

Pre-analysis Plan:
The Impact of Social Media Comments on Beliefs, Attitudes, and
Behavior

Dante Donati and Lena Song*

February 6, 2026

*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 an online survey experiment, we vary exposure to the comment section and estimate its impact on beliefs, attitudes, and behavior.

2 Experimental Design

2.1 Recruitment

We will recruit around 6,000 participants from Prolific. Before starting, eligible participants are shown a study description that says “This survey takes approximately 10 minutes. You may also earn a bonus of up to \$1 and be entered into a lottery for a chance to win \$100.”

The inclusion criteria are: US resident, age 18-64, not a pilot participant, completed at least 20 previous submissions on Prolific and has an approval rate of at least 95 (out of 100). The sample will be stratified to achieve roughly equal number of self-identified Conservatives, Moderates, and Liberals (subject to recruitment constraints on Prolific), and will be gender balanced.¹ This allows us to study heterogeneous treatment effects across political ideology and gender.

2.2 Survey Experiment Overview

The survey begins with a screener and informed consent, followed by baseline measures of political attitudes, views on racial justice, social media use, demographics, and attention checks.

Participants are then shown multiple social media posts promoting racial justice. These posts are drawn from a related field experiment in collaboration with Color of Change, a leading racial justice advocacy organization in the U.S.

To expose participants to the content while minimizing the perception of being persuaded, we ask them to react to each post as if they came across it on their Facebook feed.

Exposure to the comment section is randomized across participants. In the control condition, participants view posts without a comment section. In the Supportive condition, each post includes two comments that express support for racial justice. In the Opposing condition, each post include two comments that express opposition.

After reacting to the posts, participants complete a set of post-exposure measures capturing beliefs, attitudes, and behavioral outcomes. These include incentivized, high-stakes measures such as an opportunity to donate real money to Color of Change and the option to sign up for the organization’s mailing list. We also measure second-order beliefs, attention, cognitive responses, and

¹At the time of the study, fewer Prolific participants self-identify as conservative or moderate than as liberal. If it is not feasible to recruit equal numbers of conservatives, moderates, and liberals, we will instead aim to recruit a sample composed of around 50% conservatives and moderates combined and 50% liberals.

emotional responses to examine potential mechanisms through which comment sections influence online engagement. Finally, we ask questions that directly assess perceptions of the comment section.

2.3 Intervention Design and Randomization

The intervention randomizes the presence and stance of the comment section. We randomize participants to one of three conditions with equal probability: a control condition with no comment section, a Supportive condition featuring comments supportive of racial justice, and an Opposing condition featuring opposing comments.

All participants see posts about several sub-issues related to racial justice. Within each condition, the comment section associated with each post has the same visibility and stance. For example, in the Supportive condition, each post displays two supportive comments. The posts are otherwise identical across conditions.

The comments used in the survey are real comments collected from Facebook users during a related field experiment. To protect user privacy, we replace user’ names and profile pictures with generic placeholders, using one male and one female profile.

2.4 Outcome Variables

In this section, we describe both primary and secondary outcome variables.

We group some outcome variables into families, as described below. When multiple outcomes fall within the same family, we construct a summary index to combine them.

To construct each index, we first re-orient the outcome variables within each family so that higher values reflect greater alignment with the racial justice post or organization. We then standardize each variable using the control group mean and standard deviation. Finally, we calculate the average of the outcome variables, weighted by the inverse covariance matrix, following Anderson (2008).

2.4.1 Primary Outcomes

Newsletter Sign-up We measure participants’ willingness to sign up for the Color of Change newsletter. As a high-stakes behavioral outcome, participants are asked to provide their email address if they wish to subscribe. Participants are informed that their email address will be used to automatically register them for the Color of Change mailing list.²

²This is the only point in the study at which personally identifiable information is collected, if the participant chooses to sign up. Although willingness to sign up is assessed immediately after the intervention, the email address is requested only at the end of the survey in order to minimize any potential influence on responses to earlier questions.

Donations As a high-stakes, incentivized behavioral outcome, we elicit participants’ willingness to donate to Color of Change. Participants are automatically enrolled in a lottery and are asked to commit in advance to a donation amount, if any, before learning whether they have won. If selected as a lottery winner, the committed donation amount is deducted from their prize and transferred directly to the organization. We will analyze both the extensive margin (whether a participant donates) and the intensive margin (the donation amount) as outcome measures.

Post-exposure attitudes We measure attitudes using several post-exposure questions. These include attitudes toward the organization, willingness to engage in discussions with progressive and conservative people, perceived importance of specific sub-issues within racial justice (e.g., equity in access to education and environmental justice), and opinions toward Black or African Americans. For the issue-importance measures, we focus on sub-issues explicitly covered in the posts shown during the intervention. We aggregate these measures into a post-exposure Racial Attitudes index.

Beliefs about Others To measure the effect of comments on second-order beliefs, we use an incentivized question that elicits participants’ perceptions of the percentage of people holding conservative attitudes, using a quadratic scoring rule. Participants can earn up to \$1 based on the accuracy of their responses, a nontrivial incentive given the short length of the survey and the baseline participation payment.

Attention to the Post We measure attention using the time participants spend on the social media posts. We aggregate time spent across posts at the participant level and apply a $\log(1 + \text{time})$ transformation to reduce the influence of outliers.

Cognitive Responses We use a Likert-scale measure to assess participants’ interest in seeing additional comments on the post, which we interpret as a measure of curiosity or cognitive response to the content. We also ask participants how thought-provoking they find the post itself. We combine these two measures to construct a cognitive response to post index.

For participants in the Supportive and Opposing conditions only, we also ask, “How thought-provoking do you find these comments?” When comparing the two treatment arms, we combine this measure with the cognitive response to post index to construct an overall cognitive response index capturing the extent to which the comment section encourages deeper thinking or curiosity.

Emotional Responses For participants in the Supportive and Opposing conditions, we also ask to what extent the comments make them feel angry and, separately, annoyed. We combine responses to these Likert-scale questions to construct an emotional response index.

2.4.2 Secondary Outcomes

Post-exposure attitudes As secondary outcomes, we examine whether exposure to comments on Color of Change posts affects attitudes toward related organizations and issues. In particular, we analyze attitudes toward Black Lives Matter, as well as the perceived importance of racial justice sub-issues that are not directly covered in the intervention posts.

Cognitive Responses As secondary outcomes, we examine two additional measures of cognitive response: (i) the extent to which the comments increase curiosity about the topic, and (ii) the extent to which the comments increase curiosity about the organization. These measures allow us to assess which aspects of the content—the underlying topic or the organization itself—are most affected by exposure to different comment sections.

Beliefs about Commenters We measure beliefs about who is more likely to comment on the post, separately by political ideology and gender. Specifically, we ask participants: “In your opinion, which group is more likely to comment on this post?”

2.4.3 Open-response questions

We ask an open-response questions “What do you think is the most important issue facing Black people today?” We will analyze the response length, its topic, and sentiment as secondary outcomes.

We will also analyze “What are the main reasons you read or look at comments on social media?” to provide descriptive evidence on why and how individuals engage with comment sections.

2.5 Power calculations

We plan to recruit approximately 6,000 participants to obtain a final analysis sample of about 4,800 high-quality complete responses.³ This sample size provides 80% power to detect a minimum detectable effect (MDE) of 0.10 standard deviations.

2.6 Dealing with Experimenter Demand Effect

We take several steps to minimize experimenter demand effects. First, immediately after screening, we highlight the non-partisan nature of the study. In designing the survey instruments, we include statements reflecting both progressive and conservative positions on racial issues to avoid signaling a preferred viewpoint.

³We assume that approximately 20% of responses will not be included in the final analysis sample due to attrition or failing attention or data quality checks.

Second, we do not collect names or any identifying contact information, ensuring a high degree of anonymity and privacy. This design choice is intended to reduce social desirability concerns and encourage truthful reporting.

Third, participants in all experimental conditions are exposed to racial justice posts. The main post content is identical across conditions, and only the presence and stance of the comment section varies. The comments are designed to resemble a natural component of a social media post, further reducing the salience of the experimental manipulation.

3 Empirical Strategy

3.1 Empirical Strategy

To estimate treatment effects, define Y_i as an outcome for participant i . Y_i could represent an outcome variable measured directly from the survey or an index constructed from multiple survey items. Define G_i as the treatment group indicator. Define $\mathbf{X}_i$ as a vector of covariates measured in the main survey before the intervention. We estimate treatment effects using the following specification:

$$Y_i = \beta G_i + \boldsymbol{\tau} \mathbf{X}_i + \varepsilon_i \quad (1)$$

The covariates include baseline measures of racial attitudes and beliefs, including responses on equal rights, attitudes toward public attention to the history of slavery and racism, self-reported position on racial issues (conservative, moderate, or progressive), and beliefs about the percentage of others holding conservative and, separately, progressive views.

We potentially control for education, gender, race, age, political party affiliation, and location if any covariates are unbalanced. These controls are coded as indicator variables for education (college degree or higher), gender (male, female, other), race (White, Black, other), age (above the sample median), and political party affiliation (Republican, Democrat, other).

Our primary analysis focuses on comparing the Supportive and Opposing conditions. This comparison isolates the effect of the stance of the comment section on outcomes measured among participants in the treatment conditions.

For outcomes measured for all participants, we will analyze the effect of the presence of the comment section by pooling the Supportive and Opposing conditions and comparing them to control. Finally, to decompose the effects of supportive and opposing comments, we separately compare each treatment condition to the control condition.

To examine heterogeneous treatment effects, we estimate treatment effects separately for racial progressives and for racial moderates and conservatives combined. If we recruit a sufficient number of racial conservatives, we additionally estimate effects separately for racial moderates and racial

conservatives (as identified by the response to the equal rights question or an index constructed from baseline racial attitude measures). We will also analyze heterogeneity by party affiliation, age, and gender.

References

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American Statistical Association*, 103(484):1481–1495.