The Consumer Welfare Effects of Online Ads: Evidence from a 9-Year Experiment

Erik Brynjolfsson
Stanford University and NBER

Avinash Collis
Carnegie Mellon University

Daniel Deisenroth
Meta

Haritz Garro
Meta

Daley Kutzman
Meta

Asad Liaqat
Meta

Nils Wernerfelt*
Northwestern University

Abstract

Research on the effects of online advertising on consumer welfare is limited due to challenges in running large-scale field experiments. We analyze a long-running field experiment on Facebook in which a random subset of users received no ads in their newsfeeds. Using an incentive - compatible deactivation experiment, we find no significant differences in welfare gains from Facebook across a representative sample of 53,083 Facebook users in the ads and no ads groups. Our sample size allows for precise estimates, suggesting that either the disutility of ads is relatively small or that there are offsetting benefits, such as product discovery.

* All authors are listed alphabetically. A.C. and N.W. led the writing of the paper. D.K. led the analysis and survey development, with key support from A.L. and H.G.. E.B., A.C., D.D., and N.W. conceptualized the research and methodology. D.D. led the project internally at Meta. A.C., E.B., and N.W. are the corresponding authors. The study was not reviewed by an Institutional Review Board.

Competing interests: A.L., D.K., H.G., and D.D. are employees of Meta Platforms and hold a financial interest in Meta. N.W. was an employee of Meta while this research was conducted, although he no longer holds a financial interest in the company. E.B. and A.C. declare no competing interests, although they have previously been awarded unrestricted gifts from Meta.

The survey conducted in 2022 was registered in the AEA Registry (AEARCTR-0008990). The experiment began in 2013 and was not pre-registered.

1. Introduction

Digital advertising now represents a majority of worldwide advertising spending (Cramer-Flood, 2021). The empirical literature on digital advertising has grown concurrently. However, much of it has focused on the firm side; for example, estimating how effective digital advertising is for acquiring new customers or shifting consumer attitudes (e.g., Johnson, Lewis, and Nubbemeyer (2017); Gordon, Moakler, and Zettelmeyer (2023); Wernerfelt et al. (2024); Athey et al. (2023)). Much less is known, however, about the consumer side and the utility or disutility that consumers obtain from digital ads.

Understanding the latter is important from a policy standpoint as digital ads grow more widespread and regulatory interest in privacy and data collection increases around the world. For example, privacy advocates have argued that digital ads impose substantial privacy costs on consumers due to their use of detailed user data (e.g., Zuboff (2019)). Similarly, regulations on consumer data collection and use have been proposed or passed in recent years in the US, Europe, and elsewhere (e.g., GDPR in the EU, CCPA in California), and a focus of many of these regulations has been the data used in digital ads by online platforms. Estimates of how much (dis)utility consumers incur from digital ads would help inform policymakers as they aim to regulate the online advertising ecosystem.

At the same time, measuring consumers' utility from digital advertising is challenging for several reasons. First, many outcomes that one may consider using as proxies – e.g., purchasing behavior, time spent, or engagement – have ambiguous welfare predictions in theory.[1] Second, in digital advertising contexts, there is concern that the assumptions that underlie quasi-experimental approaches may often fail, leading to biased estimates without carefully controlled experiments (Gordon, Moakler, and Zettelmeyer 2023). Third, to rule out meaningful differences in utility, such experiments would require large sample sizes, much larger than are typically available in lab or university settings. Finally, there is evidence that short- and long-term consumer responses to digital advertising may differ (e.g., Goli et al., 2024).

In this study, we overcome these challenges by leveraging a long-running field experiment on Facebook in which 0.5% of the user base is randomly assigned to receive no ads on their Facebook newsfeed.[2] We recruit a representative sample of a total of 53,083 users from the ads and no ads groups and solicit data on their valuations of Facebook through an incentivized online choice experiment (Brynjolfsson, Collis, and Eggers 2019; Brynjolfsson et al. 2023). Comparing the valuations across the two groups allows us to estimate the consumer welfare effects of digital ads.

---

[1] For example, if ads cause an increase in purchases, that may reflect either lowered search costs for consumers (Stigler, 1961) or that the ads persuaded the consumer to purchase something they did not need (Galbraith, 1958). See Bagwell (2007) for an overview of relevant economics literature; Becker and Murphy (1993) is also a classic advertising theory citation that is particularly relevant in online contexts where ad consumption is often a choice variable for consumers.

[2] "Newsfeed" is the primary interface for Facebook and refers to the ranked set of content that users see upon opening their Facebook mobile app. Users in the "no ads" condition may still see ads on Facebook surfaces outside of newsfeed (e.g., banner ads on the desktop version of Facebook), but these represent a very small fraction of overall ad impressions. Going forward when we refer to the "ads" and "no ads" conditions we are referring to users who see ads in their newsfeed or not.

Our core result is a relatively precisely estimated null effect: we find that the median valuations of Facebook are not significantly different across the treatment and control groups. Specifically, we find that the median amount a user would be willing to accept to agree to deactivate their Facebook profile for a month is $31.95 in the no ads group and $31.04 in the ads group.[3] The difference is not significant and tightly estimated. Given the sample size, our minimum detectable effect is $3.18/month: i.e., if the true difference in the median valuations across the groups is more than $3.18/month, we would be able to detect a statistically significant difference with a probability higher than 80%. This suggests that were there to be a difference in valuations across the two arms, it is likely small in magnitude compared to the baseline values.

To bolster our main result, we further provide several checks on our methods. For example, we reassuringly find a downward-sloping demand curve as the offer value to deactivate Facebook increases, consumer valuations increase with various measures of engagement with Facebook, and we replicate our findings at more local levels within the geographies we surveyed.

This paper provides, to our knowledge, the largest experimental study of the consumer welfare of digital ads on a major platform. As noted earlier, work on this topic has faced constraints in their measurement, sample, or design that have made addressing our question a challenge. On the measurement side, the ability to directly solicit consumers' valuations through Facebook's survey infrastructure enabled us to avoid relying on indirect measures of utility (e.g., increased platform usage in Sahni and Zhang (2024); long-term purchasing behavior in Wernerfelt et al. (2024)). Similarly, past work on consumer welfare online has often relied on smaller-scale settings, potentially posing a threat to external validity (e.g., Goldstein et al. (2014); Goldfarb and Tucker (2011); Korganbekova and Zuber (2023)). Finally, Facebook's no ads experiment affords us a clean between-subjects design that enables us to measure consumer valuations without risking bias from the privacy paradox (e.g., Athey, Catalini, and Tucker, 2017).

Our results come with caveats. First, our estimates are by construction partial equilibrium. Were all users to stop receiving digital ads on Facebook, the set of products available to consumers, product prices, content available to be shown on Facebook, and many other relevant aspects of the platform would change that could affect consumers' utilities.[4] As such, our counterfactual is best thought of as what a user would experience if they unilaterally deviated from receiving ads to not receiving ads. A broader, general equilibrium-based counterfactual is beyond the scope of our paper, though we believe the partial equilibrium result is still informative and note that at least some of these general equilibrium effects would likely increase consumers' utilities from

[3] Other studies have estimated similar quantities using different methods and populations. For example, the median monthly willingness to accept to deactivate Facebook amongst populations drawn from the US is $48 in Brynjolfsson, Collis, and Eggers (2019), $64 in Sunstein (2020), and $100 in Allcott et al. (2020). For the US, the corresponding number for our sample is $49. Ours is on the lower end of this range, which we suspect is largely due to our broad sample; Sunstein (2020) and Allcott et al. (2020) relied on workers from Amazon Mechanical Turk and users recruited through engagement with Facebook ads, respectively. In contrast, given the research partnership with Meta, we could sample more uniformly across our population of interest.

[4] See, e.g., Deisenroth et al. (2024), who find evidence a negative shock to digital advertising effectiveness led to increases in market concentration and product prices in the US. Bronnenberg, Dubé, and Joo (2022) also suggest digital advertising may have helped the expansion of the US craft beer industry. Other papers that consider partial equilibrium effects of experiments on Facebook include Allcott et al. (2020) and Wernerfelt et al., (2024); see also Bursztyn et al. (2024) who estimate welfare implications of collective vs. unilateral social media deactivations.

digital ads. Second, our results rely on the assumption that the weighted responses from our incentivized choice experiment represent valid measures of consumer welfare for the populations of Facebook users who receive and do not receive ads in their newsfeed. To the extent there is error in either the measurement or representativeness (Groves et al., 2009) there could be bias in our estimates. To mitigate these concerns we use constructs from established work (Brynjolfsson, Collis, and Eggers, 2019), demonstrate robustness of our main results with and without weighting, and use a between-subjects research design that helps minimize the risk of bias in our main statistic of interest. Finally, our experiment was run on Facebook and thus we cannot directly speak to the external validity of our results on other platforms. Facebook is still an economically meaningful sample itself, and there are several other major platforms where users experience ads in an arguably similar, feed-based way. We leave a deeper exploration of the generalizability of our results to other platforms for future work.

The rest of the paper is organized as follows: Section 2 contains institutional details relevant for understanding and interpreting our results; Section 3 describes the valuation methodology and analysis plan in more depth; Section 4 provides details on our sample; Section 5 has our main results; and Section 6 provides a brief summary and conclusion.

2. Institutional Details

In this section, we outline how users experience ads on Facebook, how holdout experiments are created, and how on-platform surveys are delivered to users. This background knowledge will prove useful in understanding our experiment and interpreting the results.

How users experience ads. The main way users experience Facebook is through their newsfeed. This is a personalized list of items for each user that contains posts, videos, and other content from friends, businesses, and other actors in the Facebook ecosystem. The content appears vertically on mobile devices, and users can scroll through their newsfeed at their leisure.

For the duration of the experiment, ads on Facebook are interjected within the user's normal newsfeed and appear as in Figure 1. This format is in contrast to ads on other platforms, for example, Google or Amazon, where ads often appear in search results or banners. At a high level, ads are matched to consumers via an auction where bids are scaled by estimates of how likely users are to engage with the ad; these weights are computed using available data on the consumers (see Gordon, Moakler, and Zettelmeyer (2023) or Wernerfelt et al. (2024) for more discussion of the advertising auctions). Both the fact that ads appear alongside non-ad content within the newsfeed and that match quality is explicitly taken into account in determining which ads get shown may mitigate potential downsides from ads for consumers.

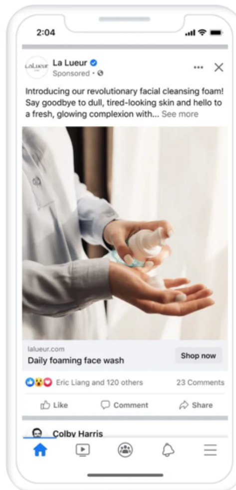


Figure 1: An example ad on newsfeed on mobile (Facebook Ads Guide, 2024).

Holdout experiments. There are many ways holdout experiments are created at Meta and tech companies more broadly. This holdout is created as follows. When a user joins Facebook, she is given a unique user ID. At the same time, a hash of that user ID will randomly determine whether that user is in the no ads or ads condition. Whatever the assignment is, the user will stay in that condition for the entire duration of her time on Facebook. This means that the users in our no ads condition have never seen ads in their newsfeed. Further, as the Facebook population grows, new users are continually added to both arms of the experiment. This helps minimize the potential for differential attrition, while still maintaining a large pool of users with a long tenure in the experiment.[5]

On-platform surveys. Facebook routinely surveys its users to solicit responses on anything from product feedback to offline details of their business (e.g., Alekseev et al., 2023). On-platform surveys are voluntary and the user can leave at any time. Our survey invitation appeared at the top of users' newsfeed when they logged in and would appear as in Figure 2 below for selected English-speaking users.

[5] We note that this way of generating the holdout means the tenure in the experiment varies across users; in Section 5 we repeat our analysis for different tenure terciles and similarly find no significant difference across users in the ads vs. no ads group within each tenure bucket.

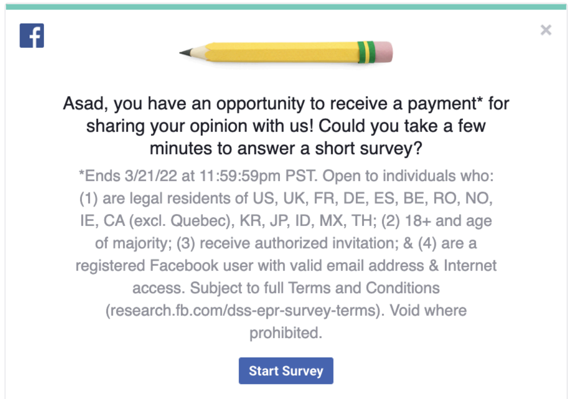


Figure 2: Sample recruitment prompt.

The text of these prompts was translated into the language the targeted user had set for their Facebook. For example, users who had selected their settings to use the Norwegian-language version of Facebook would see the prompt in Norwegian.

Importantly, as the prompt mentions, our survey was incentivized and users had the opportunity to earn real money in exchange for deactivating their Facebook account for a month. We include details on the flow and questions in Section 3 and Appendix A.3.

3. Valuation Methodology and Analysis Plan

As mentioned above, our high-level experimental design was to compare valuations for Facebook for users who see ads versus those who do not see ads in their newsfeed. In this section we go into more detail on how exactly we measure users' valuations and our analysis methodology.

3.A Valuation approach and incentive compatibility

Our key survey question solicits information on how much users would be willing to accept (WTA) to give up Facebook for one month. We use an incentive-compatible single binary discrete choice experiment (Carson et al. (2014); Brynjolfsson, Collis, and Eggers (2019)).

To recruit users, we sent on platform surveys out to a large sample of users. (Section 4 contains details on the sampling frame, response rates, and so forth.) If a user agreed to participate in the survey, they would be offered a specified amount of money and given a take it or leave it offer to stop using Facebook for one month. A mockup of the survey question in English is in Figure 3:

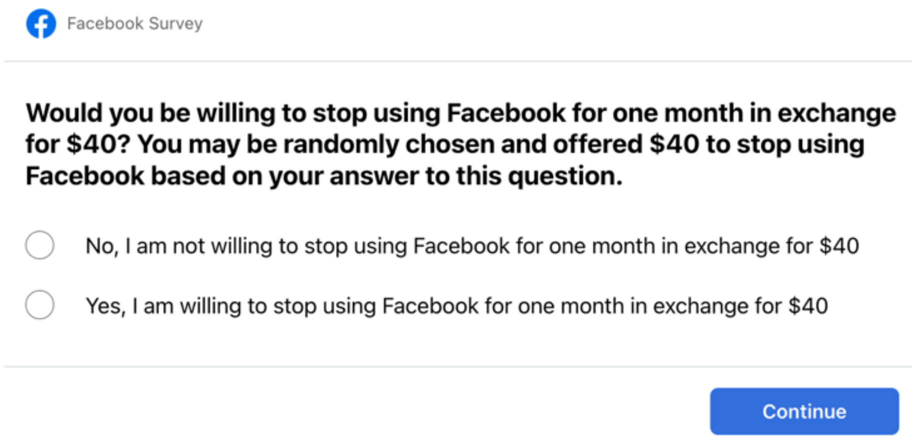


Figure 3: Main survey question with an example offer.

Offer values were randomly assigned across users, with uniform probability assigned to each of {$5, $10, $20, $30, $40, $50, $65, $80, $100} in USD. For countries that do not use USD, these dollar amounts were converted into the local currency and rounded to create a more intuitive user experience.[6]

Users were told that a random sample of them would be selected to respond with real money at stake. Specifically, if a user rejected the offer, they would never receive an offer, but amongst those who accepted the offer, a random fraction of them would be emailed later, offering the amount indicated in exchange for deactivation. This probability was not known to respondents as is best practice in the literature (e.g., Allcott et al., 2020).[7] Upon receiving their offer, deactivation was monitored on the backend by Meta; users would receive payment only if they did not reactivate their account for a month.

Similar to many other deactivation studies (e.g., Allcott et al. (2020)), we could not prevent users from either creating new accounts or reactivating before the month was up. Hence, the WTA we elicit is a lower bound on the true valuation of a user completely stopping Facebook usage for a month; rather, our elicited WTA measures users' valuation to attempt deactivation of their focal account for a month. We mention this as a caveat to the interpretation of our results. We note, however, that assuming this would affect valuations equally across users in the ads versus no ads group, it would only shift the baseline levels for the two groups as opposed to the difference, which is our primary interest.

---

[6] The amount of rounding depended on the level. Local currency amounts below 25 were rounded to the nearest integer; between 25 and 100 were rounded to the nearest 5; amounts between 100 and 500 were rounded to the nearest 25; between 500 and 1,000 to the nearest 50; between 1,000 and 5,000 to the nearest 100; between 5,000 and 10,000 to the nearest 500; between 10,000 and 100,000 to the nearest 5,000; and values about 100,000 to the nearest 20,000.

[7] This design is incentive compatible under the assumption that users know their WTA. If a user knows her valuation, the stated offer can either be higher or lower. If the offer is higher, then she should accept since there is a chance she will be offered the money and thus benefit; if the offer is lower, then she should reject as otherwise she may get the (slight) hassle of an irrelevant offer in the future. To the extent users may be uncertain about their WTA we must caveat our results.

Finally, after collecting each user's response, we asked additional questions related to basic demographics and other research projects at Meta; we provide more detail on the survey structure and payment flow in Appendix A.3.

3.B Analysis Plan

We are interested in estimating the median WTA for users in each of the ads and no ads conditions. Given the user-level response data, we start by running a weighted binary logit model to predict the probability that an individual rejects their randomly assigned offer (see below for details on the weights). This probability is a function of the log of the offer itself (*offer*, in USD), an indicator variable equal to 1 for users who see ads on Facebook (*ads*), and the interaction between these two variables.

$$Pr(Offer\ is\ rejected) = logit(\beta_0 + \beta_1 ln(offer) + \beta_2 ads + \beta_3 ln(offer) * ads) \tag{1}$$
$$= \frac{e^{\beta_0+\beta_1 ln(offer)+\beta_2 ads+\beta_3 ln(offer)*ads}}{1+e^{\beta_0+\beta_1 ln(offer)+\beta_2 ads+\beta_3 ln(offer)*ads}}$$

Given the parameters estimated by this model, to estimate the median WTA from this model, we calculate the value of *offer* at which the predicted probability of rejection is 0.5. For those users who receive ads, this calculation reduces to:

$$Median\ WTA\ in\ logistic\ model\ for\ users\ receiving\ ads = exp\left(\frac{-(\beta_0+\beta_2)}{\beta_1+\beta_3}\right)$$

Similarly, for users in the no ads group, the median WTA collapses to

$$Median\ WTA\ in\ logistic\ model\ for\ users\ not\ receiving\ ads = exp\left(\frac{-\beta_0}{\beta_1}\right)$$

Hence, the difference in median WTA between the two groups is:

$$Difference\ in\ median\ WTA = exp\left(\frac{-(\beta_0+\beta_2)}{\beta_1+\beta_3}\right) - exp\left(\frac{-\beta_0}{\beta_1}\right)$$

We estimate standard errors throughout by bootstrapping with replacement.[8]

4. Sample

[8] In some of our analyses we incorporate controls – in that case we demean each control variable using its weighted mean in the estimation sample. Since we use the estimated intercept of the model to calculate median WTA, this ensures the intercept continues to represent the average individual of the sample (rather than an individual with 0 for each included control).

Recruitment details. The survey was conducted from March 25 to April 7, 2022 in thirteen countries (US, Mexico, UK, France, Germany, Canada, Spain, Japan, Romania, South Korea, Belgium, Ireland, and Norway).[9]

Our sampling frame consisted of all monthly active users from those countries who were in the ads or no ads arm of the experiment. We further limited the sampling frame to users who were older than 18, not categorized as advertisers, and had accounts older than 30 days. We also oversampled users in both the ads and no ads groups in smaller countries to ensure sufficient sample size by as many geographies as we could.

Selected users were sent a prompt if they logged in to participate in a survey. If they agreed to be surveyed, we asked them whether they would be willing to participate in a deactivation study; if yes, we then asked them to consent to the Terms and Conditions in their local country. If the user consented, they were given the take it or leave it offer described earlier.[10]

In total, we began with 9,623,178 users across our thirteen countries who were sent survey invitations. Out of these, 1,227,191 users saw the invitation to participate in the survey, and 108,679 started the survey. A total of 63,790 users answered the terms and conditions question, with 57,179 agreeing to them. Finally, 53,166 users responded to the randomized offer question regarding their willingness to stop using Facebook for a month. Due to data logging issues with 83 respondents, our final sample size for the analysis is composed of 53,083 users. The completion rates through the survey funnel are similar to or better than those of other on platform surveys at Meta (e.g., Alekseev et al. (2023)).[11]

A small random sample of respondents were chosen to receive their actual offers. Ultimately, 381 users were emailed with their offer, 170 accepted, and 113 completed one month of deactivation and were paid.

There are several reasons why many of those we emailed may not have accepted and then why those who accepted did not complete the month of deactivation. On the former, Facebook notified selected users via their email, and that channel may have broken down for a variety of reasons (e.g., spam folder, stale address, user not reading). On the latter, going through the steps of deactivating one's Facebook profile involves a multi-step process that may have exceeded the anticipated costs. It may also be that some users underestimated their WTA when responding to the offers. We note this will not influence our estimate of the difference of medians across groups if it affected those with and without ads similarly, but is a caveat to consider in terms of our baseline estimations.

[9] These countries were selected based on legal constraints which determined where we could offer cash to respondents to deactivate Facebook as well as countries where we could obtain a sufficient number of completed survey responses in the treated group. Data from this survey were also used in Brynjolfsson et al. (2023).
[10] See Appendix A.3 for more detail on the survey flow.
[11] The response rate (survey completions from impressions) in the no ads group was 3.1% and it was 2.4% in the ads group, with the latter being significantly higher ($p < 0.01$). Both of these numbers, however, are within the normal range this kind of survey gets on platform.

Weighting. Following common survey science methods, we use inverse probability weighting models to make our estimates representative of the populations of Facebook users in each country we surveyed (see, e.g., Valliant, Dever, and Kreuter (2013)).

In our case, we employ a three-step approach that consists of design, user non-response, and question non-response weights. Specifically, these account for (i) differential probabilities of being selected to be eligible for the survey by country and experimental group, (ii) selection into starting the survey conditional on being eligible, and (iii) selection into answering the Facebook valuation question conditional on starting the survey. Our procedure follows Brynjolfsson et al. (2023) as well as the internal procedures Meta uses to analyze its own surveys.

The details of how we calculate each of these weights are provided in Appendix A.4. We note that our main results are consistent with or without weighting.

Sample Balance. As a check on the experiment randomization and our sample recruitment, we analyzed whether our unweighted sample was balanced across treatment and control for a large set of demographics and pre-treatment characteristics. The output is provided in Table A.1. We also analyzed our sample balance after weighting, and the output is provided in Table A.2.

For the unweighted sample, while we find evidence that many of the demographics are statistically significantly different, the magnitude of the differences is quite small. This is similar to past large-scale experiments on Meta where users have been recruited via survey (e.g., Wernerfelt et al (2024)). Across the 18 characteristics, the average difference between the two groups as a percent of the ads baseline is only 2.6%. Further, this is partially driven by the share of users who are 65+, who ultimately only represent 2.8% of our sample. (Removing that category, the average difference drops to 2.0%.) After weighting, only two characteristics remain statistically significantly different, and again with small differences in magnitude.[12] Collectively, we take this as evidence that our samples for the ads and no ads groups are comparable.

One concern in particular we want to address is differential attrition. For example, if the presence of ads drove some users off of Facebook, they would be less likely to be in our survey sample, and thus we could be missing the population of users who have the greatest disutility of ads. In our setting there are several pieces of evidence that point toward this not being a major issue.

First, in contrast to concerns that ads may be driving users off the platform, we find the average tenure in the ads group is *higher*. Second, the magnitudes of the tenure differences across the arms are also small. For example, the unweighted medians differ by 11 days, where the median tenure for each group is more than 11 years. Third, rerunning all our analyses controlling for tenure does not meaningfully change the results (Appendix A.2). Fourth, rerunning our analysis by tenure terciles (Figure 5b), we replicate our main result within each tenure tercile. Finally, even if we added enough high-tenure users to the no ads groups to equate the medians across our two conditions and we assumed that all of these users had the lowest (or highest) valuations of

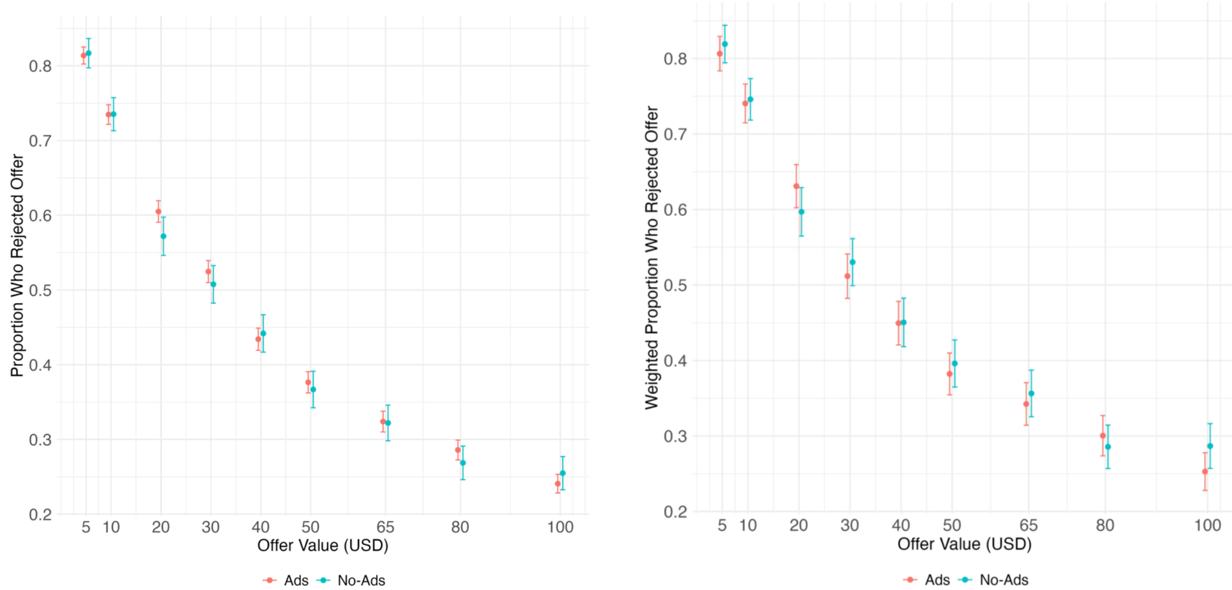
[12] Rerunning all our analyses controlling for these two variables yields no material changes to our results.

Facebook, the median valuations between the ads and no ads groups would still not be significantly different (p = 0.64, 0.27, respectively).[13]

5. Results

5.A Main results

We start by simply plotting the unweighted and weighted proportions of the sample who would reject a given offer value in Figure 4. We note that the demand curves are downward sloping for both the ads and the no ads groups, the confidence intervals overlap at every offer value, and there is no clear trend in the relative values of the ads vs. no ads point estimates across offer values. Further, the results are consistent both in the weighted and unweighted data. (Our results going forward all refer to the weighted data, but we show the unweighted results here for robustness.)



(a) (b)

Figure 4a, b: Unweighted (a) and weighted (b) proportions of sample that would have rejected our offer at each offer value.

[13] The p-values correspond to whether the extra individuals were assumed to have the lowest or highest valuations of Facebook. Given the no ads group has a higher estimated median valuation of Facebook, for differential attrition of that group to make our nonsignificant result less likely, the missing users would have to either have high enough valuations so that the median no ads WTA increases further, or low enough valuations so that it greatly decreases below the median ads WTA.

Focusing just on the median user, we estimate that the median WTA in the no ads group is \$31.95 and that of the ads group is \$31.04, with 95% confidence intervals of [\$30.26, \$33.63] and [\$29.53, \$32.56], respectively.[14] The difference in median WTA between the two groups is ninety cents and is statistically non-significant ($p = 0.43$). The 95% confidence interval for the difference (no ads - ads) is -\$3.06 to \$1.25.

The minimum detectable difference (MDE) in valuation between users who receive ads and users who do not is 9.95% (\$3.18/month): were there to be an actual difference in median WTA between users who receive and do not receive ads, that difference is likely to be less than this amount. To compute the ex-post MDE, we calculated the standard error of the difference in valuations across the ads and no ads groups via bootstrap and multiplied this standard error by 2.8 (e.g., Ioannidis, Stanley, and Doucouliagos (2017) and McKenzie and Ozier (2019)). This means that if the true difference in valuation across the two groups were higher than \$3.18, we would have been able to detect a statistically significant valuation difference across the groups with a higher than 80% probability.

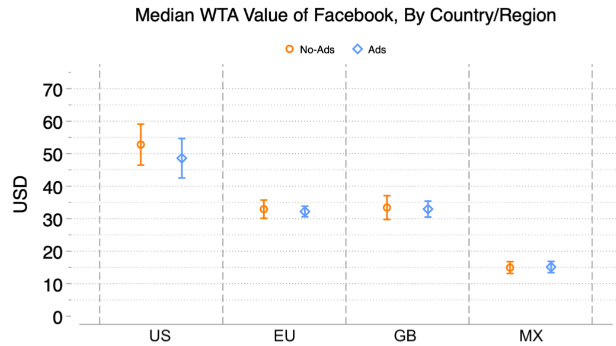
5.B Additional results

We report on two sets of analyses here. First, we analyze heterogeneity in our findings by region, tenure on the platform, and time spent on the platform. Second, we look at the effect of ads on time spent itself. We run the former analysis to explore the robustness of our WTA metric and our findings across different subgroups. We run the latter analysis since time spent is a common metric researchers and firms consider when evaluating user engagement; as a separate contribution of this paper, we wanted to highlight how it is affected by ad load on Meta and how it correlates with our consumer welfare metric.

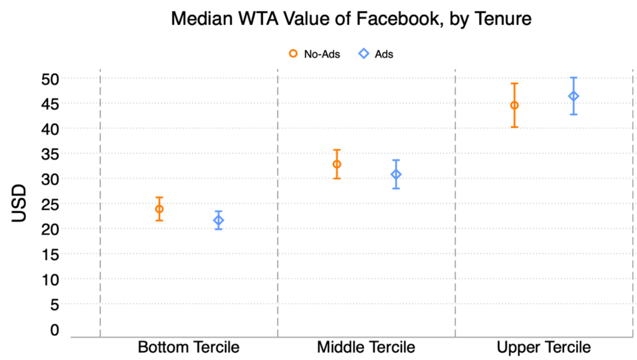
Heterogeneity. We explore heterogeneity based on region, tenure on the platform, and time spent on the platform (Figure 5). Our heterogeneity results broadly support our main finding in that (i) we observe no significant difference across median WTA in each region we surveyed, (ii) we similarly observe no significant difference within each tercile of tenure and time spent on the platform, and (iii) as we would expect, users who have invested more in the platform (as evidenced by tenure and time spent) have higher WTAs.[15] Hence, we view these heterogeneity results as supportive evidence of the quality of our measurement and the robustness of our main results. Results are provided in Figure 5, with exact numbers, as well as the output of the logistic regressions, in Appendix A.2.

---

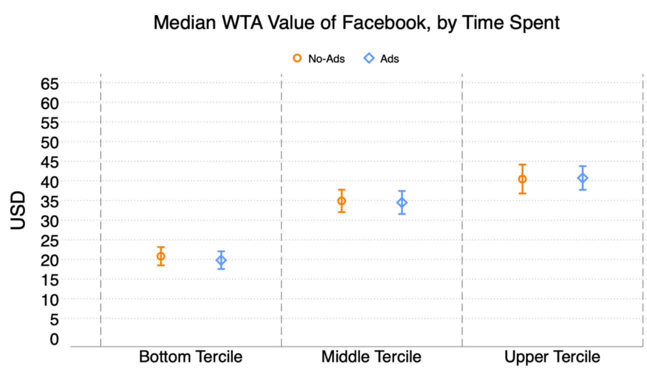
[14] As a comparison, publicly released Meta earnings data from Q1 2022 indicate that Facebook obtained average revenues of \$3.18 per user per month worldwide (\$16.1 in the US and Canada, and \$5.12 in Europe). These numbers are much smaller than the baseline valuations we estimate for both ads and no ads users.
[15] We note that the finding that the holdout itself reduces time spent and yet WTA increases across time terciles is not incompatible. Rather the time spent terciles represent substantial differences in time spent on the platform whereas the marginal change in time spent from the loss of ads is relatively small.



(a) Heterogeneity by region.



(b) Heterogeneity by tenure



(c) Heterogeneity by time spent

Figure 5a, b, c: Estimate of the median WTA for ads and no ads users by (a) region, (b) tenure, and (c) time spent. For (a) we reran our main analysis in each major region/country in our sample and within each country again see no significant difference in the median WTA across the ads and no ads groups (note: 'EU' refers to Germany, France, Spain, Romania, Belgium, Ireland, and Norway). We do not include Canada, Korea, or Japan, given small sample sizes in those countries, but we also find null results in each. For (b) and (c), we divided our sample into terciles by days since joining Facebook and time spent on the platform in the 7 days before the survey fielded (March 17-24, 2022). Users who have been on the site longer and who spend more time on the site value it more, as we would expect, whereas within each tercile we again observe no significant difference across groups. Point estimates and confidence intervals are in Table A.5.

Time spent. We also compare the relative time spent on Facebook for users in the ads vs. no ads group. We find the average time spent in the ads condition is 9.4% lower ($p < 0.001$) than for users in the no ads group. (Numbers are normalized for confidentiality.) We note that estimating how Meta ads affect time spent on the platform is a novel contribution to the literature itself.[16]

The combined null effect on WTA and the reduced effect on time spent highlights several potential mechanisms that could be at play with users' engagement with Facebook. For example, it could be that ads divert users away from Facebook to other sites; alternatively, ads could be crowding out more time-engaging content that would have appeared in their place; or, finally, it could also be that ads represent simply more recommended content for users and losing them reduces the interesting inventory for users. Our experimental design captures the net effect of removing ads – we cannot disentangle mechanisms behind the time spent reduction without additional experiments and analyses beyond our current scope. We leave that to future research.

Finally, we note that our WTA measure relies on the assumption that the stated preferences of users over our online survey capture their true preferences. In contrast, the time spent metric is logged and represents revealed preference behavior over engagement with Facebook versus the users' outside options. We mention this as a final caveat to our results - there are tradeoffs with these metrics, and some may feel more comfortable with one versus the other. We think our WTA measure is a better proxy here for true consumer welfare given the incentivized survey design and analysis, but we want to be upfront with the limitations.

6. Conclusion

There is a growing debate in the policy community around the relative societal costs and benefits of digital ads. While ads may benefit platforms and advertisers, critical questions in these debates are whether ads create disutility for consumers and if so how much? In theory, the effect on consumer welfare could go either way: for example, on one hand, ads may annoy users with irrelevant content or persuade them to buy products they do not need. On the other hand, ads may help users find valuable products and services, especially when the ads are well-targeted.

As shown in the present analysis, we find no evidence of disutility from ads on Facebook. In particular, we estimate a relatively tight null effect on the difference in consumer value derived from Facebook between users who receive targeted ads as usual and users who receive no ads. There are a number of mechanisms through which this effect could occur - for example, ads could be replaced by lower quality organic content, reductions in search costs from well-targeted ads and irritation from poorly targeted ads could cancel out, and so forth. Different mechanisms have different policy implications, and we leave a disentangling of the effects to future work.

[16] Note we are not measuring the utility consumers may derive from other activities they spend time on given the reduction in time on Facebook. Rather we focus specifically on the utility from Facebook in our analysis. A broader analysis of utility from the activities users substitute toward is outside our scope.

Relatedly, we see a number of promising paths for future work. First, for this study, we only vary the presence of ads: we do not vary the amount and kind of data used in the targeting or personalization of the ads at issue. Many policy debates today are over what kinds of data can or cannot be used for digital advertising - a robust understanding of this point would be very valuable. Second, apart from varying the level of personalization, little is known about how consumers respond to changes in ad load. This is important not just from the consumer side, but also for platforms and advertisers as they set ad frequency targets. Finally, given the number of regulations that are being considered or have already passed in this space, ex-post policy evaluations could help not only understand the general equilibrium effects but also what the tradeoffs are with different kinds of policies in this space. We believe the intersection of digital advertising and public policy will be a fruitful research area for years to come.

References

Alekseev, G., Amer, S., Gopal, M., Kuchler, T., Schneider, J. W., Stroebel, J., & Wernerfelt, N. (2023). The effects of COVID-19 on US small businesses: evidence from owners, managers, and employees. *Management Science*, *69*(1), 7-24.

Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, *110*(3), 629-676.

Athey, S., Catalini, C., & Tucker, C. (2017). *The digital privacy paradox: Small money, small costs, small talk* (No. w23488). National Bureau of Economic Research.

Athey, S., Grabarz, K., Luca, M., & Wernerfelt, N. (2023). Digital public health interventions at scale: The impact of social media advertising on beliefs and outcomes related to COVID vaccines. *Proceedings of the National Academy of Sciences*, *120*(5), e2208110120.

Bagwell, K. (2007). The economic analysis of advertising. *Handbook of Industrial Organization*, *3*, 1701-1844.

Becker, G. S., & Murphy, K. M. (1993). A simple theory of advertising as a good or bad. *The Quarterly Journal of Economics*, *108*(4), 941-964.

Bronnenberg, B., Dubé, J. P., & Joo, J. (2022). Millennials and the takeoff of craft brands: Preference formation in the US beer industry. *Marketing Science*, *41*(4), 710-732.

Brynjolfsson, E., Collis, A., & Eggers, F. (2019). Using massive online choice experiments to measure changes in well-being. *Proceedings of the National Academy of Sciences*, *116*(15), 7250-7255.

Brynjolfsson, E., Collis, A., Liaqat, A., Kutzman, D., Garro, H., Deisenroth, D., Wernerfelt, N., & Lee, J. J. (2023). *The digital welfare of nations: New measures of welfare gains and inequality* (No. w31670). National Bureau of Economic Research.

Bursztyn, L., Handel, B. R., Jimenez, R., & Roth, C. (2023). *When product markets become collective traps: The case of social media* (No. w31771). National Bureau of Economic Research

Carson, R. T., Groves, T., & List, J. A. (2014). Consequentiality: A theoretical and experimental exploration of a single binary choice. *Journal of the Association of Environmental and Resource Economists*, *1*(1/2), 171-207.

Cramer-Flood E (2021) Worldwide digital ad spending 2021. Accessed December 27, 2024, https://content-na1.emarketer.com/worldwide-digital-ad-spending-2021.

Deisenroth, D., Manjeer, U., Sohail, Z., Tadelis, S., & Wernerfelt, N. (2024). *Digital Advertising and Market Structure: Implications for Privacy Regulation* (No. w32726). National Bureau of Economic Research.

Facebook Ads Guide (2024). Update to Meta Ads Manager objectives. Accessed December 27, 2024, https://www.facebook.com/business/ads-guide/update/image

Galbraith, John Kenneth. *The affluent society*. Houghton Mifflin Harcourt, 1958.

Goldfarb, A., & Tucker, C. (2011). Online display advertising: Targeting and obtrusiveness. *Marketing Science*, *30*(3), 389-404.

Goldstein, D. G., Suri, S., McAfee, R. P., Ekstrand-Abueg, M., & Diaz, F. (2014). The economic and cognitive costs of annoying display advertisements. *Journal of Marketing Research*, *51*(6), 742-752.

Goli, A., Huang, J., Reiley, D., & Riabov, N. M. (2024). Measuring Consumer Sensitivity to Audio Advertising: A Long-Run Field Experiment on Pandora Internet Radio. *arXiv preprint arXiv:2412.05516*.

Gordon BR, Moakler R, Zettelmeyer F (2023) Close enough? A large-scale exploration of non-experimental approaches to advertising measurement. Marketing Sci. 42(4):768–793.

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011). *Survey methodology*. John Wiley & Sons.

Ioannidis, J. P., Stanley, T. D., & Doucouliagos, H. (2017). The power of bias in economics research.

Johnson G, Lewis RA, Nubbemeyer E (2017) The online display ad effectiveness funnel & carryover: Lessons from 432 field experiments. Preprint, submitted December 11, https://dx.doi.org/10.2139/ssrn.2701578

Korganbekova, M., & Zuber, C. (2023). Balancing user privacy and personalization. *Working Paper*.

McKenzie, D., & Ozier, O. (2019). Why ex-post power using estimated effect sizes is bad, but an ex-post MDE is not. *World Bank Development Impact Blog*.

Sahni, N. S., & Zhang, C. (2024). Are consumers averse to sponsored messages? The role of search advertising in information discovery. *Quantitative Marketing and Economics*, *22*(1), 63-114.

Stigler, G. J. (1961). The economics of information. *Journal of Political Economy*, *69*(3), 213-225.

Sunstein, C. R. (2020). Valuing facebook. *Behavioural Public Policy*, *4*(3), 370-381.

Valliant, R., Dever, J. A., & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples* (Vol. 1). New York: Springer.

Wernerfelt N, Tuchman A., Shapiro BT, and Moakler R. (2024). Estimating the value of offsite tracking data to advertisers: Evidence from Meta. *Marketing Science*.

Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power, edn. *Public Affairs, New York*.

Appendix

This Appendix contains four sections: first, we provide evidence on covariate balance across our treatment and control sample; second, we show the output behind our results in the main text and perform robustness checks; third, we provide additional details on the survey flow and execution; and finally, we discuss more detail on how we weighted our results.

A.1 Covariate Balance

Below we provide evidence on whether the experiment randomization and sample recruitment led to users in our treatment and control samples being comparable. We present two balance tables: Table A.1 contains the data from our raw, unweighted survey sample from users who were either in the ads or no ads condition; Table A.2 is the reweighted version of that.

We include 18 demographics and pre-determined characteristics in our data, showing the average difference between the two groups, significance level of the difference, the difference as a percent of the ads baseline (in absolute value), and the share missing. We present data in this format for confidentiality reasons.

As mentioned in the main text, we highlight the magnitude of the differences across the demographics is small for both the unweighted and weighted sample. Separately, we note that after reweighting, only female and Facebook tenure are significantly different, and the magnitudes of the differences are very similar (2.4% average difference as a percent of the ads baseline vs. 2.6% unweighted). Rerunning all our analyses controlling for both of these variables yields no material difference in the results.

Table A.1: Covariate balance table (53,083 survey respondents).

	Difference (Ads - No Ads)	p-value	Difference as fraction of Ads Baseline (abs. value)	% Missing
Age (years)	0.3768	0.003***	0.0104	0%
Age 18-24	-0.003	0.424	0.0176	0%
Age 25-34	-0.0096	0.046**	0.0268	0%
Age 35-44	0.0016	0.714	0.0064	0%
Age 45-54	0.0072	0.032**	0.0529	0%
Age 55-64	-0.0001	0.955	0.0022	0%
Age 65+	0.0039	0.014**	0.1340	0%
Female	-0.0282	<0.001***	0.0739	0%
Homeowner	-0.0007	0.906	0.0014	26%
Finished HS	-0.0319	<0.001***	0.0438	23%
Finished college	0.0232	<0.001***	0.0478	23%
Has profile photo	-0.0155	<0.001***	0.0162	0%
Primary Phone OS is iOS	0.0008	0.868	0.0020	0%
Primary Phone OS is Android	-0.0013	0.794	0.0021	0%
Facebook tenure (days)	72.9261	<0.001***	0.0196	0%
Contact email confirmed	0.0088	1.000	0.0095	0%
Day of week born	0.0553	0.536	0.0035	3%
Month born	-0.0298	0.395	0.0046	3%

Note: This table shows the unweighted comparison between the ads and no ads groups. All variables are binary except age (years), Facebook tenure (days), day of week born (1-7), and month born (1-12). The variables Homeowner, Finished HS (High School), and Finished college were self-reported in our survey, while the other variables were obtained from platform data. The first numeric column shows the raw difference between the averages in the two groups; the second numeric column shows the p-value on the difference; the third numeric column shows the raw difference as a fraction of the ads group baseline; and the last column shows the share of our data that were missing entries for that demographic. ** denotes $p < 0.05$, *** denotes $p < 0.01$.

Table A.2: Reweighted covariate balance table (53,083 survey respondents).

	Difference (Ads - No Ads)	*p*-value	Difference as fraction of Ads Baseline (abs. value)	% Missing
Age (Years)	-0.157	0.473	0.0042	0%
Age 18 -24	0.008	0.141	0.0497	0%
Age 25-34	-0.007	0.321	0.0211	0%
Age 35-44	-0.002	0.735	0.0091	0%
Age 45-54	-0.002	0.694	0.0118	0%
Age 55-64	0.007	0.127	0.0897	0%
Age 65+	-0.004	0.266	0.0976	0%
Female	0.039	<0.001***	0.0816	0%
Homeowner	-0.001	0.897	0.0020	29%
Finished HS	0	0.922	0.0000	23%
Finished College	-0.007	0.401	0.0158	23%
Has a profile picture	0	0.906	0.000	0%
Primary Phone OS is iOS	-0.006	0.447	0.0132	0%
Primary Phone OS is Android	0.004	0.567	0.0074	0%
Facebook tenure (days)	107.736	<0.001***	0.0307	0%
Contact email confirmed	0.003	0.531	0.0033	0%
Day of week born	0.024	0.414	0.0060	3%
Month born	-0.025	0.628	0.0038	3%

Note: This table shows the comparison between the ads and no ads groups, now weighted as discussed in the main text and Appendix A.4. *** denotes $p < 0.01$.

A.2 Main Output and Robustness Checks

In this section, we report the output from our main specification (Equation 1) for the total sample and each subgroup we consider (Table A.3). We also report the numbers behind Figure 5a, b, and c in the main text (Table A.5).

Finally, we redo all those analyses in Table A.4 and Table A.6 where we add controls for tenure on Facebook. We find no material differences in the results across the tables.

Table A.3: Logistic regression output for the whole sample and each subgroup. Dependent variable is a user-level rejection dummy.

	Total	US	EU	GB	MX	Bottom TS Tercile	Middle TS Tercile	Upper TS Tercile	Bottom Tenure Tercile	Middle Tenure Tercile	Upper Tenure Tercile
Offer	-0.863***	-0.870***	-0.880***	-1.091***	-0.984***	-0.914***	-0.923***	-0.824***	-0.823***	-0.884***	-0.940***
	(0.0278)	(0.0573)	(0.0483)	(0.0752)	(0.0554)	(0.0542)	(0.0488)	(0.0438)	(0.0451)	(0.0485)	(0.0510)
Control	0.0355	-0.261	0.0231	0.134	0.286	-0.290	0.0235	0.388*	0.109	-0.252	0.189
	(0.138)	(0.307)	(0.201)	(0.338)	(0.278)	(0.252)	(0.246)	(0.224)	(0.220)	(0.243)	(0.261)
Control X Offer	-0.0175	0.0478	-0.0123	-0.0439	-0.102	0.0806	-0.00952	-0.103*	-0.0624	0.0568	-0.0404
	(0.0382)	(0.0823)	(0.0550)	(0.0916)	(0.0804)	(0.0717)	(0.0678)	(0.0612)	(0.0619)	(0.0670)	(0.0709)
Constant	2.988***	3.454***	3.075***	3.834***	2.662***	2.780***	3.277***	3.050***	2.612***	3.084***	3.572***
	(0.1000)	(0.214)	(0.177)	(0.277)	(0.191)	(0.189)	(0.176)	(0.160)	(0.160)	(0.175)	(0.187)
N	53,083	5,378	24,960	5,473	6,099	17,690	17,691	17,691	17,678	17,690	17,709

Table A.4: Rerunning Table A.3 by including user-level control for tenure on Facebook.

	Total	US	EU	GB	MX	Bottom TS Tercile	Middle TS Tercile	Upper TS Tercile	Bottom Tenure Tercile	Middle Tenure Tercile	Upper Tenure Tercile
Offer	-0.874***	-0.873***	-0.881***	-1.095***	-0.995***	-0.926***	-0.928***	-0.840***	-0.823***	-0.892***	-0.941***
	(0.028)	(0.058)	(0.048)	(0.075)	(0.056)	(0.055)	(0.049)	(0.045)	(0.045)	(0.049)	(0.051)
Control	0.015	-0.262	0.019	0.138	0.298	-0.300	0.010	0.358	0.109	-0.274	0.187
	(0.139)	(0.308)	(0.201)	(0.338)	(0.278)	(0.253)	(0.247)	(0.226)	(0.220)	(0.243)	(0.261)
Control X Offer	-0.016	0.047	-0.011	-0.046	-0.104	0.081	-0.011	-0.100	-0.063	0.062	-0.040
	(0.038)	(0.082)	(0.055)	(0.092)	(0.081)	(0.072)	(0.068)	(0.062)	(0.062)	(0.067)	(0.071)
Constant	3.036***	3.467***	3.078***	3.848***	2.692***	2.831***	3.289***	3.123***	2.673***	2.825***	3.641***
	(0.101)	(0.215)	(0.177)	(0.278)	(0.191)	(0.191)	(0.176)	(0.163)	(0.166)	(0.184)	(0.231)
Tenure control	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	53,077	5,375	24,957	5,473	6,099	17,689	17,689	17,688	17,678	17,690	17,709

Table A.5: Numeric values behind Figure 5. (Note: TS refers to 'Time Spent'; sample sizes as per the columns in Table A.3)

	Total	**US**	**EU**	**GB**	**MX**
Ads	\$31.04 (\$29.53, \$32.56)	\$48.53 (\$42.76, \$54.30)	\$32.23 (\$30.62, \$33.83)	\$32.98 (\$30.53, \$35.42)	\$15.10 (\$13.47, \$16.74)
No Ads	\$31.95 (\$30.26, \$33.63)	\$52.9 (\$47.12, \$58.69)	\$32.95 (\$29.85, \$36.05)	\$33.57 (\$30.10, \$37.04)	\$14.95 (\$13.09, \$16.81)
Δ Ads - No Ads	-\$0.90 (-\$3.16, \$1.35) p = 0.43	-\$4.38 (-\$12.70, \$3.95) p = 0.30	-\$0.73 (-\$4.26, \$2.80) p = 0.69	-\$0.59 (-\$4.89, \$3.70) p = 0.79	\$0.15 (-\$2.35, \$2.66) p = 0.90

	Bottom Tenure Tercile	**Middle Tenure Tercile**	**Upper Tenure Tercile**
Ads	\$21.64 (\$19.76, \$23.51)	\$30.71 (\$28.00, \$33.41)	\$46.29 (\$42.62, \$49.96)
No Ads	\$23.92 (\$21.72, \$26.12)	\$32.78 (\$29.91, \$35.66)	\$44.65 (\$40.50, \$48.80)
Δ Ads - No Ads	-\$2.29 (-\$5.20, \$0.63) p = 0.12	-\$2.08 (-\$6.10, \$1.95) p = 0.31	\$1.64 (-\$3.95, \$7.23) p = 0.57

	Bottom TS Tercile	**Middle TS Tercile**	**Upper TS Tercile**
Ads	\$19.86 (\$17.73, \$21.99)	\$34.40 (\$31.52, \$37.28)	\$40.75 (\$37.78, \$43.71)
No Ads	\$20.96 (\$18.72, \$23.20)	\$34.78 (\$31.73, \$37.84)	\$40.47 (\$36.94, \$43.99)
Δ Ads - No Ads	-\$1.10 (-\$4.15, \$1.95) p = 0.48	-\$0.38 (-\$4.47, \$3.70) p = 0.85	\$0.28 (-\$4.39, \$4.95) p = 0.91

Table A.6: Numeric values behind Figure 5, adding control for tenure (sample sizes as per the columns in Table A.4).

	Total	**US**	**EU**	**GB**	**MX**
Ads	\$30.73 (\$29.20, \$32.25)	\$48.40 (\$42.88, \$53.93)	\$32.21 (\$30.79, \$33.64)	\$32.97 (\$30.48, \$35.47)	\$15.16 (\$13.44, \$16.89)
No Ads	\$32.20 (\$30.54, \$33.85)	\$53.14 (\$46.92, \$59.36)	\$32.96 (\$30.11, \$35.82)	\$33.61 (\$30.02, \$37.20)	\$14.95 (\$13.14, \$16.76)
Δ Ads - No Ads	-\$1.47 (-\$3.69, \$0.76) p = 0.197	-\$4.74 (-\$12.63, \$3.16) p = 0.239	-\$0.75 (-\$3.93, \$2.44) p = 0.646	-\$0.64 (-\$5.07, \$3.80) p = 0.778	\$0.22 (-\$2.27, \$2.70) p = 0.864

	Bottom Tenure Tercile	**Middle Tenure Tercile**	**Upper Tenure Tercile**
Ads	\$21.63 (\$19.89, \$23.38)	\$30.61 (\$27.76, \$33.46)	\$46.30 (\$42.38, \$50.22)
No Ads	\$23.92 (\$21.71, \$26.14)	\$32.85 (\$29.91, \$35.80)	\$44.63 (\$40.71, \$48.55)
Δ Ads - No Ads	-\$2.29 (-\$5.05, \$0.47) p = 0.104	-\$2.24 (-\$6.30, \$1.81) p = 0.278	\$1.67 (-\$3.81, \$7.15) p = 0.549

	Bottom TS Tercile	**Middle TS Tercile**	**Upper TS Tercile**
Ads	\$19.81 (\$17.81, \$21.80)	\$34.04 (\$31.29, \$36.78)	\$40.27 (\$37.05, \$43.49)
No Ads	\$21.10 (\$18.78, \$23.42)	\$35.16 (\$32.14, \$38.18)	\$40.77 (\$37.09, \$44.45)
Δ Ads - No Ads	-\$1.29 (-\$4.40, \$1.81) p = 0.414	-\$1.12 (-\$5.23, \$2.98) p = 0.592	-\$0.50 (-\$5.57, \$4.56) p = 0.846

A.3 Additional Survey Details

In this section we describe more details on the overall flow of the experiment on the user side.

For a user who was eligible to receive a survey and who logged in during our experimental window, the flow they would go through would be as follows: (i) they would be shown a prompt at the top of their newsfeed upon logging in asking them if they would like to participate in the survey; (ii) if they agreed, they would be asked a set of initial questions to determine if they would be willing to participate in a deactivation study; (iii) if yes, they would be asked our main valuation question for Facebook; (iv) after that, they were asked some additional questions related to other research projects; and (v) finally, if they were randomly selected, they would be emailed their offer and be able to receive compensation for following through on the deactivation.

Below, we provide more detail on (ii) and (iv) as the others were covered in more detail in the main text. We note that part of the data from this survey was also used in Brynjolfsson et al. (2023), and hence, the structure of the survey reflects multiple research projects.

Initial questions. If the user clicked ‘Start Survey,’ they would next be asked if they would be willing to participate in a deactivation study. Specifically, we asked: “Thinking about all the ways you use Facebook, would you be willing to stop using Facebook for one month if you were offered money in return?” Those who said “no” received a follow-up question asking why and then received a final opportunity to participate (“Would you like to learn more about the payment opportunity?”). Those who indicated they are willing to forego Facebook in exchange for money, or are at least willing to learn more in the follow-up question then proceeded to the Terms and Conditions.

The terms by which incentivized lotteries can be offered vary across countries. For example, some jurisdictions require different minimum ages or basic skills checks. After consenting and agreeing to the Terms and Conditions, the user was sent to the incentivized experiment.

Additional questions. After soliciting users’ answers to our incentivized experiment, we asked additional questions for other research projects and on basic demographics. These all occurred after we elicited users’ answers to our Facebook valuation question, and users were not told about any subsequent questions, so we believe it is very unlikely these questions could have affected the earlier responses.[17]

[17] After accepting or rejecting their first offer, we did ask users to accept or reject a separate offer. (For example, if a user rejected $30, they could be asked next for $50.) In the original pre-registration (AEARCTR-0008990) that was done for this survey, we described analyzing the data using a double-bound dichotomous choice design that would leverage the additional data from the second question. In practice, we found the data from the second offer produced worrying results – for example, the median valuations estimated using both offers was substantially higher than that estimated using the first offer, which, given the responses, is consistent with users’ anchoring on the value they were offered first. This anchoring bias with double-bound methods has also been reported in several other papers (e.g., Cameron & Quiggin, 1994; Herriges & Shogren, 1996; Bateman et al., 2001; DeShazo, 2002; Carson et al., 2003) and violates the fundamental requirement for procedural invariance. Consequently, we focus on the single-bound dichotomous choice results based on the first offer. We note, however, that our single-bound approach is incentive

A.4 Weighting Details

Our weighting strategy consists of three key building blocks: (1) design weights to account for differential probability of inclusion into the sample by country and experimental group, (2) user non-response weights to account for the probability of a user responding to the survey, (3) question non-response weights to account for the probability of a user who started the survey responding to the survey item in question. Below we go through how we calculated each of these for the respondents in our final dataset.

Design weights. For the design weights, the weight for user *i* from country *c* and experimental group g is given by

$$w_{cg}^{design} = \frac{\#\ monthly\ active\ Facebook\ users\ in\ country\ c\ and\ group\ g}{\#\ users\ from\ country\ c\ and\ group\ g\ that\ were\ eligible\ to\ receive\ the\ survey}$$

As the standard definition of design weights, they capture the number of units of our population of interest that each unit of our sample – those eligible to receive the survey – represents.

User non-response weights. Conditional on being eligible to receive a survey, only a subset of users actually started it. Our user non-response weights model the probability of starting the survey given a user was eligible to receive the survey as a function of observables. Specifically, we let the user non-response weight for user *i* in country *c* and experimental group *g* be:

$$w_{icg}^{user\ non-response} = 1 / \text{P}(i \text{ started the survey} \mid i \text{ was eligible to receive the survey})$$

where P(i started the survey) is estimated separately for each county and experimental group and is a function of gender, primary phone operating system, whether the user has a profile picture, age (quartile bins), friend count (quartile bins), the number of days within the last 28 days that the user was active, an indicator for whether the user was active for all days within the last 28 days, and days since confirmed (i.e., tenure, in quartile bins). Following Sarig, Galili, and Eilat (2023), we estimate these probabilities using a regularized logistic regression with LASSO.

Question non-response weights. Finally, to correct for question non-response, we similarly inverse probability weight by the estimated probability a user who started the survey answered the Facebook valuation question. For user *i* in country *c* and experimental group *g* this is defined as:

$$w_{icg}^{question\ non-response} = 1 / \text{P}(i \text{ answered the Facebook valuation question} \mid i \text{ started the survey})$$

Similar to the user non-response weights, we estimate these weights separately for each country and experimental group and use the same logistic specification and set of right hand side variables. We include on the right hand side, however, an indicator for if they had answered they

compatible since respondents were not told they would receive a second offer and analyzing the data using the double-bound approach also yields a tight null for the median valuations across our two groups.

would be willing to stop using Facebook for one month if they were offered money in return (an earlier question on the survey).

Finally, for estimates at the country-level, we multiply user and question non-response weights to obtain weights for each respondent, whereas for estimates where we pool data across countries, we multiply the three weights to obtain weights for each respondent.

References

Bateman, Ian J., Richard T. Carson, Brett H. Day, W. Michael Hanemann, Nick Hanley, Tannis Hett, Michael Jones-Lee, et al., Economic Valuation with Stated Preference Techniques: A Manual (Northampton, MA: Edward Elgar, 2002).

Cameron, Trudy Ann, and John Quiggin, "Estimation Using Contingent Valuation Data from a Dichotomous Choice with Follow-Up Questionnaire," Journal of Environmental Economics and Management 27:3 (1994), 218–234.

Carson, Richard T., Robert C. Mitchell, W. Michael Hanemann, Ray J. Kopp, Stanley Presser, and Paul A. Ruud, "Contingent Valuation and Lost Passive Use: Damages from the Exxon Valdez Oil Spill," Environmental and Resource Economics 25:3 (2003), 257–286.

DeShazo, Jay R., "Designing Transactions without Framing Effects in Iterative Question Formats," Journal of Environmental Economics and Management 43:3 (2002), 360–385.

Herriges, Joseph, and Jay Shogren, "Starting Point Bias in Dichotomous Choice Valuation with Follow-Up Questioning," Journal of Environmental Economics and Management 30:1 (1996), 112–131.

Sarig, T., Galili, T., & Eilat, R. (2023). balance--a Python package for balancing biased data samples. *arXiv preprint arXiv:2307.06024*.