Does Clickbait Lead to More Clicks?

Hsuanyi Cheng, Patrick Kim, Kayla Wopschall, Ziwei Zhao

8/12/2021

Introduction

Background

Research Question

Hypothesis

Facebook Ads Experiment

Design

We chose Facebook advertising as our preferred platform for the experiment for several reasons. First, it is one of the most widely used social media platforms, which would allow flexibility in choosing the sample population and ensure that we have enough subjects in the experiment given our limited budget. Secondly, Facebook has an intuitive A/B testing feature that allows us to easily design ad campaigns with a control ad and a treatment ad that never get shown to the same Facebook user.

We conducted two pilot experiments. In each experiment, we use one existing news article and design two ads for it, each with a headline and an image. For the ad that gets shown to the control group, both the headline and image are rather neutral. They state facts without a call to action or question for the reader, and the image should not elicit strong emotional reactions. For the ad shown to the treatment group, both the headline and image are "clickbait" - the headline contains a question that entices the reader to click on the article to find the answer, and the image portray more dynamic action that might draw attention.

Figure 1 shows the two ads used for the first pilot experiment. It uses the same NPR article about traveling overseas during the pandemic. The control group sees the headline "Traveling overseas - everything you need to know" with an image of a woman hiking in an idyllic nature scene. The treatment group sees the headline "Traveling overseas? It might not be worth it" with an image of travelers getting their temperatures checked in an airport security line.

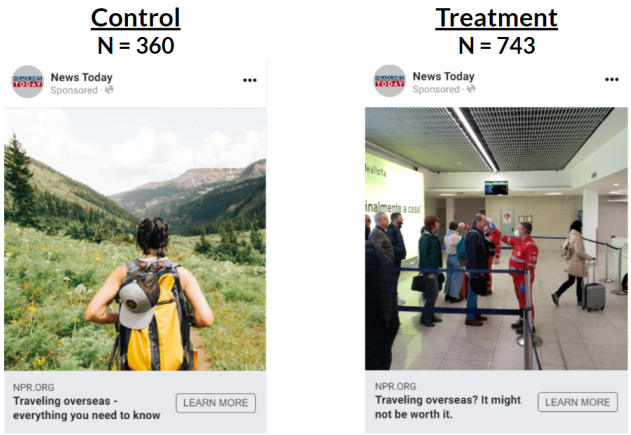

Figure 1: Facebook Pilot Experiment 1

Figure 2 shows the two ads used for the second pilot experiment. It uses an New York Times article about the drought on the West Coast. The control group sees the headline "What to understand about the drought in the west" with an image of a red fiery sky. The treatment group sees the headline "Will you run out of water?" with a farmer walking on barren land in front of a few cows.

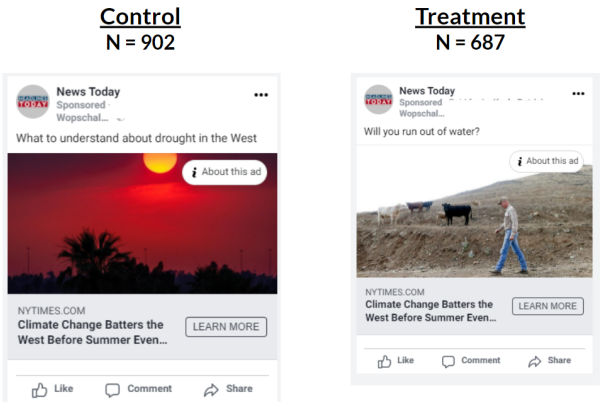

Figure 2: Facebook Pilot Experiment 2

Each experiment got to run for about two days. Because the business page from which we created ads is brand new, and we did not have much, if any, previously advertising records on Facebook, our respective advertising accounts got disabled by Facebook. It was eye opening to see efforts being made by social media platforms to limit "fake news", even though it was a big roadblock for our experiment. However, we were able to extract data from these pilot studies and see some meaningful results.

Analysis

Regression Model

The first regression analysis is linear regression model of the variable click on the variable treated - both indicator variables. Click indicates whether a Facebook user clicked through to the article after seeing the ad, and treated indicates if a user is in the treatment group. Table 1 shows the results. The first column shows the results from the first pilot study using the travel ads. The coefficient on treated is 0.122 with a 0.00 p-value, meaning that people in the treatment group are 12.2% more likely to click on the ad. The second column shows results from the second pilot study using the drought ads. The coefficient on treated is 0.020

with a p-value of 0.0691. People in the treatment group are 2.0% more likely to click on the ad, but it is only only statistically significant at the 10% level, compared to the 1% level in the first pilot study.

Table 1: Facebook experiment - click regressed on treated

	Dependent variable:	
	click	
	travel ads	drought ads
	(1)	(2)
treated	0.122^{***}	0.020^{*}
	(0.016)	(0.011)
Constant	0.031^{***}	0.040^{***}
	(0.009)	(0.007)
Observations	1,103	1,589
R^2	0.033	0.002
Adjusted R^2	0.032	0.001
Residual Std. Error	0.311 (df = 1101)	0.215 (df = 1587)
F Statistic	37.035^{***} (df = 1; 1101)	3.308^{*} (df = 1; 1587)
Note:		*p<0.1; **p<0.05; ***p<0.01

Covariate Check

Since Facebook's algorithms randomized our populations for the experiments and their business goals are presumed to be focused on revenue from advertisers, we conducted covariate checks to ensure that randomization was done properly. We conducted a regression of treated on female to see if gender predicted which group a user would fall in. The female coefficient is not statisticaly significant in the travel ads experiment, but it is significant at the 10% level in the drought ads experiment.

Table 2: Facebook experiment - covariate check - gender

	Dependent variable:	
	treated	
	travel ads	drought ads
	(1)	(2)
female	0.037	-0.042^{*}
	(0.029)	(0.025)
Constant	0.650^{***}	0.453^{***}
	(0.024)	(0.018)
Observations	1,103	1,589
R^2	0.001	0.002
Adjusted R^2	0.001	0.001
Residual Std. Error	0.469 (df = 1101)	0.495 (df = 1587)
F Statistic	1.600 (df = 1; 1101)	2.897^{*} (df = 1; 1587)
Note:		*p<0.1; **p<0.05; ***p<0.01

We ran a similar regression of treated on age. Age was provided from facebook in buckets: 18-24, 25-34, 35-44, 45-54, 55-64 and 65+. The regression results effectively compare each age group to the 18-24 group. For both the travel ads experiment and the drought ads experiment, all coefficients on age groups are positive. None of the coefficients are statistically significant in the travel ads experiment, but they are all statistically significant at different levels for the drought experiment. Without other covariates or a deeper understanding of Facebook's algorithm, it is difficult for us to tell why these the drought experiment's randomization may not have been done properly.

Table 3: Facebook experiment - covariate check - age

	Dependent variable:	
	treated	
	travel ads	drought ads
	(1)	(2)
as.factor(age)25-34	0.260	0.061**
	(0.713)	(0.027)
as.factor(age)35-44	0.297	0.227**
	(0.709)	(0.091)
as.factor(age)45-54	0.182	0.249*
	(0.708)	(0.147)
as.factor(age)55-64	0.197	0.526***
	(0.708)	(0.055)
as.factor(age)65+	0.126	0.570***
	(0.707)	(0.032)
Constant	0.500	0.366***
	(0.707)	(0.016)
Observations	1,103	1,589
R^2	0.010	0.085
Adjusted R^2	0.006	0.082
Residual Std. Error	0.468 (df = 1097)	0.475 (df = 1583)
F Statistic	2.279** (df = 5; 1097)	29.540*** (df = 5; 1583)
Note:		*p<0.1; **p<0.05; ***p<0.01

Survey Experiment

Because we spent so much time trying to do the Facebook experiment, we unfortunately had to redesign, deploy, and analyze a new experiment shortly before the deadline. We chose to utilize SurveyMonkey because their platform allows for surveying users cheaply and efficiently with little danger of being locked out of the platform. The respondent is shown two image ads and asked to indicate which ad they would be more likely to click on. The survey targeted 400 respondents from across the United States using the broadest demographic available on SurveyMonkey. The full dataset contains 437 responses as SurveyMonkey provided these extra responses for free.

Design

SurveyMonkey contains an A/B testing feature that allows for different text and images to be shown to respondents at random. We utilize this feature to show two different images to the respondent and create many sampling pairs. First, two ads with tame language and normal stock images were created by the team. For each ad, one version is created with extreme language in the title, another version with a shocking image, and a final version with both modified at once. This means that there are four versions of each article, and each version has a 25% chance of being displayed to the respondent according to SurveyMonkey's randomization procedure.

For each ad:

1. Control
2. Text Treatment
3. Image Treatment
4. Both

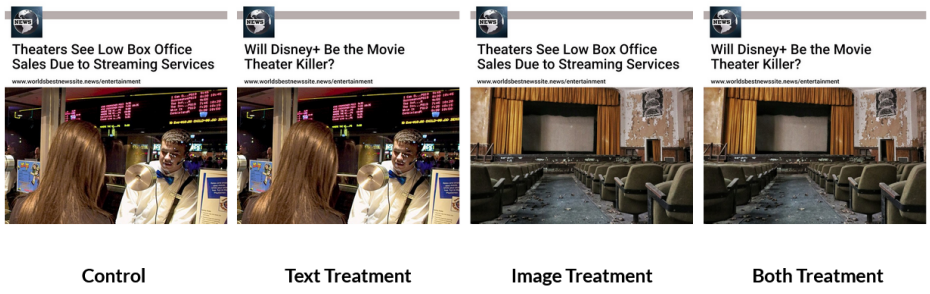

Figure 3: Article 1

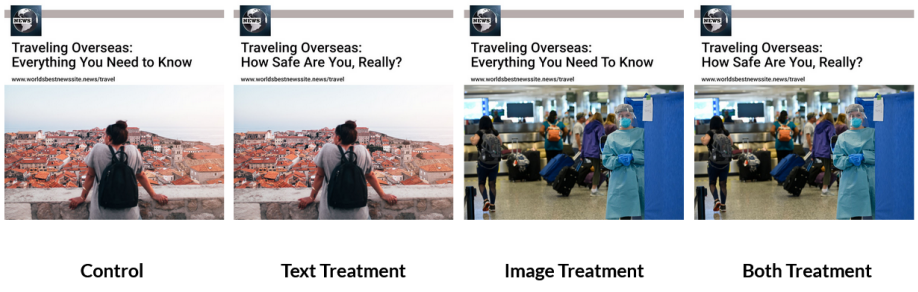

Figure 4: Article 2

Since there are two ads and four possible versions of each ad, this creates 16 possible pairings, meaning that each pairing has a 1 in 16 chance of appearing in a respondent's survey.

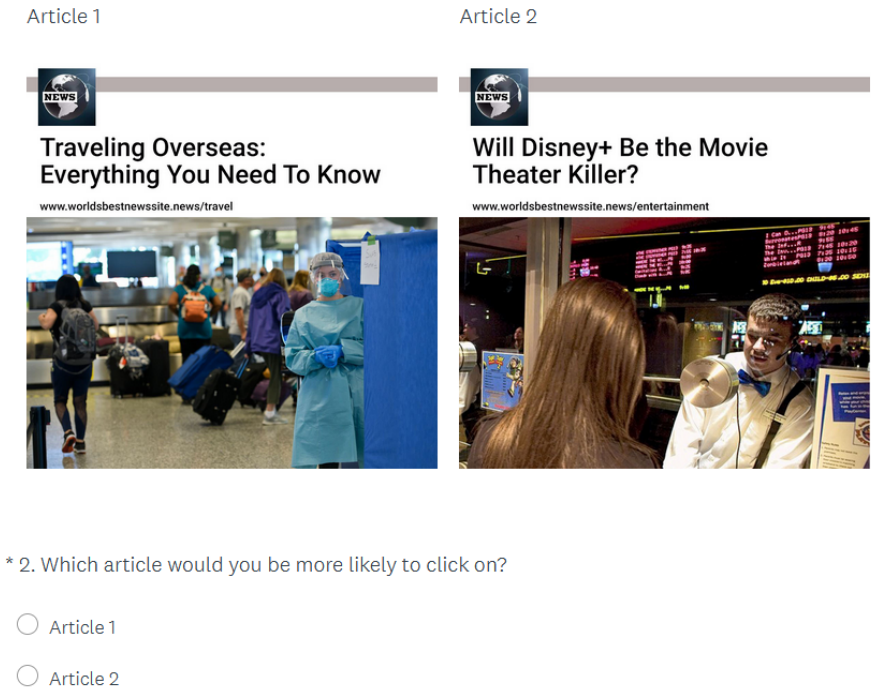

Figure 5: Survey Sample

Analysis

Respondents only ever see one version of each ad, which means that the potential outcomes of a respondent observing a treated ad and a control ad are never observed. To overcome this problem, a test of proportions is used as a basic check to determine whether the difference in ads has a significant effect on respondent choices. For each treatment type, we compare the number of times that ad version was selected by the respondents between both ads, and the p-values obtained are outlined below.

Test of Proportion P-values:

- Control group: 0.6510766
- Text treatment group: 1
- Image treatment group: 0.3319755
- Both treatment group: 0.4414183

Evidence of a difference in outcomes based on the ad is weak according to our dataset.

Within test

Between test

Conclusions

Overall conclusions