Social media advertising loads as prices

George Beknazar-Yuzbashev, Columbia University. Email: gb2683@columbia.edu
Rafael Jiménez-Durán, Bocconi University. Email: rafael.jimenez@unibocconi.it
Andrey Simonov, Columbia University. Email: as5443@gsb.columbia.edu
Mateusz Stalinski, University of Warwick. Email: Mateusz.Stalinski@warwick.ac.uk

Sample

We anticipate recruiting approximately 1,500 participants using Facebook ads. The ad campaigns are selected to optimize conversions, which we measure as participants finishing the intake survey that opens after installing our browser extension. We are targeting desktop ads to US-based adults with UK or US English language who use Chrome to access Facebook and exclude those who use Firefox, Safari, or Edge to access Facebook. At the start of recruitment, the extension does not work for Edge, but we are planning to extend support to that browser, at which point we will remove Edge from the exclusion criteria.

We exclude from the sample those participants who:

- Do not have English set as their interface language on Facebook;
- Are not using Chrome or Edge (the latter – if we manage to support it);
- Appear to be using bots or create multiple accounts in order to game the system and get more rewards than one per individual (for example, during piloting, we noticed a single individual with browser language set to “zh-CN” who has managed multiple consecutive installations and survey completions from different IP addresses).
- Uninstall the extension upon installing it, before the intervention period begins.
- Do not use their Facebook account for at least 1 minute per week during the baseline period.

At the time of pre-registration we are planning to apply for different grants requesting additional funding. If the grants get approved, we will expand our sample size by approximately 600 more participants (this number depends on the price of recruitment ads on Facebook, which are hard to predict in an election cycle).

Treatment

Once participants install the browser extension, we randomize them uniformly across four treatment arms, which (possibly) manipulate their ad experience on Facebook. The functionality of the extension is based on two key functions with respect to ads that appear in the feed: (i) hiding ads, and (ii) replacing ads with other ads (to reduce the degree of ad targeting). In order to equate the potential (minimal) latency issues as well as CPU speed differences for all groups, we run code for both functions for all participants. Different treatment groups then see different actions taken as a result of these functions. The intervention starts after a baseline period of six weeks from the day of installing the browser extension and has a duration of six weeks. We do not alert participants when the intervention starts.

- Quantity reduction group. In this treatment group, we hide (remove) all ads from participants' Facebook feeds (the only exception are ads that appear in the viewport when Facebook loads).
- Targeting reduction group. In this treatment group, we replace each eligible ad (see below) from participants' Facebook feeds with an English-language ad randomly selected from the pool of ads shown to all of our participants in the previous day. We eliminate the possibility to comment, view comments, or share the ad.
- Passive Control group. This control group does not experience any modification to the ads.
- Replacement Control group. For participants in this control group, we substitute eligible ads with themselves. The purpose of this control group is to test whether the feasibility limitations that shape the targeting reduction group (e.g., the fact that we eliminate the possibility to share) have an effect on the outcomes of interest.

Eligible ads are all ads that appear in the feed, except:

- Type "carousel";
- Ads that appear in the viewport when Facebook loads;
- Video ads;
- Facebook forms.

Outcomes

Time. The main outcome of interest is the amount of active time spent on Facebook and on two different types of content displayed on Facebook: ads and organic posts. Active time spent is calculated as described in Beknazar, Jiménez-Durán, McCrosky, and Stalinski (2022).

- Another main outcome of interest is the time substitution to Twitter and Reddit, where we can also measure active time, as well as YouTube, where we only measure the approximate time the browser window was open.
- Besides these websites, a secondary outcome of interest is the time spent on other social media platforms as in Beknazar, Jiménez-Durán, McCrosky, and Stalinski (2022). Moreover, we consider the substitution in terms of time spent on the top-10 websites visited by each participant during baseline.

Engagement with ads. Another main outcome of interest is the level of consumer engagement and activity of consumers with the ad content. Our main measure is an index (Kling, Liebman, and Katz, 2007) constructed with the number of impressions (views) and clicks (although we will also report the effect on each of these outcomes separately).

- We will also report as a secondary outcome of interest the time spent on the advertiser's website (capped at 5 minutes for each visit), and "bounce backs" (returns in 30 seconds or less) to Facebook.

- Besides this, another secondary outcome of interest is user reactions to ads (i.e., likes/shares), in the treatment arms where we can collect these.

Valuation. Another main outcome of interest is the post-intervention vs. pre-intervention change in the (incentivized) Willingness-to-Accept (WTA) to deactivate social media for four weeks. We will elicit the WTA in the baseline survey and in the endline survey. The question is open-ended, and we will truncate the values at $1000.

- As a secondary outcome of interest, we will elicit the WTP to keep the browser extension installed for four more weeks. We will elicit it using a take-it-or-leave-it approach, randomly varying the amount that we offer to participants.

Other secondary outcomes of interest are the amount of content consumed and produced, and the toxicity of content produced. Besides these outcomes, we will report auxiliary outcomes to understand Facebook's response to the intervention and to rule out potential confounders. For example, we will compute the ad load "offered" (before hiding) to participants and a measure of how targeted the ads offered are. We also have access to participants' Twitter handles (collected using the browser extension) and will compute, using the API, the number of posts they produce and people they follow, to measure substitution patterns more precisely. Along these lines, we will measure the change in the percent of social media use on browser vs mobile, which we will elicit in the baseline and endline surveys. We will measure time spent on mobile using screenshots from users' phones. We will also ask for screenshots of targeting data on Facebook.

PE Outcomes. We will also have a set of secondary outcomes related to political economics. First, we will elicit the propensity to share fake news during the intake and endline surveys. We will also extract urls to websites shared and observed by the participants to examine their interaction with non-fake-news websites (we will use the list of news websites from Gregory Martin, Andrey Simonov, and Shoshana Vasserman. "Beyond the Paywall: Measuring Supply and Demand for Online News in a Rapidly Changing News Environment", work in progress) as well as fake-news websites (we will use the list from Melnikov, 2021 and Lasser and Rupp, 2022) and the propensity to share political information. Lastly, using the available information about our participants we will match them to voter records to obtain a measure of turnout.

Heterogeneity

We expect platforms to optimally set ad loads depending on 1) the elasticity of a user with respect to their ad load, 2) the network externalities that the user imposes on others, and 3) the value of targeting the user with ads. Based on this intuition, we will report the average ad load by:

1. Users with above vs. below median elasticity (predicted elasticity). We will predict the elasticity using our ads reduction intervention and demographics/baseline variables.

2. Users with above vs. below median network effects. Our measure of network effects will be the self-reported number of friends on Facebook, but we will also use other measures such as the number of posts produced or the composition of the feed.
3. Users with above vs. below value of ads. The value of ads is driven both by the probability that a user engages with the ad (e.g. clicks) and the value of the click for the advertiser. We will measure engagement through users' actions (e.g., clicks) and the time they spend viewing ads. We also obtain a measure of the cost-per-click of targeting users with ads based on the predicted number of clicks from the Facebook Ads manager, targeting on gender and age.

Besides these variables, we will conduct more systematic approaches to understand which variables predict variability in ad loads.

First, we will conduct a variance decomposition exercise to test whether variation between individuals is a bigger source of variation than that of within individuals.

We will also analyze which variables predict cross-sectional differences in ad-loads. We will use OLS and machine-learning methods, including as predictors: baseline activity on and off the platforms, demographics, baseline content produced, baseline toxicity produced and consumed, predicted ad elasticity, measures of network effects interacted with ad engagement by users (the share of time allocated to ads versus other content by user), the percent of content that individuals see from friends vs from non-friends.

Lastly, we will analyze heterogeneous treatment effects using the same predictors and the Generalized Random Forest approach (Athey, Tibshirani, & Wager, 2019).

Empirical Analysis

To increase power, we will exploit both between and within variation, taking advantage that for most individuals we will have at least six weeks of baseline period without intervention and six weeks of intervention. In our main analyses we will run daily difference-in-difference two-way fixed effect regressions. We will also report specifications that are robust to the staggered nature of our treatment.

We will exploit that we have a large number of periods and hence will use Driscoll and Kraay (1998) standard errors to increase power. For those outcomes for which we do not have a long enough time dimension, we will also report standard errors clustered at the individual level. We will also report event-study estimates.

We will report two main specifications: 1) comparing the quantity reduction treatment arm with the passive control group, and 2) comparing the targeting reduction group with the replacement control group. We will also report a third version, which compares both control groups with each other. For some outcomes for which we do not have daily observations (e.g., valuation outcomes), we will report OLS regressions controlling for baseline outcomes, demographics, and enrollment and intervention start dates.

In terms of time frame, our main specification will include up to six weeks of intervention period. However, we will also report longer-run estimates.

In terms of attrition, we expect based on our previous work to have a low attrition rate (less than 15% over the main period of interest). We also expect this attrition rate to not be differential across treatment groups.

Based on our estimates, we will report “price” elasticities and diversion ratios.

Power

We conducted a power analysis using pilot data (from 31 users in the quantity reduction condition, 28 users in the replacement control condition, 31 in the targeting reduction condition, and 120 in the pure control condition). This intervention lasted for 13 days and had a baseline of 14 days.
We conducted 100 simulations at different sample sizes of our main empirical specification (diff-in-diff) and a more conservative specification that relies only on difference in means (cross-sectional specification). We did this for both the quantity reduction and the targeting reduction interventions. Based on these simulations, we require a sample of at least 1,200 participants to have 80% power with conventional size. For this reason, we pre-register a sample of around 1,500 participants. Note also that we expect the effect sizes to be larger during the main intervention, which will last longer than the pilot intervention.

However, the smallest effect size of the cross-sectional difference in means is 0.11 SD, which would require a sample size of 2,500 in a design with two treatment arms. Even if the difference in means is not our main specification, we will attempt to expand our number of observations, which will be contingent on having a grant approved.

Table 1: Pilot results

			Pilot sample				Simulation sample		
Outcome	Specification	Treatment	Effect (SD)	Effect (Coef)	t-statistic	p-value	% Significant simulations	% Positive simulations	Min. sample size for 80% power
Ad clicks	Cross-Section	Quality	-0.35	-0.55	-2.22	0.03	89%	2%	100
	Cross-Section	Quantity	-0.12	-0.19	-0.94	0.35	86%	2%	1200
	DID	Quality	-0.14	-0.22	-1.16	0.25	80%	10%	100
	DID	Quantity	-0.41	-0.64	-3.79	0.00	100%	0%	100
Pings	Cross-Section	Quality	-0.18	-98.23	-2.95	0.00	99%	0%	100
	Cross-Section	Quantity	0.11	57.59	2.86	0.00	82%	95%	100
	DID	Quality	0.04	21.54	1.04	0.30	98%	100%	500
	DID	Quantity	0.06	32.71	1.17	0.24	83%	91%	500

Previously collected sample

Recruitment started on June 17, 2024 and proceeds on a rolling basis. Before August 14, 2024, we registered 1002 installations of our extension. We used part of this sample for the pilot (90 participants), but all remaining users remained in an untreated state (and were used as a control for the pilot). Note that the intervention for this sample of users will start after pre-registration, and we plan to include them in our main specifications to increase power. For each participant, the intervention will start at the latest of August 22, 2024 or six weeks from recruitment date (this date means that some of the participants recruited earlier will have a benchmark period slightly longer than 6 weeks, and this is due to a delay in the final version of the extension clearing the Chrome Store).

References

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests.

Beknazar-Yuzbashev, G., Jiménez Durán, R., McCrosky, J., & Stalinski, M. (2022). Toxic content and user engagement on social media: Evidence from a field experiment. Available at SSRN 4307346.

Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental analysis of neighborhood effects. Econometrica, 75(1), 83-119.

Melnikov, N. (2021). Mobile internet and political polarization. Available at SSRN 3937760.