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

Erik Brynjolfsson
Stanford University and NBER

Avinash Collis
Carnegie Mellon University

Asad Liaqat
Meta

Daley Kutzman
Meta

Haritz Garro
Meta

Daniel Deisenroth
Meta

Nils Wernerfelt
Northwestern University

Abstract

We analyze a long-running field experiment of online advertising on Facebook in which a random 0.5% subset of all users are assigned to a group that does not ever see ever ads. We recruit a representative sample of Facebook users in the ads and no-ads groups and estimate their welfare gains from using Facebook using a series of incentive-compatible choice experiments. We find no significant differences in welfare gains from Facebook. Our findings suggest that either the disutility of ads for consumers is relatively small or that there are offsetting benefits, such as helping consumers find products of interest.

Corresponding authors listed alphabetically. Email: erikb@stanford.edu, avinashcollis@cmu.edu

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 though he no longer holds a financial interest in the company. E.B., and A.C. declare no competing interests, although they have in the past been awarded unrestricted gifts from Meta.

This RCT was registered as AEARCTR-0008990.

Introduction

Advertisements may create utility or disutility for users depending on the context. Marotta et al. (2021) argue that advertising can reduce the welfare of consumers, particularly when it is poorly targeted, by exposing them to irrelevant content. This is a costly use of their time, and they would prefer to avoid such ads (Goldfarb and Tucker, 2011). In contrast, Stigler (1961) argues that ads match buyers and sellers and reduce search

costs, thereby improving the utility of users. Here, advertising is seen as playing an informational role.

In the context of modern digital platforms with targeted advertising, it is unclear which of these forces plays a stronger role and whether ads create utility or disutility on average. The causal effects of online advertising on consumer welfare effects is ultimately an empirical question, but one with little clear evidence to date.

There is also a broader literature in economics, marketing, and information systems studying the impact of online advertising on consumer behavior and advertiser, publisher, and platform revenues (Evans (2009), Goldfarb (2014)). Kaldor (1950) argues that advertising may play a persuasive role and is anti-competitive. Therefore, it may increase industry concentration and prices for consumers. Many studies have looked at the effectiveness of online advertising for advertisers by conducting field experiments (e.g. Lewis and Rao (2015), Tadelis et al. (2023)). Several recent studies have also focused on the value of user data and cookies to publishers by studying how ad prices and publishers' revenues change when policy changes make it harder for platforms to better target users for online advertising (see Johnson (2022) for a literature review and Wernerfelt et al. (2024) for a recent field experiment on Facebook).

In contrast, research on measuring the welfare effects of online ads on the consumer side is limited due to the challenge of running field experiments where online ads are experimentally varied across users. Furthermore, short-run effects may differ from long-run effects. The goal of this paper is to estimate the long-run welfare effects of online ads. Prior studies looking at the welfare effects of advertising on the consumer side have looked at the cognitive costs of annoying and obtrusive display ads (Goldfarb and Tucker (2011), Goldstein et al. (2014)). Sahni and Zhang (2024) analyzed a field experiment on a search engine where treated users were assigned to a lower level of ad load (by 17%). They found that users assigned to the version of the search engine that showed them more ads expressed higher satisfaction with the search engine compared with users who got the version with fewer ads. In the context of audio advertising, Huang et al. (2018) conducted a field experiment on Pandora where treated users were exposed to different levels of audio ads over a period of 21 months. They find that the relationship between ad load and the number of hours spent on the platform is negative and linear and that increasing ad load causes users to switch to paid subscriptions. Moreover, they find that measuring the long-term effect is crucial since the ad load sensitivity for the period they estimate is three times as much as what they would have estimated with a month-long experiment. This provides further motivation for our work, where we analyze a very long-term experiment lasting over nine years.

In our study, we analyze data from a long-running field experiment on Facebook, launched in 2013 and still ongoing, in which 0.5% of the entire user base is assigned to a no-ads group while the rest of the user base experiences their usual levels of advertising. We believe this is the largest field experiment of its kind involving all of Facebook's user base (2.9 billion users at the time our study was conducted in 2022). We recruit a representative sample of 53,166 users from the ads and no-ads groups and invite them to take part in a survey. The survey includes an incentivized online choice experiment where we solicit users' valuations of Facebook.

We find that users' valuations of Facebook are not statistically different from each other in the treatment and control groups, both in the overall sample and in individual geographies, indicating that users do not experience a significant disutility from current levels of advertising. Our analyses further indicate that the minimum detectable effect is $3.18/month: i.e. if the true valuation difference across the groups were differ by more than $3.18/month, we would be able to detect a statistically significant difference with a probability higher than 80%.

Our study contributes to the current regulatory and policy debates around online advertising. Many regulations have been proposed or passed in recent years in the US, Europe, and elsewhere (e.g., GDPR in the EU, CCPA in California), and several of these regulations focus on the ability of online platforms to show targeted ads to consumers. A full welfare analysis of the value of online ads would analyze the benefits and costs to users, publishers, platforms (advertising intermediaries), and advertisers. An important underlooked area here is the long-run utility effects of online ads to users. Our study provides the empirical evidence needed to help quantify the potential disutility of online ads in dollars.

Study

Digital platforms including Facebook routinely conduct A/B tests to measure the causal impact of various design changes and policies on outcomes of interest. Many of these platforms have a holdout group where users are not exposed to any of the experiments. This is to measure the cumulative impact of multiple experiments or other changes over the longer term (Kohavi et al. (2020)). For advertising-based platforms such as Facebook, this holdout group usually consists of a randomly selected group of users who never see any ads.

Since 2013, Facebook has maintained an ongoing holdout group of users (0.5 percent of the total at the time in which we ran this study) who receive no ads except in a few

surfaces with low ad loads[1]; this is to enable the company to monitor platform functionality and ensure that systems are running as expected. Users are randomly assigned to this holdout group when they first sign up for a new Facebook account. As a result, users in this group have *never* been exposed to ads. The random distribution of users includes a small percentage (but large numbers) of users in each country and demographic group.

We leverage this holdout group in our current study. We consider users in the no ads group as our control group and the remainder of the users as our treated group. To measure the utility or disutility of online ads, we can look at the differences in users' valuation of Facebook across these two groups. To measure user valuations, we apply the previous literature measuring valuations of Facebook and digital goods more broadly using incentive-compatible online choice experiments (Brynjolfsson et al. (2019), Allcott et al. (2020), Brynjolfsson et al. (2023)).

More specifically, we select a representative sample of users in control and treatment groups and show them an invitation to take part in an online survey at the top of their news feed. The survey sample for each component was weighted to be representative of the population of "monthly active users" of Facebook in each country (i.e., users who have been active on Facebook within the last 30 days) for both the treatment group (ads) and the control group (no ads). Using a sample that is representative of Facebook users in each country is a stark improvement over existing estimates of valuations of Facebook and other digital platforms that were based on laboratory experiments or off-platform surveys of the general public.

A key question in the survey solicits their valuation for Facebook using an incentive-compatible single binary discrete choice experiment, as used in Brynjolfsson et al. (2019). We ask users if they would continue to keep access to Facebook or give it up for one month in exchange for a certain amount of cash. Users were asked: "Would you be willing to stop using Facebook for one month in exchange for X?", where X was chosen randomly from a set of 9 monetary values from $5 to $100 in the US (and equivalent amounts converted to local currencies in other countries in our sample). A given user is only shown a single valuation question ("take it or leave it" offer). Users are told that they could be randomly selected for their choices to be fulfilled, and if so, they would actually be eligible to receive the offer amount if they deactivated their Facebook account for a month. Figure 1 shows screenshots from the survey.

[1] Users in the holdout may see video ads or display ads on the web.

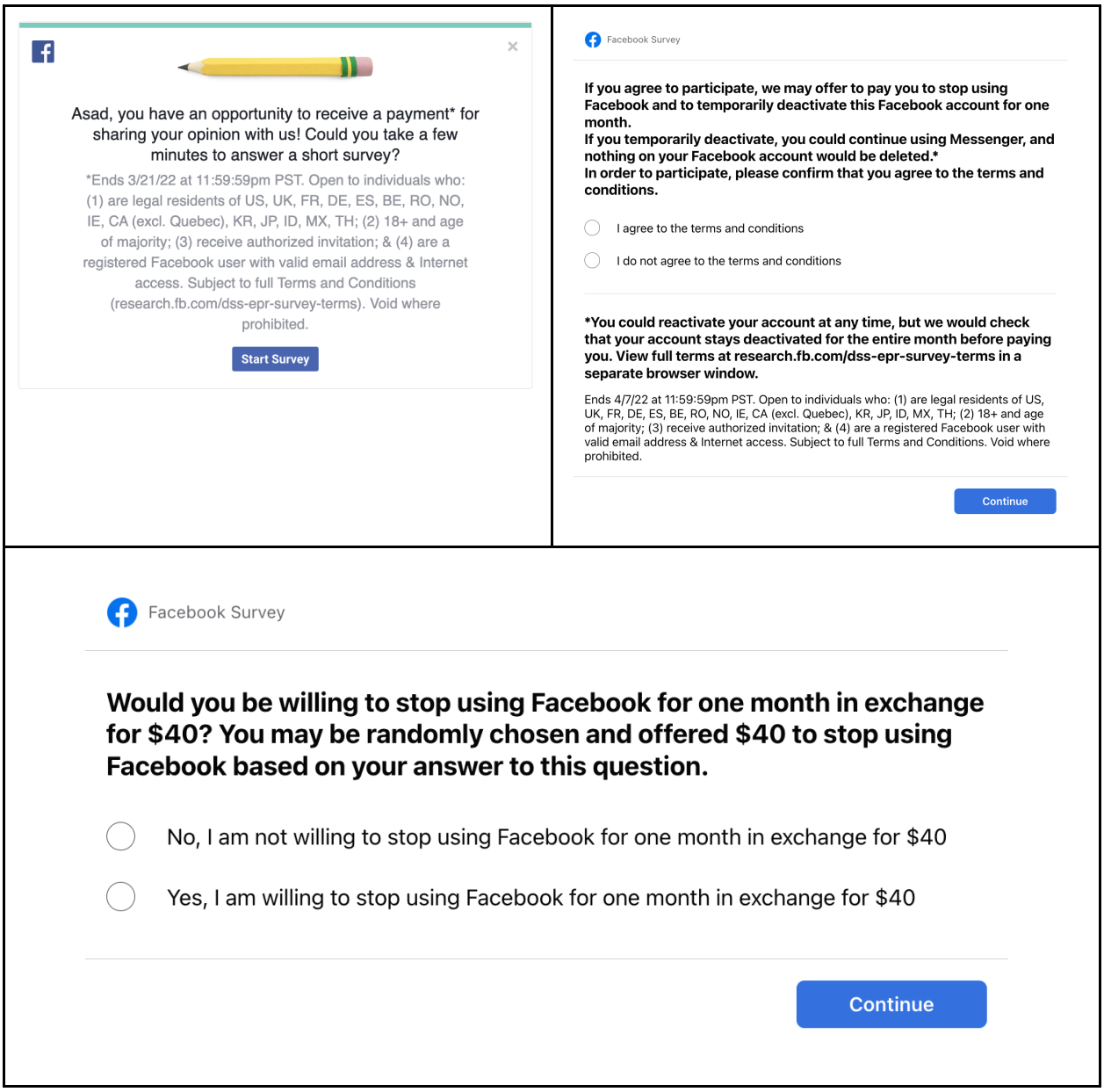


Figure 1: Screenshots from the survey fielded to users in the treatment (ads) and control (no-ads) groups

We recruited users from 13 countries for our survey: US, Mexico, UK, France, Germany, Canada, Spain, Japan, Romania, South Korea, Belgium, Ireland, and Norway. 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. Respondents were given offers in their own currency. For instance, if a respondent in France was chosen to receive an offer equivalent to US $50, they were given an offer of

45 Euros, which was approximately equivalent to US $50 at the time the survey was conducted. These offers were incentivized, and participants had to agree to a set of terms and conditions drafted by legal experts to be in compliance with local laws in each country. Each choice made by respondents of whether to accept or reject an offer had a 2% probability of being selected for fulfilment. This probability was not known to respondents as is best practice in the literature (Allcott et al. (2020)).

The survey was conducted from March 25 to April 7, 2022. We obtain complete responses from 39,717 users in the treatment (ads) group and 13,449 users in the control (no-ads) group. After applying weights, these users are representative of the overall control and treatment groups in the 13 selected countries. Immediately after the survey concluded, 2% of sampled users (381) were randomly selected and offered their choice. 170 users preferred their cash option over keeping access to Facebook and were offered cash for deactivating their Facebook accounts. 113 users successfully stopped using Facebook and were paid cash totaling over $14,400. There could be several reasons why not everyone complied with the deactivation. Facebook notified selected users via their email and some users may not have checked their email, or it may have gone to their spam folder. For users in our study who were chosen for deactivation, we were able to observe activity on Facebook once the study was completed. We find a small 0.2 percentage point difference comparing the percentage of monthly active users who deactivated their accounts versus all remaining survey respondents. This suggests that our consumer welfare estimates reflect consumers' rational choices and are aligned with their revealed preferences and not simply stated preferences.

Note that users are not aware of the fact that they may be in the ads or no-ads group. Moreover, users in these groups experience their assigned conditions over the entire duration of their tenure on Facebook. Therefore, any differences in user valuations should reflect the utility or disutility of ads in the long term.

Results

Appendix Table 1 presents median willingness to accept valuations for giving up Facebook for 1 month for different subgroups in our sample. We first investigate differences in time spent on our platform between users who receive ads (treatment) and no ads users (control). As shown in Figure 2 below, average time spent in the week preceding the survey (March 17-24 2022) for surveyed users in the ads group is 9.4 percent lower ($p<0.001$) than for users in the no ads group. This difference could be attributable to ads diverting user attention away from the Facebook platform. That is, time spent on Facebook is likely not a good proxy for the impact of ads on utility / disutility.

Brynjolfsson et al. (2023) find that time spent and consumer welfare measures are only weakly correlated.

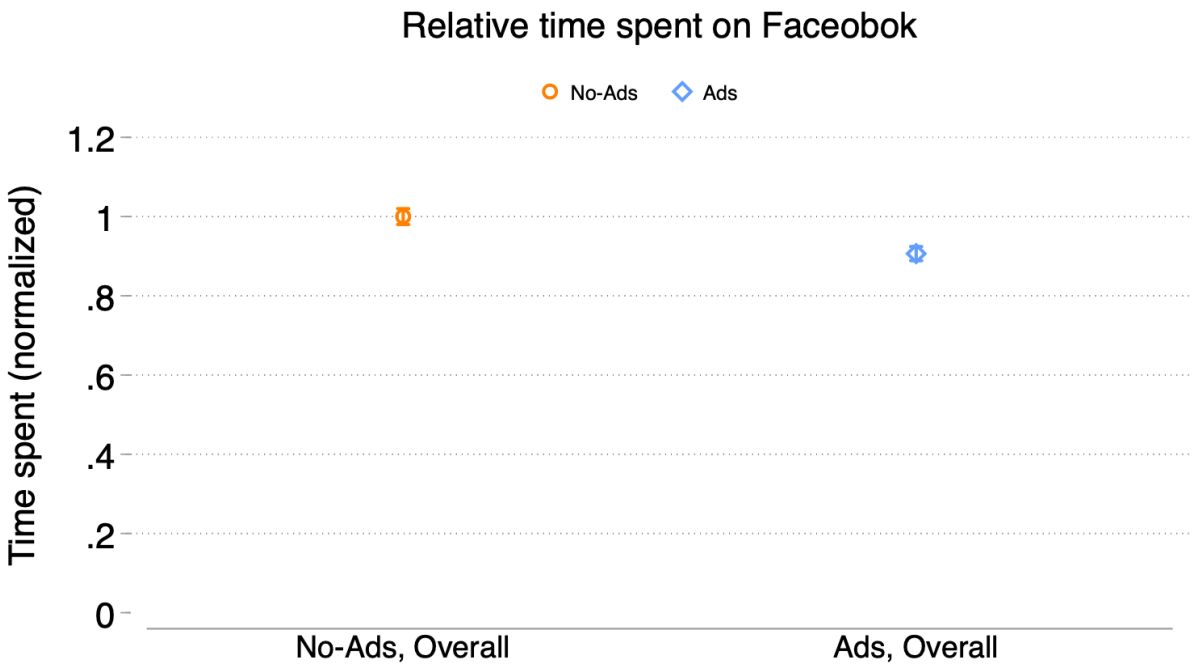


Figure 2: Average time spent on Facebook by users in the no-ads holdout vs. regular users in the week leading up to the WTA measurement. The values have been rescaled such that the no-ads mean time spent is 1. 95% CIs also presented.

Next, we explore differences in estimated willingness to accept to stop using Facebook for 1 month (WTA) between the two surveyed groups. This is a better metric for utility/disutility impact of ads, because it compares users' all-in valuation of access to the Facebook platform between groups where the only difference is that one receives ads and one does not (Brynjolfsson et al. (2019)). Crucially, we do *not* directly "prompt" people to think about ads per se and instead focus on their (incentivized) valuations absent any specific prompting that could bias their responses. Based on a standard logit regression specification for contingent valuation, Figure 3 below demonstrates that the median WTA in the no-ads (control) and ads (treatment) groups is $31.95 and $31.04, respectively. The difference in median WTA across treatment and control is statistically indistinguishable from zero, with the 95% confidence interval of the difference ranging from -$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. 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.[2]

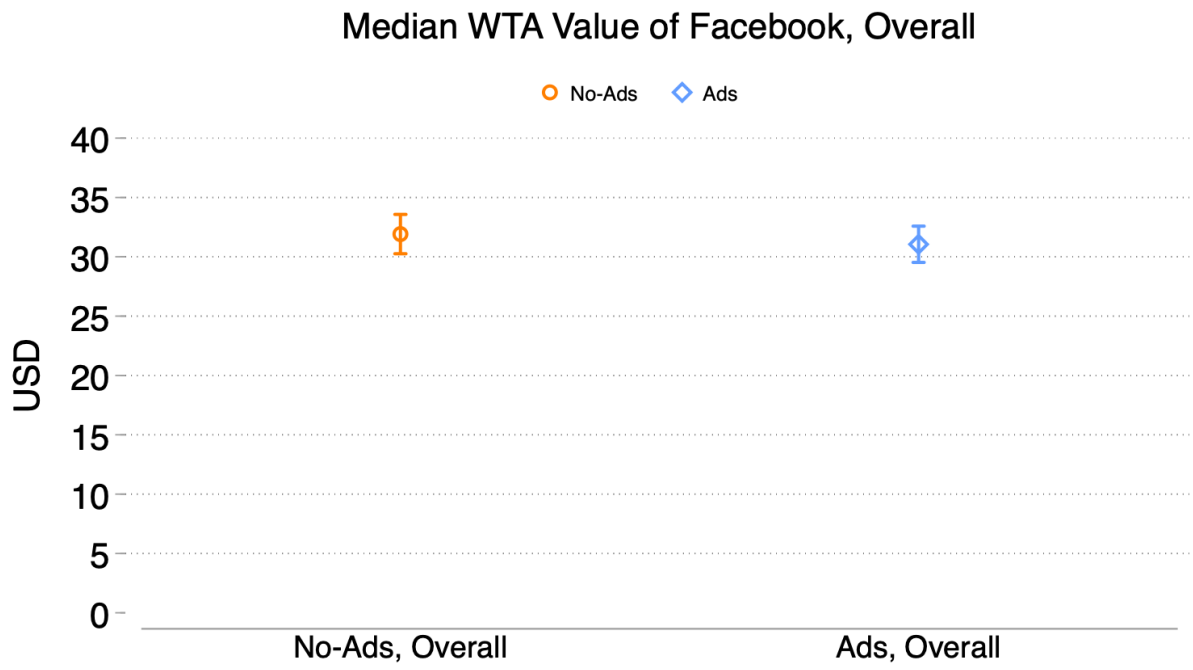


Figure 3: Overall Median WTA estimates for no-ads holdout vs. regular users

We also test whether time spent on Facebook influences the relative utility and/or disutility of ads. Figure 4 below provides the median WTA point estimates and associated confidence intervals for users in the lowest, middle, or highest tercile of time spent on Facebook, respectively. In both the ads and no-ads groups, median WTA is higher among users who spend more time on the platform. We detect no differences in median WTA across the ads and no-ads groups in any time-spent tercile. As shown in Appendix Figure 1, for each of the three time-spent terciles, the 95% confidence interval of the difference in median WTA includes zero.

[2] See more on ex-post MDE in Ioannidis, Stanley and Doucouliagos (2017) and here: https://blogs.worldbank.org/en/impactevaluations/why-ex-post-power-using-estimated-effect-sizes-bad-ex-post-mde-not

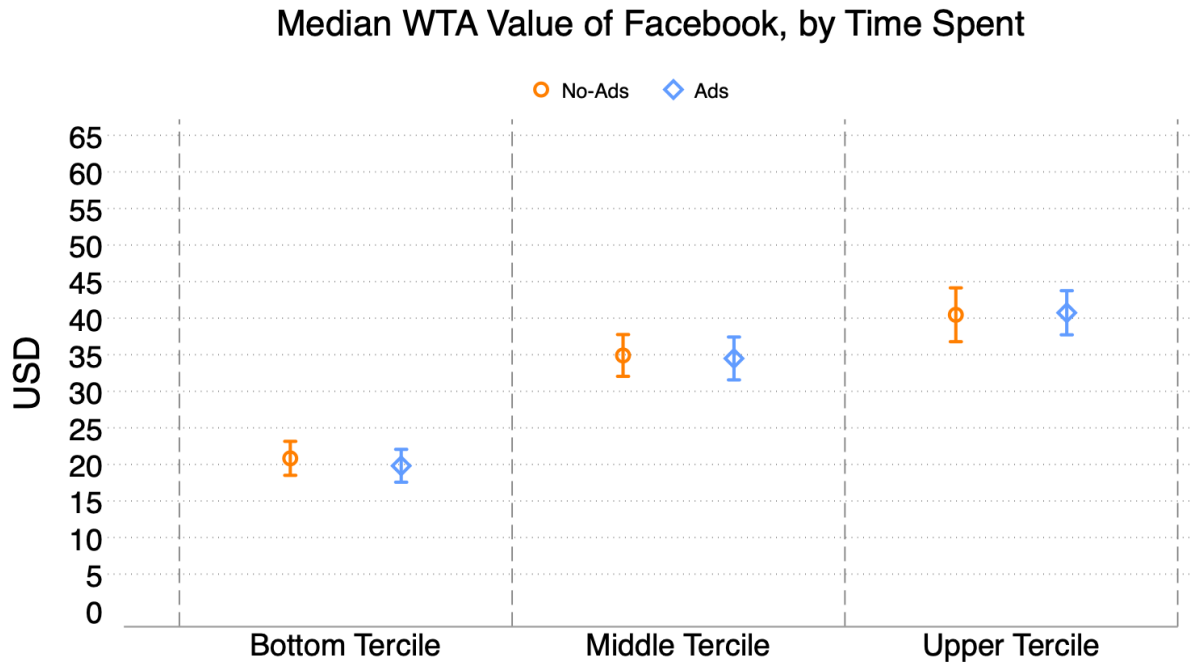


Figure 4: Median WTA estimates for no-ads holdout vs. regular users, by terciles of time spent on the platform

Figure 5 below shows the results for users with varying lengths of tenure on the Facebook platform. Median WTA is higher among users with higher tenure. Crucially, we fail to detect any differences in median WTA across the ads and no-ads groups in any of the tenure terciles. As shown in Appendix Figure 1, the 95% confidence interval of the difference in median WTA includes zero, for each of the tenure terciles.

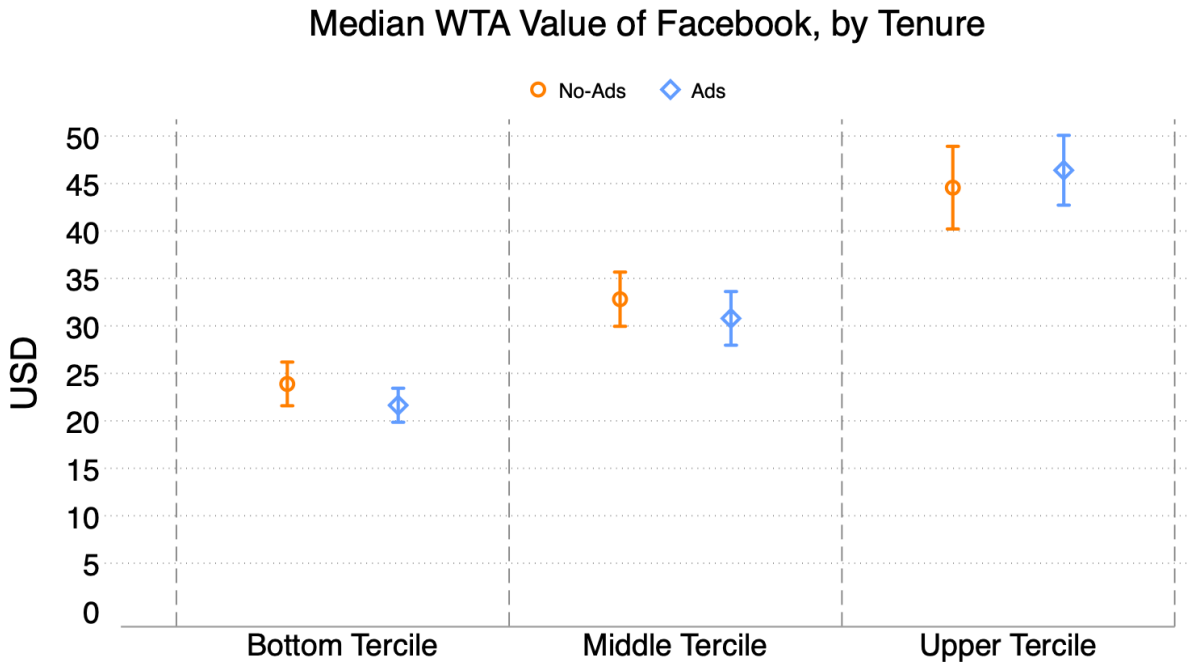


Figure 5: Median WTA estimates for no-ads holdout vs. regular users, by terciles of how long users have been on the platform

Finally, we investigate whether there are detectable differences in utility/disutility from ads within individual countries or regions (Appendix Figure 1). Here, it should be acknowledged that sample sizes are necessarily smaller and thus the differences we are powered to detect are necessarily larger; we thus aggregate certain regions to facilitate tractable comparisons. The results are nonetheless the same: we compare median WTA between users who receive ads and users who don't in the US, the EEA (7 countries in our sample), Great Britain and Mexico. In any of these regions, we do not detect any difference in median WTA, as shown in Appendix Figure 1.

Conclusion

There is a growing debate in the policy community around the relative societal costs and benefits borne by digital ads. While ads may benefit platforms and advertisers, a critical question in these debates is whether ads create disutility for consumers. In theory, the effect on consumer welfare could go either way: on one hand, ads may annoy users with irrelevant content. On the other hand, they may help them find useful products and services, especially when the ads are well-targeted.

As shown in the present analysis, we find no evidence of disutilty from advertising. In particular, we find no statistically significant difference in consumer value derived from

Facebook between users who receive targeted ads as usual and users who receive no ads. This could be due to the beneficial effects of some well-targeted ads canceling out the annoyance of irrelevant ads. Or perhaps both effects are simply relatively small. In either case, our findings suggest that a no-ads environment may not be particularly beneficial for consumers.

This study also represents an important contribution to the academic literature. Whereas previous studies leveraged off-platform survey experiments, the present analysis studies users who *never* received ads, including users with longer tenure on the Facebook platform. In other words, our study provides the first ecologically-valid estimate of the utility/disutility of ads in dollar terms.

There are several important limitations to this analysis that must be noted. First, the study only evaluates the utility/disutility of ads on Facebook—we do not explore the impact on other platforms such as Google, X, TikTok, etc. Second, within Facebook, the no ads holdout does not apply to all ads on all surfaces. It covers the wide majority of ads a user would ever see on the platform, but there are small set of other surfaces such as video ads or display ads on the web where a "no ads" user may still see a few of ads on the platform.

Finally, there are two dimensions of ads that this study is unable to investigate. The first is the amount of ads—control group users receive almost no ads, while treatment group users receive the normal amount of production ads, and there are no intermediate treatments, nor are their treatments with much higher levels of ads. Thus our results should not be assumed to apply to all possible levels of ad load. The second is that, for this study, we do not vary the amount of data used in the targeting or personalization of the ads at issue. In theory, this can make a significant difference to the utility or disutility of ads. This is an important topic of study as limitations in this dimension may represent the frontier for putative policy changes under consideration in this domain.

References

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

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., ... & Lee, J. J. (2023). *The digital welfare of nations: New measures of welfare gains and inequality* (No. w31670). National Bureau of Economic Research.

Evans, D. S. (2009). The online advertising industry: Economics, evolution, and privacy. *Journal of economic perspectives*, *23*(3), 37-60.

Goldfarb, A. (2014). What is different about online advertising?. *Review of Industrial Organization*, *44*, 115-129.

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.

Huang, J., Reiley, D., & Riabov, N. (2018). Measuring consumer sensitivity to audio advertising: A field experiment on pandora internet radio. *Available at SSRN 3166676*.

Ioannidis, J.P.A., Stanley, T.D., & Doucouliagos, H. (2017). The power of bias in economics research. *Economic Journal, 605(127), 236-265*

Johnson, G.A. (2022). Comment: FTC APNR on “Commercial Surveillance and Data Security”. Available at https://downloads.regulations.gov/FTC-2022-0053-0680/attachment_1.pdf.

Kaldor, N. (1950). The economic aspects of advertising. *The review of economic studies*, *18*(1), 1-27.

Kohavi, R., Tang, D., & Xu, Y. (2020). *Trustworthy online controlled experiments: A practical guide to a/b testing*. Cambridge University Press.

Lewis, R. A., & Rao, J. M. (2015). The unfavorable economics of measuring the returns to advertising. *The Quarterly Journal of Economics*, *130*(4), 1941-1973.

Marotta, V., Wu, Y., Zhang, K., & Acquisti, A. (2022). The welfare impact of targeted advertising technologies. *Information Systems Research*, *33*(1), 131-151.

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.

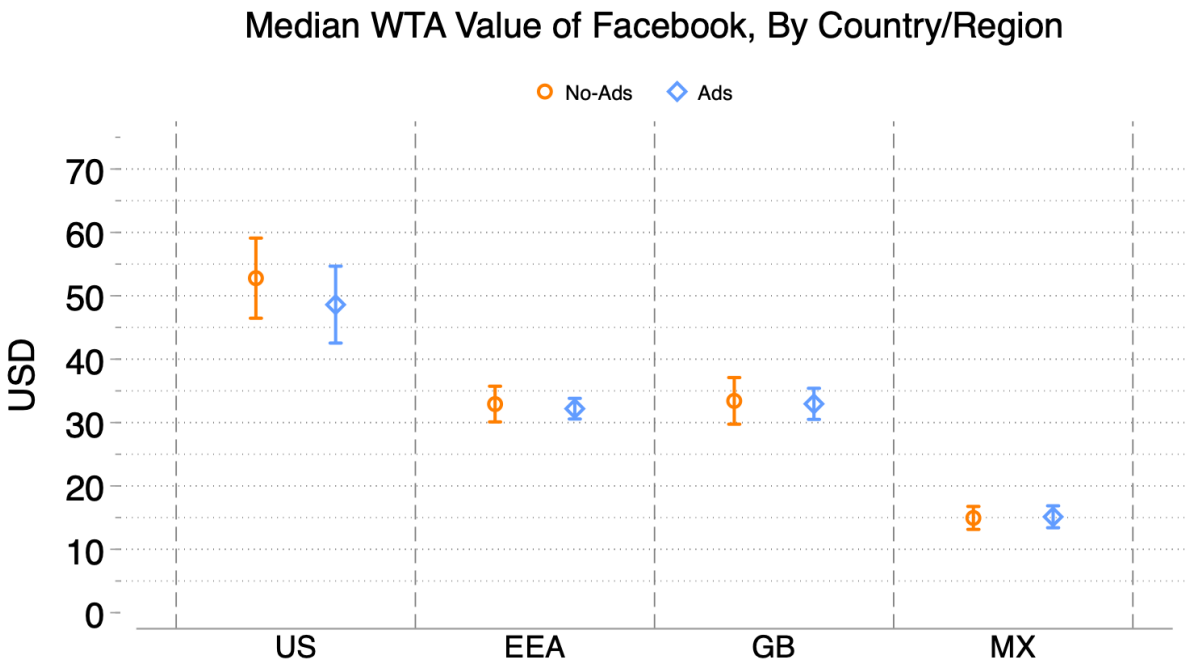
Tadelis, S., Hooton, C., Manjeer, U., Deisenroth, D., Wernerfelt, N., Dadson, N., & Greenbaum, L. (2023). *Learning, sophistication, and the returns to advertising: Implications for differences in firm performance* (No. w31201). National Bureau of Economic Research.

Appendix

Sample	Statistic	Estimate	Lower bound of 95% CI	Upper Bound of 95% CI
Overall	No-Ads (C)	31.95	30.36	33.53
	Ads (T)	31.04	29.55	32.54
	Delta (T-C)	-0.90	-3.06	1.25
By Country/Region				
US	No-Ads (C)	52.90	47.08	58.73
	Ads (T)	48.53	42.89	54.16
	Delta (T-C)	-4.38	-12.82	4.07
EEA	No-Ads (C)	32.95	29.94	35.97
	Ads (T)	32.23	30.72	33.73
	Delta (T-C)	-0.73	-4.06	2.60
GB	No-Ads (C)	33.57	29.91	37.23
	Ads (T)	32.98	30.74	35.22
	Delta (T-C)	-0.59	-4.85	3.66
MX	No-Ads (C)	14.95	13.29	16.61
	Ads (T)	15.10	13.42	16.79
	Delta (T-C)	0.15	-2.26	2.57
By Tenure				
Bottom Tenure Tercile	No-Ads (C)	23.92	21.66	26.18
	Ads (T)	21.64	19.79	23.48
	Delta (T-C)	-2.29	-5.25	0.68
Middle Tenure Tercile	No-Ads (C)	32.78	29.95	35.61
	Ads (T)	30.71	27.90	33.52
	Delta (T-C)	-2.08	-6.16	2.01

Upper Tenure Tercile	No-Ads (C)	44.65	40.62	48.68
	Ads (T)	46.29	42.36	50.22
	Delta (T-C)	1.64	-4.04	7.31
By Time Spent				
Bottom TS Tercile	No-Ads (C)	20.96	18.75	23.16
	Ads (T)	19.86	17.81	21.91
	Delta (T-C)	-1.10	-4.10	1.91
Middle TS Tercile	No-Ads (C)	34.78	31.59	37.98
	Ads (T)	34.40	31.62	37.18
	Delta (T-C)	-0.38	-4.66	3.90
Upper TS Tercile	No-Ads (C)	40.47	37.22	43.72
	Ads (T)	40.75	37.73	43.76
	Delta (T-C)	0.28	-4.10	4.67

Appendix Table 1: Median WTA for the no-ads holdout vs. regular users. This table shows the point estimates and 95% confidence intervals for each of the sample cuts included above in Figures 2-5, along with the point estimate and 95% confidence interval for the difference between median WTA for the no-ads holdout and regular users.



Appendix Figure 1: WTA estimates for no-ads holdout vs. regular users, by country/region. The estimates for EEA (European Economic Area) include users in our sample from Germany, France, Spain, Romania, Belgium, Ireland, and Norway.