

Pre-analysis Plan:
The Impact of Comments on Social Media Campaigns

Dante Donati and Lena Song *

May 14, 2025

*Donati: Columbia Business School. dd3137@gsb.columbia.edu. Song: University of Illinois Urbana-Champaign. lenasong@illinois.edu.

1 Introduction

We measure the causal impact of comment sections on social media posts promoting racial justice. Using online field experiments on Facebook, we first collect and analyze comments to understand how political ideology influences engagement with posts. We then evaluate how exposure to different comment narratives affects subsequent engagement (comments, views), intentions (clicks), and off-platform behaviors (email sign-ups, donations, petitions).

2 Experimental Design

2.1 Experiment Overview

The experiment is structured around three main components:

- Content design and pre-testing: this involves creating the content for the study and conducting a pre-test to select content to be used in the main studies.
- Study 1 - engagement with posts: this study focuses on documenting participants' engagement with racial justice posts and how this varies with political ideology.
- Study 2 - the effect of the comment section: in this study, we will study how the presence and narratives of a comment section influence participants' responses and engagement with the content.

2.2 Content design and pre-testing

To create professional content (ad creatives and copy), we hired four graphic designers, assigning each to focus on 2–3 key issues. These issues, identified in collaboration with our partner organization, Color of Change, included voter suppression, environmental justice, criminal justice and police reform, education reform, and technology fairness.¹ Each designer developed multiple concepts and taglines for their assigned ads. These were reviewed by the partner organization, which approved, revised, or rejected the proposals. Approved concepts were then finalized by our designers, adhering to the branding guidelines of the organization.

Figure A1 in the Online Appendix displays the final graphics (a total of ten, two for each issue) and their corresponding headlines exactly as they would appear to Facebook users. Each visual is designed to tell a compelling story about its assigned issue, emphasizing the urgency and importance of taking action. The content aims to engage users by sparking curiosity and encouraging further exploration of each topic. The use of bold imagery and compelling design elements is intended to capture attention while users browse their Facebook feed.

¹The Online Appendix provides a detailed description of these issues.

We pre-tested the banners to identify which ones, within each issue, were most likely to generate a high number of clicks. This was done using Facebook’s A/B testing tool across all 10 banners. For the test, we specified the audience (ZIP codes with a high share of progressive populations), budget (\$50 per banner), duration (1 week), and optimization goal (reach), without imposing a frequency cap. The click-through rates (CTRs) for this test (Pre-test A)—as measured by the ratio Link Clicks/Reach—are displayed in Table 1. These vary between 0.12% to 0.25%. For each issue, the banners with higher CTR are displayed on the right in Figure A1.

Table 1: Results from Pre-test A

Issue	Ad Name	Link Clicks	Reach	CTR (%)
technology	pixels	15	6115	0.245
technology	man	13	6683	0.195
voting	lady	14	5777	0.242
voting	flag	11	5792	0.190
police	lady	13	6146	0.212
police	hands	13	6315	0.206
environment	kid	13	6235	0.209
environment	street	11	6290	0.175
education	future	9	6216	0.145
education	history	6	4884	0.123

Table 2: Results from Pre-tests B and C

Issue	Ad Name	Link Clicks	Reach	CTR (%)
voting	lady	32	26794	0.119
environment	kid	33	28351	0.116
police	hands	32	27771	0.115
technology	pixels	29	28208	0.103

We conducted two additional tests, Pre-tests B and C, where we capped the frequency at one impression per person. Test B was conducted with a large potential audience (approximately 200,000 users per banner), while Test C targeted a smaller audience (approximately 6,000 users per banner). These adjustments were made to simulate a campaign that closely resembles the one planned for Study 2. The results of Pre-tests B and C, combined with insights from the analysis of comments collected in Study 1, will inform the selection of the banner(s) to be used in Study 2.

Table 2 presents the aggregate results for Pre-tests B and C.² Notably, the CTRs for Pre-tests B and C are significantly lower than those reported for Pre-test A. This discrepancy arises because

²The issue on education was excluded from these tests due to its low performance in Pre-test A and budgetary considerations.

Pre-test A did not impose a frequency cap, allowing users to see each banner an average of two times, which increased the likelihood of clicks. For the Study 2 campaign, we will aim to saturate small audiences while maintaining a frequency cap of one impression per user. While this approach may result in lower CTRs, it is essential to mitigate potential divergent delivery bias in ad delivery caused by the ad platform’s algorithm (as discussed later).

2.3 Audience characteristics

We rely on zip codes to target ads to different audiences. The zip code characteristics were collected from various sources:

- **Meta Audience Estimates:** We use information on audience size provided by Meta, which reports the estimated number of Accounts Center users advertisers could potentially reach over a given period.³ This data was collected through the Marketing API for each zip code on October 15, 2024.
- **Voting Behavior:** To proxy political preferences and ideology at the zip code level, we rely on the 2020 voting results. These come at the precinct level and were obtained from The Upshot.⁴ We assigned each precinct to its nearest zip code according to Euclidean distance of the centroids using GIS software, and then aggregated voting information across all precincts matched with the same zip code.
- **Population and Racial Composition:** We use 2020 Census data to obtain information on the share of the Black population and the total population in each zip code.⁵

Once this information was assigned to each zip code, we sampled and assigned them to either Study 1 or Study 2. Zip codes were first categorized into three ideology groups based on the Republican vote share: low (below 30%), medium/mixed (between 45% and 55%), and high (above 70%). Each ideology group was further divided into two subgroups based on racial composition (low and high share of Black population relative to the group median). Hence, a total of 6 ideology-race strata were created.

- **Study 1 Sampling:** Zip codes were randomly sampled to form 60 clusters – 10 clusters in each ideology-race stratum – while ensuring an average total potential audience of 500,000 users per cluster.⁶ Within each cluster, we will use Facebook’s A/B testing tool to randomly split the audience into five arms (five ad banners), with 100,000 users in each arm. We exclude zipcodes used in pre-tests.

³<https://www.facebook.com/business/help/1665333080167380?id=176276233019487>

⁴<https://github.com/TheUpshot/presidential-precinct-map-2020>

⁵<https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2020&layergroup=ZIP+Code+Tabulation+Areas>

⁶To achieve this, we excluded the largest 1% of zip codes, removing those with audiences exceeding 75,000 users.

- **Study 2 Sampling:** Zip codes were randomly sampled to form clusters with an average potential audience reach of 50,000 users per cluster.⁷ As in Study 1, these clusters were sampled within each ideology-race stratum. The goal is to effectively reach 2,500,000 users. We will pick as many clusters as necessary to meet this goal, ensuring an equal number of clusters from each ideology-race stratum. We exclude zip codes used in pre-tests and Study 1.

2.4 Study 1 Design

In Study 1, we run ad campaigns on five banners – one for each issue – that achieved the highest CTR in our pre-tests. The campaign is set up with the objective of maximizing engagement (including likes, comments, and share) with posts that promote racial justice. This allows us to examine engagement with divisive content across various domains, and collect comments from Facebook users that will then be used in Study 2.

To provide evidence on differences in engagement across audience and ad characteristics, we set up hundreds of ads targeted at different audience groups (as described in the next section) and run A/B tests across different ad banners within each audience group. That is, within each audience group, users are divided into seeing an ad on one of five issues (education, environment, police, technology, or voting).

2.5 Study 2 Design

In Study 2, we manipulate the comments that participants see below our ads in a new ad campaign, using Meta A/B testing tool and an automated pipeline that hides new comments from users.

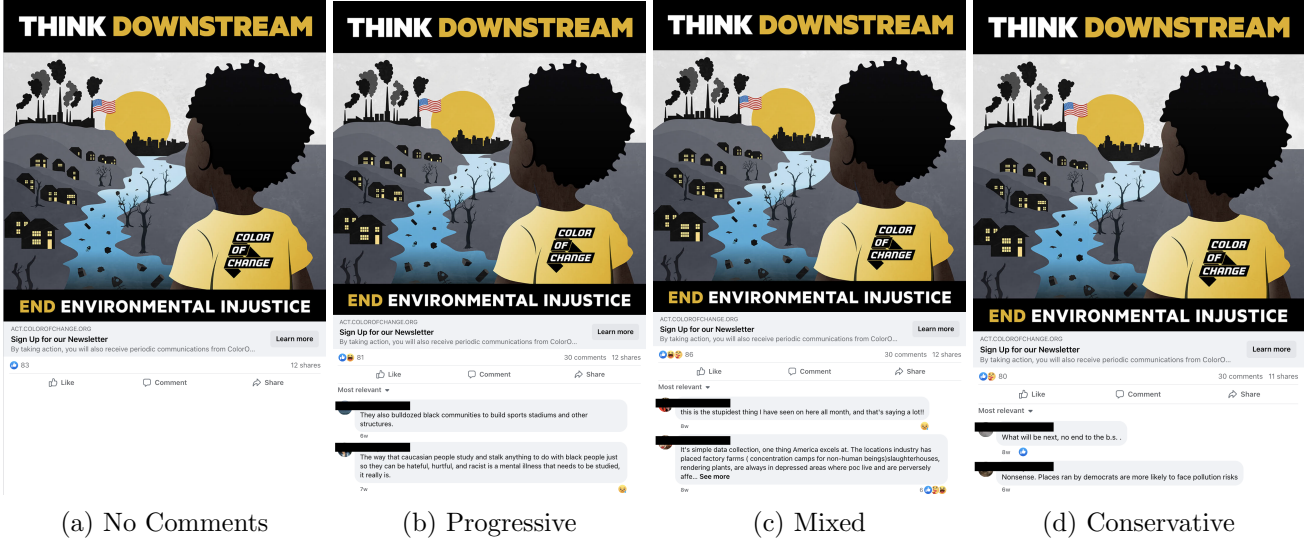
We investigate how exposure to the comment section and the different narratives expressed in the comment section of a post (collected in Study 1) affects individuals’ subsequent engagement with that post (comments/views), as well as their intentions (clicks) and off-platform behaviors. We are interested in the average treatment effect, as well as in the heterogeneous effects across individuals in different audience groups.

2.5.1 Intervention Design

Within each audience group, we use A/B testing to randomly split its population into four conditions: no visible comments (No Comments), visible comments that include both conservative and progressive comments (Mixed Comments), conservative comments only (Conservative Comments), progressive comments only (Progressive Comments). Figure 1 provides an example of what the users can see, depending on the random assignment for their audience group.

⁷To achieve this, we excluded zip codes with audiences exceeding 50,000 users, and further removed clusters with a composite audience outside the 45,000-55,000 range.

Figure 1: Study 2 Experimental Arms



To preserve the internal validity of the study, we adopt several additional precautionary measures. First, all new comments posted by the users will be automatically hidden from everybody using an automatic pipeline.⁸ This ensures that the No Comment condition displays no comments. In addition, it minimizes interactions between users who are exposed to the same post, and help address potential SUTVA violations arising from the social nature of the comment section. Second, to keep posts in treatment conditions similar in terms of total number of comments, the research team regularly monitor the comment section across posts, and add additional comments (which are immediately hidden), so the number of comments –i.e., the total count displayed to the users– stays similar across conditions. Third, to ensure comparability within each ZIP code, we set up the A/B testing so reach and frequency of the ads are constant across conditions. Specifically, in each condition, we max out spend while setting a frequency cap, so that most users are exposed to the ad exactly once. To minimize interference of the Meta ad delivery algorithm, we set the campaign objective to maximize reach (instead of engagement or traffic) and try to show ads to as many people as possible in the ZIP code. These precautionary measures, along with the fact that the post content is the same across conditions, allow us to isolate the effect of the comment section.

Topic Selection: We will select 1-2 topics from Study 1 to focus on, to maximize statistical power. We will choose topic(s) that are timely and relevant at the time of the experiment, with ads showing above-average performance in terms of engagement during Study 1.

⁸The commenter will not know that their comment has been hidden; the comment remains visible to the commenter as well as to the researcher.

Post Selection: We will use posts from Study 1 and show them to new audience groups.

To select the posts to use for the other conditions, we will identify triplets of posts in the following procedure. As part of the analysis in Study 1, comments are classified using a 5-point scale for political ideology: Strongly progressive or left-leaning, Slightly or moderately progressive, Centrist/unclear or no explicit stance, Slightly or moderately conservative, or Strongly conservative or right-leaning. To select posts for the Mixed, Conservative, and Progress Comments conditions, we will focus on posts that have at least two progressive and two conservative direct comments (i.e., comments that are directly responding to the post, rather than comments that reply to another comment). For each triplet of posts that have similar number of reactions, we assign them into the Mixed Comments, Conservative Comments, or the Progressive Comments condition.⁹

For the No Comment condition, we select posts that are not used in any other conditions and have a number of reactions similar to the average number of reactions in each triplet group. We then delete all comments on these posts, ensuring that the post has zero or a very small number of comments when it is first delivered to the audience in Study 2.

Comment Selection To select comments for the Conservative or Progressive Comments conditions, we will choose two comments that align with the corresponding ideology. For the Mixed Comments condition, we will select one conservative comments and one progressive comments. These selected comments will remain visible to the future audience, while all other direct comments will be hidden.

To better isolate the effect of the number of comments versus the narrative of the comments on outcomes, we manipulate the number of comments in the posts displayed to audience groups in Study 2. In the No Comment condition, we delete as many comments as possible so that the comment counter is much smaller than the counter in treatment conditions.¹⁰ In the triplet of posts used for treatment conditions, we equalize the number of comments across each condition by adding and/or deleting comments.

2.6 Outcome Variables

We examine a range of outcomes by measuring engagement with our posts.

2.6.1 Primary Outcomes

Users exposed to the post can perform different actions, both on and off-platform, which constitute our primary outcomes. These are: the number of additional likes on the post, the number of

⁹In addition, we equalize the number of shares across posts by sharing them ourselves.

¹⁰Some comments, such as those violating Facebook policy, are not visible to the research team and therefore cannot be deleted. As a result, for some comment sections, it may not be possible to reduce the comment counter to zero.

reactions to the post, the number of positive reactions to the post, the number of additional comments to the post, intent to sign up for a newsletter (click-through rate on the post and landing-page-views rate), and actual newsletter sign-ups as recorded by the landing website of the organization (Yes/No). We will use Facebook pixel to track outcomes across platforms.¹¹ By exposing all users to the same ad creatives, while varying the comment section below them, we can identify the causal impact of the comment narratives on the above outcomes.

2.6.2 Secondary Outcomes

We will also consider several secondary outcomes, categorized in several families.

To measure engagement, we will look at:

- average length of comments
- average log token count for comments
- average number of sentences for comments
- average sentence length of comments
- the average depth of comments, where we calculate the depth of a comment as the number of replies it has

For both Study 1 and 2, we will also consider several dimensions of comment content:

- political stance
- sentiment
- informativeness
- (dis)agreement with the post
- offensiveness
- toxicity
- Simpson's index

¹¹If feasible, we will also run a version of the experiment in which we expose existing Color of Change audiences to a post that encourages them to donate to a related cause or to sign a petition, depending on ongoing campaigns at Color of Change. We will measure intent to donate to/support the cause (click-through rate on the post) and actual donations/pledges as recorded by the landing website of the organization (both Yes/No and dollar amount donated for donations, and Yes/No for petitions). In this experiment, we will run a simple A/B test with 5 arms (No Comments, Progressive, and 3 Conservative) on the existing audience using a similar specification as outlined below, without clustered errors.

- variance in political stance

To measure overall conversation quality, we will examine several measures based on theories of deliberative democracy:

- reciprocity, where deliberative quality is high in conversations where participants maintain a consistent topic, without significantly altering the discourse
- justification, where deliberative quality is high in conversations where each participant justifying their position or viewpoint provides evidence or reasoned arguments
- respect, where deliberative quality is rated high in conversations where participants engage with each other respectfully and civilly

3 Empirical Strategy

3.1 Empirical Strategy

Given that Study 1 provides descriptive evidence, this section will focus on the empirical strategy for Study 2. In Study 2, we will test the causal effects of different comment sections on user engagement and behavior. The analysis will primarily assess how the presence and narrative of the comment section affect subsequent engagement with the posts, as well as off-platform actions such as clicks, email sign-ups, and donations. The focus will be on comparing the different comment conditions (No Comments, Mixed Comments, Conservative Comments, and Progressive Comments) and examining how variations in the comment section influence users' interactions with the posts.

For the outcomes listed above, we will estimate the effects of treatment assignment using both tests for comparisons of means and regression analysis. For each social media post, the primary outcomes are observed at the aggregate level (e.g., clicks, comments, sign-ups, etc.). Since we know the number of individuals reached by each post and the number of those who took an action, we can infer outcomes Y_i for each individual i as a realization of $Y \sim \text{Ber}(p)$, where p represents the proportion of positive outcomes (e.g., clicks, comments, sign-ups) at the post level at the end of the study.

To estimate the treatment effects of exposure to the comment section, we will compare the means of outcomes Y across individuals in the four conditions. We will rely on the χ^2 test for equality of proportions with continuity correction to compare the different groups.

Moreover, we will use regression analysis to compare the average treatment effects across conditions accounting for individual and post-level covariates such as age and gender. Specifically, define $Y_{i,p}$ as an outcome for individual i on post p at the end of the study. Define $G_{i,p}$ as the treatment group indicator, which is the same for all individuals assigned to post p . Define $\mathbf{X}_{i,p}$ as a vector of covariates such as the political ideology-racial composition of the audience of the post, or age

groups and gender of the individuals. We will estimate the treatment effects using the following regression:

$$Y_{i,p} = \beta G_{i,p} + \tau \mathbf{X}_{i,p} + \varepsilon_{i,p}. \quad (1)$$

We will cluster the standard errors at the post level p , to account for intra-cluster correlation of the observations among individuals exposed to the same post.

To analyze the effect of the number and presence of comments, we will group the treatment conditions together and compare them to No Comment condition. To study the effect of the narrative of the comment section, we will compare across the Mixed Comments, Conservative Comments, and Progressive Comments conditions.

To study heterogeneous treatment effects, we will estimate the regression separately for different moderators, or include interaction terms. We will analyze heterogeneity by political ideology, racial composition, age, and gender.

Finally, we will also analyze the outcomes aggregated at the post level, through both comparison of means and regression analysis as described above. Post-level analysis also allows us to compare social media activity for secondary outcomes (e.g., engagement and dimensions of comments and conversations) across conditions. In this case, the analysis will be conducted both at the post- and the comment-level.

3.2 Pre-Analysis Plan

The pre-analysis plan specified the intervention, the content creation process, estimating equations, moderators, and variable construction.

Online Appendix: Not for Publication

The Impact of Comments on Social Media Campaigns

Dante Donati and Lena Song

A Experimental Design Appendix

List of Issues and Description

- **Voter Suppression:** Black communities face deliberate barriers like restricted polling access, strict ID laws, and voter roll purges, aimed at limiting their voting power. Misinformation campaigns also target Black voters to reduce turnout, undermining fair representation. Breaking down these barriers is crucial to ensure Black voices are heard in democratic processes.
- **Environmental Justice:** Black communities often live near pollution sources like factories and highways, leading to higher rates of health issues such as asthma. These neighborhoods are frequently overlooked in clean-up efforts and lack green spaces. Environmental justice aims to provide Black communities with clean air, safe water, and healthy environments.
- **Criminal Justice or Police Reform:** Black communities experience disproportionate police violence, profiling, and harsher sentencing. This systemic bias erodes trust in law enforcement and perpetuates disadvantages. Police reform is essential for fair treatment, accountability, and ensuring Black communities feel protected, not targeted, by the justice system.
- **Education Reform:** Black students often attend underfunded schools with fewer resources, larger classes, and limited access to advanced courses. These disparities create achievement gaps and limit future opportunities. Education reform seeks equitable funding and support to provide Black students with the quality education they deserve.
- **Technology fairness:** Black communities face systemic biases in technology, from algorithmic discrimination in hiring and lending to facial recognition tools that disproportionately misidentify Black individuals. These inequities perpetuate existing racial disparities and limit opportunities. Ensuring technology fairness involves designing inclusive systems, addressing bias in algorithms, and creating tools that serve all communities equitably.

Figure A1: Ad Banners and Headlines



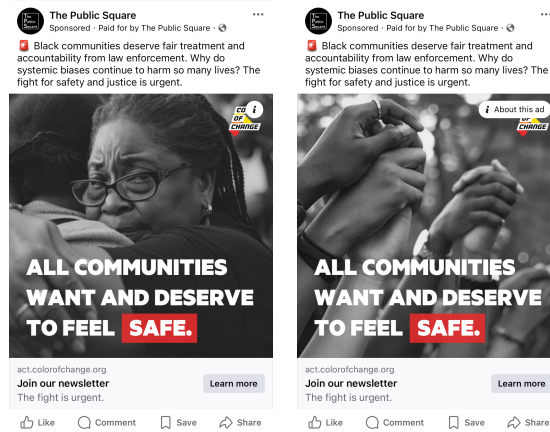
(a) Education

(b) Environment



(c) Voting

(d) Technology



(e) Police