Should I Stay or Should I Go? Social Media Advertising in Times of Boycott

Rémi Devaux
Institut Mines-Télécom Business School

Cite as:
Devaux Rémi (2025), Should I Stay or Should I Go? Social Media Advertising in Times of Boycott. *Proceedings of the European Marketing Academy*, 54th, (125392)

Paper from the 54th Annual EMAC Conference, Madrid, Spain, May 25-30, 2025

Should I Stay or Should I Go?
Social Media Advertising in Times of Boycott

immediate

Abstract

In the summer of 2020, more than 1,000 well-known brands suspended their advertisements from Facebook and Instagram as the platforms hosted an increasing number of 'unsafe' content for advertisers such as hate speeches and political controversies. Using data from a brand that did not boycotted Facebook Ads we document that during the boycott period, Facebook advertisements, experienced a drop in advertising effectiveness and per-impression prices compared to ad published on a similar ad network. Further analyses show that (i) ads displayed inside the news feed experienced a stronger decrease in price and effectiveness, suggesting negative spillovers from unsafe content to ads and (ii) a targeting mismatch delivered ads were to non-relevant demographic groups during the boycott. We also find that Facebook's deletion of controversial accounts has mitigated the crisis.

JEL Classification: L82, L86, M37

Keywords— Online Advertising, Content Moderation, Externalities, Boycott

Track: Advertising & Marketing Communications

1 Introduction

Advertising boycotts are common in situations where brands seek to avoid displaying ads in a controversial context. Example includes the recent departure of many brands from X (formerly Twitter) after Elon Musk took over the platform and its content moderation policies. In practice, 'ad-unfriendly' content may take different forms (e.g. racism, violence, hate speeches, sexual allusion, extremism, political controversies), which we will regroup under the term 'unsafe content'.

In June 2020, a coalition of civil rights associations named #StopHateForProfit (hereafter, #SHFP) urged advertisers to suspend their campaigns on Facebook and Instagram, as the platforms hosted an increasing number of controversial content. In its open letter to advertisers, #SHFP argued that "*advertisements are running alongside divisive, hateful and conspiratorial content – not something that most companies want*" (SHFP, 2020). As a result, a thousand of well-known consumer brands joined #SHFP (among which The North Face, Ben & Jerry's, Verizon, Unilever or Coca-Cola) and suspended their advertising campaigns on Facebook and Instagram for at least the month of July 2020. At first, Facebook did not not change its moderation policy, triggering sustained complaints from #SHFP. The platform eventually responded to the changes asked by the movement when it announced on August 19$^{\text{th}}$, that thousands of groups, pages, ads and accounts accused of spreading hateful content and conspiracy theories have been suspended across Facebook and Instagram (Facebook, 2020). On September 17$^{\text{th}}$, Facebook announced a policy update enforcing stricter content moderation for users posting inside Facebook groups (Facebook, 2020).

We assess the consequences on this boycott on Facebook Ads outcome using data from a non-boycotting US brand in the skincare industry, which maintained its campaigns on Facebook. Our results show that ad effectiveness and prices decreased during the Facebook Ad Boycott. These results are mainly explained by (i) negative spillovers from unsafe content to ads and (ii) a mismatch of the Facebook Ads targeting algorithm which delivered the ads to non-relevant groups of users during the boycott. We also show that Facebook's decision to bans thousands of controversial pages, groups and accounts has offset the negative effect from the boycott. This paper stems from several streams of the economics and marketing literature:

Boycott and Brand Activism. Recent empirical works argue that consumers not only buy product but also 'consume values' and that political stance hence influence demand (Conway and Boxell, 2024). From the consumer side, recent studies highlight the consequences of online consumers boycott for firms (Liaukonytė et al., 2022; Winkler et al., 2024).

Context Effects in Online Advertising. Marketing scholars show that online content and website credibility impact advertising effectiveness in a significant manner. For advertisers, media content is an externality: it can impact the return-of-investments of their ads (e.g. sales, brand image) in a positive or a negative way. These context effects are highlighted in recent empirical works (Shehu et al., 2021; Simonov et al., 2024).

Content Moderation & Platform Strategies. Theoretical models show that ad-funded platforms have conflicting incentives regarding unsafe content and moderation. On the one hand, Beknazar-Yuzbashev

et al. (2024) shows that ad-funded platforms have incentives to display harmful content to users, even when they dislike it, as it increases engagement, time spend on the platform and thus revenues. On the other hand, platforms also have incentives to engage in moderation efforts whenever controversial content represent too much of a nuisance for advertisers (Liu et al., 2022).

2 Data and Descriptive Evidences

We used data from a partner firm operating in the American skincare industry. The brand mainly commercializes skin care products: moisturizing cream, scar gel, matches to hide and heal cold sores and cream for stretch marks. Products are sold online through Amazon, and offline through bick-and-mortar retailers.

Facebook Ads Data The Facebook data consists of an ad-week panel between June 2019 and December 2020. In Facebook data, observations are grouped at the Ad Set $\times$ Platform $\times$ Placement $\times$ Demographic group level. An Ad Set is a group of ads that share the same parameters of diffusion (e.g. ad objective, geo-targeting...). Platform indicates whether the ad is distributed on Facebook or Instagram, while the placement indicates on which part of the website/app the ad is displayed to users (e.g. feed, stories, Facebook marketplace or Facebook search pages). Finally, demographic group include both the gender and the age groups to which the ad has been displayed. Our Facebook ads data contain 1,396 ad creatives over the 2019-2020 period, for a total of 15,160 observations.

For each ad and week, the data provides quantitative metrics like the cost of the ad as well as many response indicators: impressions, reach, clicks, engagement with the brand's page, user's reactions (likes, comments and share on the sponsored posts).

Microsoft Advertising Data Our focal brand was also investing continuously on Microsoft Advertising. Microsoft Ads include search ads in the form of text and display banners appearing on Bing and other Microsoft properties such as MSN, Yahoo! and partner publishers. This type of ads is a relevant control for our study, as the brand continuously invested on these ads during the same period as Facebook Ads. For each campaign-week, Microsoft data provides costs, impressions and click recorded on the ads.

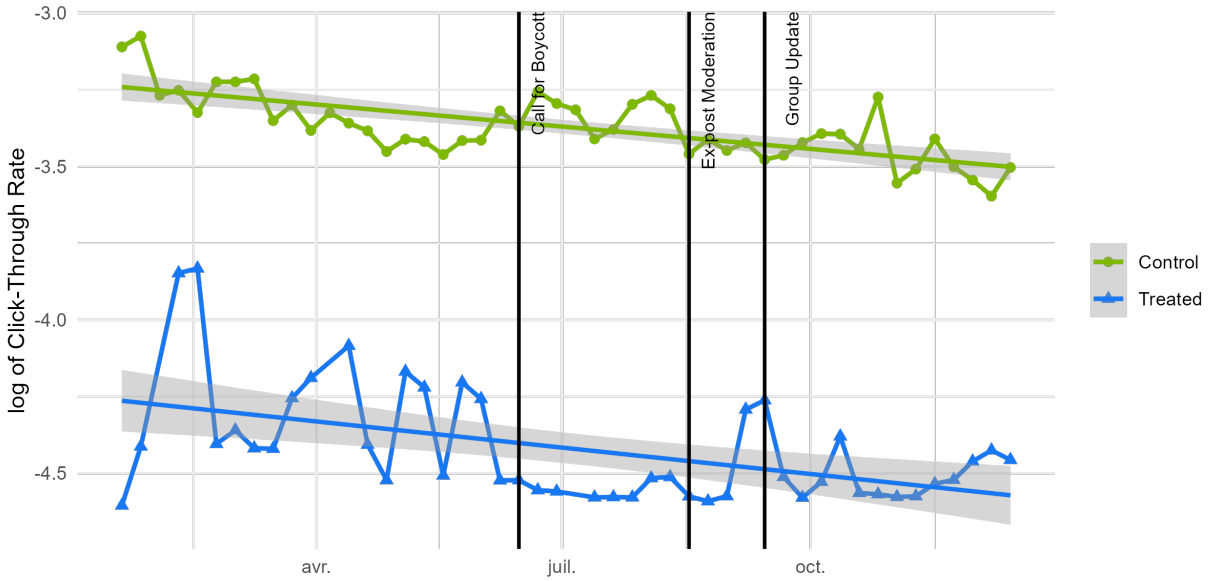
Over the June 2019 - December 2020 period, Microsoft ads recorded 4.2M impressions for a budget of 72,100 USD. If the level of spend and impressions are different across Facebook and Microsoft ads, the key metrics followed similar trends previous to the boycott, as we will see later on. Table A.1 in Appendix presents descriptive statistics for the main costs and performance variables in the data.

Our empirical exploits the quasi-experimental shock created by the Facebook Ad Boycott in the summer of 2020. In order to compare Facebook and Microsoft data, we will aggregate Facebook Ads data at the weekly level. We will exploit the granularity of Facebook Ads data later on.

Descriptive evidences in appendix shows that the brand did not reduced its ad investment on the platform after the #SHFP open letter in June. We also show that the brand did not changed the budget and ad

Figure 1: Evolution of Click-Through Rates (CTR)

The control group includes ads from Microsoft Advertising, while the Treated Group includes Facebook and Instagram ads. The blue line represents a linear time trend estimated by regression with 95% confidence intervals in gray.



impressions allocated to Facebook and Microsoft Ads in a singnificant manner.

The Click-through rate (CTR) – the ratio of ad clicks over ad impressions – is a key indicator used in the industry to gauge ad effectiveness. Figure 1 shows the evolution of the CTR across Facebook and Microsoft ads. On both networks, the CTR exhibits a similar trend before the boycott. During the boycott, Facebook Ads' CTR significantly decrease while Microsoft Ads' CTR remain similar to its trend during and after the boycott. After August's moderation update, the CTR increases again on Facebook Ads' while we see no trend change on Microsoft. Additional descriptive statistics (that can be provided by the author) show that we observe similar variations on Facebook ads' the cost-per-1,000 impressions (CPM). Facebook ads' cost-per-click (CPC), a proxy for advertising's cost-effectiveness increased during the boycott, suggesting it was more expensive for the brand to get clicks during the boycott.

3 Econometric Analysis

3.1 Impact of th Boycott

We take advantage of the natural experiment setting of the Facebook Ad Boycott by specifying the following Differences-in-Differences (DiD) regression for all ads from June 2019 to December 2020:

$$\ln(Y_{it}) = \tau(F_i \times \mathbb{1}\{\text{Boycott}\}_t) + \alpha_i + \gamma_t + \varepsilon_{it}. \tag{1}$$

In the equation above, Y_{it} is the outcome of ad i in week t. The treatment group dummy F_i equals one for ads i displayed on Facebook or Instagram, while the boycott dummy $\mathbb{1}\{\text{Boycott}\}_t$) equals one between week of June, 15$^{\text{th}}$ 2020 and the last week of August. To control for unobserved effects at the platform level, we implement a platform and week fixed effects α_i. The coefficient of interest is τ and it measures the effect of advertising on Facebook and Instagram from the beginning of the boycott to the end of August.

Table 1: Differences-in-Differences Results

	(1) ln(CTR)	(2) ln(CPM)	(3) ln(CPC)
Boycott x Meta	-0.208^{***}	-0.666^{***}	0.594^{***}
	(0.031)	(0.090)	(0.186)
Two-way Fixed Effects	✓	✓	✓
Observations	225	225	219
R^2	0.048	0.072	0.022

*p<0.1; **p<0.05; ***p<0.01

Note: Newey-West robust standard errors displayed in parentheses. A small $\Delta =$ has been added to CTR to avoid log() issues. Regression on Cost-per-Clicks (CPC) excludes six observations for which the number of clicks were null

Results from the model (1) are reported in Table 1. Consistent with our descriptive evidences, CTR decreased on Facebook Ads (-18.8%) during the boycott. The CPM decreased sharply on both platforms (-48.6% in average). Treatment effect is also positive and significant on the CPC metric: in other words, during the boycott, it had been more costly to get effective impressions compared to the control group.

In appendix, we show that Facebook suffered more from the #SHFP movement. Because Facebook is naturally more prompt to political debates and controversies, unsafe content may have circulated more sharply on Facebook than Instagram.

3.2 Effect of Facebook's Response

As stated before, assessing when the boycott and its impact stopped is challenging. Regarding the action taken by Facebook to counter unsafe content on the platform, two dates are important: August 19th (ex-post content moderation actions) and September 17th (policy update). To study whether the boycott impact changed after these dates, we employ a modified specification of equation (1) :

$$\begin{aligned}\ln(Y_{it}) =& \tau_1(F_i \times \mathbb{1}\{\text{After boycott}\}_t) + \tau_2(F_i \times \mathbb{1}\{\text{After Bans}\}_t) + \tau_3(F_i \times \mathbb{1}\{\text{After Update}\}_t) \\ &+ \alpha_i + \gamma_t + \varepsilon_{it}.\end{aligned} \tag{2}$$

Here, the treatment variables are staggered. The boycott dummy equals 1 from June 17th to the last observed period (end of 2020). The After Bans indicator takes value 1 after August 19th and the After

Table 2: Differences-in-Differences Results

	(1) ln(CTR)	(2) ln(CPM)	(3) ln(CPC)
Meta × After Boycott	-0.0025**	-0.6924***	1.040***
	(0.0012)	(0.1547)	(0.3004)
Meta × After Bans	0.0051***	0.9392***	0.2963
	(0.0011)	(0.3027)	(0.4991)
Meta × After Update	0.0013	-0.4192	-1.026
	(0.0020)	(0.2801)	(0.6807)
Two-way Fixed Effects	✓	✓	✓
Observations	217	217	215
R^2	0.93	0.92	0.87

*p<0.1; **p<0.05; ***p<0.01

Note: Newey-West robust standard errors displayed in parentheses. Regression on Cost-per-Clicks (CPC) excludes six observations for which the number of clicks were null. A small $\Delta = 1$ has been added to the CTR to avoid ln(0).

Update variable after September 17th. (2), we assume that boycott can impact Facebook Ads effectiveness until December 2020 (end of the period covered by the data). This specification will allow us to separate the effect of the boycott (τ_1) and the impact of Facebook's moderation policies (τ_2 and τ_3)

Results are shown in Table 2. Consistent with our previous estimates, the boycott dummy has a negative impat on CTR and CPM. However, Facebook's intervention seems to have mitigated the boycott's negative effect, with CTR and CPM increasing on both Facebook and Instagram after the massive bans on August 19th. Finally, we find no impact on Facebook's group update on ad effectiveness.

These estimates all suggest that the boycott did impacted the ad effectiveness and cost. More importantly, it sheds light on the effect of content moderation. While the ban of hateful groups and pages countered the negative effect on clicks and ad prices, the policy update on September seems to have no effect on the trend of CTR, CPM and CPC. In appendix A.8, we show that the negative effect of the boycott impact ads during the first two weeks of August (before the ban happens), and find no impact after the ban.

Our interpretation is that, while the bans in August where heavily covered by the media, it was not the case of September's update policy. Demonstrating moderation efforts was key for Facebook to reinstate trust among users and advertisers on its platforms. In this perspective, massive and publicized bans may have been more effective in addressing the crisis in the short-run than changes in the functioning of groups. Especially since the impact of implementing new group rules on the production of hateful content might only be visible on the long-run.

3.3 Robustness Checks

We performed several robustness checks, all available upon request.

Using Google Search Data as a Control During the period, the brand also advertised on Google Search by placing ads on branded and generic keywords. We run the same Differences-in-Differences analysis as in equation (1) using Google Ads as the control group and find the same results.

Using Facebook Ads Outcomes Outside of the US as a Control Group In the context of #SHFP, the boycott took place in the context of US political climate. It was initiated by American associations and brands. We leverage Facebook Ads outcomes in another country (Italy) as the control group and employ the same DiD approach in as equation (1). Results from the differences-in-differences show that Ads outcomes decreased more sharply during the boycott in the US than in Italy.

Treatment timing We show that our results are robust when we specify different treatment timing for the the boycott (June-and-July and July-only). We also perform a placebo test by showing no negative effect from advertising on Facebook in May and in June prior to the boycott. Finally, we perform two regressions with a timing on August only and shows that Facebook ads are negatively impacted in the first two week on August, but not after August 17th's week, consistent with the fact that content moderation have mitigated the negative effect of the boycott.

4 Disentangling the Impact of the Boycott on Ad Effectiveness

4.1 Evidence of Heterogeneity across Placement and Gender

We previously showed that ad value dropped during the Facebook Ad Boycott. Two reasons may explain our results. First, ad effectiveness may have dropped because of the negative spillover imposed from unsafe content to ads. Because of controversial and hateful content, users may pay less attention to ads, or associate them with the unsafe content. Second, the sudden suspension of billions of ad impressions might have prevent the targeting algorithm to work as usual. We explore both of these hypotheses by exploiting the information about the 1,369 ads in our data.

On Facebook and Instagram, ads can appear in different sections of the site: in the feed (groups feed, video feed and news feed), inside the stories, inside the "explore" page of Instagram, in the search results of Facebook search engine, on the Facebook Marketplace. Ads placed inside the news feed are more closely associated to the content produced by users on the platform. During the boycott, ads that appeared inside the news feed should have been displayed along unsafe content that spread on the platform. Table 3 shows that ads placed inside the feed during the boycott are associated with a lower performance by looking at CTR and engagement rate (share of social interaction with the ad). Since ads in the feed are more exposed to content posted on Facebook, our estimates strongly suggests that the reduction in advertising value is significantly driven by the presence of unsafe content on the platform. For simplicity

Looking at demographics, Table A.2 in appendix shows that the amount of ad impressions addressed to male Facebook was multiplied by 10 during the boycott, going from 0.7% to 7.8%. A sudden change is

Table 3: Effects of Boycott per Ad Placement

	log(CTR) (1)	log(Engagement Rate) (2)	log(CPC) (3)
Boycott	0.0040***	0.0233***	0.0673
	(0.0011)	(0.0038)	(0.0998)
Boycott × inside News Feed	-0.0035***	-0.0479***	0.1805*
	(0.0013)	(0.0047)	(0.1040)
Fixed Effects			
Week-year (γ)	✓	✓	✓
Ad Set (θ)	✓	✓	✓
Additional Controls			
Ad Objective	✓	✓	✓
Platform	✓	✓	✓
Log of ad budget	✓	✓	✓
Observations	8,562	8,562	4,530
R^2	0.13419	0.56348	0.72073

*p<0.1; **p<0.05; ***p<0.01

Note: Robust standard errors displayed in parentheses, clustered at the ad-level. A small $\Delta = 1$ has been added to the CTR and Engagement rate to avoid discarding zero-outcome observations (e.g. ads that are not clicked).

strategy is not plausible, since women is the primary target of the focal brand on social media. Moreover, from September to December 2020 (after the boycott), the share of male users targeted by the brand decrease again to 0.5%, a level similar to the non-boycott average. Regression results show that the CTR decreased for ads targeted to male users during the boycott. Overall, the decrease in ad effectiveness across men users is consistent with our intuition that targeting such demographic groups during the boycott is an abnormal decision which drives lesser value for the brand.

Table 4: Effects of Boycott per Gender

	log(CTR) (1)	log(Engagement Rate) (2)	log(CPC) (3)
Boycott	-0.0017	-0.0720**	0.2839***
	(0.0012)	(0.0344)	(0.0996)
Boycott × Male	-0.0053***	-0.0104	0.0348
	(0.0016)	(0.0176)	(0.0590)
Fixed Effects			
Week-year (γ)	✓	✓	✓
Ad Set (θ)	✓	✓	✓
Additional Controls			
Log of ad budget	✓	✓	✓
Observations	8,562	8,562	4,530
R^2	0.13419	0.56348	0.72073
			*p<0.1; **p<0.05; ***p<0.01

Note: Robust standard errors displayed in parentheses, clustered at the ad-level. A small $\Delta = 1$ has been added to the CTR and Engagement rate to avoid discarding zero-outcome observations (e.g. ads that are not clicked).

5 Concluding Words

Our results suggest that brands advertising on social media are better-off suspending their campaigns when platforms encounter crisis related to polemical content. For platforms, our results emphasizes the importance of content moderation. Advertisers are sensitive to 'unsafe' content (violent, hateful, controversial and polemical messages) with good reasons. We show that content moderation is likely to have mitigated the drop in ad effectiveness and prices in the case of Facebook and Instagram. Hence, platforms must take into account the sensitivity of advertisers to unsafe content when designing and updating content moderation policy. This paper is a case study limited to a particular case: a brand operating in the skincare industry. The same study may have resulted in different findings in other industries or competitive environment. For example, in certain context a brand may benefit from a controversial endorsement – see for example the case of Goya in Liaukonytė et al. (2022) – or act strategically to take advantage of a boycott movement.

REFERENCES

Alison, T. (2020). Our latest steps to keep facebook groups safe. https://about.fb.com/news/2020/09/keeping-facebook-groups-safe/.

Beknazar-Yuzbashev, G., Jiménez-Durán, R., and Stalinski, M. (2024). A model of harmful yet engaging

content on social media. In *AEA Papers and Proceedings*, volume 114, pages 678–683. American Economic Association.

Conway, J. and Boxell, L. (2024). Consuming values. *Available at SSRN 4855718*.

Facebook (2020). An Update to How We Address Movements and Organizations Tied to Violence — Meta — about.fb.com. https://about.fb.com/news/2020/08/addressing-movements-and-organizations-tied-to-violence/. [Accessed 16-11-2024].

Greenblatt, J. (2020). An Open Letter to the Companies that Advertise on Facebook. https://www.adl.org/news/letters/an-open-letter-to-the-companies-that-advertise-on-facebook.

Liaukonytė, J., Tuchman, A., and Zhu, X. (2022). Frontiers: spilling the beans on political consumerism: do social media boycotts and buycotts translate to real sales impact? *Marketing Science*.

Liu, Y., Yildirim, P., and Zhang, Z. J. (2022). Implications of revenue models and technology for content moderation strategies. *Marketing Science*.

Shehu, E., Abou Nabout, N., and Clement, M. (2021). The risk of programmatic advertising: Effects of website quality on advertising effectiveness. *International Journal of Research in Marketing*, 38(3):663–677.

Simonov, A., Valletti, T., and Veiga, A. (2024). Express: Attention spillovers from news to ads: Evidence from an eye-tracking experiment. *Journal of Marketing Research*, page 00222437241256900.

Winkler, D., Wlömert, N., and Liaukonyte, J. (2024). Separating the artist from the art: Social media boycotts, platform sanctions, and music consumption. *Cornell SC Johnson College of Business Research Paper*.