Social media advertising loads as prices*

George Beknazar-Yuzbashev[†] **Rafael Jiménez-Durán**[‡]
Andrey Simonov[§] **Mateusz Stalinski**[¶]

Preliminary draft – please do not cite or circulate

October, 2025

Abstract

Most digital platforms are funded through advertising rather than direct payments. We argue that three factors explain this prevalence: algorithmic curation enables ad-load personalization and thus price discrimination; microtargeting improves the match quality between users and ads; and users are more elastic to monetary prices than to ad loads. We conduct a field experiment on Facebook to measure the relative efficiency of ad-based funding and to decompose the sources of this efficiency. We ask 2,000 users to install a browser extension that (i) hides nearly all ads or (ii) replaces micro-targeted ads with generic ones. We find that two-thirds of ad-load variation occurs across users, consistent with price discrimination. Removing targeting sharply reduces ad clicks and long-term engagement, indicating that targeting increases the match quality between users and ads. Hiding 86% of ads increases time on the platform by 6%, showing that users are inelastic to ad loads. Counterfactual simulations reveal that an ad-based model yields higher profits and consumer surplus than a subscription model, primarily due to the low elasticity of engagement to ads. Consequently, policies that ban microtargeting could reduce platform and consumer surplus.

Keywords: social media platforms, online advertisement, user engagement, field experiment

*This research is funded by the University of Chicago's Becker Friedman Institute, Columbia Business School's Digital Future Initiative, Columbia Business School's Jerome A. Chazen Institute for Global Business, George Mason University Antonin Scalia Law School, Law & Economics Center's Program on Economics & Privacy, and Marketing Science Institute. We thank Tesary Lin, Nils Wernerfelt, Ali Goli, Yufeng Huang, JP Dubé, Lena Song, Marina Rizzi, seminar participants at VQMS, Economics of Advertising Workshop in Tallinn, Marketing Science Conference 2025. The study was approved by the Columbia University Institutional Review Board (IRB-AAAU8500). The project is logged as externally approved by the Humanities and Social Sciences Research Ethics Committee at the University of Warwick under the reference number HSSREC 264/23-24. It was pre-registered in the American Economic Association Registry for randomized control trials under trial numbers AEARCTR-0014227.

[†]Columbia University, gb2683@columbia.edu.

[‡]Bocconi University, IGIER, and Chicago Booth Stigler Center, rafael.jimenez@unibocconi.it.

[§]Columbia University, as5443@gsb.columbia.edu.

[¶]University of Warwick and CAGE, mateusz.stalinski@warwick.ac.uk.

1 Introduction

Competition authorities around the world increasingly regulate social media companies, due to their influence on online discourse and offline outcomes. Some of these regulations target the core of social media platform business model: the ability to monetize through personalized ads. In the EU, the Digital Markets Act (DMA) requires user consent to combine and cross-use personal data for advertising,[1] and the Digital Services Act (DSA) requires real-time explanations of ad targeting rules.[2] While social media companies are trying to adapt and comply with new regulations, the transition period proves to be challenging. For instance, Meta is facing a $1 billion fine for alleged violations of the DMA and failing to give consumers a less-personalized alternative to their ad-funded product.[3] The ad-funded model is under pressure due to privacy-related product changes by leading tech firms. One example is Apple's App Tracking Transparency (ATT), which reduced Meta's ad effectiveness and has lead to $10 billion revenue loss by restricting Meta's access to third-party data on consumers and the ability to target ads (e.g. Aridor et al. 2024b; Deisenroth et al. 2024).

In this paper, we evaluate the ad-funded business model of social media platforms under the current and counterfactual targeting and monetization policies. The key object of interest are advertising "loads", the share of content that users see that corresponds to ads. We describe how social media platforms set their ad loads, whether these ad loads serve as relevant prices for consumers, and how this will change as the ability of social media platforms to personalized ads to consumers deteriorates. We further evaluate counterfactual scenarios where Facebook changes its monetization strategy from an ad-funded to subscription model under different ability to personalize ads.

We ground our analysis in a theoretical framework of consumer time allocation and platform ad load choice. Users decide how much time to spend on a platform, taking as given both the ad load and the average quality of ad matches. Users may derive either positive or negative utility from viewing ads, but as long as the marginal utility of a minute spent watching ads remains below the opportunity cost of time, higher ad loads reduce engagement. This makes ad loads functionally equivalent to prices: they follow the law of demand and increase ad revenue, holding engagement fixed. We allow the platform to engage in price

[1] See Chapter 3, Article 5: https://eur-lex.europa.eu/eli/reg/2022/1925/oj/eng.

[2] See Section 2, Article 26: https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng.

[3] See https://digital-markets-act.ec.europa.eu/commission-sends-preliminary-findings-meta-over-its-pay-or-consent-model-breach-digital-markets-act-2024-07-01_en, https://winbuzzer.com/2025/03/25/meta-faces-1-billion-eu-fine-over-pay-or-consent-model-xcxwbn.

discrimination by personalizing both ad quantity and quality. The model predicts that the optimal ad load is lower for users whose engagement is more inelastic with respect to ads, who generate positive network externalities, or who belong to high-value demographics from an advertiser's perspective. Furthermore, the platform chooses to increase match quality when it raises the user's marginal utility from ad exposure and when advertisers' willingness to pay is sensitive to targeting precision. As a result, the second prediction of the model is that user engagement should decline when the match quality of ads falls.

To evaluate whether ad loads serve as prices for social media consumers, we conduct a pre-registered field experiment. In the experiment, users install a custom browser extension through which we can vary both the quantity of advertisement and the degree of their microtargeting. We use Facebook ads to recruit almost 2,000 US-based adult desktop users to install the browser extension and complete a baseline survey. Participants then enter a six-week baseline period, during which the extension does not alter any content but passively collects data on user interactions with social media platforms and the browser. At the end of the baseline period, participants are randomly assigned to one of four treatment groups: *Hide*, *Replace*, *Pure Control*, and *Keep Control*. These groups determine the intervention applied during the subsequent six-week intervention period.

In *Hide*, almost all Facebook ads are seamlessly hidden. *Pure Control* serves as the corresponding control condition, with no intervention applied. In *Replace*, ad microtargeting on Facebook is effectively eliminated by replacing user-targeted ads with generic ads (i.e., ads targeted at other users). Finally, *Keep Control* serves as an active control condition, in which ads are replaced with themselves—maintaining a similar user experience as in the *Replace* while varying only the degree of microtargeting.

We measure the impacts of the interventions on both advertisement engagement and platform engagement, as well as on time spent on non-treated platforms to capture possible substitution effects. Following the approach proposed by McKenzie 2012, we leverage both within-subject and between-subject variation to maximize statistical power in estimating treatment effects. The six-week baseline and six-week intervention periods allow us to employ a difference-in-differences approach at the day level, yielding a long panel with 134 daily periods (19.1 weeks). In addition, we collect supplementary survey outcomes—such as willingness to pay to temporarily deactivate social media—to capture effects on platform valuation.

Our first result is that platforms engage in substantial price discrimination by varying ad

loads across individuals. On average, across days and users, ad load—defined as the number of ads shown to the user divided by the number of posts and comments—was approximately 13% on Facebook, or roughly one ad per eight pieces of content. The average ad load on Twitter was similar, at 12%. Importantly, these averages mask substantial heterogeneity: 66% of the daily variation in ad load on Facebook, and 67% on Twitter, occurs between individuals rather than within individuals, suggesting that platforms systematically treat users differently.

Next, we turn to the experimental results. The hiding intervention reduced ad load on Facebook by 86%, a marked reduction relative to the ad load experienced by the Pure Control group. Since browser extensions operate only on desktop devices, this could potentially limit the strength of the intervention. However, even accounting for the desktop share of Facebook consumption in our sample (51%, as reported in the baseline survey), the intervention induced a substantial 44% reduction in ad load across all devices.

The hiding intervention increased content consumption on Facebook by 9% and active time on the platform by 6% over a six-week period. The pre-registered index of platform engagement rose by 0.03 standard deviations (SD). Our long-term estimates, based on more than 19 weeks of intervention, indicate that the engagement effects of ad load are stable over time, with a treatment effect on the platform engagement index of 0.02 SD. In addition to the baseline specification, we estimate two alternative models to address data skewness and the influence of outliers (super-users). In the first, we trim the top 1% of observations for each outcome; in the second, we apply an inverse hyperbolic sine (asinh) transformation to the data. Both approaches increase the precision of our estimates and yield significant positive engagement effects of hiding ads in both the short and long run. Across all specifications, however, users appear highly *inelastic* with respect to ad load, with an average elasticity below 0.2.

Our replacement intervention, in turn, significantly decreased the targeting fit of ads. To quantify how well an ad "fits" a particular user, we calculate the Euclidean distance between the user's demographic characteristics and the average demographic profile targeted by that advertiser. This measure shows that the intervention exposed participants to ads that are typically not directed at their demographic.

The replacement intervention reduced the index of ad engagement by 0.05 SD relative to the Keep Control group. This effect was driven primarily by a sizeable decline in the likelihood of clicking on ads—a reduction of 0.25 SD. Moreover, the intervention significantly

decreased active time on Facebook by 13% and content consumption by 13% in the long term, leading to an overall decline of 0.03 SD in the platform engagement index. These findings are robust to alternative specifications (asinh and trimming), both of which increase the precision of our estimates. The platform engagement effects also strengthened over time—short-term estimates showed negative but smaller and statistically insignificant impacts on engagement. Overall, the results suggest a degree of complementarity between user engagement and advertisement microtargeting, indicating possible benefits of microtargeting both to the platform and the users.

We report several supplementary results. First, we find no evidence of an algorithmic response by Facebook to the hiding intervention. Specifically, the intervention did not alter the average ad load supplied by the platform prior to any filtering, with a treatment effect of only 0.3 pp over the six-week intervention period. Moreover, no algorithmic response is detected in the long run—the nineteen-week treatment effect is just 0.2 pp. In contrast, Facebook's algorithm appears mildly responsive to the replacement intervention: the average ad distance of supplied content (before any intervention was applied) increased by 1.5%. This pattern may suggest that the algorithm has internalized the inelasticity of user engagement with respect to ad load, while recognizing that users' preferences regarding product categories are more malleable.

Second, we find little evidence of substitution between Facebook and other digital platforms in response to the interventions. The hiding intervention, however, increased time spent on users' top ten favorite domains (as defined by baseline data), potentially reflecting a "linking effect," whereby greater time on Facebook translates into higher engagement with websites frequently linked in users' feeds. Finally, we find no evidence that either intervention affected willingness to pay to deactivate social media, a proxy for users' valuation of the platform.

We conduct counterfactual analysis to understand why the microtargeting ad-based business model is so prevalent and the drivers of the profitability of this business model. We decompose how much of the status quo's profitability compared to a subscription-based model comes from targeting, ad-load discrimination, and the relative elasticity of demand with respect to monetary prices as opposed to ad loads. Using our experimental estimates, we simulate four regimes: (i) no microtargeting with today's ad load, (ii) no microtargeting with a platform-wide ad load chosen to maximize profits, (iii) a paywall with no ads, and (iv) no ads and no paywall. The exercises reveal that the bulk of value arises from having

an ad-based monetization technology at all: turning off microtargeting or ad-load discrimination moves outcomes modestly, whereas switching to a subscription reduces both profits and user welfare. These findings help rationalize the prevalence of ad-based business models on social media.

We address potential concerns with our approach. One might worry that the results reflect differences in user experience induced by the extension, rather than variation in ad load (hiding intervention) or ad microtargeting (replacement intervention). We took several steps to preempt such issues. The extension holds each batch of content until all posts, comments, and ads are processed—including any intervention—before displaying it to the user. As an additional safeguard, we never hide or replace ads that are in the visible portion of the page (viewport) when loaded.[4] As a result, users never observe the hiding or replacement actions, and the extension operates seamlessly. In the replacement intervention, we further control for potential user experience effects by measuring treatment effects relative to an active control group in which ads are replaced with themselves; thus, the only difference between groups is the degree of microtargeting.

Another potential concern is differential attrition. Using a conservative measure of survival, we find that more than 75% of participants in all treatment groups remained in the study for 12 weeks, covering both the baseline and intervention periods. Survival rates do not vary by treatment group. This result holds in specifications that control for survival propensity—based on individual characteristics—and its interaction with the treatment indicator. Thus, our findings are unlikely to be driven by differential attrition. This conclusion also applies to our long-term results—survival likelihood likewise does not vary by treatment.

Theoretically, the idea that ad loads behave as prices appeared at least since Anderson and Coate 2005 (see Calvano and Polo 2021 for a review and Anderson and Peitz 2020 for a more recent framework). Empirical research has documented a negative relationship between ad loads and user engagement on traditional media (Wilbur 2008) and on digital platforms such as Pandora and e-commerce websites (Moshary 2021; Goli et al. 2025a; Goli et al. 2025b). In the context of social media, Brynjolfsson et al. 2024 report results from a 9-year long field experiment on Facebook and find an elasticity of time spent with respect to ad loads of 0.094. We contribute to this literature by varying not only the quantity but

[4]This safeguard helps in cases where the extension fails to hold the entire batch before processing. Moreover, since most batch content is off-screen, this precaution has little impact on the strength of the intervention.

also the quality of ads, provide evidence of price-discrimination, and explain why ad models are such a prevalent business model.

Our work also relates to growing subset of the literature on the economics of social media (Aridor et al. 2024a) that focuses on platform business models. This work highlights the impact of ad-driven incentives on the proliferation of harmful content (Liu et al. 2022; Acemoglu et al. 2024; Beknazar-Yuzbashev et al. 2024). Ad loads are central to the antitrust questions related to social media (e.g. see Katz and Allcott 2025). We contribute to this literature by providing a rationale for why advertising-based business models are so prevalent in the context of digital platforms.

Lastly, we contribute to the literature on the economics of privacy (Acquisti et al. 2016; Lin 2022; Goldfarb and Que 2023; Dubé et al. 2025), which has shown that limiting tracking can reduce platform revenue (Alcobendas et al. 2023; Aridor et al. 2023; Aridor et al. 2024b; Johnson 2024). We contribute to this literature by isolating the user-engagement channel from advertiser-side effects by reducing micro-targeting for users.

The paper proceeds as follows. Section 2 presents a conceptual framework. Section 3 presents our experimental design. Section 4 presents our results, and Section 6 concludes.

2 Conceptual Framework

We now present a model where users spend time on a platform. The platform can personalize both the quantity of ads it serves to the users and their average match quality.

Users. There is a set of users $\mathcal{I}$ who decide how much time to spend on a platform to maximize their utility. Users derive utility from two sources: from the amount of time they spend consuming content and from the amount of ads they consume. Consider a user $i \in \mathcal{I}$ who allocates time $t^i \geq 0$ to consume content on the platform. The user takes as given the (potentially personalized) ad load denoted by $a^i \in [0, 1]$, which denotes the amount of time spent consuming ads for each unit of time spent consuming content. Thus, the total amount of ads that the user consumes is $A^i = a^i t^i$. The user also takes as given the average ad quality (match quality) denoted by $q^i \in [0, 1]$.

The total amount of time spent on the platform is $t^i(1 + a^i)$. Each unit of time spent carries an opportunity cost w^i, reflecting either the value of the user's time or their implicit cost of sharing personal data through engagement.

The user's optimization problem, considering the utility from content consumption, ad consumption, and the opportunity cost of time, is:

$$\max_{t^i \geq 0} \; U^{i,c}(\mathbf{t}) + U^{i,a}(t^i a^i, q^i) - w^i t^i (1 + a^i),$$

where $U^{i,c}$ is the utility of content consumption, which depends on the vector of engagement from all users, $\mathbf{t}$, to capture direct network effects from other users' engagement. We assume that the marginal utility of time spent is positive, $U^{i,c}_{t^i} > 0$. The utility from consuming ads is $U^{i,a}$, which depends both on the quantity and quality of ads consumed. The marginal utility from ads, represented by $U^{i,a}_{A^i}$, can be positive or negative, but we restrict it to be lower than the marginal utility of engagement: $U^{i,a}_{A^i} < U^{i,c}_{t^i}$ in other words, users prefer to spend time consuming content vs. consuming ads. The quality of ads has a "vertical" nature, in the sense that it increases the marginal utility of ads, $U^{i,a}_{A^i q^i} \geq 0$. However, users can still derive positive or negative utility from consuming ads with a better match quality; for example, reflecting that they might have some instrumental utility from privacy (Lin 2022), which would result in users experiencing a negative marginal utility from a better match quality. Lastly, to ensure that second-order conditions hold, we assume that both utilities are concave in t^i. The solution to this optimization problem is $T^i(\mathbf{t}^{-i}, a^i, q^i)$, where $\mathbf{t}^{-i}$ is the vector of engagements of the other users.

Platform. The platform chooses ad loads and content targeting to maximize ad revenue. Following (Weyl 2010), we assume that the platform follows an "insulating" tariff strategy, that allows it to, effectively, select its users' engagement levels. Define the inverse ad demand for user i as $a^i(\mathbf{t}, q^i)$. Let $\mathbf{q}$ denote the vector of ad qualities for all users. Additionally, let $p^i(a^i, q^i)$ denote the price (in dollars per unit of time) that the platform charges to advertisers. We allow this price to depend on the ad load, to capture potential information overload (Anderson and De Palma 2012), and on the match quality, to capture that advertiser willingness to pay increases with a better match quality. Lastly, we assume that there is a small but non-zero marginal cost c per unit of time spent on the platform.

The platform's optimization problem is then to choose engagements ($\mathbf{t}$) and ad match qualities ($\mathbf{q}$) to maximize its profits:

$$\max_{\mathbf{t},\mathbf{q}} \sum_i \left[p^i(a^i(\mathbf{t}, q^i), q^i) \times a^i(\mathbf{t}, q^i) - c \right] t^i.$$

Rearranging the first-order condition with respect to t^i, we get a formula for the optimal advertising load of the platform:

$$a^i = \frac{c}{p^i}\left[1 - \frac{1 + \varepsilon_{p^i,a^i}}{|\varepsilon_{t^i,a^i}|} + \sum_{j \neq i} \frac{s^j}{s^i} \frac{\varepsilon_{t^j,t^i}}{|\varepsilon_{t^j,a^j}|}\right]^{-1}, \tag{1}$$

where $\varepsilon_{x,y}$ denotes the elasticity of x with respect to y. This expression characterizes the optimal ad load that the platform should set. Concretely, a user's advertising load should be lower when:

1. All else equal, the user's engagement is elastic with respect to ad loads; that is, when $|\varepsilon_{t^i,a^i}|$ is large.
2. All else equal, the user's engagement brings positive network effects on average; that is, when $\sum_{j \neq i} \frac{s^j}{s^i} \frac{\varepsilon_{t^j,t^i}}{|\varepsilon_{t^j,a^j}|}$ is large.
3. All else equal, and provided that network effects are small enough, the advertising price p^i is higher.

In terms of the optimality condition with respect to q^i, our assumptions imply that the platform will reach a corner solution and maximize the match ad quality.

3 Experiment Design

3.1 Study Flow

Figure 1 summarizes the study flow. The interventions are delivered via a browser extension, described in detail in Section 3.2. We recruited participants through social media advertisements targeting English-speaking, US-based adults who primarily use desktop computers. This approach ensures that participants conduct a substantial share of their social media activity on desktop, enabling effective intervention via a browser extension (which does not affect mobile usage). Prospective participants who clicked on a recruitment ad were directed to the study website, where they could learn more about the project.[5] Participants were then asked to provide their email address, install the browser extension via the *Chrome Web Store*, and complete the onboarding process, including a baseline survey. After a six-week baseline period during which the extension passively collected data, participants were randomized into one of four treatment groups, and the intervention period commenced.

[5]The study website is available at https://business.columbia.edu/social-media-platform-study. We hosted the site on an official Columbia University domain to enhance trust among prospective participants.

During the six-week intervention period, the browser extension modified participants' Facebook ad experience according to their assigned experimental group. In the *Hide* group, nearly all Facebook ads were seamlessly hidden. The *Pure Control* group experienced no changes to their ad experience. In the *Replace* group, most microtargeted ads were seamlessly replaced with generic ads (i.e., ads targeted at other users). The *Keep Control* group served as an "active" control condition, in which ads were replaced with the same ads originally shown—thus matching the replacement process of the Replace group while preserving ad content. After the intervention period, participants were invited to complete an endline survey, distributed via email and browser notifications.

3.2 Browser Extension

Prospective participants install a browser extension called *Social Media Research*. The extension was built for *Google Chrome* browser, the market leader with a 65% market share.[6] Browser extensions are commonly used by Internet users. For example, *Grammarly* can boast of 47 million extension users,[7] and *Google Translate* extension is not far behind with 38 million active users.[8] Furthermore, relying on browser extensions to collect research data and conduct randomized controlled trials has become increasingly popular (Levy 2021; Beknazar-Yuzbashev et al. 2025; Farronato et al. 2024; Beknazar-Yuzbashev and Stalinski 2022; Aridor 2022).

Upon installation, all participants complete onboarding, including the consent procedure. Any other extension functionality is not enabled until the process is finalized and the participant expresses affirmative consent. Overall, the extension performs two key functions: it can modify participants' Facebook page and it collects data about their browsing activity, especially their interactions with social media websites. We will explain the scope of each function in turn.

3.2.1 Ad hiding and ad replacement actions

The extension modifies the advertising experience for Facebook users by identifying ads in their feeds. It then applies one of two actions: (i) seamlessly *hiding* the ad with probability

[6] https://gs.statcounter.com/browser-market-share, accessed 15-09-2024.

[7] https://chromewebstore.google.com/detail/grammarly-ai-writing-and/kbfnbcaeplbcioakkpcpgfkobkghlhen?hl=en, accessed 15-09-2024.

[8] https://chromewebstore.google.com/detail/google-translate/aapbdbdomjkkjkaonfhkkikfgjllcleb?hl=en-GB, accessed 15-09-2024.

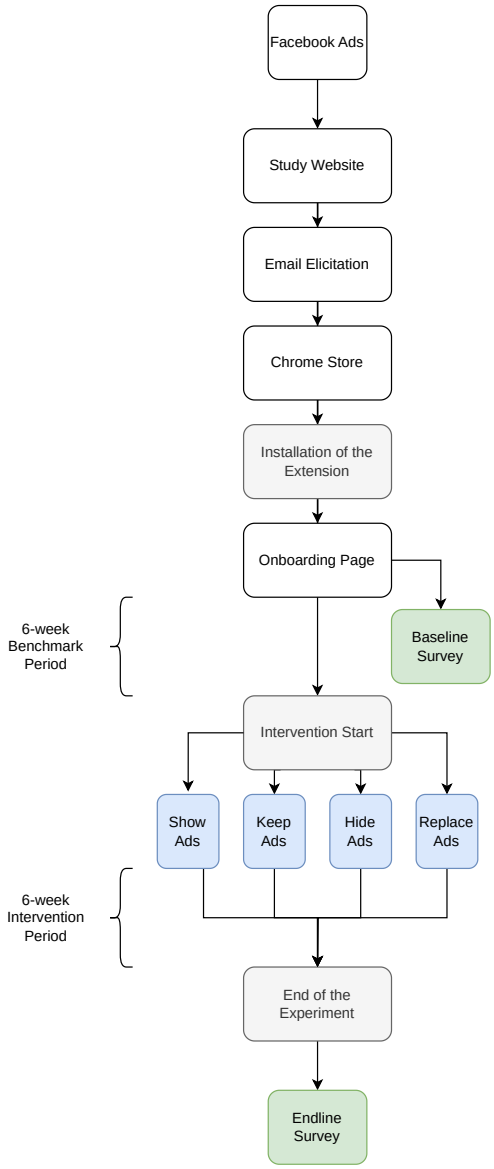


FIGURE 1: STUDY FLOW

one, or (ii) seamlessly *replacing* the ad with another ad with probability one.

We "process" ads regardless of the group to address potential latency differences, but only implement the effect associated with the group.[9] Additionally, the extension holds each

[9]This feature is important because replacement requires additional calls to the database to retrieve data about the replacement ad.

batch of content until all elements within it have been processed and acted upon (including any hiding or replacement), ensuring that all users experience a consistent delivery of content and minimizing any differences.

Hiding Action. Figure A1 in the appendix illustrates the effects of the ad hiding action. The left panel shows an unmoderated feed with a non-ad post, followed by an ad from *Spoke*, a company offering custom-fit garments, and another non-ad post. The ad is framed in red for demonstration purposes. In the right panel, we display the moderated feed, after the ad hiding action has been applied. The *Spoke* ad is removed, and the second non-ad post is pulled upward, along with the content below it. As a result, a third non-ad post, from *Time Out London*, becomes visible.

We take great care to ensure that the ad hiding action is seamless. First, the moderated feed is formatted naturally, with no gaps or unusual spacing. Second, as previously mentioned, the extension holds each batch of content until all elements within it are processed and acted upon, and only then displays the batch to the user. This ensures that the viewer does not see any elements disappearing abruptly. However, since this process may occasionally not work during the initialization of the Facebook page (affecting the first batch of content loading on the page), we introduce an additional fail-safe. Specifically, the extension does not hide ads that are in the viewport (i.e., the visible area of the page) at the point of detection. These features are crucial because, often, when Facebook is first loaded, the second post happens to be an ad fully visible from the user's perspective.

Replacement Action. The ad replacement action means that an ad that Facebook intends to display to a user is replaced with a different ad. In such case, we change features of the original (replaced) ad with those of the replacement ad. The change encompasses all features except for the call to action button and includes the text, image, and URL of the ad. The extension can replace an ad with a randomly selected ad from a pool in our database. The ads in our database are English-language Facebook ads recently collected from feeds of *all* users, which means that the replacement ad is not microtargeted at the current user.

Figure A2 in the appendix shows an example of an ad that replaced one that Facebook intended to display. The left panel of the figure demonstrates an image of the replacement ad—it is formatted in the same way as any other Facebook ad. The right panel of the figure showcases possible interactions with the replacement ad. Users can like the ad or

react to it using any of the available emojis, and the reaction will be visible in a natural way, as displayed in the figure. Furthermore, if a user clicks on the ad, its image or the call to action button, it will take them to the advertiser's chosen website, which adds to the naturalness of user experience. If someone attempts to comment on or share the ad, a small pop-up displays that commenting or sharing is disabled (as shown in the figure).[10] In general, authors of Facebook posts can opt to disable commenting, so encountering ads with commenting disabled is not unusual. Moreover, it is rare for users to attempt commenting on or sharing ads.

As with the hiding action, the replacement action is not performed for ads in the viewport. Overall, the same measures that ensure the seamlessness of the hiding action are also applied to the replacement action. Additionally, we do not replace ads containing videos, forms, or carousel ads, as we cannot ensure that these will be replaced with the same type of ad. Instead, we replace image-based ads with other image-based ads.

3.2.2 Data collection

On top of enabling the experimental intervention, the extension collects data on user interactions with social media websites, and the browser more generally. In particular, the extension collects information on all Facebook ads, posts, and comments displayed to the user. This includes author's name, a link to their profile, the text of the element, as well as its image. Furthermore, for ads, the extension records the link to the advertiser's preferred website (on click of the ad), the text of the call to action button, and the type of ad (image, video, carousel). This allows us to have a broad pool of potential Facebook replacement ads for the replacement action (as described in Section 3.2.1). The extension also records user interactions with Facebook content and the interaction's timestamp. Covered interactions include viewing, clicking, commenting on, sharing, and reacting to content.

In addition to Facebook, the extension records all such data for Reddit and X (formerly Twitter). Furthermore, the extension periodically (approximately every minute) records the domain and subdomain names of the active tab of the browser.

[10] This is a necessary technical limitation, because otherwise all comments and shares would apply not the replacement ad but the original one, which would be visible to the user's friends.

3.3 Recruitment and Sample

Recruitment took place between June 18, 2024, and December 24, 2024. Figure A3 in the appendix provides an example of our recruitment advertisements. The ad emphasizes that participants can earn $21 for completing the study, which involves installing a browser extension and completing two surveys. We describe the study as cutting-edge academic research on social media but intentionally avoid disclosing specific information about potential experimental treatments—we do not mention any interventions on advertisements. Throughout recruitment and onboarding, we present the extension as a tool designed to enhance user experience, while noting that it may modify social media content. This framing minimizes the risk of self-selection based on participants' advertisement preferences.

We construct our main experimental sample in the following way. As specified in the pre-registration, we include all individuals who installed the browser extension and used their Facebook account for at least one minute per week, on average.[11] Further following the pre-registration, we exclude people who did not use Chrome or Edge browsers, those whose language preference on Facebook was not set to English, and those who uninstalled the extension before the intervention period began.[12] Lastly, in an attempt to eliminate users who relied on bots or exhibited unusual behavior, we exclude individuals for whom any of the outcomes are more than 30 standard deviations from the mean—in our case, only one person was dropped this way. After applying the above inclusion and exclusion criteria, we arrive at the main experimental sample of 1,810 users.

3.4 Randomization and Treatments

At the end of the six-week baseline period, we randomized participants into one of four experimental conditions—*Hide*, *Replace*, *Keep Control*, and *Pure Control*. In the *Hide* group, the extension automatically hides all Facebook ads that are outside the viewport. This hiding mechanism is described in detail in Section 3.2.1. The corresponding control group, *Pure Control*, experiences no modifications to the user experience. Thus, between these two groups, we exogenously vary the ad load—effectively altering the implicit price of Facebook consumption. In the *Replace* condition, we replace ads in participants' Facebook feeds

[11] Activating the intervention required that users have the latest version of the extension installed at the start of the intervention period. Hence, only the users with the latest version at the time of the intervention start are included in the sample.

[12] Uninstallations are based on the extension recording a specific uninstallation event. We do not impute uninstallations from the lack of user activity.

with randomly selected ads from our ad database. The replacement procedure is described in detail in Section 3.2.1. The associated active control group, *Keep Control*, receives the same replacement treatment, except that eligible ads are replaced with themselves. The replacement process—including ad formatting and possible user interactions—is held constant across both groups. Consequently, the only exogenous variation between the groups is the degree of ad microtargeting.

Sample Size. Out of the 1,810 users in our sample, 414 were assigned to Pure Control, 413 to Hide, 445 to Keep Control, and 538 to Replace. We initially used equal probabilities for group assignment. However, given the technical challenges associated with replacing ads (higher likelihood of encountering errors), we ended up oversampling Replace condition. Our empirical strategy, outlined in Section 3.6, accounts for the oversampling by using inverse probability weights.

Sample Balance. Table B1 in the appendix presents sample balance for the Show vs. Pure Control conditions, while Table B2 reports balance for Replace vs. Keep Control. In both cases, we assess balance on twelve covariates, including user demographics (age, sex, ethnicity, income, education, and political affiliation) and social media usage variables (number of Facebook friends and baseline engagement measures). We find no evidence of imbalance in either comparison—none of the covariates differs significantly by treatment at the 5% level and only one covariate is significant at the 10% level (in Table B2).

3.5 Outcome Variables

We collect two types of outcomes: browser-based measures—mostly recorded by the browser extension to document people's social media activity and their interactions with content—and survey measures—providing further insight into participants' attitudes and preferences. The survey measures are collected during the baseline and endline surveys. We discuss each type of outcomes in turn.

3.5.1 Browser-Based Measures

Advertising Loads. We collect advertising load data at the user-day-platform level. Ad load is defined as the number of ads shown to the user divided by the total number of posts and comments shown on a given platform, yielding a value between 0 and 1. We refer to

this measure as the *experienced* ad load, as it reflects the share of ads actually viewed by the user. However, because platforms may endogenously adjust ad delivery in response to user behavior altered by the treatments, we also construct a second measure: the *supplied* ad load. This captures the share of ads the platform originally intended to show—i.e., before any hiding or replacement occurred as part of the intervention.

Time on Social Media. Our primary measure of user engagement is time spent on social media platforms. First, we are interested in the direct effects of the treatments on active time spent on Facebook, which is the platform where we intervene. Our definition of active time follows Beknazar-Yuzbashev et al. 2025, and counts only time during which the user actively loads new elements on the page (which requires activity such as scrolling) and a short period after.[13] Furthermore, we document the substitution effects to Reddit and X (formerly Twitter), reporting active time measured in the same way.

To further understand any substitution effects, we provide additional measures of time spent on other platforms. Specifically, we report the approximate time a website was open in the active browser tab for YouTube, as well as the combined time spent on a set of other social media and related websites listed in Beknazar-Yuzbashev et al. 2025. We will also discuss time spent on the top-10 websites visited by each participant during their baseline period. These time measures are less precise than the active time reported for Facebook, X, and Reddit. Collecting active time requires the browser extension to carefully identify DOM elements on each specific platform, and is very costly to implement and maintain.

Platform and Ad Engagement. We collect additional measures of platform engagement. First, we consider the total number of posts and comments shown to the user per day, excluding advertisements, as their quantity is affected by the hiding intervention. Second, we construct an index of platform engagement on Facebook, based on active time and content consumption, following the index construction method of Kling et al. 2007. To complement our analysis of user engagement, we examine users' direct interactions with Facebook ads. Specifically, we construct an index based on the number of ad impressions and ad clicks per day. We also report ad impressions and ad clicks as separate outcomes.

[13] Any time on the website following more than three minutes of inactivity is considered idle.

Short-Term vs. Long-Term Effects. After the end of the six-week pre-registered intervention period, we continued the intervention and monitored user engagement. This allows us to report both short-term (six weeks) and long-term effects of the intervention. The latter are based on the intervention duration of 134 day-periods, which is equivalent to 19.1 weeks. Despite the extended observation period, we find no evidence of differential attrition (see Section 4.3.1 for a discussion of attrition).

3.5.2 Survey Measures

Platform Valuation. We seek to understand the impact of the treatments on social media valuation by users. To that end, we collect, as a primary outcome, the willingness to accept (WTA) to deactivate social media for four weeks.

3.5.3 Heterogeneity

One of the main objectives of this paper is to document heterogeneity in ad loads across social media platforms. We report the distribution of (1) daily ad loads per user by platform and (2) average ad loads per user by platform. To further explore the sources of variation, we conduct a variance decomposition exercise to assess whether between-user variation exceeds within-user variation. Separately, we analyze heterogeneity of treatment effects of the hiding and replacement interventions with respect to baseline levels of engagement.

3.6 Empirical Strategy

To maximize statistical power, our experimental design relies on within-subject variation. For outcomes continuously measured by the browser extension, we use a difference-in-differences approach over a six-week baseline period (treatments disabled) and a six-week intervention period (treatments enabled), with the individual-day as the unit of observation.[14]

Specifically, we estimate the following two-way fixed effects (TWFE) regression:

$$Y_{i,t} = \alpha_i + \delta_{t,C_i} + \beta D_{i,t} + Errors_{i,t} + \varepsilon_{i,t}, \tag{2}$$

where $Y_{i,t}$ is the outcome for user i in period t (measured relative to the intervention start date), α_i represents individual fixed effects, δ_{t,C_i} are cohort $\times$ period fixed effects, $D_{i,t}$ is

[14]For outcomes not continuously measured by the extension but collected in the baseline and endline surveys, we estimate regressions of the outcome on a treatment group indicator, an endline survey indicator, and their interaction, including individual fixed effects.

an indicator variable equal to 1 if the user was treated, and $Errors_{i,t}$ consists of binary variables indicating that individual i experienced a particular type of a browser extension error in period t and their interactions with treatment.[15]

Depending on which treatment effect we want to estimate and against what counterfactual, $D_{i,t} = 1$ corresponds to the Hide or Replace condition, whereas $D_{i,t} = 0$ pertains to Keep Control or Pure Control. We include cohort $\times$ period fixed effects, as opposed to just period fixed effects, to ensure that our estimation is robust to the staggered roll-out of the treatment—see Cengiz et al. 2019 for an application of this approach. User's cohort C_i is determined by their intervention start date.

We exploit the opportunity presented by having a large number of daily time periods in the study (Cameron and Miller 2015), and use Driscoll and Kraay 1998 standard errors, which account for both serial and cross-sectional dependence, to increase power—Hoechle 2007 provides an analysis of their finite sample performance. Regarding treatment effects, we are primarily interested in two main comparisons. We intend to compare the Hide group to Pure Control and the Replace group to Keep Control. All regressions estimating treatment effects use inverse probability weights based on the probability of group assignment for individuals recruited on a given day. This accounts for the relative oversampling of the Replace condition (see Section 3.4).

Lastly, we estimate heterogeneity of treatment effects with respect to baseline characteristics, such as baseline active time on Facebook, using the following specification. Let M_i be an indicator equal to 1 if, for individual i, the variable is above the median:

$$Y_{i,t} = \alpha_i + \delta_{t,C_i} + \beta D_{i,t} + \phi\left(D_{it} M_i\right) + Errors_{i,t} + \varepsilon_{i,t}. \tag{3}$$

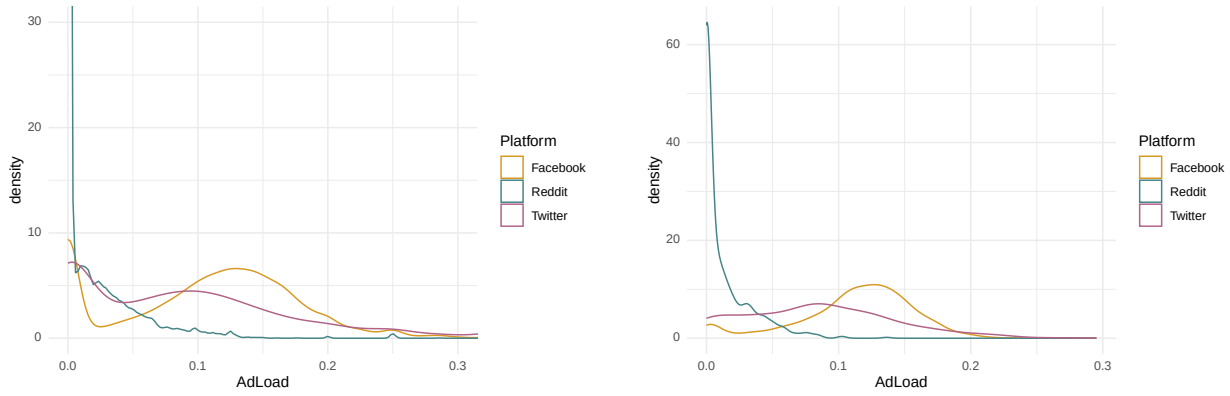
[15] We consider three types of possible errors: (i) faulty hiding—the hiding intervention was not operational, (ii) DOM update—Facebook made a sudden change to the DOM structure of the page, which resulted in issues such as ad names displaying incorrectly, (iii) faulty replacement—eligible ads in the Replace group were incorrectly replaced. Depending on whether the error affected all treatment groups, occurred for different individuals on different days, or occurred during both the baseline and intervention periods, some variables in $Errors_{i,t}$ may be absorbed by the intervention start $\times$ period fixed effects.

4 Results

4.1 Ad Loads Heterogeneity

Figure 2(a) illustrates the substantial heterogeneity in daily advertising loads experienced by users on Facebook, Reddit, and Twitter during the baseline period. In contrast, Figure 2(b) presents the distribution of individual ad loads, abstracting away from temporal variability and focusing instead on person-level averages.

In this individual-level distribution, Facebook's ad load is centered around a mean ad load of 0.128, with a standard deviation of 0.111, and displays thick tails, reflecting significant dispersion in users' exposure to ads. Twitter's distribution exhibits even greater dispersion, with observations less tightly clustered around its mean of 0.121. Reddit stands out from the other platforms with a notably lower average ad load of 0.013.



(a) Experienced Ad Load

(b) Individual Experienced Ad Load

FIGURE 2: DISTRIBUTION OF AD LOAD BY PLATFORM

Note: **Panel A** presents the distributions of the experienced advertising load per person-day on Facebook, Reddit, and Twitter, based on the baseline period data. **Panel B** presents the equivalent distributions per person. The ad load is defined as the number of ads as a share of total posts and comments shown to the user on each platform. The figure is based on data from the baseline period. For the Pure Control group, we also include the intervention period data. The x-axis for Panel A is truncated 0.3 and the y-axis is truncated at 30. Panel B relies on the subsample of users who viewed at least 50 posts and comments throughout the relevant period.

One possible explanation for the wide distribution of individual ad loads is heterogeneity in user consumption patterns—specifically, variation in the time users spend in comment sections versus the main feed. For instance, users who engage more with comments may encounter fewer ads, which are typically concentrated in the feed. To investigate this, Figure 3 presents an alternative ad load measure: the number of ads as a share of *feed-only* content. This contrasts with our baseline definition, which includes both feed and comment section

content. Despite focusing solely on feed content, we still observe substantial heterogeneity in individual ad loads on all three platforms. This suggests that differences in engagement with comment sections are unlikely to be the primary driver of the dispersion observed in Figure 2. Instead, the evidence points to genuine, *time-invariant* heterogeneity in ad exposure across users—likely shaped by the platform's underlying content delivery algorithms.

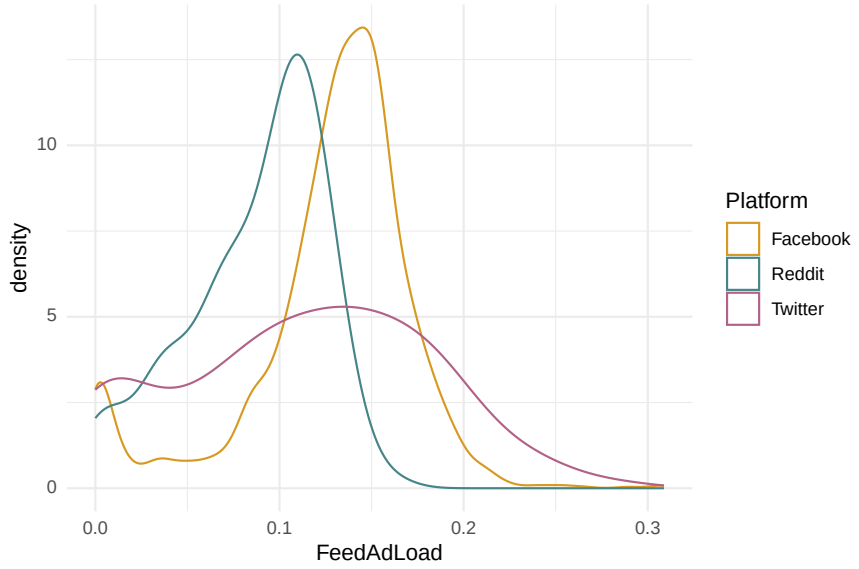


FIGURE 3: DISTRIBUTION OF INDIVIDUAL FEED AD LOAD BY PLATFORM

Note: The figure presents the distributions of the experienced feed advertising load per person on Facebook, Reddit and Twitter, based on the baseline period data. The feed ad load is defined as the number of ads as a share of total content in the feed shown to the user on the platform. The browser extension can directly identify feed elements on Reddit on Twitter. On Facebook, we consider an element to be in the feed unless it is a part of a sequence of at least four consecutive comments—comments in the feed do not occur in groups larger than three. The figure is based on data from the baseline period. For the Pure Control group, we also include the intervention period data. The figure relies on the subsample of users who viewed at least 50 posts and comments throughout the relevant period.

Variance Decomposition. To formalize our investigation into the sources of heterogeneity in ad loads, we perform a variance decomposition analysis for three social media platforms. Figure A4 in the appendix presents results from a two-way fixed effects specification, where daily ad load $a_{i,t}$ is decomposed into individual fixed effects α_i and time fixed effects β_t. We find that individual fixed effects account for 66.1% of the variance in daily ad loads on Facebook, 66.8% on Twitter, and 59.0% on Reddit. These results reinforce our earlier conclusion that a substantial share of the variation in ad exposure is driven by individual-level differences. Such heterogeneity is consistent with algorithmic personalization and potentially reflective of price discrimination strategies.

Ad Load Predictors. We also analyze which variables contribute most to predicting advertising loads. Figure 4 reports general dominance for a broad set of covariates, i.e., the average incremental R^2 a variable contributes when considering all possible subset models. We include demographic characteristics such as race, education, income, and age, as well as information about IT preferences (operating system, use of social media on desktop).

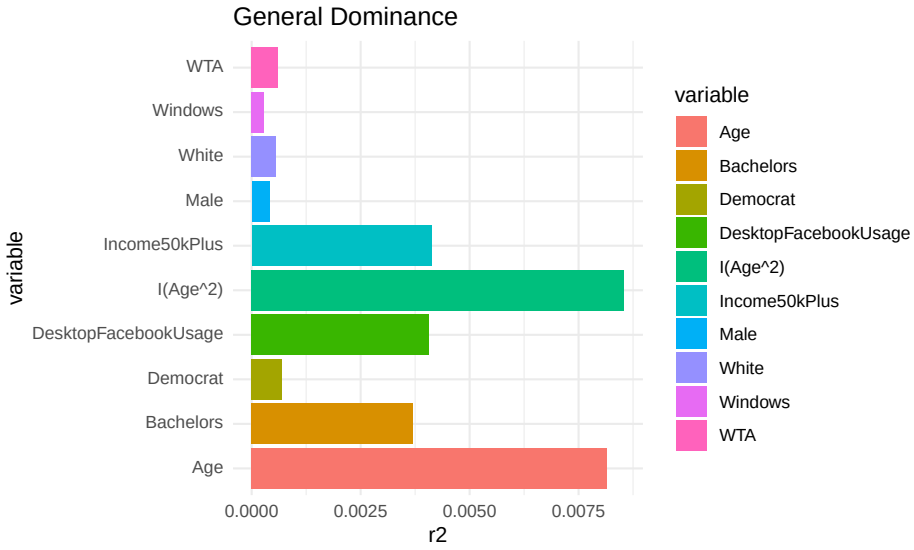


FIGURE 4: GENERAL DOMINANCE OF AD LOAD PREDICTORS ON FACEBOOK

Note: The figure reports general dominance for a broad set of covariates predicting advertising loads on Facebook, i.e., the average incremental R^2 a variable contributes when considering all possible subset models. The included variables are as follows: (1) age in years, (2) age squared, (3) an indicator for having at least a bachelor's degree, (4) an indicator for being a Democrat, (5) desktop share of Facebook usage (from the baseline survey), (6) an indicator for having household income above $50,000, (7) an indicator for being male, (8) an indicator for identifying as white/Caucasian, (9) an indicator for using Windows operating system, (10) willingness to accept to deactivate Facebook for four weeks during the baseline survey (WTA).

From Figure 4, we infer that age is the strongest predictor—*age* and age^2 have the first- and second-highest general dominance. This suggests possible price discrimination by age. Other important predictors include income, the share of Facebook usage on desktop devices, and education (completion of a bachelor's degree).

4.2 Impact Evaluation

4.2.1 Intervention Strength

Hiding Intervention and Ad Load. Figure 5 shows the average ad load experienced by Facebook users over the duration of the study, disaggregated by treatment condition. In the pre-intervention period, the average ad load was 0.128, with no meaningful differ-

ences across treatment groups. Immediately following the intervention, ad load in the Hide condition dropped sharply and remained stable near the post-intervention mean of 0.018, representing an 86% reduction relative to baseline. Ad loads in the Pure Control condition stayed close to their baseline levels, averaging 0.137 in the intervention period, indicating that the intervention affected only the intended group.

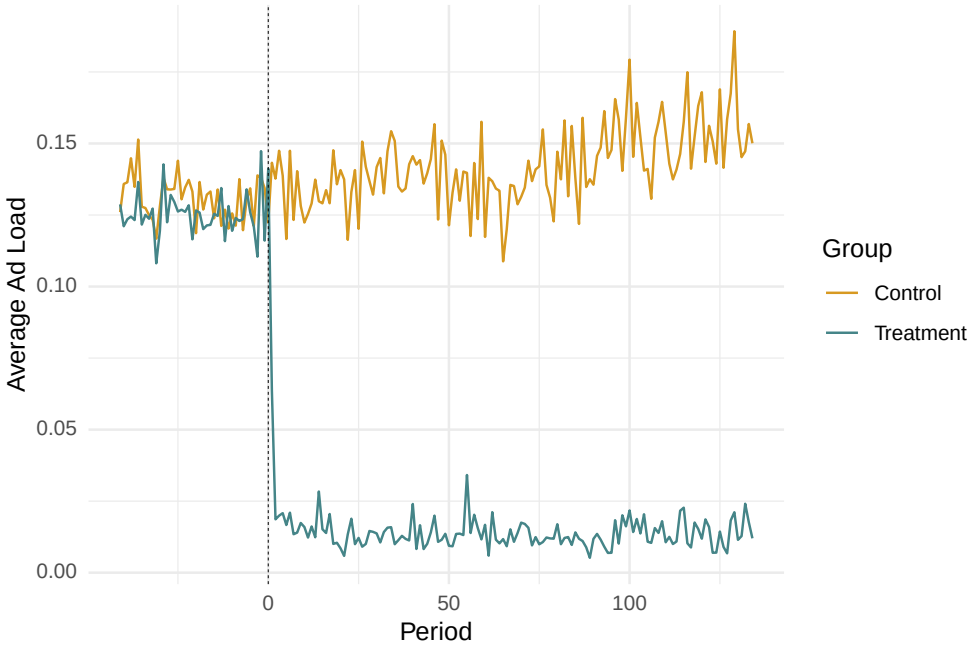


FIGURE 5: AVERAGE ADVERTISING LOADS ON FACEBOOK

Note: The figure shows the average daily advertising load on Facebook for the Hide condition and the Pure Control condition. Advertising load is defined as the number of ads shown as a share of all Facebook posts and comments viewed by the user. The x-axis represents study periods, measured in days relative to each individual's intervention start date. The vertical dashed line marks the final day of the pre-intervention period (period 0).

These results confirm that our hiding intervention successfully introduced substantial exogenous variation in Facebook ad exposure. Accounting for the average desktop share of Facebook usage—reported as 50.9% in the intake survey—we estimate an overall ad load reduction of 44% across all devices.[16]

Table B3 in the appendix provides the associated regression analysis. The difference-in-differences estimates show that the hiding intervention reduced the average ad load *experienced* by users in the Hide condition by 11.3 pp or 1 SD (column 1, panel A) relative to Pure

[16]This calculation assumes that the ad load experienced via Facebook's mobile app remained at the baseline mean, as measured by the browser extension. This assumption is plausible in light of our results that Facebook did not react to the intervention by targeting users with different ad loads.

Control. At the same time, the intervention had no impact on the average ad load *supplied* by the platform (column 3, panel A), i.e., ad load before any filtering was applied, with the treatment effect between the Hide and Pure Control conditions of only 0.3 pp. The latter result suggests the lack of response of Facebook's algorithm to the hiding intervention—users were still targeted with the same share of ads.

Replacement Intervention and Microtargeting. Figure 6 shows average ad distance, our measure of how well ads matched user characteristics, over the duration of the study. While overall ad distance fluctuated during the baseline period, there were no discernible differences in trends between the Replace condition and Keep condition. Immediately after the intervention, ad distance in the Replace condition increased relative to the other groups and remained elevated for the rest of the study. We conclude that the replacement intervention successfully introduced exogenous variation in the degree of ad microtargeting.

Table B3 in the appendix formalizes this conclusion. The intervention increased average ad distance by 11% (column 2, panel B), a result significant at the 1% level. Unlike in the case of the hiding intervention, here we observe some indication of algorithmic response—Facebook changed the composition of ads supplied to users, which resulted in a 1.5% increase in average ad distance (column 4, panel B). While the effect size is small compared to the primary effect of the intervention, we conclude that Facebook responded by supplying *less* microtargeted ads. This may reflect the high weight the algorithm places on recent ad interactions—as users clicked on and reacted to different ad content, it internalized the change in preferences promptly. The presence of an algorithmic response here, but not in the case of the hiding intervention, can be reconciled by assuming that the algorithm considers the relative stability of ad load elasticity (users are inelastic) compared to the more changeable preferences regarding the types of ads users are interested in.

4.2.2 Own-Platform Engagement: Hiding Intervention

Ad Engagement. Figure 7(a) presents the short-run treatment effects of the hiding intervention on engagement (see Panel A of Table B4 in the appendix for full regression tables). The hiding intervention strongly reduced ad engagement. According to the difference-in-differences estimates, ad impressions fell by 95% or 0.28 standard deviations (SD) and ad clicks decreased by 58% or 0.11 SD. The overall index of ad engagement fell by 0.19 SD, a result significant at the 1% level. Figure 7(b) presents the long-run treatment effects, based on

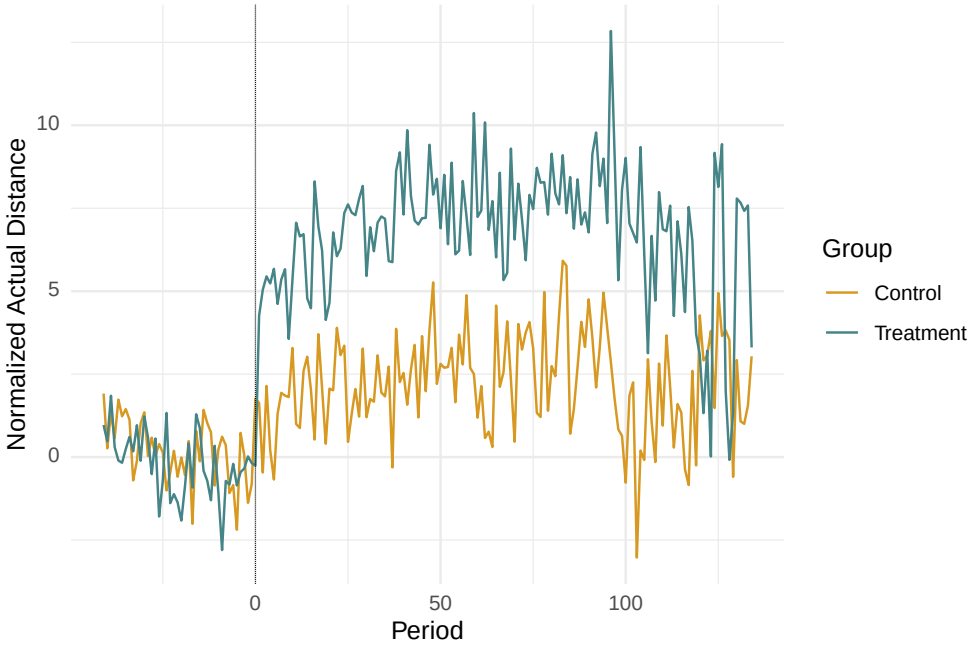


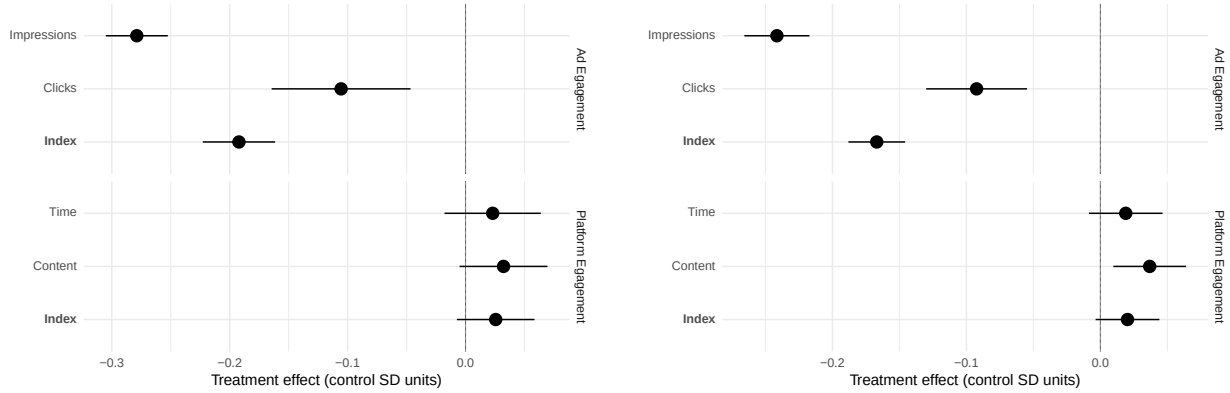
FIGURE 6: NORMALIZED AD DISTANCE ON FACEBOOK

Note: The figure displays the average daily ad distance on Facebook, disaggregated by treatment condition (Replace vs. Keep conditions). Ad distance is defined as the Euclidean distance between a user's characteristics and the average demographic profile targeted by that advertiser, based on baseline data. The values on the y-axis are normalized using the mean and standard deviation from the baseline period. The x-axis represents study periods, measured in days relative to each individual's intervention start date. The vertical dashed line marks the final day of the pre-intervention period (period 0).

19 weeks of intervention (see Panel A of Table B5 in the appendix for full regression tables). The long-term effects of the hiding intervention on ad engagement are both directionally and quantitatively consistent with the short-run effects.

Short-Term Effects on Platform Engagement. While ad engagement declined, overall engagement with the platform increased as a result of the hiding intervention. The short-term effects estimated using our pre-registered specification in levels are modest (see Figure 7(a) or Panel A of Table B4). The intervention increased content consumption by 5.6 posts and comments per day, an effect equivalent to 9% or 0.03 SD. This result is significant at the 10% level. We also observe a directionally consistent, though statistically insignificant, increase in the active time spent on Facebook—by 0.65 minutes per day.

The precision of our estimates is affected by outliers and data skewness (see Section 4.2.6 for more discussion). To address that, we estimate two additional specifications: (i) applying the inverse hyperbolic sine (asinh) transformation to the data and (ii) maintaining the spec-



(a) Short-Term Effects

(b) Long-Term Effects

FIGURE 7: OWN-PLATFORM ENGAGEMENT EFFECTS OF HIDING INTERVENTION

Note: The figure displays estimated treatment effects of the hiding intervention on Facebook using Equation 2 and our main experimental sample. The dependent variables are: (i) number of ad impressions, (ii) number of ad clicks, (iii) an ad engagement index based on (i) and (ii), (iv) active time spent on the platform, (v) number of posts and comments shown to the user, and (vi) a platform engagement index based on (iv) and (v). **Panel A** shows results based on the six-week intervention period, whereas **Panel B** shows results based on a longer intervention period (134 day-periods or 19.1 weeks). The unit of observation is the individual-day, measured relative to the intervention date. We include 95% confidence intervals based on Driscoll-Kraay standard errors. All regressions were estimated using inverse probability weights reflecting the probability of group assignment for individuals recruited on a given day.

ification in levels but trimming the top one percentile of observations for each outcome. The results are presented in Tables B6 and B8 in the appendix, respectively. Both specifications indicate significant positive effects of the intervention on the active time spent on Facebook and content consumption. The effect sizes on the index of platform engagement range from 0.06 SD in the trimming specification to 0.07 SD in the asinh specification, both significant at the 1% level.

Long-Term Effects on Platform Engagement. The long-term results based on the pre-registered specification are presented in Figure 7(b) and Panel A of Table B5. The hiding intervention increased content consumption by 0.04 SD, a result significant at the 1% level. The observed increase in active time on the platform, equivalent to 0.57 minutes per day, is statistically insignificant. The overall platform engagement index rose by 0.02 SD, a result significant at the 10% level.

Similar to our short-term analysis, we estimate the additional specifications (i)-(ii) to increase the precision our estimates in light of data skewness and outliers. The results are presented in Tables B7 and B9 in the appendix. Both specifications reveal positive effects of the hiding intervention on active time spent on Facebook in the range of 1.6 minutes per

day—the results are significant at the 1% level. The point estimates are higher than in the baseline level specification, showcasing the importance of addressing outliers. The overall impact of the intervention on platform engagement is equal to 0.05 SD for both the trimming and asinh specifications.

Taking both the short-term and long-term analyses together, we conclude that (i) the hiding intervention increases platform engagement across a range of metrics, (ii) the effect sizes are modest to moderate depending on the specification, and (iii) the effects emerge early and remain stable over time.

Elasticities. It is helpful to express the engagement effects in the form of an elasticity. Specifically, we define elasticity as the ratio of the percentage change in active time spent on Facebook to the percentage change in advertising load. To compute these changes, we rely on difference-in-differences estimates of the effects of the hiding intervention on ad load (Table B3) and active time (Table B4). This yields a short-run elasticity estimate of:

$$\text{Elasticity} = \frac{\%\Delta\text{Active Time}}{\%\Delta\text{Ad Load}} = -0.07.$$

We can repeat this exercise using the alternative specifications for platform engagement effects. For instance, employing the asinh specification (Table B6), we obtain an elasticity of –0.16.[17] Overall, we conclude that the elasticity of user engagement with respect to advertising load is below 0.2 in absolute value, indicating that consumer demand is highly price inelastic. Our estimates are consistent with Brynjolfsson et al. 2024, who reach a similar conclusion.

4.2.3 Own-Platform Engagement: Replacement Intervention

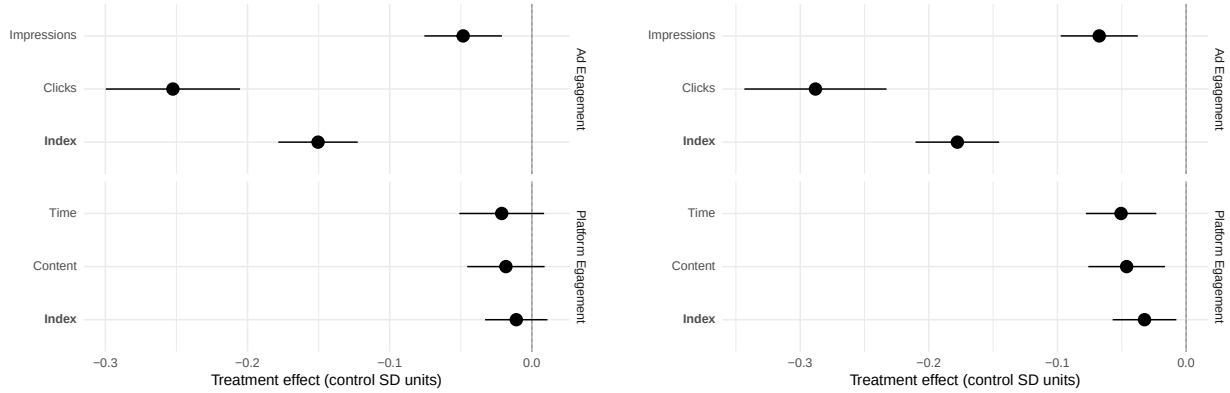
Ad Engagement. Figure 8(a) presents the short-term treatment effects of the replacement intervention on engagement relative to Keep Control (see Panel B of Table B4 for full regression results). Similar to the hiding intervention, ad replacement reduces engagement

[17] To express treatment effects estimated on inverse hyperbolic sine (asinh) transformed outcomes as percentage changes in the original (level) units, we use the following formula:

$$\%\Delta Y = \frac{\sinh(m+\beta) - \sinh(m)}{\sinh(m)} \times 100,$$

where β is the estimated treatment effect and m is the mean of the asinh-transformed outcome variable. This expression maps the estimated effect back to the original scale and expresses the change as a percentage relative to the mean.

with ads. The induced weakening of microtargeting led to a 0.05 SD decrease in ad impressions and a 0.25 SD decline in ad clicks. The combined ad engagement index fell by 0.15 SD, a result significant at the 1% level. The long-term effects (see Figure 8(b) or Panel B of Table B5) are in line with the short-term results—the replacement intervention reduced the index of ad engagement by 0.18 SD. Overall, we conclude that microtargeting is a key driver of ad engagement, underscoring its value to the platform.



(a) Short-Term Effects

(b) Long-Term Effects

FIGURE 8: OWN-PLATFORM ENGAGEMENT EFFECTS OF REPLACEMENT INTERVENTION

Note: The figure displays estimated treatment effects of the replacement intervention on Facebook relative to Keep Control using Equation 2 and our main experimental sample. The dependent variables are: (i) number of ad impressions, (ii) number of ad clicks, (iii) an ad engagement index based on (i) and (ii), (iv) active time spent on the platform, (v) number of posts and comments shown to the user, and (vi) a platform engagement index based on (iv) and (v). **Panel A** shows results based on the six-week intervention period, whereas **Panel B** shows results based on a longer intervention period (134 day-periods or 19.1 weeks). The unit of observation is the individual-day, measured relative to the intervention date. We include 95% confidence intervals based on Driscoll-Kraay standard errors. All regressions were estimated using inverse probability weights reflecting the probability of group assignment for individuals recruited on a given day.

Platform Engagement. The replacement intervention negatively affected not only ad engagement but also overall platform engagement. Using our pre-registered specification, we detect a long-term decline in engagement (see Figure 8(b) and Panel B of Table B5). Weaker microtargeting reduced active time on the platform by 1.7 minutes per day (13% or 0.05 SD) and content consumption by 9.4 posts and comments per day (13% or 0.05 SD), both results are significant at the 1% level. Moreover, the platform engagement index decreased by 0.03 SD, significant at the 5% level. These findings are robust to alternative specifications addressing outliers and data skewness (see Panel B of Tables B7 and B9 in the appendix). The estimated treatment effects on the platform engagement index are 0.04 SD under trimming and 0.06 SD under the asinh specification, both significant at the 1% level.

The short-term point estimates from our baseline specification are directionally consistent with the long-term results but statistically insignificant (see Figure 8(a) and Panel B of Table B4). This pattern holds for the trimming specification (see Panel B of Table B8 in the appendix) with the asinh specification yielding a marginally significant negative effect (Panel B of Table B6). These findings suggest that users may update their beliefs about the platform experience gradually, underscoring the need for long-duration interventions to inform policy. Taken together, the effects on ad and platform engagement imply that both consumers and the platform may benefit from ad microtargeting.[18]

4.2.4 Substitution Effects

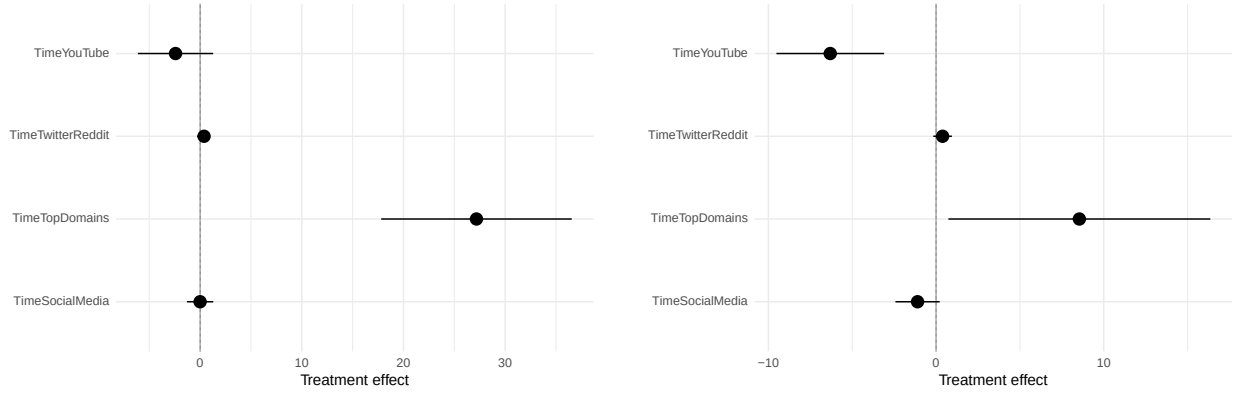
Figure 9(a) presents the effects of the hiding intervention on engagement with non-treated platforms. Overall, we find little evidence of substitution: the intervention does not significantly affect active time spent on Twitter and Reddit, YouTube, or the time spent on the combined set of other pre-registered social media platforms. At the same time, the intervention leads to a considerable increase in total time spent on the user's top 10 most-visited domains. This pattern may reflect a "linking effect," where increased time on Facebook leads to greater engagement with websites frequently linked in the user's feed.

Figure 9(b) shows the equivalent effects of the replacement intervention. We find limited evidence that varying the degree of ad microtargeting on one platform affects engagement with other platforms. The treatment marginally increased the active time spent on Twitter and Reddit, which are most similar to Facebook, the platform where we intervened. Additionally, we observe a positive effect on the time spent on user's top 10 most-visited domains. However, the time spent on YouTube went down and the effect on the time spent on the set of non-treated social media platforms defined in Beknazar-Yuzbashev et al. 2025 is insignificant. Taken together, there is no clear substitution pattern caused by the replacement intervention.

4.2.5 Platform Valuation

Figure 10 presents the results of both interventions on users' valuation of the platform. We find that neither the hiding intervention nor the replacement intervention significantly affects

[18] While a simple revealed preference argument would suggest that microtargeting improves user welfare, our analysis does not account for potential negative externalities such as impulse buying or post-purchase regret that may arise from ad microtargeting.



(a) Hiding Intervention

(b) Replacement Intervention

FIGURE 9: IMPACTS ON ENGAGEMENT WITH NON-TREATED PLATFORMS

Note: **Panel A** reports estimated treatment effects of the hiding intervention on Facebook, relative to the Pure Control group, based on Equation 2 and the main experimental sample. **Panel B** presents treatment effects of the replacement intervention, relative to the Keep Control group. The dependent variables are: (i) time spent on YouTube; (ii) combined active time on Twitter and Reddit; (iii) time spent on the top 10 domains as determined from baseline data; and (iv) time spent on a set of non-treated social media platforms defined in Beknazar-Yuzbashev et al. 2025. All outcome variables are inverse hyperbolic sine (asinh) transformed. The unit of observation is the individual-day, indexed relative to the intervention start date. Regressions include 95% confidence intervals based on Driscoll-Kraay standard errors. All estimates are weighted using inverse probability weights reflecting group assignment probabilities conditional on recruitment day.

the willingness to accept (WTA) deactivating social media for four weeks, which serves as our proxy for the platform's value. The null result for the hiding intervention is consistent with Brynjolfsson et al. 2024, who find no effect of removing online Facebook ads on users' platform valuation.[19] We extend this result by showing that reducing ad microtargeting (replacement intervention) also has no effect on platform valuation.

4.2.6 Heterogeneity of Engagement Effects

Heterogeneity by Baseline Engagement. Figure 11 presents the treatment effects of the hiding intervention, estimated separately for individuals whose baseline active time on Facebook was above or below the median. Figure 11(a) reports the short-term effects, while Figure 11(b) shows the long-term effects. As expected, the negative effects of the intervention on ad engagement, including ad clicks, are primarily driven by the most active users—for whom the ad engagement index declines by 0.35–0.40 SD. The effects among below-median users are smaller but remain statistically detectable.

The effects on platform engagement present a more nuanced picture. First, the confi-

[19]We note our estimates of the effects of the interventions on WTA are less precise than Brynjolfsson et al. 2024 due to a smaller sample size.

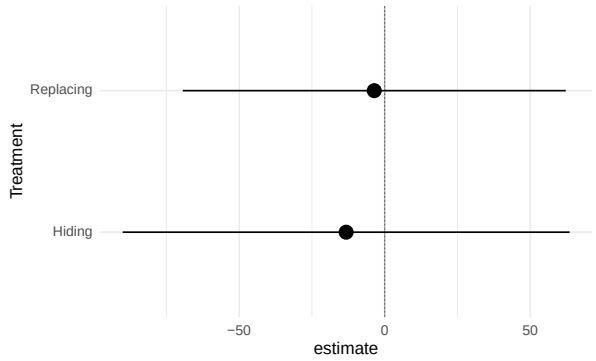


FIGURE 10: EFFECTS OF INTERVENTIONS OF PLATFORM VALUATION

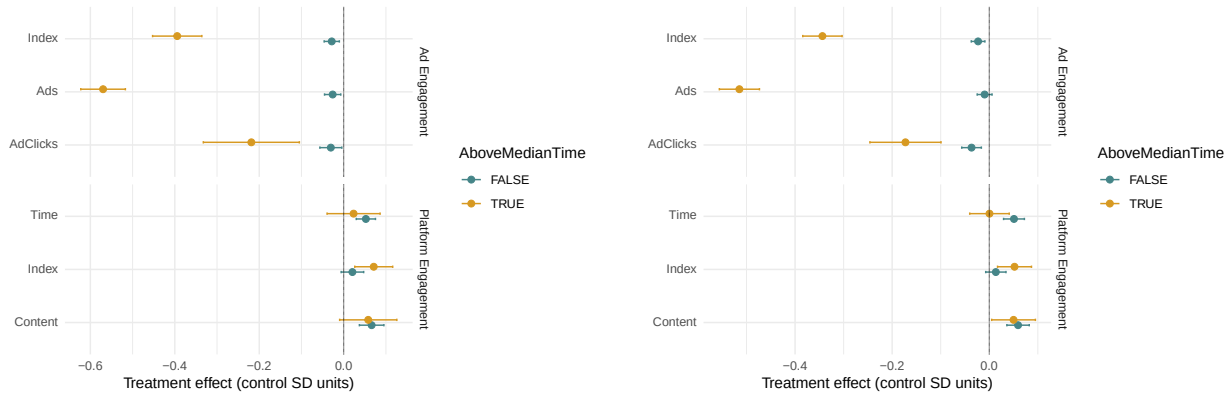
Note: The figure displays the estimated treatment effects of the hiding and replacement interventions on individuals' willingness to accept (WTA) the deactivation of social media for four weeks. WTA is measured in both the baseline and endline surveys. The estimates are based on a regression of WTA on the treatment group indicator, the endline survey indicator, and their interactions, with individual fixed effects. The unit of observation is the individual, and 95% confidence intervals are provided, based on standard errors clustered at the individual level. All regressions were estimated using inverse probability weights reflecting the probability of group assignment for individuals recruited on a given day.

dence intervals for the above-median users are notably wide, with the estimates influenced by outliers (super-users). For this reason, in Sections 4.2.2 and 4.2.3, we complemented the baseline specification in levels with alternative specifications designed to address data skewness and the presence of outliers. Despite the noise, we find that the positive effects on platform engagement are driven by the most active users—the platform engagement index increases significantly for above-median individuals (by approximately 0.05 SD), while no significant effect is observed for below-median individuals.

Figure 12 presents the corresponding analysis for the replacement intervention. The negative treatment effects on ad engagement are strongly driven by individuals with above-median baseline engagement. For them, the induced weaker microtargeting reduces ad clicks by more than 0.5 SD. Consistent with the results presented in Section 4.2.3, we do not detect significant short-run effects of the replacement intervention for users with high activity at baseline. However, we observe clear negative effects in the long-term, with active time and content consumption on Facebook falling by almost 0.1 SD for the above-median users.

4.3 Robustness and Potential Concerns

This section discusses potential concerns and robustness checks. Differential attrition poses a major threat to identification in studies like ours. Section 4.3.1 provides evidence suggesting that our results are unlikely to be driven by attrition. Additionally, in Section 4.3.2, we



(a) Short-Term Effects

(b) Long-Term Effects

FIGURE 11: HETEROGENEITY OF ENGAGEMENT EFFECTS: HIDING

Note: The figure displays estimated treatment effects of the replacement intervention on Facebook relative to Pure Control separately for users whose baseline active time on Facebook was above and below the median. The estimation is based on Equation 3, with $\beta + \phi$ representing the above-median coefficient and β representing the below-median coefficient. The regression relies on the main experimental sample. The dependent variables are: (i) number of ad impressions, (ii) number of ad clicks, (iii) an ad engagement index based on (i) and (ii), (iv) active time spent on the platform, (v) number of posts and comments shown to the user, and (vi) a platform engagement index based on (iv) and (v). **Panel A** shows results based on the six-week intervention period, whereas **Panel B** shows results based on a longer intervention period (134 day-periods or 19.1 weeks). The unit of observation is the individual-day, measured relative to the intervention date. We include 95% confidence intervals based on Driscoll-Kraay standard errors. All regressions were estimated using inverse probability weights reflecting the probability of group assignment for individuals recruited on a given day.

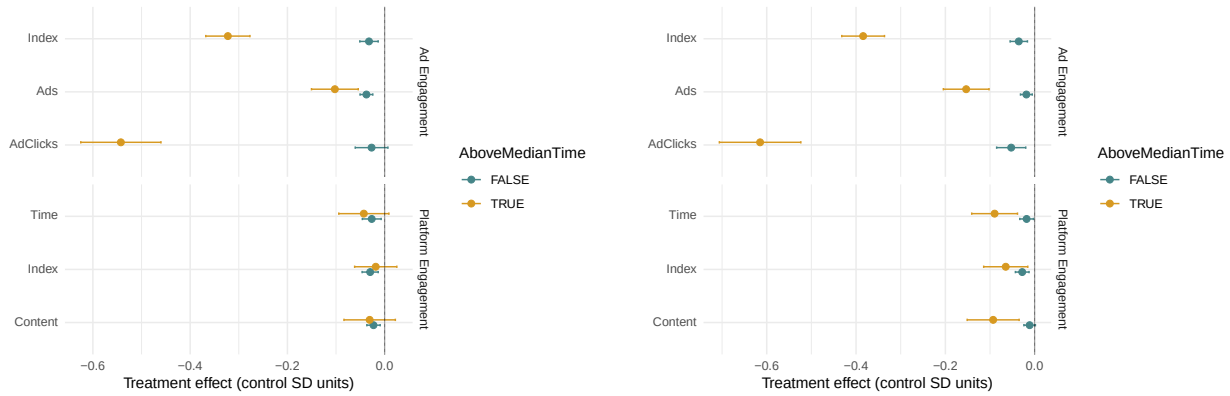
perform a robustness analysis by considering alternative specifications.

4.3.1 Attrition

Figure A5 in the appendix shows the survival rate in the experiment by treatment group over time. We use a conservative approach to measuring attrition, considering individual i as having survived until period t if they were active on the day corresponding to period t or later. It is likely that some individuals remained in the study after their last activity, meaning the reported survival rates represent lower bounds, and the attrition rates—upper bounds.

The survival rates at the end of the six-week intervention period, relative to the beginning of the baseline period, exceed 75% in all treatment groups: 76.8% in Pure Control, 76.9% in Keep Control, 76.3% in Hide, and 76.1% in Replace. Furthermore, attrition rates by treatment did not vary over time, as shown in Figure A5. Thus, we conclude that our results are unlikely to be driven by differential attrition.

Table B10 in the appendix presents the associated regression analysis, supporting our claim. First, regressing survival at the end of the intervention period on treatment indica-



(a) Short-Term Effects

(b) Long-Term Effects

FIGURE 12: HETEROGENEITY OF ENGAGEMENT EFFECTS: REPLACEMENT

Note: The figure displays estimated treatment effects of the hiding intervention on Facebook relative to Keep Control separately for users whose baseline active time on Facebook was above and below the median. The estimation is based on Equation 3, with $\beta + \phi$ representing the above-median coefficient and β representing the below-median coefficient. The regression relies on the main experimental sample. The dependent variables are: (i) number of ad impressions, (ii) number of ad clicks, (iii) an ad engagement index based on (i) and (ii), (iv) active time spent on the platform, (v) number of posts and comments shown to the user, and (vi) a platform engagement index based on (iv) and (v). **Panel A** shows results based on the six-week intervention period, whereas **Panel B** shows results based on a longer intervention period (134 day-periods or 19.1 weeks). The unit of observation is the individual-day, measured relative to the intervention date. We include 95% confidence intervals based on Driscoll-Kraay standard errors. All regressions were estimated using inverse probability weights reflecting the probability of group assignment for individuals recruited on a given day.

tors shows no differential attrition in pairwise comparisons between Hide and Pure Control, Replace and Keep Control, and Keep Control and Pure Control (columns 1-3). Second, we extend the regressions by including survival propensity, based on individual characteristics, and its interaction with the treatment dummy (columns 4-6). For all three pairwise comparisons, we find no evidence that the treatment group moderates the association between survival propensity, based on individual characteristics, and actual survival. Finally, we obtain identical conclusions when we repeat the analysis from columns 1-6, but with the last period of recorded activity, which we use as a proxy for the last period in the study, as the outcome variable (columns 7-12). The dependent variable is capped at 42, which indicates the last period of the six-week intervention.[20] In this case, we find that treatment does not impact dropout time, regardless of whether survival propensity is included.

Lastly, since we report long-term outcomes, we also consider long-term attrition. Figure A5 extends beyond the end of the six-week intervention period (the dashed vertical line) and shows that the dropout rate did not vary by treatment until the end of the study. Moreover,

[20]Note that the intervention starts in period 1, with baseline periods represented by negative numbers and zero.

Table B11 in the appendix presents regressions of the last period in the study on treatment indicators, with the last period now being uncapped. Once again, we find no evidence of differential attrition for any of the pairwise comparisons. The result remains robust to the inclusion of the survival propensity, based on user characteristics, and its interaction with the treatment indicator is insignificant for all pairwise comparisons. We conclude that, even with an extended observation window, there is no evidence of differential attrition by treatment condition.

4.3.2 Alternative Specifications

Skewness and Outliers. Throughout Section 4.2, we reported the robustness of our treatment effects to two alternative specifications designed to address data skewness and the presence of outliers. Tables B6 and B7 present results based on the inverse hyperbolic sine (asinh) transformation. Tables B8 and B9 report estimates from the specification in levels with the top 1% of observations trimmed for each outcome. For both specifications, we find improved precision and consistent evidence that the hiding intervention increased platform engagement in both the short and long term, while the replacement intervention reduced platform engagement in the long term.

Alternative Standard Errors. Table B12 in the appendix presents the short-term engagement results with standard errors clustered at the individual level, as opposed to Driscoll-Kraay standard errors. The effects of the hiding intervention on the index of ad engagement remain significant at the 1% level, while the corresponding effect of the replacement intervention remains significant at the 10% level. The platform engagement results are no longer statistically significant. Table B13 in the appendix provides the corresponding analysis for the long-term outcomes, yielding identical insights. Clustering at the individual level is conservative and does not account for the rich within-variation provided by a large number of time periods per person, which we exploit in our baseline specification with Driscoll-Kraay standard errors. The lack of robustness to conservative clustering is consistent with our interpretation that user engagement is very *inelastic* to ad loads and that ad microtargeting has positive but small effects on engagement.

Tables B14 and B15 in the appendix provide the corresponding analysis for regressions based on the asinh transformation. Thanks to the reduced noise, the positive effects of the hiding intervention on platform engagement and the long-term negative effects of the

replacement intervention on platform engagement are robust to clustering standard errors at the individual level.

5 Counterfactual Monetization Schemes for Social Media

5.1 Monetization Schemes

This section uses experimental estimates to measure welfare effects of moving away from today's microtargeted, ad-funded, and individualized ad-load regime. To shed light on this, we construct a set of counterfactual exercises under different ad load and targeting scenarios, which help understand why the microtargeting ad-based business model is so prevalent and the drivers of the profitability of this business model.

We compare the current status quo (implemented by Facebook in the US) to four alternatives:

1. No microtargeting and fixed ad loads (i.e., set to the current level);
2. No microtargeting and optimal ad loads;
3. Paywall (no ads at all, but access gated behind a subscription fee);
4. No ads (and no paywall).

5.2 Construction of the Counterfactuals

Let $(a_i)_i$ be the vector of ad loads for each individual, $(q_i)_i$ be the vector of ad match qualities for each individual, p denote the subscription charged for using the platform, and $\pi((a_i)_i, (q_i)_i, p)$ denote the profits that the platform gets under given ad load, micro targeting, and subscription policies.

We compare the status quo profits under personalized ad loads, microtargeting, and no subscription,

$$\pi((a_i)_i^*, (q_i)_i^*, 0)$$

to the profits under a pure subscription-based business model;

$$\pi(0, \overline{q}, p^*)$$

and decompose the difference

$$\pi((a_i)_i^*, (q_i)_i^*, 0) - \pi(0, \overline{q}, p^*)$$

as the addition of:

1. $\pi((a_i)_i^*, (q_i)_i^*, 0) - \pi((a_i)_i^*, \overline{q}, 0)$ – the additional profits earned due to microtargeting, all else equal;
2. $\pi((a_i)_i^*, \overline{q}, 0) - \pi(a^*, \overline{q}, 0)$ – the additional profits earned due to ad-load discrimination; and
3. $\pi(a^*, \overline{q}, 0) - \pi(0, \overline{q}, p^*)$ – the additional profits earned due to the change in business model from ads to subscriptions.

To generate the first component, we simply report estimates of the "replace" vs. "keep" treatment arms, which remove ad targeting while keeping ad load constant. To construct profits, we use predicted ad impressions times the cost per impression that we obtained from the Facebook advertising manager data.

To generate the second component, we turn off targeting and then set a single platform-wide ad load to (approximately) the profit-maximizing level, and propagate its effect to outcomes. Concretely, in the "replace" arm and post period we estimate user-level CATEs of ad load on outcomes via an instrumental forest and combine them with observed costs per impression. We approximate set optimal ad load to equal sample mean post-period ad load in the replacement arm as a conservative calibration. We then predict the incremental change for each outcome from moving each user from their post-period average ad load to this common level using their estimated slope.

To generate the third component, we eliminate ads and choose a flat subscription price to maximize fee revenue from the empirical WTA distribution; subscribers generate fee revenue and keep using the platform, non-subscribers churn. Crucially, we allow individuals' WTA to depend on the aggregate market share of the platform, to allow for network effects.[21] Selection into paying is implemented by comparing a scaled WTA measure to the fee, setting post-period time and profits to zero for non-subscribers, and subtracting the fee from subscribers' WTA (consumer surplus). We multiply individuals' WTA times 0.2 to convert it to a WTP, using the average gap reported by Sunstein 2020.

[21] We use data from Bursztyn et al. 2023 to compute the elasticity of WTA with respect to network size.

5.3 Simulation Results

Figure 13 reports, for each outcome and counterfactual regime, the mean treatment effect across the $B = 200$ bootstrap replications, together with percentile 95% confidence intervals. Three findings emerge clearly.

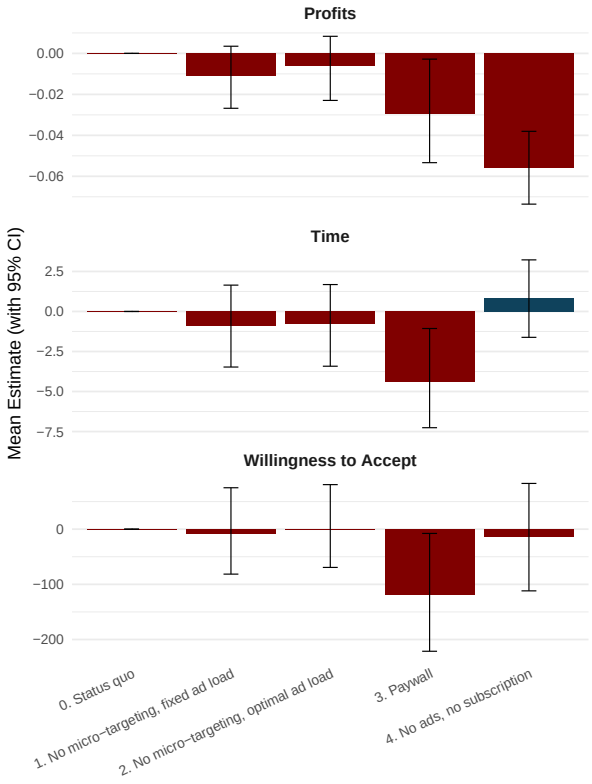


FIGURE 13: COUNTERFACTUAL MONETIZATION SCENARIOS

Note:

First, the status quo delivers the highest value along all three margins — profits, time, and WTA. This is not automatic: in principle, a platform could choose a policy that raises engagement at the expense of revenue, or vice versa. In our data, the observed policy $\pi((a_i)_i^*, (q_i)_i^*, 0)$ sits on the joint frontier: it is the most profitable regime, and it also keeps users on the platform the longest and preserves the highest valuation measure. This is consistent with the earlier impact estimates, which showed that users are only weakly elastic to the ad load actually implemented in the experiment.

Second, removing microtargeting is costly. The "no microtargeting, fixed ad load" counterfactual (1) produces profits, time, and WTA estimates that are slightly below the status

quo benchmark. This is not fixed by allowing the platform to adapt by picking a new optimal ad load in the "no microtargeting, optimal ad load" counterfactual (2). Economically, this says that the two margins we switched off here (showing the right ad and choosing the right ad quantity for a given user) are *second-order* relative to the ability to monetize attention with ads at all.

Third, the business-model switch is first-order. Once ads are eliminated and access is gated by a subscription ("paywall"), profits fall sharply and by much more than the loss from turning off targeting or ad-load discrimination. This happens despite choosing the subscription price to maximize revenue off the empirical WTA distribution and despite letting subscribers keep their usage paths. The reason is mechanical: at that price a non-trivial share of users does not select into paying, and the network-effect correction we apply further depresses participation. The fall in time mirrors this logic: fewer people remain active, so aggregate time drops.

The final counterfactual – no ads and no subscription – highlights the same point from the other direction. Here profits are hit the hardest of all regimes, because we have removed both the ad margin and the subscription margin. At the same time, time increases slightly and WTA falls only moderately. This mimics what the experimental evidence would lead us to expect: removing ads tanks profits with while making the service only marginally more pleasant to the users.

Overall, the counterfactuals therefore point to a simple conclusion: most of the profitability of the current model comes from the *presence of an ad-based monetization technology*, not from the *fine-tuning* of targeting or ad-load schedules. The latter margins help and are empirically detectable, but the large, robust, and negative movements appear only when we take the platform off ads.

6 Conclusion

As new privacy regulations are rolled out, social media will have to re-evaluate and potentially adjust their business models. In this paper, we evaluate the ad-funded business model of social media platforms under alternative data availability. Using a custom browser extension, we show that ad loads serve as prices on social media. Less personalized ads make the effective ad load "price" higher for consumers, suggesting that value from targeted advertising is greater than disutility from targeted ads due to intrinsic privacy tastes. Interestingly,

platforms price discriminate using ad loads: ad loads vary a lot across consumers, with the coefficients of variation close to one, and up to two-thirds of the daily variation in ad loads can be explained by variation between users. This suggests that tech regulations targeting the ability of platforms to personalize ads will have large distributional effects on consumers. Counterfactual simulations suggest that an ad-based business model gives both higher profits and higher consumer surplus than a subscription-based one, which helps rationalize the prevalence of ads on social media.

References

Acemoglu, Daron, Daniel Huttenlocher, Asuman Ozdaglar, and James Siderius (2024). *Online Business Models, Digital Ads, and User Welfare*. Tech. rep. National Bureau of Economic Research.

Acquisti, Alessandro, Curtis Taylor, and Liad Wagman (2016). "The economics of privacy". In: *Journal of economic Literature* 54.2, pp. 442–492.

Alcobendas, Miguel, Shunto Kobayashi, Ke Shi, and Matthew Shum (2023). "The impact of privacy protection on online advertising markets". In: *Available at SSRN 3782889*.

Anderson, Simon P and Stephen Coate (2005). "Market provision of broadcasting: A welfare analysis". In: *The Review of Economic Studies* 72.4, pp. 947–972.

Anderson, Simon P and André De Palma (2012). "Competition for attention in the information (overload) age". In: *The RAND Journal of Economics* 43.1, pp. 1–25.

Anderson, Simon P and Martin Peitz (2020). "Ad clutter, time use, and media diversity". In.

Aridor, Guy (2022). "Drivers of Digital Attention: Evidence from a Social Media Experiment". In: *Available at SSRN 4069567*.

Aridor, Guy, Yeon–Koo Che, and Tobias Salz (2023). "The Effect of Privacy Regulation on the Data Industry: Empirical Evidence from GDPR". In: *RAND Journal of Economics* 54.4, pp. 695–730. DOI: 10.1111/1756-2171.12455. URL: https://doi.org/10.1111/1756-2171.12455.

Aridor, Guy, Rafael Jiménez-Durán, Ro'ee Levy, and Lena Song (2024a). "The economics of social media". In: *Journal of Economic Literature* 62.4, pp. 1422–1474.

Aridor, Guy et al. (2024b). "Evaluating the impact of privacy regulation on e-commerce firms: Evidence from apple's app tracking transparency". In.

Beknazar-Yuzbashev, George, Rafael Jiménez-Durán, Jesse McCrosky, and Mateusz Stalinski (2025). "Toxic content and user engagement on social media: Evidence from a field experiment". In.

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

Beknazar-Yuzbashev, George and Mateusz Stalinski (2022). "Do social media ads matter for political behavior? A field experiment". In: *Journal of Public Economics* 214, p. 104735.

Brynjolfsson, Erik et al. (2024). *The Consumer Welfare Effects of Online Ads: Evidence from a 9-Year Experiment*. Tech. rep. National Bureau of Economic Research.

Bursztyn, Leonardo, Benjamin R Handel, Rafael Jimenez, and Christopher Roth (2023). *When product markets become collective traps: The case of social media*. Tech. rep. National Bureau of Economic Research.

Calvano, Emilio and Michele Polo (2021). "Market power, competition and innovation in digital markets: A survey". In: *Information Economics and Policy* 54, p. 100853.

Cameron, A Colin and Douglas L Miller (2015). "A practitioner's guide to cluster-robust inference". In: *Journal of human resources* 50.2, pp. 317–372.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer (2019). "The effect of minimum wages on low-wage jobs". In: *The Quarterly Journal of Economics* 134.3, pp. 1405–1454.

Deisenroth, Daniel et al. (2024). *Digital advertising and market structure: Implications for privacy regulation*. Tech. rep. National Bureau of Economic Research.

Driscoll, John C and Aart C Kraay (1998). "Consistent covariance matrix estimation with spatially dependent panel data". In: *Review of economics and statistics* 80.4, pp. 549–560.

Dubé, Jean-Pierre et al. (2025). "The intended and unintended consequences of privacy regulation for consumer marketing". In.

Farronato, Chiara, Andrey Fradkin, and Chris Karr (2024). *Webmunk: A new tool for studying online behavior and digital platforms*. Tech. rep. National Bureau of Economic Research.

Goldfarb, Avi and Verina F Que (2023). "The economics of digital privacy". In: *Annual Review of Economics* 15.1, pp. 267–286.

Goli, Ali, Jason Huang, David Reiley, and Nickolai M Riabov (2025a). "Measuring consumer sensitivity to audio advertising: a long-run field experiment on Pandora internet radio". In: *Quantitative Marketing and Economics*, pp. 1–31.

Goli, Ali, David H Reiley, and Hongkai Zhang (2025b). "Personalizing ad load to optimize subscription and ad revenues: Product strategies constructed from experiments on pandora". In: *Marketing Science* 44.2, pp. 327–352.

Hoechle, Daniel (2007). "Robust standard errors for panel regressions with cross-sectional dependence". In: *The stata journal* 7.3, pp. 281–312.

Johnson, Garrett A (2024). "4. Economic Research on Privacy Regulation: Lessons from the GDPR and Beyond". In: *The Economics of Privacy*. University of Chicago Press, pp. 97–126.

Katz, Justin and Hunt Allcott (2025). *Digital Media Mergers: Theory and Application to Facebook-Instagram*. Tech. rep. Working paper.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz (2007). "Experimental Analysis of Neighborhood Effects". In: *Econometrica* 75.1, pp. 83–119.

Levy, Ro'ee (2021). "Social media, news consumption, and polarization: Evidence from a field experiment". In: *American Economic Review* 111.3, pp. 831–70.

Lin, Tesary (2022). "Valuing intrinsic and instrumental preferences for privacy". In: *Marketing Science* 41.4, pp. 663–681.

Liu, Yi, Pinar Yildirim, and Z John Zhang (2022). "Implications of revenue models and technology for content moderation strategies". In: *Marketing Science* 41.4, pp. 831–847.

McKenzie, David (2012). "Beyond baseline and follow-up: The case for more T in experiments". In: *Journal of development Economics* 99.2, pp. 210–221.

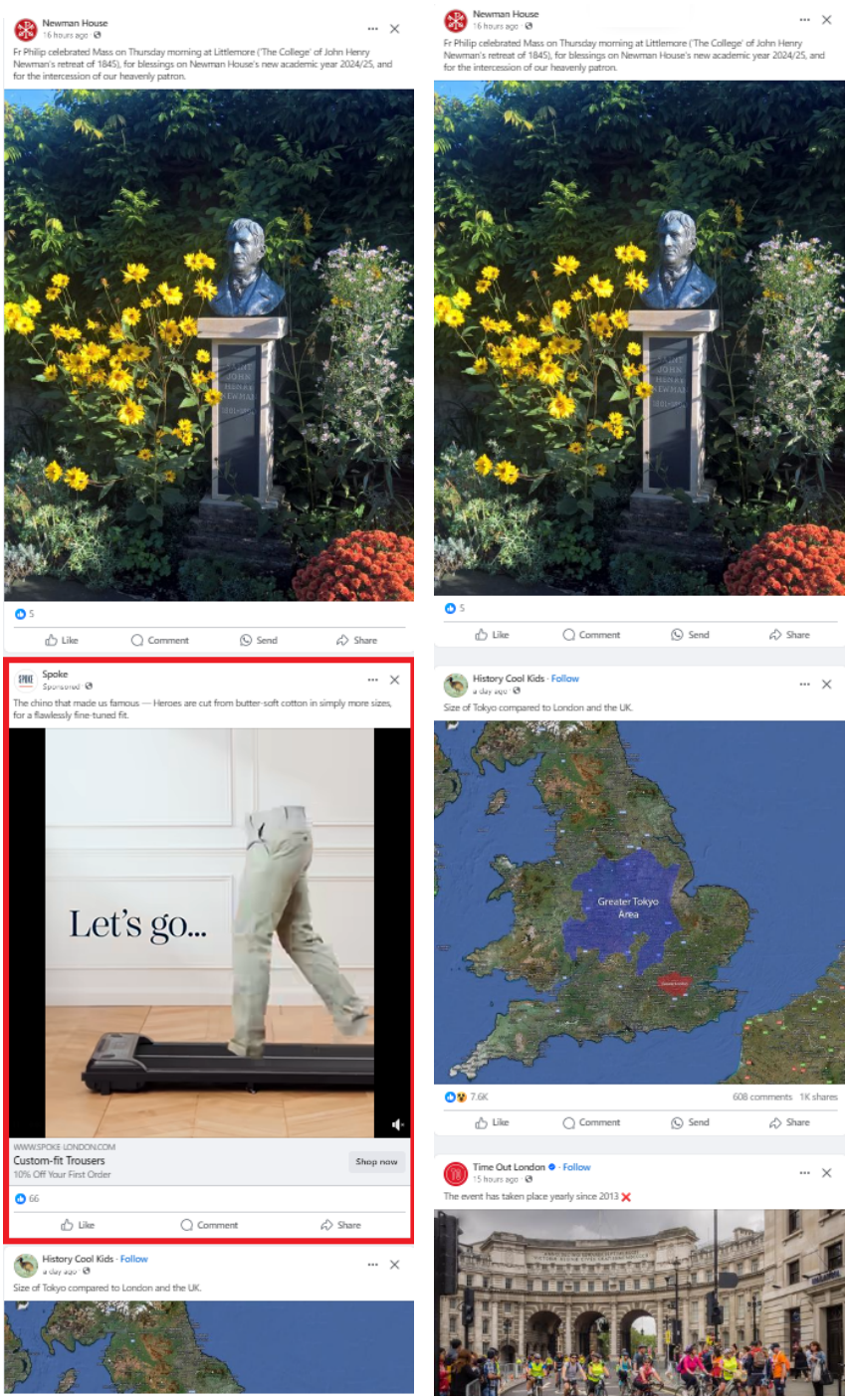
Moshary, Sarah (2021). "Sponsored search in equilibrium: evidence from two experiments". In: *Available at SSRN 3903602*.

Sunstein, Cass R (2020). "Valuing facebook". In: *Behavioural Public Policy* 4.3, pp. 370–381.

Weyl, E Glen (2010). "A price theory of multi-sided platforms". In: *American Economic Review* 100.4, pp. 1642–72.

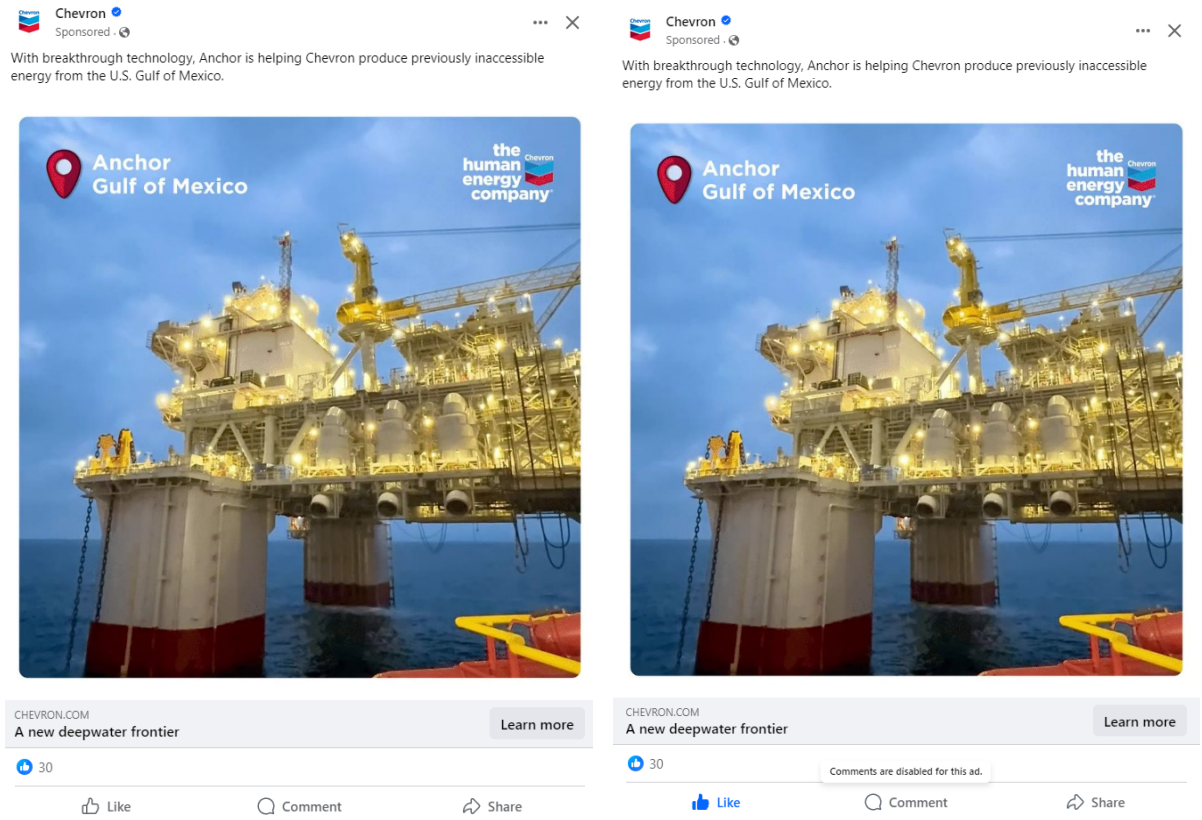
Wilbur, Kenneth C (2008). "A two-sided, empirical model of television advertising and viewing markets". In: *Marketing science* 27.3, pp. 356–378.

A Figures



(a) Original Feed (b) Moderated Feed

FIGURE A1: HIDING INTERVENTION



(a) Replacement Ad

(b) Possible Interactions

FIGURE A2: REPLACEMENT INTERVENTION

Note: The **left panel** of the figure shows an example of a replacement ad used in the replacement intervention. The **right panel** illustrates potential interactions with the replacement ad. Users can like the ad or react to it using any available emoji, and their reaction will be displayed naturally, as shown in the figure. If someone attempts to comment on or share the ad, a small pop-up appears, informing them that commenting and sharing are disabled (as depicted in the figure).

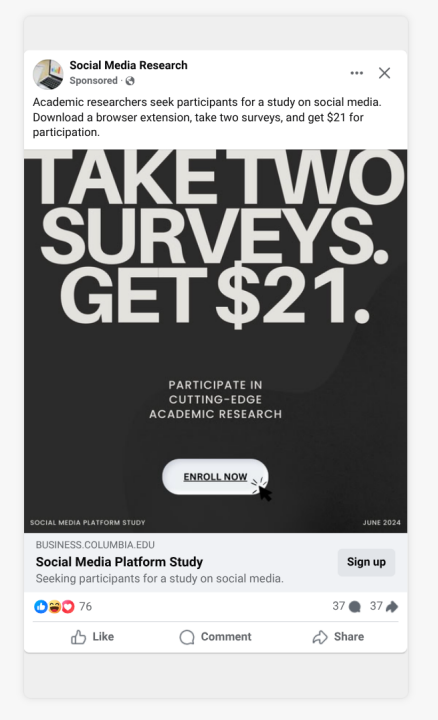


FIGURE A3: EXAMPLE OF A RECRUITMENT AD

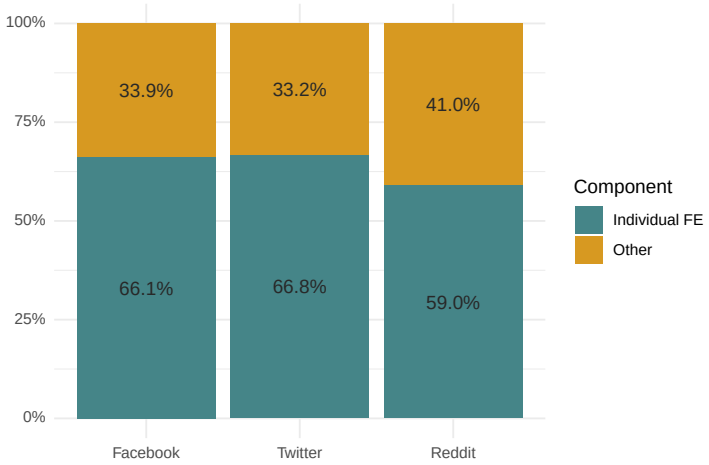


FIGURE A4: AD LOADS: VARIANCE DECOMPOSITION

Note: The figure reports the results of a variance decomposition exercise, showing the share of variance in experienced ad load attributable to individual fixed effects for each social media platform (Facebook, Reddit, and Twitter). Estimates are obtained from a two-way fixed effects regression that includes both individual and time fixed effects. The decomposition is based on feed-only ad load. The figure is based on data from the baseline period. For the Pure Control group, we also include the intervention period data. The figure relies on the subsample of users who viewed at least 50 posts and comments throughout the relevant period.

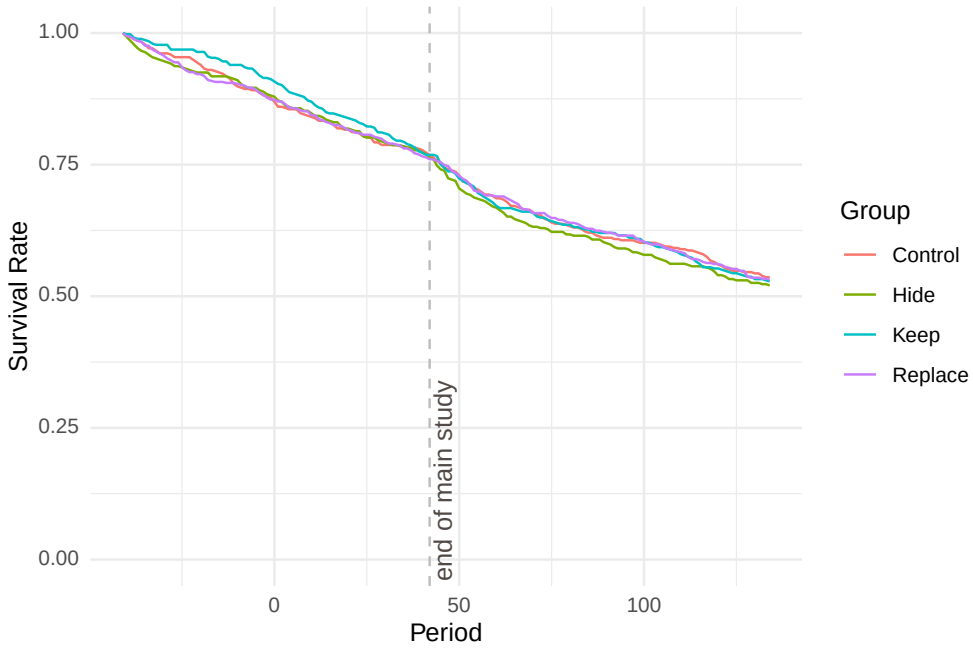


FIGURE A5: SURVIVAL RATE IN THE STUDY

Note: The figure presents the share of users enrolled in the study who survived until a particular study period, measured as the number of days relative to the intervention start date (which starts in period 1). We say that a particular user survived until period t if they were active on their browser on the day corresponding to period t or later. The survival shares are split by treatment conditions.

B Tables

Table B1: Balance Table: Hide vs. Pure Control

	Control (N=414)		Treatment (N=413)			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff. in Means	p
Age	43.9	13.0	44.6	13.4	0.7	0.446
Male	0.4	0.5	0.3	0.5	0.0	0.829
White	0.7	0.5	0.7	0.5	0.0	0.913
Bachelors	0.4	0.5	0.3	0.5	0.0	0.261
Income 50k+	0.4	0.5	0.4	0.5	0.0	0.399
Democrat	0.6	0.5	0.6	0.5	0.0	0.926
Facebook Desktop Usage	52.0	23.2	49.4	23.1	−2.6	0.121
Baseline Time on FB	14.3	22.5	15.8	34.4	1.4	0.474
Baseline Content on FB	73.4	132.7	80.9	185.0	7.4	0.507
Baseline Ads on FB	8.6	17.2	8.1	16.6	−0.4	0.715
Baseline Ad Clicks on FB	0.2	0.4	0.1	0.3	0.0	0.191

Note: This table compares user characteristics between the Hide treatment group and the Pure Control group in the main experimental sample. It reports group means, standard deviations, differences in means, and p-values from two-sided tests of differences in means. The variables are reported in the following order: (1) age (in years), (2) an indicator for male, (3) an indicator for identifying as white/Caucasian, (4) an indicator for having at least a bachelor's degree, (5) an indicator for having household income above $50,000, (6) an indicator for being a Democrat, (7) desktop share of Facebook usage (from the baseline survey), (8) number of Facebook friends (from the baseline survey), (9) baseline active time on Facebook (minutes per day), (10) baseline number of posts and comments consumed per day on Facebook, (11) baseline number of ad impressions per day on Facebook, and (12) baseline number of ad clicks per day on Facebook. The p-values are based on regressions estimated using inverse probability weights reflecting the probability of group assignment for individuals recruited on a given day.

Table B2: Balance Table: Replace vs. Keep

	Control (N=445)		Treatment (N=538)			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff. in Means	p
Age	44.0	12.8	45.5	13.2	1.5	0.073
Male	0.3	0.5	0.3	0.5	0.0	0.374
White	0.7	0.4	0.8	0.4	0.0	0.756
Bachelors	0.4	0.5	0.4	0.5	0.0	0.414
Income 50k+	0.4	0.5	0.4	0.5	0.0	0.560
Democrat	0.6	0.5	0.6	0.5	0.0	0.428
Facebook Desktop Usage	50.8	22.7	51.3	24.1	0.5	0.764
Baseline Time on FB	15.7	27.3	15.6	29.6	−0.1	0.961
Baseline Content on FB	83.3	180.9	82.3	165.1	−1.0	0.927
Baseline Ads on FB	9.8	24.1	8.9	19.3	−0.9	0.524
Baseline Ad Clicks on FB	0.1	0.5	0.2	1.0	0.1	0.253

Note: This table compares user characteristics between the Hide treatment group and the Pure Control group in the main experimental sample. It reports group means, standard deviations, differences in means, and p-values from two-sided tests of differences in means. The variables are reported in the following order: (1) age (in years), (2) an indicator for male, (3) an indicator for identifying as white/Caucasian, (4) an indicator for having at least a bachelor's degree, (5) an indicator for having household income above $50,000, (6) an indicator for being a Democrat, (7) desktop share of Facebook usage (from the baseline survey), (8) number of Facebook friends (from the baseline survey), (9) baseline active time on Facebook (minutes per day), (10) baseline number of posts and comments consumed per day on Facebook, (11) baseline number of ad impressions per day on Facebook, and (12) baseline number of ad clicks per day on Facebook. The p-values are based on regressions estimated using inverse probability weights reflecting the probability of group assignment for individuals recruited on a given day.

Table B3: Effect of Interventions on Facebook Advertising Loads

	User-Side Outcomes		Platform-Side Outcomes	
	Experienced Ad Load	Experienced Advertiser Distance	Supplied Ad Load	Supplied Advertiser Distance
Panel A: Hide vs Pure Control				
Treated	-0.113***	-0.070***	0.003	-0.016*
	(0.003)	(0.011)	(0.002)	(0.008)
Mean	0.137	0.997	0.137	0.997
SD	0.113	0.233	0.113	0.233
N	27334	16736	27335	18999
Panel B: Replace vs Keep Control				
Treated	0.000	0.109***	-0.000	0.015**
	(0.004)	(0.007)	(0.004)	(0.006)
Mean	0.137	0.999	0.137	0.999
SD	0.124	0.233	0.124	0.233
N	32344	21273	32344	21257

Note: **Panel A** presents estimated treatment effects of the hiding intervention on Facebook (relative to Pure Control), based on Equation 2 and our main experimental sample. **Panel B** reports corresponding estimates for the replacement intervention (relative to Keep Control). The dependent variables are: (i) experienced ad load, advertisement as a share of posts and comments viewed by the user, (ii) experienced advertising distance, (iii) supplied ad load, advertisement as a share of posts and comments that a user was targeted with, before any filtering was applied, and (iv) supplied advertiser distance, based on ads that the user was targeted with, before any filtering was applied. Ad distance is defined as the Euclidean distance between a user's characteristics and the average demographic profile targeted by that advertiser, based on baseline data. The unit of observation is the individual-day, where day is measured relative to the intervention date. Driscoll-Kraay standard errors are parenthesized. All regressions were estimated using inverse probability weights reflecting the probability of group assignment for individuals recruited on a given day. $*$, $**$, and $***$ denote significance at the 10%, 5%, and 1% levels, respectively.

Table B4: Short-Term Effects of Interventions on Engagement

	Platform Engagement			Ad Engagement		
	Index	Time	Content	Index	Impressions	Clicks
Panel A: Hide vs Pure Control						
Treated	0.026	0.646	5.641*	-0.192***	-7.049***	-0.038***
	(0.017)	(0.576)	(3.282)	(0.015)	(0.334)	(0.011)
Mean	0.00	11.26	59.93	0.00	7.44	0.066
SD	0.85	28.08	175.10	0.808	25.28	0.357
N	69468	69468	69468	69468	69468	69468
Panel B: Replace vs Keep Control						
Treated	-0.011	-0.674	-3.738	-0.150***	-1.586***	-0.093***
	(0.011)	(0.475)	(2.797)	(0.014)	(0.450)	(0.009)
Mean	0.00	12.37	64.52	0.00	8.30	0.067
SD	0.852	31.74	204.40	0.796	32.80	0.37
N	82572	82572	82572	82572	82572	82572

Note: **Panel A** presents estimated treatment effects of the hiding intervention on Facebook (relative to Pure Control), based on Equation 2 and our main experimental sample. **Panel B** reports corresponding estimates for the replacement intervention (relative to Keep Control). The dependent variables are: (1) a platform engagement index, constructed from active time spent on the platform and the number of posts and comments shown to the user; (2) active time spent on the platform; (3) number of posts and comments shown to the user; (4) an ad engagement index, constructed from the number of ad impressions and the number of ad clicks; (5) number of ad impressions; and (6) number of ad clicks. The unit of observation is the individual-day, measured relative to the intervention date. Driscoll-Kraay standard errors are reported in parentheses. All regressions were estimated using inverse probability weights reflecting the probability of group assignment for individuals recruited on a given day. $*$, $**$, and $***$ denote significance at the 10%, 5%, and 1% levels, respectively.

Table B5: Long-Term Effects of Interventions on Engagement

	Platform Engagement			Ad Engagement		
	Index	Time	Content	Index	Impressions	Clicks
Panel A: Hide vs Pure Control						
Treated	0.020*	0.572	6.632***	-0.167***	-6.397***	-0.039***
	(0.012)	(0.420)	(2.476)	(0.011)	(0.326)	(0.008)
Mean	0.00	11.90	63.08	0.00	7.89	0.073
SD	0.85	30.16	180.03	0.813	26.50	0.422
N	116515	116515	116515	116515	116515	116515
Panel B: Replace vs Keep Control						
Treated	-0.032**	-1.701***	-9.401***	-0.178***	-2.033***	-0.109***
	(0.013)	(0.467)	(3.073)	(0.016)	(0.457)	(0.011)
Mean	0.00	13.09	69.67	0.00	8.90	0.067
SD	0.864	33.66	203.49	0.808	30.11	0.38
N	139275	139275	139275	139275	139275	139275

Note: **Panel A** presents the estimated treatment effects of the hiding intervention on Facebook, relative to the Pure Control group, based on Equation 2. **Panel B** reports the corresponding estimates for the replacement intervention, relative to the Keep Control group. All regressions are based on the extended sample, which involves 19.1 weeks of the intervention period (long-term effects) as opposed to 6 weeks specified in the pre-registration (short-term effects). The dependent variables are: (1) a platform engagement index, constructed from active time spent on the platform and the number of posts and comments shown to the user; (2) active time spent on the platform; (3) number of posts and comments shown to the user; (4) an ad engagement index, constructed from the number of ad impressions and the number of ad clicks; (5) number of ad impressions; and (6) number of ad clicks. The unit of observation is the individual-day, measured relative to the intervention date. Driscoll-Kraay standard errors are reported in parentheses. All regressions were estimated using inverse probability weights reflecting the probability of group assignment for individuals recruited on a given day. $*$, $**$, and $***$ denote significance at the 10%, 5%, and 1% levels, respectively.

Table B6: Short-Term Engagement Effects (Asinh)

	Platform Engagement			Ad Engagement		
	Index	Time	Content	Index	Impressions	Clicks
Panel A: Hide vs Pure Control						
Treated	0.066***	0.117***	0.202***	-0.285***	-0.687***	-0.033***
	(0.021)	(0.038)	(0.051)	(0.013)	(0.029)	(0.004)
Mean	0.00	1.20	1.63	0.00	0.904	0.05
SD	0.976	1.82	2.48	0.841	1.59	0.241
N	69468	69468	69468	69468	69468	69468
Panel B: Replace vs Keep Control						
Treated	-0.021*	-0.043	-0.069**	-0.087***	-0.048**	-0.035***
	(0.013)	(0.026)	(0.035)	(0.012)	(0.022)	(0.004)
Mean	0.00	1.22	1.64	0.00	0.883	0.05
SD	0.978	1.85	2.49	0.844	1.59	0.242
N	82572	82572	82572	82572	82572	82572

Note: **Panel A** presents estimated treatment effects of the hiding intervention on Facebook (relative to Pure Control), based on Equation 2 and our main experimental sample. **Panel B** reports corresponding estimates for the replacement intervention (relative to Keep Control). The dependent variables are: (1) a platform engagement index, constructed from active time spent on the platform and the number of posts and comments shown to the user; (2) active time spent on the platform; (3) number of posts and comments shown to the user; (4) an ad engagement index, constructed from the number of ad impressions and the number of ad clicks; (5) number of ad impressions; and (6) number of ad clicks. All outcomes are transformed using the inverse hyperbolic sine (asinh) function; for index variables, the transformation is applied prior to standardization. The unit of observation is the individual-day, measured relative to the intervention date. Driscoll-Kraay standard errors are reported in parentheses. All regressions were estimated using inverse probability weights reflecting the probability of group assignment for individuals recruited on a given day. $*$, $**$, and $***$ denote significance at the 10%, 5%, and 1% levels, respectively.

Table B7: Long-Term Engagement Effects (Asinh)

	Platform Engagement			Ad Engagement		
	Index	Time	Content	Index	Impressions	Clicks
Panel A: Hide vs Pure Control						
Treated	0.054***	0.103***	0.175***	-0.260***	-0.640***	-0.032***
	(0.017)	(0.029)	(0.042)	(0.010)	(0.023)	(0.004)
Mean	0.00	1.22	1.72	0.00	0.951	0.053
SD	0.976	1.84	2.52	0.841	1.62	0.256
N	116515	116515	116515	116515	116515	116515
Panel B: Replace vs Keep Control						
Treated	-0.057***	-0.107***	-0.143***	-0.110***	-0.089***	-0.040***
	(0.012)	(0.023)	(0.029)	(0.013)	(0.023)	(0.004)
Mean	0.00	1.25	1.74	0.00	0.959	0.05
SD	0.978	1.87	2.53	0.839	1.64	0.243
N	139275	139275	139275	139275	139275	139275

Note: **Panel A** presents estimated treatment effects of the hiding intervention on Facebook (relative to Pure Control), based on Equation 2 and our main experimental sample. **Panel B** reports corresponding estimates for the replacement intervention (relative to Keep Control). All regressions are based on the extended sample, which involves 19.1 weeks of the intervention period (long-term effects) as opposed to 6 weeks specified in the pre-registration (short-term effects). The dependent variables are: (1) a platform engagement index, constructed from active time spent on the platform and the number of posts and comments shown to the user; (2) active time spent on the platform; (3) number of posts and comments shown to the user; (4) an ad engagement index, constructed from the number of ad impressions and the number of ad clicks; (5) number of ad impressions; and (6) number of ad clicks. All outcomes are transformed using the inverse hyperbolic sine (asinh) function; for index variables, the transformation is applied prior to standardization. The unit of observation is the individual-day, measured relative to the intervention date. Driscoll-Kraay standard errors are reported in parentheses. All regressions were estimated using inverse probability weights reflecting the probability of group assignment for individuals recruited on a given day. ∗, ∗∗, and ∗∗∗ denote significance at the 10%, 5%, and 1% levels, respectively.

Table B8: Short-Term Engagement Effects (Trimming)

	Platform Engagement			Ad Engagement		
	Index	Time	Content	Index	Impressions	Clicks
Panel A: Hide vs Pure Control						
Treated	0.057***	1.604***	11.155***	-0.277***	-6.190***	-0.055***
	(0.018)	(0.559)	(3.233)	(0.013)	(0.335)	(0.005)
Mean	0.00	10.90	54.47	0.00	6.34	0.051
SD	0.86	27.43	151.92	0.811	18.19	0.297
N	67956	68712	68712	67956	68712	68712
Panel B: Replace vs Keep Control						
Treated	0.007	-0.275	-3.326	-0.081***	-1.140**	-0.037***
	(0.012)	(0.509)	(2.466)	(0.017)	(0.434)	(0.007)
Mean	0.00	11.46	54.47	0.00	6.62	0.064
SD	0.854	29.64	159.06	0.83	22.31	0.352
N	80892	81732	81732	80892	81732	81732

Note: **Panel A** presents estimated treatment effects of the hiding intervention on Facebook (relative to Pure Control), based on Equation 2 and our main experimental sample. **Panel B** reports corresponding estimates for the replacement intervention (relative to Keep Control). All regressions are based on the sample excluding the top 1 percentile of observations—the exclusion is performed for each outcome separately. The dependent variables are: (1) a platform engagement index, constructed from active time spent on the platform and the number of posts and comments shown to the user; (2) active time spent on the platform; (3) number of posts and comments shown to the user; (4) an ad engagement index, constructed from the number of ad impressions and the number of ad clicks; (5) number of ad impressions; and (6) number of ad clicks. The unit of observation is the individual-day, measured relative to the intervention date. Driscoll-Kraay standard errors are reported in parentheses. All regressions were estimated using inverse probability weights reflecting the probability of group assignment for individuals recruited on a given day. ∗, ∗∗, and ∗∗∗ denote significance at the 10%, 5%, and 1% levels, respectively.

Table B9: Long-Term Engagement Effects (Trimming)

	Platform Engagement			Ad Engagement		
	Index	Time	Content	Index	Impressions	Clicks
Panel A: Hide vs Pure Control						
Treated	0.047***	1.631***	10.893***	-0.235***	-5.468***	-0.049***
	(0.013)	(0.385)	(2.282)	(0.010)	(0.258)	(0.004)
Mean	0.00	11.30	56.10	0.00	6.57	0.052
SD	0.855	28.93	152.01	0.819	18.74	0.307
N	113783	115231	115111	113765	115165	115169
Panel B: Replace vs Keep Control						
Treated	-0.038***	-1.199**	-9.019***	-0.090***	-1.651***	-0.037***
	(0.014)	(0.496)	(2.653)	(0.017)	(0.445)	(0.007)
Mean	0.00	12.27	63.09	0.00	7.50	0.065
SD	0.865	31.60	177.12	0.824	23.01	0.362
N	136699	137782	137859	136513	137874	137858

Note: **Panel A** presents estimated treatment effects of the hiding intervention on Facebook (relative to Pure Control), based on Equation 2 and our main experimental sample. **Panel B** reports corresponding estimates for the replacement intervention (relative to Keep Control). All regressions are based on the extended sample, which involves 19.1 weeks of the intervention period (long-term effects) as opposed to 6 weeks specified in the pre-registration (short-term effects). In addition, we exclude the top 1 percentile of observations for each outcome. The dependent variables are: (1) a platform engagement index, constructed from active time spent on the platform and the number of posts and comments shown to the user; (2) active time spent on the platform; (3) number of posts and comments shown to the user; (4) an ad engagement index, constructed from the number of ad impressions and the number of ad clicks; (5) number of ad impressions; and (6) number of ad clicks. The unit of observation is the individual-day, measured relative to the intervention date. Driscoll-Kraay standard errors are reported in parentheses. All regressions were estimated using inverse probability weights reflecting the probability of group assignment for individuals recruited on a given day. ∗, ∗∗, and ∗∗∗ denote significance at the 10%, 5%, and 1% levels, respectively.

Table B10: Main Attrition Analysis

	Survived						Last Period					
	Hiding	Replacing	Keeping	Hiding	Replacing	Keeping	Hiding	Replacing	Keeping	Hiding	Replacing	Keeping
(Intercept)	0.768***	0.769***	0.768***	−0.065	0.115	−0.065	31.502***	33.252***	31.502***	−13.681	1.813	−13.681
	(0.021)	(0.020)	(0.021)	(0.164)	(0.160)	(0.164)	(1.088)	(0.917)	(1.088)	(9.088)	(7.514)	(9.088)
GroupTreatment	−0.005	−0.008	0.000	0.010	−0.177	0.180	−0.168	−1.964	1.749	2.682	−11.561	15.494
	(0.030)	(0.027)	(0.029)	(0.233)	(0.219)	(0.229)	(1.565)	(1.341)	(1.423)	(12.552)	(11.080)	(11.793)
SurvivalScore				1.099***	0.846***	1.099***				59.655***	40.697***	59.655***
				(0.203)	(0.198)	(0.203)				(11.069)	(9.073)	(11.068)
GroupTreatment×SurvivalScore				−0.023	0.224	−0.253				−3.604	12.725	−18.959
				(0.289)	(0.270)	(0.283)				(15.240)	(13.388)	(14.313)
N	827	983	859	777	918	807	827	983	859	777	918	807

Note: Columns 1-6 report individual-level regressions of an indicator for survival until the end of the intervention period (period 42, i.e., 42 days after the start of the intervention) on the treatment indicator. Columns titled "Hiding" pertain to the comparison between Hiding and Pure Control. Columns titled "Replacing" pertain to the comparison between Replacement and Keep Control. Columns titled "Keeping" pertain to the comparison between Keep Control and Pure Control. In columns 4-6, we additionally include the survival propensity based on user characteristics—computed by regressing survival on age, male indicator, white indicator, indicator for obtaining a bachelor's degree, indicator for being a Democrat, share of desktop usage of Facebook, active time spent on Facebook during the baseline period, content consumption on Facebook during the baseline, number of ad impressions during the baseline, and baseline number of ad clicks. Columns 7-12 report analogous regressions, with the dependent variable being the last period of recorded activity, which we use as a proxy for the last period in the study. The dependent variable is capped at 42, which indicates the last period of the six-week intervention. Note that the intervention starts in period 1, with baseline periods represented by negative numbers and zero. Standard errors clustered at the individual level are reported in parentheses. ∗, ∗∗, and ∗∗∗ denote significance at the 10%, 5%, and 1% levels, respectively.

Table B11: Long-Term Attrition Analysis

	Hiding	Replacing	Keeping	Hiding	Replacing	Keeping
(Intercept)	137.751***	141.816***	137.751***	−50.386	−86.110**	−50.386
	(5.259)	(5.086)	(5.258)	(40.493)	(42.031)	(40.489)
GroupTreatment	0.089	−0.819	4.065	−11.402	−1.428	−35.724
	(7.553)	(6.943)	(7.316)	(57.815)	(56.926)	(58.370)
SurvivalScore				248.394***	296.069***	248.394***
				(52.183)	(54.613)	(52.178)
GroupTreatment×SurvivalScore				15.080	0.804	47.675
				(74.458)	(73.487)	(75.544)
N	827	983	859	777	918	807

Note: Columns 1-3 report individual-level regressions of the last period of recorded activity, which we use as a proxy for the last period in the study, on the treatment indicator. Columns titled "Hiding" pertain to the comparison between Hiding and Pure Control. Columns titled "Replacing" pertain to the comparison between Replacement and Keep Control. Columns titled "Keeping" pertain to the comparison between Keep Control and Pure Control. In columns 4-6, we additionally include the survival propensity based on user characteristics—computed by regressing survival on age, male indicator, white indicator, indicator for obtaining a bachelor's degree, indicator for being a Democrat, share of desktop usage of Facebook, active time spent on Facebook during the baseline period, content consumption on Facebook during the baseline, number of ad impressions during the baseline, and baseline number of ad clicks. The dependent variable is not capped at any value. Note that the intervention starts in period 1, with baseline periods represented by negative numbers and zero. Standard errors clustered at the individual level are reported in parentheses. ∗, ∗∗, and ∗∗∗ denote significance at the 10%, 5%, and 1% levels, respectively.

Table B12: Short-Term Engagement Effects: Clustered Errors

	Platform Engagement			Ad Engagement		
	Index	Time	Content	Index	Impressions	Clicks
Panel A: Hide vs Pure Control						
Treated	0.026	0.646	5.641	-0.192***	-7.049***	-0.038
	(0.036)	(1.283)	(7.338)	(0.051)	(1.106)	(0.027)
Mean	0.00	11.26	59.93	0.00	7.44	0.066
SD	0.85	28.08	175.10	0.808	25.28	0.357
N	69468	69468	69468	69468	69468	69468
Panel B: Replace vs Keep Control						
Treated	-0.011	-0.674	-3.738	-0.150*	-1.586	-0.093
	(0.032)	(1.278)	(7.158)	(0.082)	(1.098)	(0.057)
Mean	0.00	12.37	64.52	0.00	8.30	0.067
SD	0.852	31.74	204.40	0.796	32.80	0.37
N	82572	82572	82572	82572	82572	82572

Note: **Panel A** presents estimated treatment effects of the hiding intervention on Facebook (relative to Pure Control), based on Equation 2 and our main experimental sample. **Panel B** reports corresponding estimates for the replacement intervention (relative to Keep Control). The dependent variables are: (1) a platform engagement index, constructed from active time spent on the platform and the number of posts and comments shown to the user; (2) active time spent on the platform; (3) number of posts and comments shown to the user; (4) an ad engagement index, constructed from the number of ad impressions and the number of ad clicks; (5) number of ad impressions; and (6) number of ad clicks. The unit of observation is the individual-day, measured relative to the intervention date. Standard errors clustered at the individual level are reported in parentheses. All regressions were estimated using inverse probability weights reflecting the probability of group assignment for individuals recruited on a given day. ∗, ∗∗, and ∗∗∗ denote significance at the 10%, 5%, and 1% levels, respectively.

Table B13: Long-Term Engagement Effects: Clustered Errors

	Platform Engagement			Ad Engagement		
	Index	Time	Content	Index	Impressions	Clicks
Panel A: Hide vs Pure Control						
Treated	0.020	0.572	6.632	-0.167***	-6.397***	-0.039
	(0.031)	(1.107)	(6.242)	(0.042)	(1.019)	(0.025)
Mean	0.00	11.90	63.08	0.00	7.89	0.073
SD	0.85	30.16	180.03	0.813	26.50	0.422
N	116515	116515	116515	116515	116515	116515
Panel B: Replace vs Keep Control						
Treated	-0.032	-1.701	-9.401	-0.178*	-2.033**	-0.109*
	(0.031)	(1.162)	(6.446)	(0.093)	(0.958)	(0.066)
Mean	0.00	13.09	69.67	0.00	8.90	0.067
SD	0.864	33.66	203.49	0.808	30.11	0.38
N	139275	139275	139275	139275	139275	139275

Note: **Panel A** presents estimated treatment effects of the hiding intervention on Facebook (relative to Pure Control), based on Equation 2 and our main experimental sample. **Panel B** reports corresponding estimates for the replacement intervention (relative to Keep Control). All regressions are based on the extended sample, which involves 19.1 weeks of the intervention period (long-term effects) as opposed to 6 weeks specified in the pre-registration (short-term effects). The dependent variables are: (1) a platform engagement index, constructed from active time spent on the platform and the number of posts and comments shown to the user; (2) active time spent on the platform; (3) number of posts and comments shown to the user; (4) an ad engagement index, constructed from the number of ad impressions and the number of ad clicks; (5) number of ad impressions; and (6) number of ad clicks. The unit of observation is the individual-day, measured relative to the intervention date. Standard errors clustered at the individual level are reported in parentheses. All regressions were estimated using inverse probability weights reflecting the probability of group assignment for individuals recruited on a given day. $*$, $**$, and $***$ denote significance at the 10%, 5%, and 1% levels, respectively.

Table B14: Short-Term Engagement Effects: Clustered Errors (Asinh)

	Platform Engagement			Ad Engagement		
	Index	Time	Content	Index	Impressions	Clicks
Panel A: Hide vs Pure Control						
Treated	0.066*	0.117*	0.202**	-0.285***	-0.687***	-0.033**
	(0.037)	(0.070)	(0.092)	(0.045)	(0.076)	(0.013)
Mean	0.00	1.20	1.63	0.00	0.904	0.05
SD	0.976	1.82	2.48	0.841	1.59	0.241
N	69468	69468	69468	69468	69468	69468
Panel B: Replace vs Keep Control						
Treated	-0.021	-0.043	-0.069	-0.087**	-0.048	-0.035**
	(0.032)	(0.060)	(0.081)	(0.037)	(0.050)	(0.014)
Mean	0.00	1.22	1.64	0.00	0.883	0.05
SD	0.978	1.85	2.49	0.844	1.59	0.242
N	82572	82572	82572	82572	82572	82572

Note: **Panel A** presents estimated treatment effects of the hiding intervention on Facebook (relative to Pure Control), based on Equation 2 and our main experimental sample. **Panel B** reports corresponding estimates for the replacement intervention (relative to Keep Control). The dependent variables are: (1) a platform engagement index, constructed from active time spent on the platform and the number of posts and comments shown to the user; (2) active time spent on the platform; (3) number of posts and comments shown to the user; (4) an ad engagement index, constructed from the number of ad impressions and the number of ad clicks; (5) number of ad impressions; and (6) number of ad clicks. All outcomes are transformed using the inverse hyperbolic sine (asinh) function; for index variables, the transformation is applied prior to standardization. The unit of observation is the individual-day, measured relative to the intervention date. Standard errors clustered at the individual level are reported in parentheses. All regressions were estimated using inverse probability weights reflecting the probability of group assignment for individuals recruited on a given day. $*$, $**$, and $***$ denote significance at the 10%, 5%, and 1% levels, respectively.

Table B15: Long-Term Engagement Effects: Clustered Errors (Asinh)

	Platform Engagement			Ad Engagement		
	Index	Time	Content	Index	Impressions	Clicks
Panel A: Hide vs Pure Control						
Treated	0.054	0.103	0.175**	-0.260***	-0.640***	-0.032***
	(0.034)	(0.064)	(0.085)	(0.039)	(0.069)	(0.012)
Mean	0.00	1.22	1.72	0.00	0.951	0.053
SD	0.976	1.84	2.52	0.841	1.62	0.256
N	116515	116515	116515	116515	116515	116515
Panel B: Replace vs Keep Control						
Treated	-0.057*	-0.107*	-0.143*	-0.110***	-0.089*	-0.040***
	(0.031)	(0.058)	(0.078)	(0.038)	(0.051)	(0.014)
Mean	0.00	1.25	1.74	0.00	0.959	0.05
SD	0.978	1.87	2.53	0.839	1.64	0.243
N	139275	139275	139275	139275	139275	139275

Note: **Panel A** presents estimated treatment effects of the hiding intervention on Facebook (relative to Pure Control), based on Equation 2 and our main experimental sample. **Panel B** reports corresponding estimates for the replacement intervention (relative to Keep Control). All regressions are based on the extended sample, which involves 19.1 weeks of the intervention period (long-term effects) as opposed to 6 weeks specified in the pre-registration (short-term effects). The dependent variables are: (1) a platform engagement index, constructed from active time spent on the platform and the number of posts and comments shown to the user; (2) active time spent on the platform; (3) number of posts and comments shown to the user; (4) an ad engagement index, constructed from the number of ad impressions and the number of ad clicks; (5) number of ad impressions; and (6) number of ad clicks. All outcomes are transformed using the inverse hyperbolic sine (asinh) function; for index variables, the transformation is applied prior to standardization. The unit of observation is the individual-day, measured relative to the intervention date. Standard errors clustered at the individual level are reported in parentheses. All regressions were estimated using inverse probability weights reflecting the probability of group assignment for individuals recruited on a given day. $*$, $**$, and $***$ denote significance at the 10%, 5%, and 1% levels, respectively.