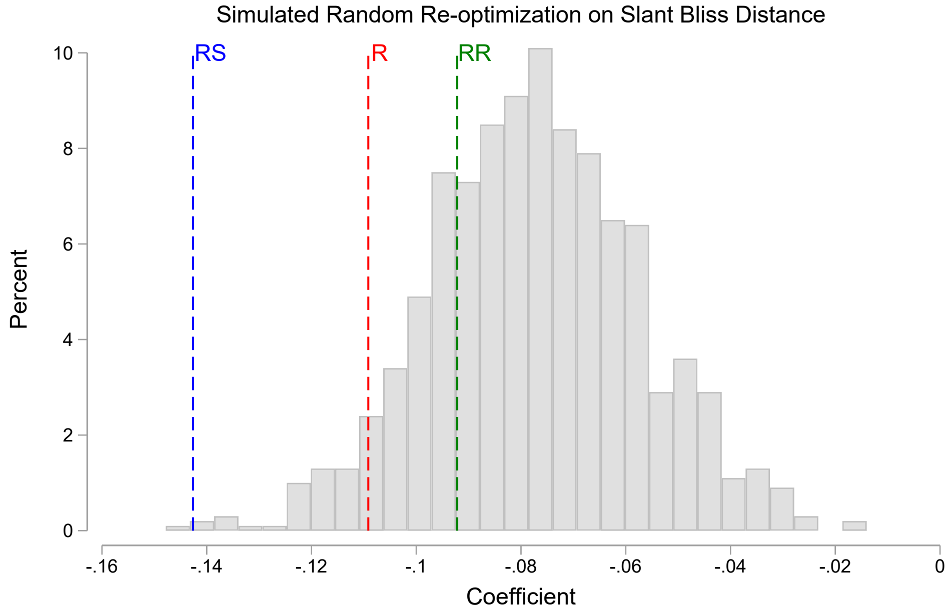
Notes: This figure explores the relationship between participants' stated bliss points on the reliability dimension and their self-reported political ideology. Specifically, the x-axis shows seven bins of self-reported political ideology, and the y-axis shown the average reliability bliss-point for participants whose self-reported ideology falls in the bin on the x-axis.

Figure A.4: CDF of portfolio changes by participants in the R treatment at midline

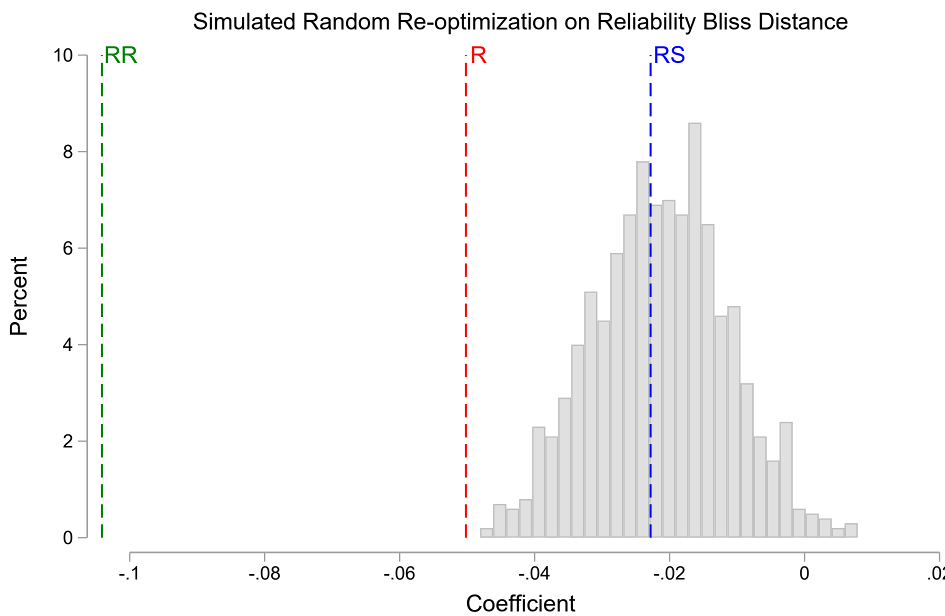


Notes: This figure shows the cumulative distribution function (CDF) of the number of total portfolio changes by participants in the Re-optimization (R) treatment during the midline survey. Following F new pages to one's portfolio and unfollowing U pages from one's original portfolio is counted as $F+U$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

Figure A.5: Simulation of Random Page Likes and Unlikes



(a) Absolute Average Slant



(b) Average Reliability

Notes: This figure presents 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 1. Each gray bar in the figure corresponds to an estimate of Equation 1 for a simulation. We perform 1000 simulations. The dotted vertical lines indicate the actual estimated coefficients for the Re-optimization (red), Re-optimization + Slant Info (blue), and Re-optimization + Reliability Info (green) treatments.

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.[32] 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.[33]

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

[33] 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. For each category, we showed participants in the re-optimization treatments the two most popular U.S.-focused outlets dedicated to hard news. If a participant already followed one of these top-ranked outlets, we offered the next most popular outlet in that category instead. We detail these categories and their rankings below; Table A.18 lists the full set of outlets potentially offered in each category.

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.18: Outlets offered

Name	Group	Rank	Reliability	Slant
The Wall Street Journal	Higher reliability conservative	1	100	0.28
The Daily Caller	Higher reliability conservative	2	100	0.87
National Review	Higher reliability conservative	3	92	0.9
Washington Examiner	Higher reliability conservative	4	92	0.82
Fox News	Lower reliability conservative	1	69	0.8
Opposing Views	Lower reliability conservative	2	75	0.27
MRCTV	Lower reliability conservative	3	62	0.86
The Political Insider	Lower reliability conservative	4	69	0.9
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
New York Post	Lower reliability moderate	2	69	0.25
AOL	Lower reliability moderate	3	75	0.01
The New York Times	Higher reliability liberal	1	100	-0.55
TIME	Higher reliability liberal	2	100	-0.33
NBC News	Higher reliability liberal	3	100	-0.27
Occupy Democrats	Higher reliability liberal	4	100	-0.9
Upworthy	Lower reliability liberal	1	75	-0.81
VICE	Lower reliability liberal	2	70	-0.63
Mother Jones	Lower reliability liberal	3	69	-0.87
Atlanta Black Star	Lower reliability liberal	4	75	-0.71

Notes: This table presents the outlets potentially offered to participants. Outlets are divided into categories based on their reliability and slant, and then ranked based on their popularity. We offered participants the first two outlets that they did not already follow in each category. 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 followed on Facebook. We collected page-follow 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 unavailable, while for the second batch, data between baseline and midline surveys is partially missing. 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. In rare cases, we retrospectively identified partial baseline data (via later snapshots), 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 conservative, as data from unaffected users indicated minimal activity between baseline and midline surveys.
- Pre-endline and post-endline follows: To identify pages followed during 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 midline survey and the snapshot.[34] 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 5.1.2, which focuses on 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, or that slant is more clustered. 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 (using mid-rank assignment for ties), then compute the average rank for each outlet across participants.

Figures A.6, A.7, A.8, and A.9 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 4, Figures A.6 and A.7 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, 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.

In contrast, as in the right panel of Figure 4, Figures A.8 and A.9 show that participants consistently underestimate the reliability of most outlets and mis-rank them relative to the truth. For

[34] 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.

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.10 confirms that perceived rankings of slant are substantially more accurate than those for reliability, consistent with the analysis in Section 5.1.2.

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 1. Column (1) reports treatment effects on slant misperceptions; column (2) on reliability misperceptions.

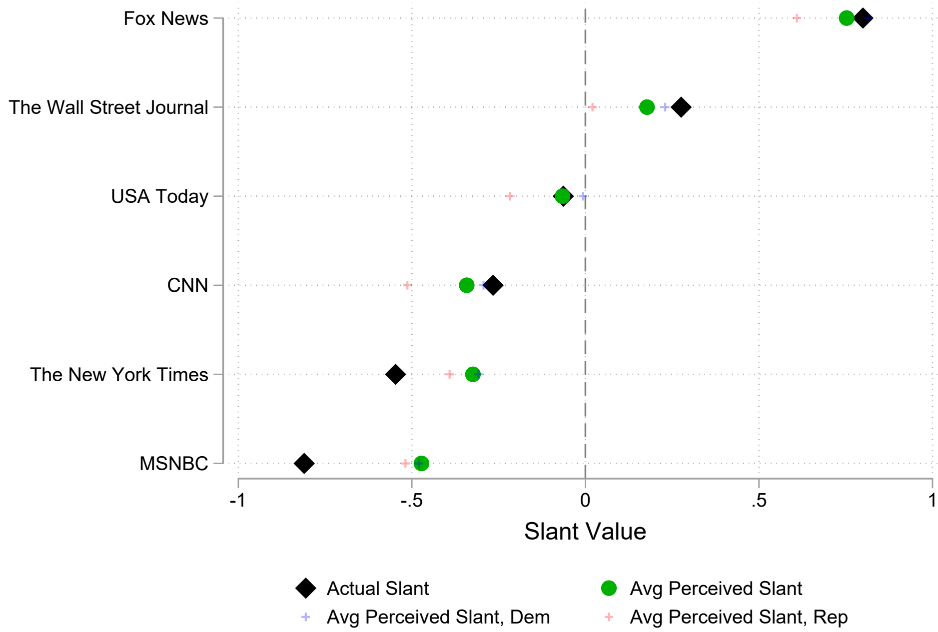
We find no detectable effect of the RS treatments on slant misperceptions. However, the RR treatment significantly reduces reliability misperceptions by 5.6 points. These findings reinforce the baseline pattern: participants already have reasonably accurate perceptions of slant, but substantial misperceptions 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.

Table A.19: ATE on Misperceptions

	(1) Average Slant Misperceptions	(2) Average Reliability Misperceptions
Re-optimization (R)	0.022	-0.028
	(0.016)	(0.947)
Re-optimization + Slant Info (RS)	-0.014	-0.331
	(0.016)	(0.935)
Re-optimization + Reliability Info (RR)	0.045***	-5.903***
	(0.016)	(0.975)
Control Mean	0.41	34.83
Control SD	0.21	12.74
Observations	1504	1512
R-squared	0.010	0.035

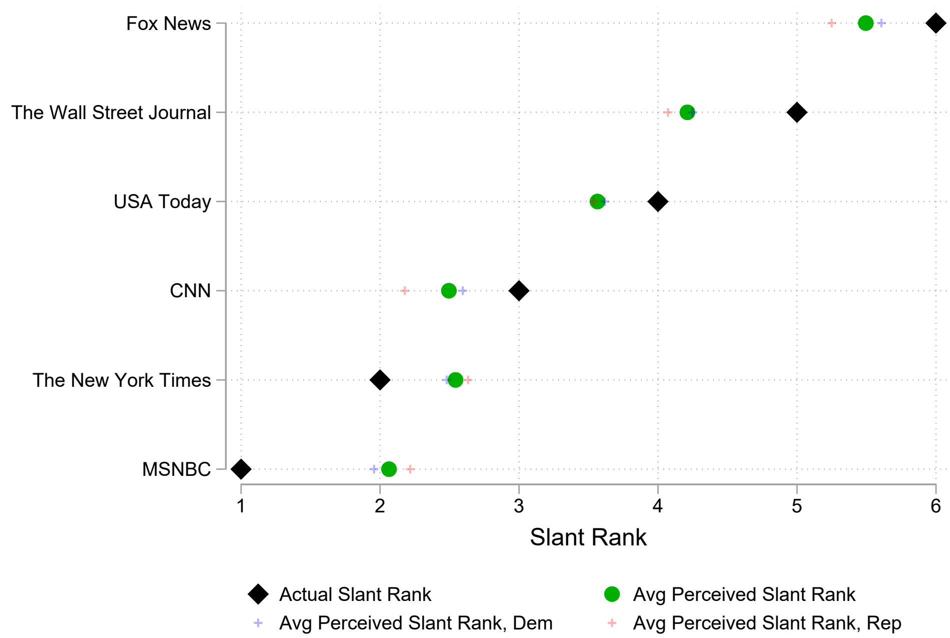
Notes: This table explores the effect of our intervention on misperceptions of slant and reliability. Specifically, it presents estimates of coefficients β_j from Equation 1, where the outcome variable is the distance between the perception and the true value, averaged over the six mainstream outlets 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.

Figure A.6: Average Perceived Slant of Mainstream Outlets



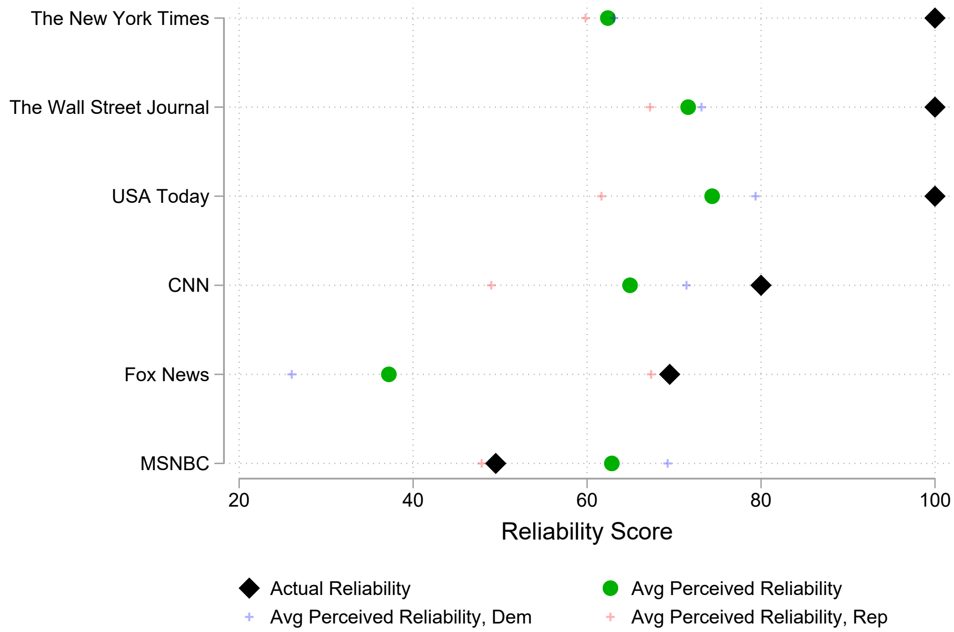
Notes: In this figure we plot the average perceived slant of six mainstream outlets (green dots) compared to the actual slant of the same outlets (black diamonds). The blue plus signs indicate the average perceived slant among self-reported Democrats, and the red plus signs indicate the average perceived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

Figure A.7: Average Perceived Slant Rank of Mainstream Outlets



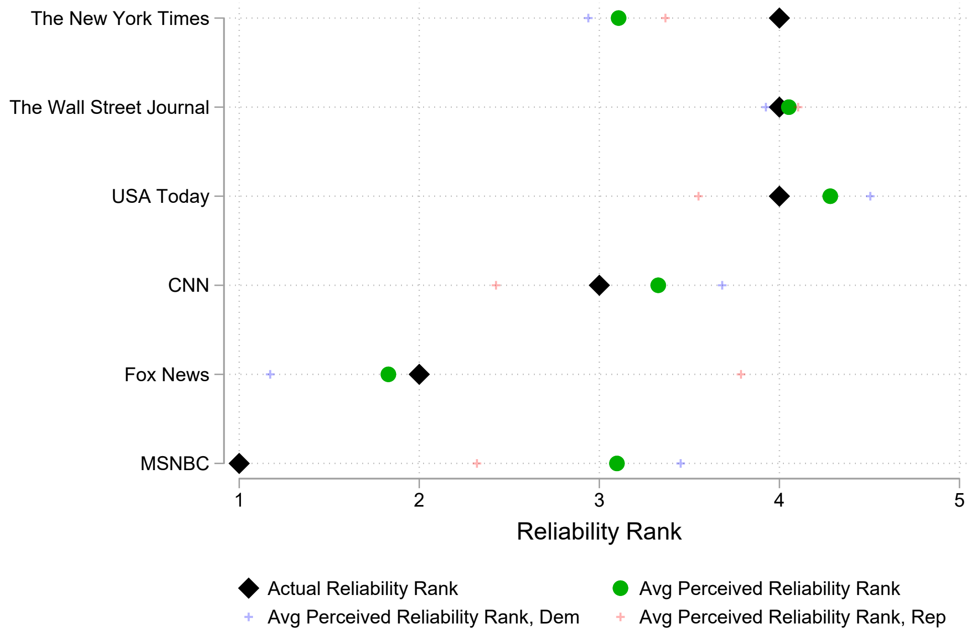
Notes: In this figure we plot the average perceived slant rank of six mainstream outlets (green dots) compared to the actual slant rank of the same outlets (black diamonds). In other words, for each individual, we create a ranking over the six outlets based on perceived slant (ties are given the same 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.8: Average Perceived Reliability of Mainstream Outlets



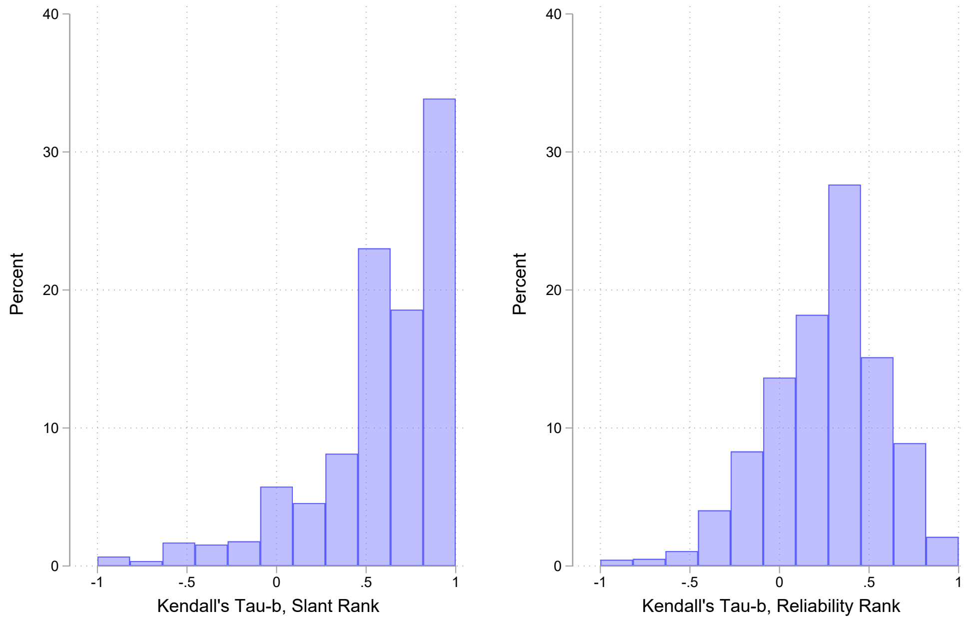
Notes: In this figure we plot the average perceived reliability of six mainstream outlets (green dots) compared to the actual reliability of the same outlets (black diamonds). The blue plus signs indicate the average perceived reliability among self-reported Democrats, and the red plus signs indicate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

Figure A.9: Average Perceived Reliability Rank of Mainstream Outlets



Notes: In this figure we plot 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.10: 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).

F. Survey Screenshots

F.1 Information Interventions

Figure A.11: 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.12: Reliability Information for RR

Here's the **average reliability** score of news outlets you currently like on Facebook:

98.5

Here is the reliability rating of each of the news outlets (ordered from highest rating to the lowest):

- The Economist: **100**
- NPR: **100**
- The New York Times: **100**
- TIME: **100**
- National Review: **92.5**

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

F.2 Like and Unlike Tables

Figure A.13: Instructions for Liking/Unliking Pages

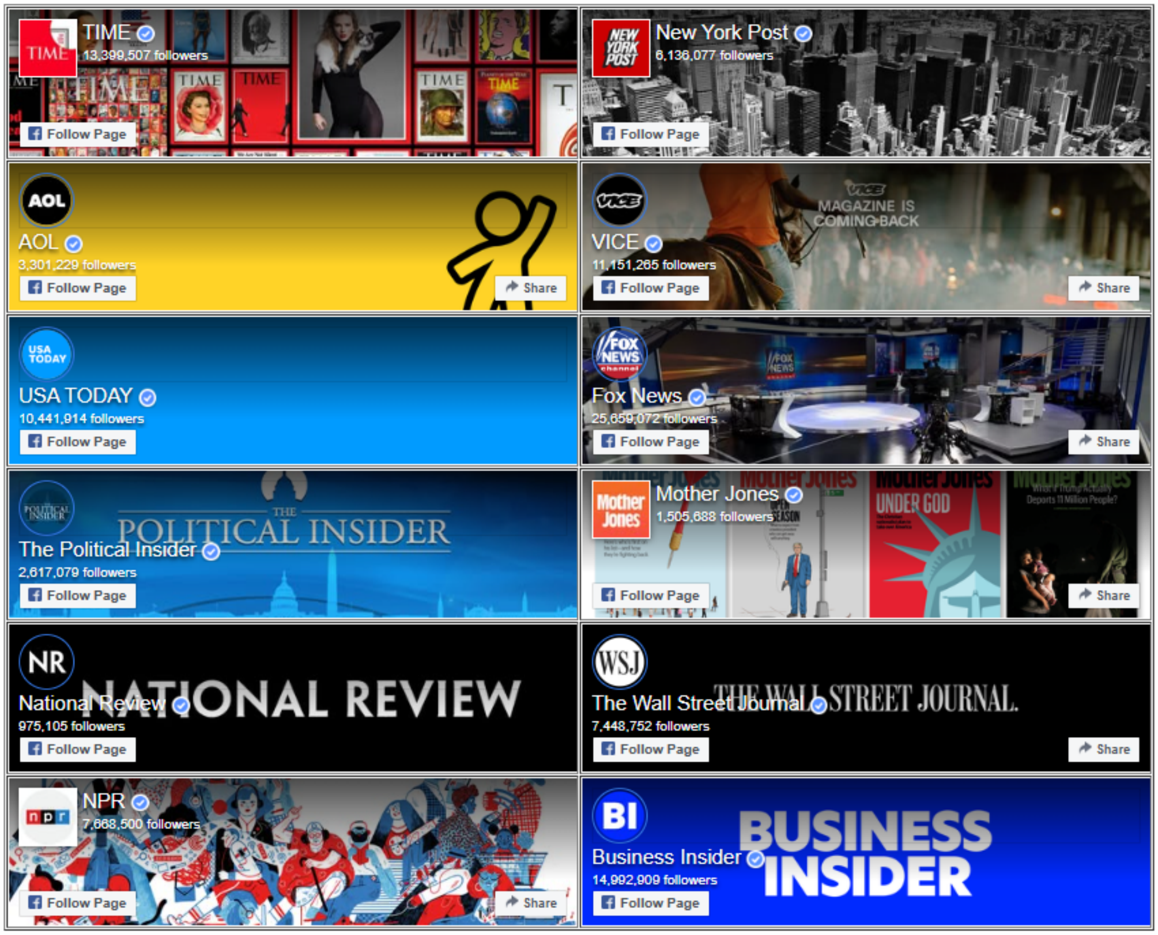
You're almost done! This (second to last) part of the survey focuses on Facebook pages. Liking the Facebook page of a news outlet is a good way to make sure articles from that outlet appear in your newsfeed. Below, we provide various options of news outlets you can select ("like" or "follow") with just one click. Pages you select are likely to start appearing in your news feed. You can also unlike or unfollow any news page you liked in the past to remove it from your newsfeed. If you are not interested in any of the pages, feel free to not select any.

The table below might take a few seconds to load, please be patient.

Please click the 'Like Page' or 'Follow Page' button on any of the pages you wish to follow.

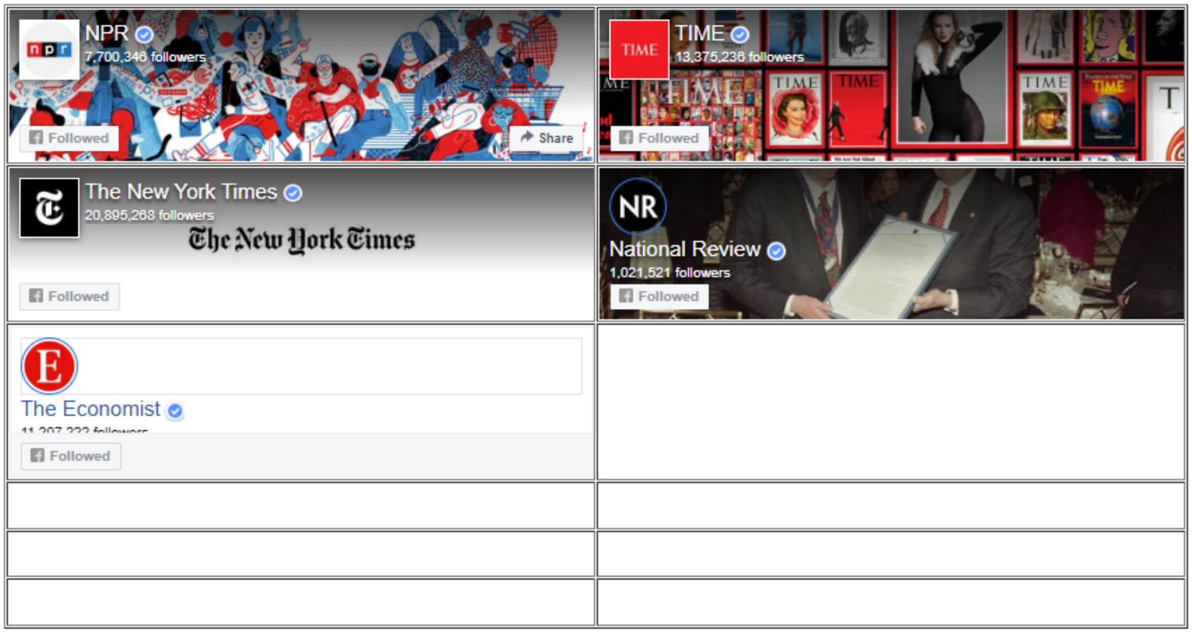
Notes: These are the instructions provided to participants regarding the re-optimization interface toward the end of the midline survey.

Figure A.14: Likes Table, R



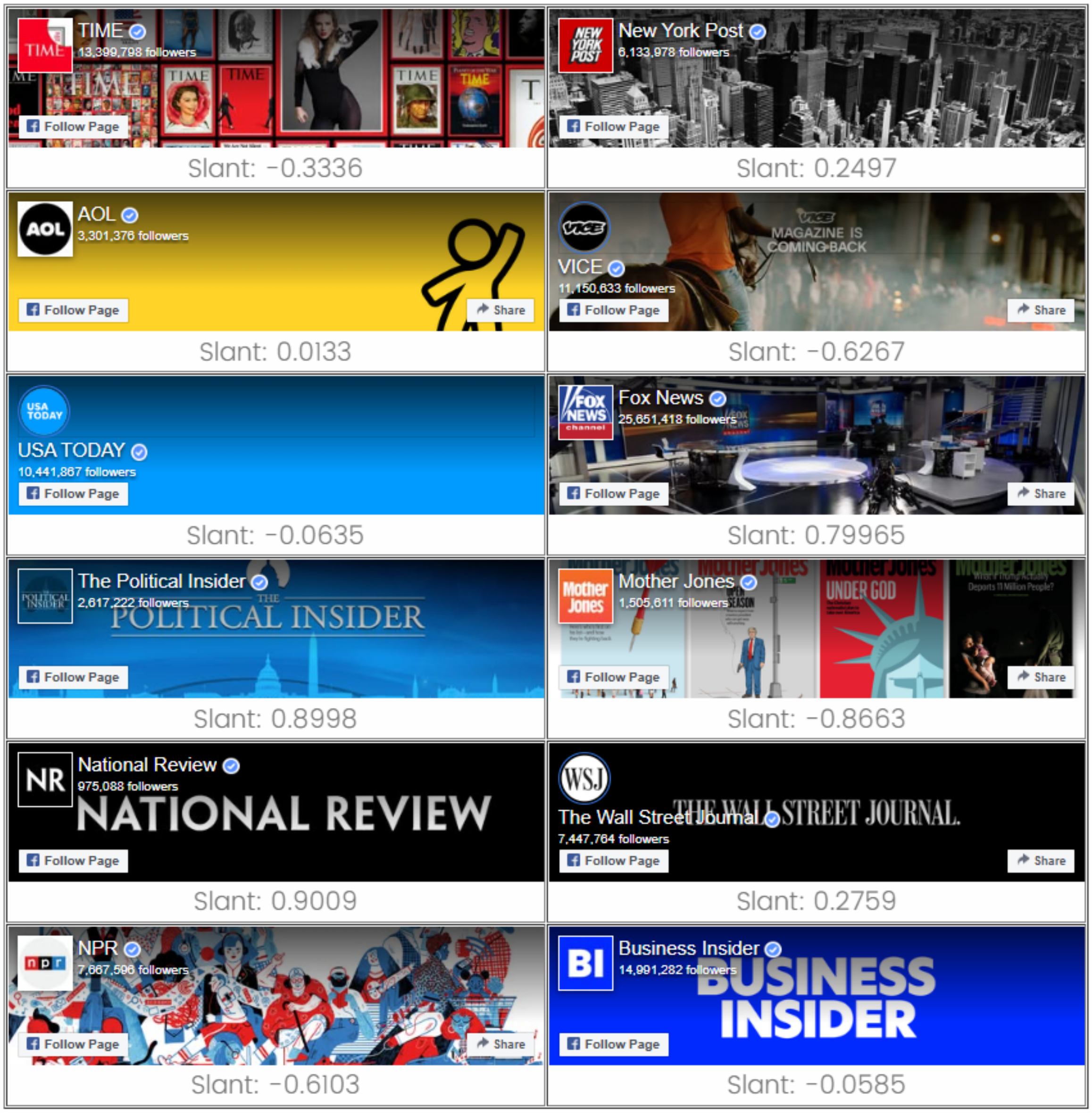
Notes: This is the first table within the re-optimization interface at the end of the midline survey. This version of the table has no labels and was shown to participants in the R treatment. Participants were able to directly clink the "Follow Page" button to follow any of the pages in the table, in which case the button would turn gray and would read "Followed". The table included 12 outlets 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.15: Unlikes Table, R



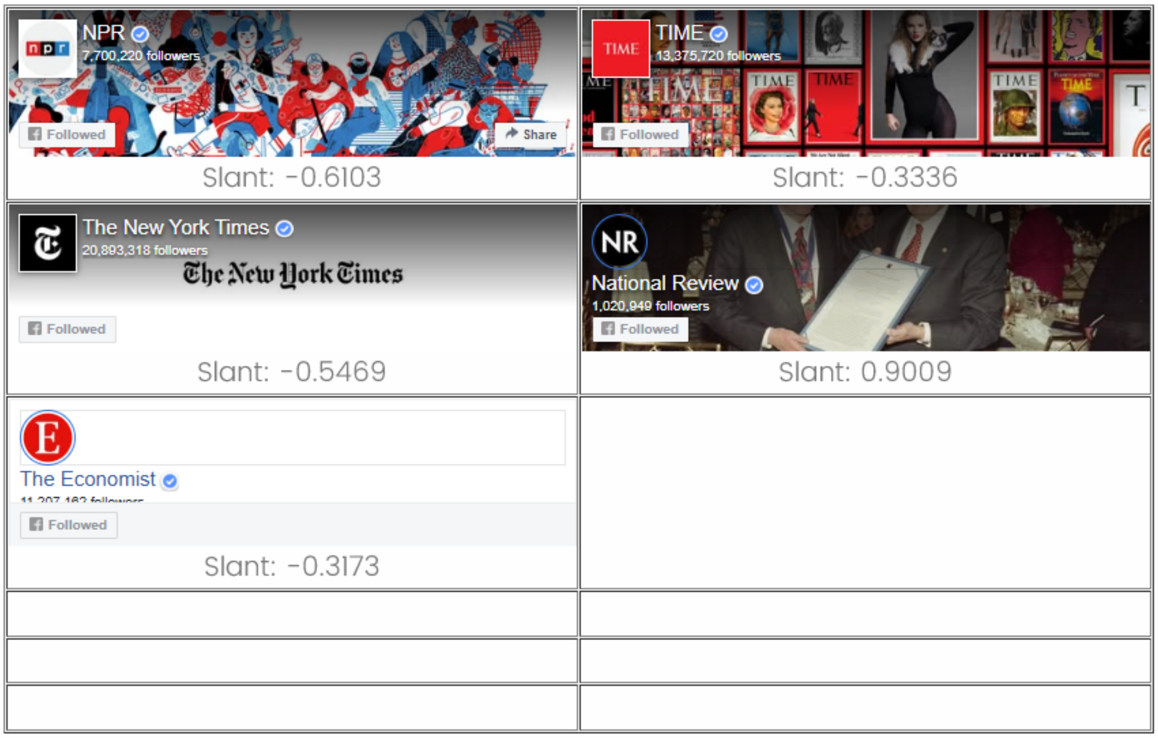
Notes: This is the second table within the re-optimization interface at the end of the midline survey. This version of the table has no labels and was shown to participants in the R treatment. Participants were able to directly click the "Followed" button to unfollow any of the pages in the table, in 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.16: Likes Table, RS



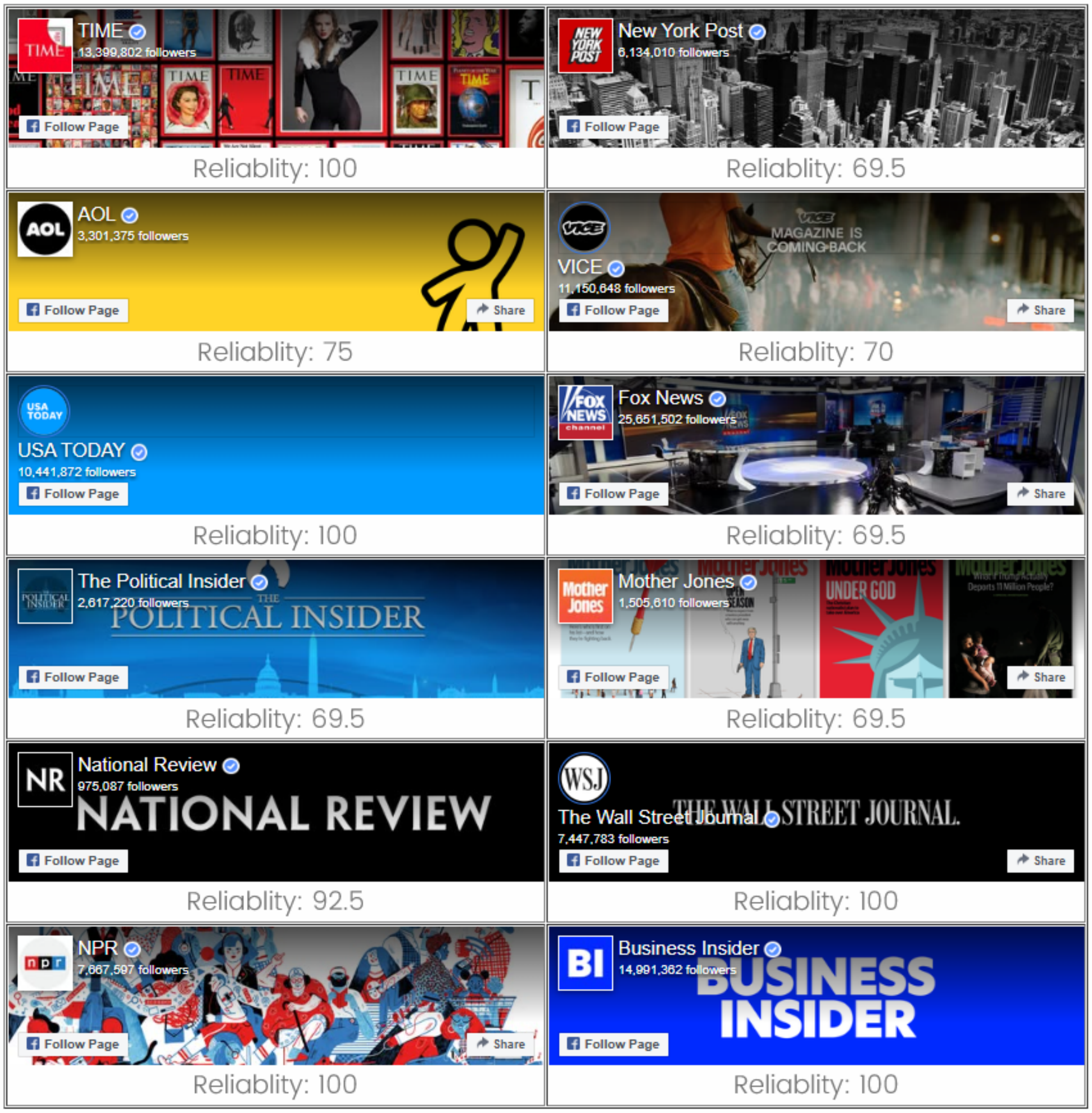
Notes: This is the first table within the re-optimization interface at the end of the midline survey. This version of the table has slant labels and was shown to participants in the RS treatment. Participants were able to directly click the ”Follow Page” button to follow any of the pages in the table, in which case the button would turn gray and would read ”Followed”. The table included 12 outlets 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.17: Unlikes Table, RS



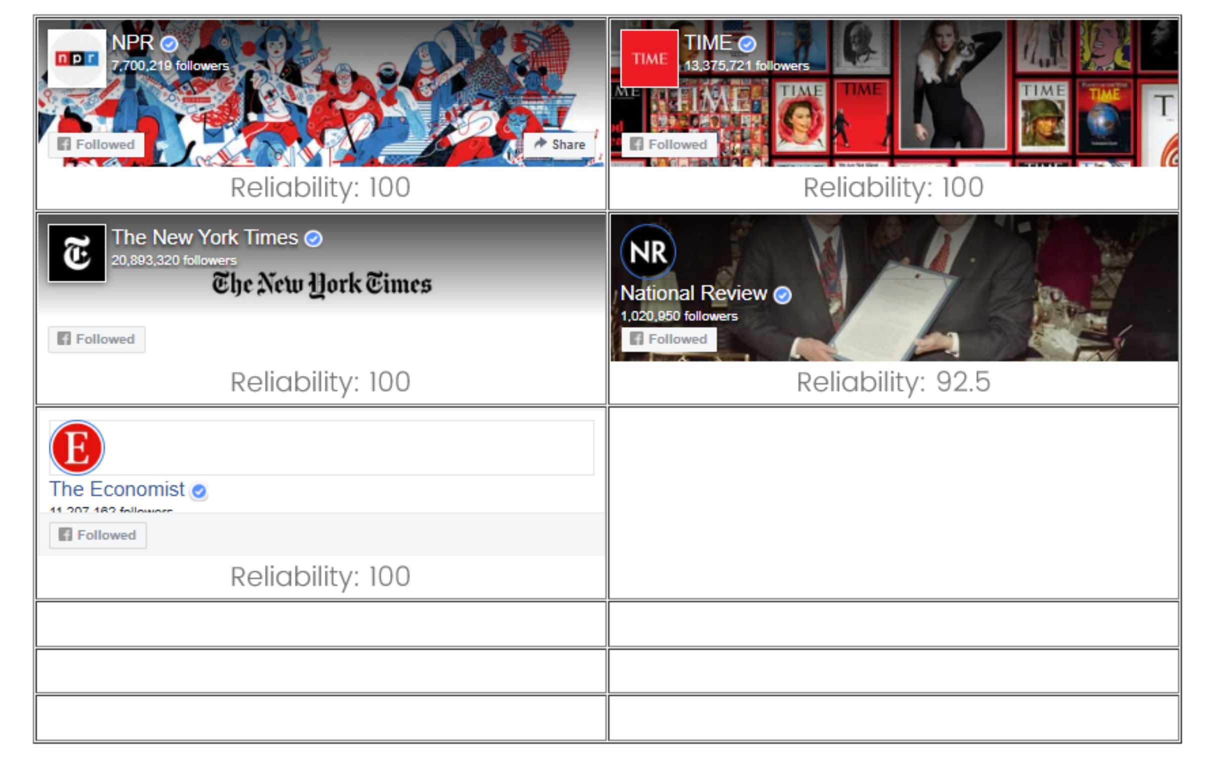
Notes: This is the second table within the re-optimization interface at the end of the midline survey. This version of the table has slant labels and was shown to participants in the RS treatment. Participants were able to directly click the "Followed" button to unfollow any of the pages in the table, 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.18: Likes Table, RR



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 clink 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 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.19: Unlikes Table, RR



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 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.

F.3 Experimenter Demand

Figure A.20: Instructions for NED (No Experimenter Demand) Group

In a previous survey we asked you for permissions to view pages you like. We no longer require these permissions and they can be removed by deleting the Making Sense of Media Study Facebook App. This would mean that we will no longer be able to see pages you follow on Facebook. You can either remove the app yourself or we can automatically remove it for you. If you prefer that we do this automatically please skip to the end of the question. If you prefer removing these permissions manually, please use the following steps:

1. Go to https://www.facebook.com/settings?tab=applications&apps_only
2. Click on "Remove" near "The Page Report"

Before proceeding to the next page, please mark whether you already removed permissions or whether we should do this automatically.

○ Please remove the permissions to observe pages I like

○ I removed these permissions manually

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 (and even if they did not we did not collect their data).

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

Figure A.21: 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.

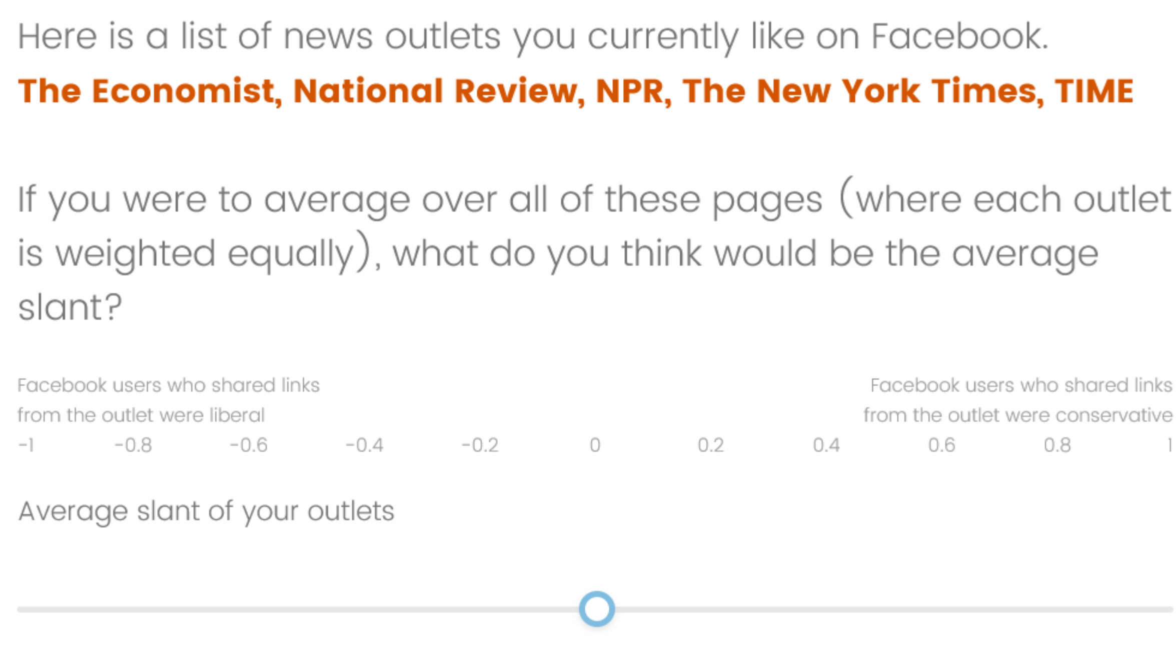
We use a measure of slant based on the ideology of people who share an outlet's content on social media. Here's how the measure works. First, it takes all of the hard news that an outlet posted on Facebook. ("Hard news" is typically used to refer to topics like politics, international affairs and business 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.

<u>The Bottom Line:</u>
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 an understanding check, which participants had to answer correctly before moving on.

Figure A.22: 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.23: Explanation of NewsGuard Measure for Reliability

We use an established measure of reliability developed by a professional organization called NewsGuard. Here's how the measure works. Each journalistic outlet receives a reliability rating between 0-100. The rating reflects the extent to which the outlet meets journalistic standards.

There are nine journalistic standards. Each standard is associated with a certain number of points. A team of analysts reviews each outlet and decides whether it meets each standard. Here are the standards:

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's 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 the standard, it gets the corresponding points. The outlet's **reliability rating** is the total points it receives. Here are a few examples.

- An outlet that meets all of the journalistic standards gets 100 points.
- An outlet that meets none of the journalistic standards gets 0 points.

The Bottom Line:

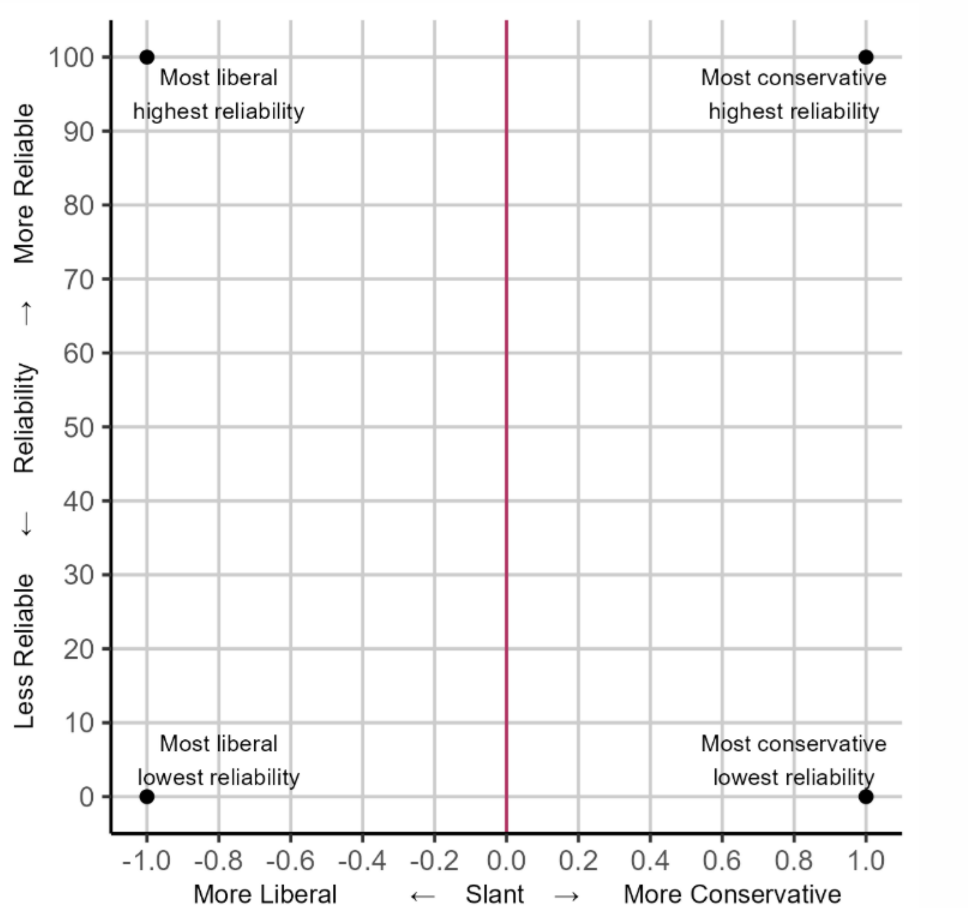
From now on, whenever we talk about the "reliability" of an outlet, we are referring to the extent to which it meets journalistic standards.

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

Figure A.24: 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.25: Elicitation of Bliss Point



Notes: This is the interface that we used to elicit the participant's "bliss point". The question accompanying the figure read: "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. Please think about the news outlets you've seen on Facebook over the last four weeks. Then, find where those outlets would fit on the figure based on how reliable and politically slanted they are, and mark that spot."