# r/Political: Does political signaling change online engagement?

Nura Hossainzadeh, Sam Kane, Ellie Parker, Evan Powell, and Paul Terrasi
DS241 Field Experiments and Casual Inference
Professor Scott Guenther
20 April 2026

## Abstract

Our study examines whether political identity signaling in usernames affects user engagement with otherwise apolitical posts on Reddit. We implement a randomized field experiment in which 180 posts are assigned to accounts with neutral (baseline), liberal-signalling (“no_kings”), or conservative-signaling (“MAGA”) usernames and posted across a set of 30 general-interest subreddits. We measure engagement using two outcomes: net score (upvotes minus downvotes) and total interactions (upvotes, downvotes, and comments), both adjusted for post view counts and transformed to address skew in the data. We do not find strong evidence that political signaling affects post-level reception as measured by net score. Although a statistically significant negative effect appears for liberal usernames in the non-attrited sample, this result does not persist across models incorporating imputed data, suggesting that it may be driven by attrition. In contrast, we find evidence across all versions of our analysis that conservative-signaling usernames receive fewer total interactions than neutral usernames. This corresponds to a conservative treatment effect coefficient of approximately -0.05, meaning that at a typical level of exposure around 30,000 views, conservative posts receive about 40% fewer interactions than otherwise similar neutral posts. However, the experiment is underpowered following outcome transformations and attrition, which limits our ability to detect smaller effects and interpret null results. Taken together, these results suggest that political identity signaling does not have a robust effect on post-level reception, as measured by net score, but may influence whether users engage with content, as reflected in total interactions.

## Introduction

Online platforms increasingly mediate how individuals interact, evaluate information, and engage with others. Even in settings where content is neutral, users may rely on observable cues about a poster’s identity when deciding whether and how to respond.

This study investigates whether political identity signaling, conveyed only through an individual’s username, causally affects engagement with otherwise neutral content on Reddit. Specifically, we ask: Do posts made by users with politically-signaled usernames generate different levels of interaction than identical posts made by users with neutral usernames? We define political

identity signaling as three conditions consisting of neutral (control) usernames, liberal-signaled usernames, and conservative-signaled usernames.

To answer this question, we must use a causal framework. Observational comparisons of engagement across users with different usernames are likely to be confounded by unobserved differences in conduct, posting style, subreddit/audience selection, or general behavior. For example, users who choose politically expressive usernames may differ systematically from others in ways that also influence engagement – perhaps they write with more passionate diction or are more likely to post controversial questions in political subreddits. As a result, simple correlations between politically signaled usernames and engagement cannot be interpreted as causal. To address this challenge and assess causality, we implement a randomized field experiment in which we intentionally vary political signaling while holding post content constant.

In the experiment, we create Reddit accounts with three types of usernames: politically neutral, liberal-signaled, and conservative-signaled. The political signaling in our treatments reflects the current U.S. political environment, including anti- and pro-Trump sentiment, which we discuss further in the Justification section below. These accounts then post carefully constructed, non-political questions across a set of general-interest subreddits. We randomize the assignment of the usernames to post within each subreddit, ensuring that each subreddit receives an equal number of posts from each treatment condition. This blocked randomization design ensures posts are comparable across treatment and control conditions, except for the political identity which is conveyed by the username, allowing us to isolate the causal effect of identity signaling on user engagement.

Our outcomes of interest capture two distinct, but complementary, dimensions of user response. We measure post-level reception by the net score (upvotes minus downvotes) within 24 hours after posting, reflecting whether responses are overall positive or negative and how strongly users respond. Overall engagement is measured by total interactions (upvotes, downvotes, and comments), which reflects the extent to which users interact with the post. In our analysis, we adjust these outcomes, first by normalizing by view count to account for algorithmic boosting of posts based on view count, and second by logging both the scores/interactions and the view counts to reduce skew in the distribution of post-level engagement. Together, these measures allow us to distinguish between how users react to the content and how much they engage with it.

## Justification

This experiment taps into longstanding and deeply researched psychological and sociological questions surrounding, on the one hand, abstract conceptions of in-group/out-group behavior and the use of identity-based heuristics to form judgments or conceptions about another person; and on the other, features of the contemporary American cultural and political landscape, which has for decades been characterized in the academic literature as polarized. This political polarization has its roots in the civil rights era (Abramowitz and McCoy, 2018), when Lyndon Johnson aligned the Democratic party with the civil rights cause and the white working class

began moving into the Republican party. This trend was reinforced in the 1980s, when Ronald Reagan opposed Roe vs. Wade and religious conservatives were pulled more into the Republican party. Since then, the Republican party has attracted citizens with more traditional, religiously-informed views on family life and other social institutions, as well as a belief in a more limited role for government in addressing socioeconomic divisions, while Democrats have placed greater emphasis on minority rights and using government resources to alleviate socioeconomic divisions.

The newest character in the saga of American political polarization is Donald Trump. Trump has drawn on pre-existing political and cultural divisions in American society (Segal, 2025), but has shifted political divisions away from traditional party lines, creating a new and serious political divide (Finkelstein, 2025) centered on support for or opposition to his "Make America Great Again" (MAGA) agenda. In other words, to a significant extent, Trump himself has become the faultline upon which Americans are divided. His agenda includes aggressive immigration enforcement, implementing tariffs to favor domestic production, and protecting traditional and conservative positions on gender and women's rights. On the other hand, his opponents argue that his policies violate the rights of women, immigrants, and sexual minorities and undermine a more inclusive conception of American polity and society. These opponents often adopt the slogan "no kings" to express their disapproval of Trump purportedly acting in a king-like manner, such as by using the levers of executive power to curb the rights of minority citizens and non-citizens as well as undermining institutions and norms for his own personal and political gain (see Lonsdorf, 2025; Campaign Legal Center, 2026; and Haddock & Roy, 2026). Crucially and perhaps tragically, running through these strong political disagreements is often a very strong personal, emotive component (Kleinfeld, 2023), where fellow citizens on each side of the political divide see each other as having fundamentally different worldviews and moralities. Citizens who view politics in this way have a difficult time imagining that there could ever be compromise or agreement across lines of political division.

In this study, we explore how the political divide centered on Trump may affect everyday interactions between individuals, even when they discuss apolitical topics. Since Americans increasingly perceive political differences as indicative of contrasting worldviews—and, to put it simply, since Americans often do not like people who disagree with them politically (Kleinfeld, 2023)—we aim to explore how personal feelings and perceptions stemming from political differences impact how Americans interact with each other. In particular, we seek to measure how Reddit users represented by usernames that affiliate with either a positive or negative stance toward Trump (either including "MAGA" or "no_kings" in the handle) respond to requests for help and advice on the platform. This is compared to a neutral username control.

Going beyond the contemporary American context, our study also relates to longstanding research on how identity markers and in-group out-group divisions affect human interaction. Researchers in psychology and communication have long studied the effects of identity markers on trust and engagement between conversants (see, for example Chaiken, 1980). These identity markers often serve as heuristics that individuals use as a basis for deciding whether and how to engage with someone, and how much trust to put in what they say. In other words,

rather than relying solely or predominantly on the substance of an individual's speech when determining how to respond, individuals often draw on identity-based heuristics. Relatedly, much research has been completed outside of the social media context on features of in-group/out-group behavior and the emotional bases of this behavior (see, for example, Wagner and Christ, 2007).

Based both on this abundant decades-old research that predates the digital age and more recent literature on discourse in social media specifically (for example, Hutchens, et. al., 2019), we anticipate that identity heuristics and in-group/out-group behavior will also be key to understanding the context we seek to study. Individuals who receive treatment—who come across questions posed by users with blatantly political handles/usernames—will rely on the username as a heuristic for deciding whether to engage with the person, even if the user is not asking a question on a political topic. Engaging in conversation with a stranger, particularly online, always entails some risk (with a central risk being that they may not respond courteously) and also requires time and energy, and individuals will likely decide to rely on cues implied by usernames to decide whether to engage in conversation. If the username evokes positive feelings—in other words, taps into a side of the culture war that the individual has sympathy towards—they will likely engage with the person, but they will make the opposite decision if the username does not evoke these feelings of safety and comfort.

On this basis, we hypothesize that signaling political identity affects how people engage, not because of disagreement with the posted content, but because users react to the identity cues associated with a poster. We expect two engagement metrics to reveal this trend: net score and total interactions, with significant differences in scores and interactions between posts published by treated and untreated usernames. More specifically, we anticipate that conservative usernames will receive less engagement since Reddit generally leans left, and Reddit users will tend to feel less comfortable engaging with conservative usernames. The Data section lists specific hypotheses and their respective testing procedures.

# Experiment Design

## Comparison of Experimental Outcome

Our subjects are drawn from the population of Reddit users. During their normal browsing habits, they will encounter a post that is a part of our experiment. We do not explicitly control which subjects see which posts, but instead randomize the subreddit, topic, and posting date such that exposure to any given treatment group is effectively random. The subject receives the treatment at the time of encountering the post; exact presentation will vary based on the Reddit interface that each subject is using, but in most cases, the user will see the post title and username simultaneously along with possibly some of the post body. Once the subject has spent enough time with this summary on their screen to read it, they have received the treatment and the view count for that post increases by one (wa2b, 2022). The subject may or may not open the post to read the full body and existing comments, but this is not recorded. Instead, our outcome of interest is whether the user chooses to interact by either upvoting,

downvoting, or commenting. Measurements are made at the aggregate level by post, looking at cumulative interactions (either vote score or total interactions) normalized by the total number of views at the 24 hour mark. We then compare these normalized interaction counts between the posts in both treatment groups and control groups.

## Randomization Process

We started our experiment with 15 Reddit usernames (5 for each treatment group), each of which would post 12 times over the course of the experiment to meet the 180 posts dictated by our power analysis. To randomize the assignment of these accounts to posts and posting days, we first created a posting pool of 60 usernames by placing each username into the posting pool 12 times, once for each time it would be used to post. With this evenly split username posting pool, there would be an equal number of posts from each account across all subreddits, and an equal posting burden across account owners.

In our randomization process, two usernames from each treatment group were randomly selected for each subreddit and assigned in random order to the subreddit posts. To spread out posting between subreddits, each account holder was randomly assigned 12 posting days from the 20-day experiment window. A more condensed posting period could increase the risk of users seeing posts from two treatment posts in the same subreddit and having their reaction to one impact their reaction to the other. Spreading out post publication times therefore decreased potential spillover from users seeing multiple experiment posts during a single Reddit-scrolling session. Each username-subreddit-post combination was then randomly assigned a posting day selected from the account holder’s assigned post days.

This randomization process was repeated until no single username was assigned to post more than once in a given subreddit, and no subreddit received more than two posts in a day. Figure 1 below shows the randomization within the experiment design.

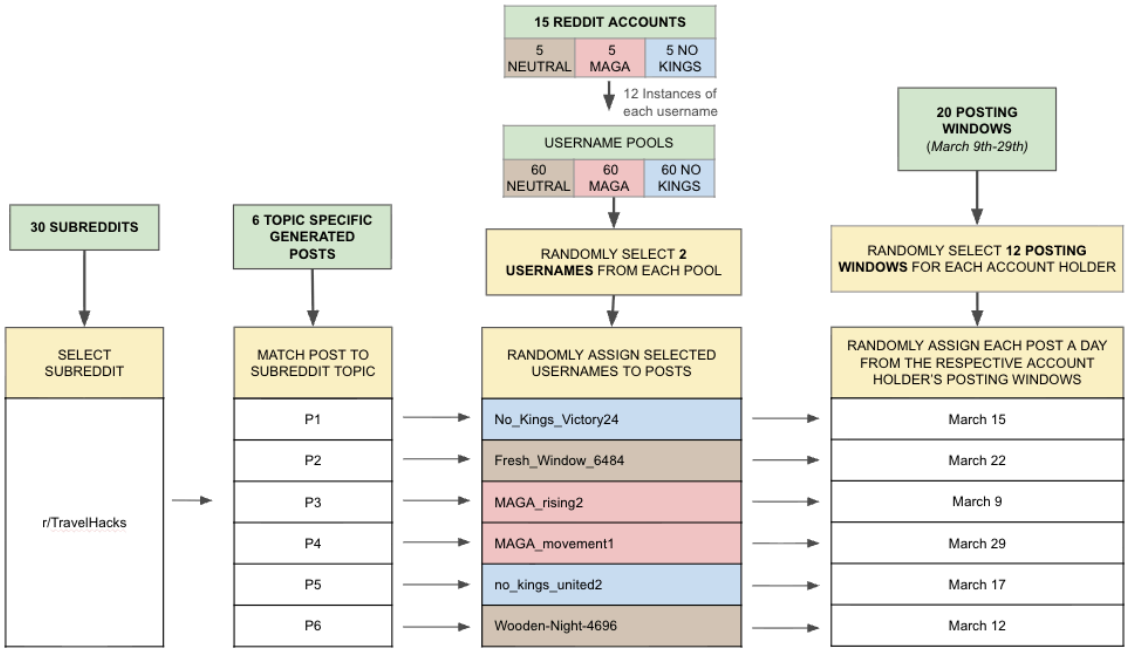



**Figure 1**: Experiment randomization process.

The random assignment of posts to usernames ensured equal distribution of post characteristics across experimental conditions, including any potential difference in interest level and political lean. These characteristics were assessed using a survey described below, which revealed only moderate differences across most posts. Appendix B details the randomization checks conducted on treatment group post assignments.

### Deviations and Solutions

Throughout the experiment, several deviations were made from the proposed randomized posting schedule. Three instances of human error resulted in swapping assigned posts between different accounts; however, in each instance we were able to adapt the remaining posts to ensure that every subreddit still received the same number of posts from each treatment category. Additionally, due to posting requirements in certain subreddits, some posts were posted at a later date than originally scheduled to allow accounts to meet required karma thresholds. However, during experiment planning we intentionally added an additional 3 days at the end of the experiment window to account for any delayed posts, ensuring that these posts were still captured in the planned measurement period.

## Treatment

We created 15 reddit accounts whose usernames represent three experimental conditions: 5 neutral usernames, 5 liberal usernames, and 5 conservative usernames. The usernames with overt political signaling represent the two treatment conditions (liberal and conservative), while the neutral usernames represent the control condition.

We used Reddit's auto-generated usernames when creating the control accounts. The treatment usernames contained the terms "no_kings" and "MAGA," to which we appended five words that evoke political action, plus a number, to keep the usernames as similar as possible aside from their political signaling. We purposefully added the five words to create slightly more variation in the usernames than if we had only appended numbers to mitigate the risk of subreddit moderators or frequent users from detecting sudden increases in posts from near-identical handles. Table 1 summarizes the handles by experimental condition.

| Control Condition: Neutral Usernames | Treatment Condition 1: Liberal Usernames | Treatment Condition 2: Conservative Usernames |
|---|---|---|
| CharacterDig2229 | no_kings_victory24 | MAGA_victory24 |
| Wooden-Night-4696 | no_kings_movement1 | MAGA_movement1 |
| Fresh_Window_6484 | no_kings_now1 | MAGA_now1 |
| Interesting-Nerve663 | no_kings_united2 | MAGA_united2 |
| Evening_Issue_8448 | no_kings_rising2 | MAGA_rising2 |

**Table 1.** Reddit handles by experimental condition.

Before starting the experiment, we deployed a survey to ensure that the usernames remained balanced in their political signaling despite the slight variation across each experimental condition's handles. The survey asked respondents to score each username's political leaning on a 5-point ordinal scale ranging from "Far Left" to "Far Right". Across all three username groups, the 57 responses overwhelmingly coincided with our intended political signaling: on average, 94% of respondents categorized the liberal usernames as leaning "left" or "far left"; 96% categorized the conservative usernames as leaning "right" or "far-right"; and 89% categorized the neutral usernames as neutral. Appendix A provides a detailed summary of the survey results.

We generated 180 posts to randomly assign to each of the 15 usernames to post across pre-selected subreddits. We constructed the content, tone, and structure of each post to remain neutral and non-political, with the subject matter relating generically to each subreddit's theme or topic. The aim was for the only difference between the conditions to be the political identity signaling conveyed by the username. This would allow us to isolate the effect of identity signaling on engagement and response behaviors. We verified that posts within each subreddit were balanced in interest level and political lean through a survey that received 24 responses, detailed in Appendix B.

## Population and Sample

In this study, Reddit posts comprise the unit of analysis. The population of interest is Redditors who read popular subreddits, which we define as subreddits with more than 200,000 weekly viewers and at least three new posts published daily, on average, in the week leading up to the experiment. We shortlisted potential experimental subreddits from a combination of subreddits

listed on the platform's 'Popular' page, search engine results for active subreddit communities, and an LLM-generated list of 50 high-activity subreddits.

To mitigate harms, we opted to use political signals that are ubiquitous and easily recognizable but neither discriminatory nor offensive. We also ensured that the post content was apolitical and respectful, and ideally posed questions that would add value to the subreddit community and not just create internet waste.

To build account credibility and meet minimum karma requirements for posting to some of our selected subreddits, we posted and commented in non-experimental subreddits during the weeks leading up to the experimental period. While attempting to publish some of the posts scheduled for the first five days of the experiment, we realized that some accounts did not have sufficient karma, so we continued to post to non-experimental subreddits to reach operable karma minimums. We ensured that all posts were politically neutral and that all comments were respectful and benign (e.g., "Amazing", "I never thought of it that way") and, if possible, genuine and helpful while maintaining a neutral, generic tone.

## Power Analysis

Prior to launching the experiment, we conducted a power analysis to identify the combination of subreddits (blocks) and posts per subreddit needed for our experiment to detect effect sizes of 5, 15, and 25 more votes, on average, by treated usernames than neutral usernames. This analysis determined that we would need 30 subreddits, each with six posts, to detect a minimum effect size of 15 more votes in either treatment condition compared to the neutral baseline with 80% confidence.

At the time of analysis, we had chosen net score as a primary outcome of interest and had not yet determined the need to log transform nor to normalize by view count. To calculate estimated power, we scraped vote metrics from 20 potential experimental subreddits. We truncated this scraped vote distribution at 100 votes, then randomly generated score values for each of our experimental usernames, with no systematic differences among experimental groups. We then ran randomization inference for each of the three effects sizes to systematically inflate scores for posts published by liberal and conservative usernames, compared to a neutral baseline. Selected test effects sizes of 5, 15, and 25 followed Taylor et. al's (Taylor. et. al. 2023) findings that identity visibility increases the likelihood that an average exposure results in an upvote by 5%, while the effect on downvotes is small and not statistically significant. As such, a 5-vote increase in net score was the baseline effect size, with medium (15) and large effects sizes (25) reflecting a three- and five-fold relative increase, respectively. Figure 2 displays the results of this initial power analysis.

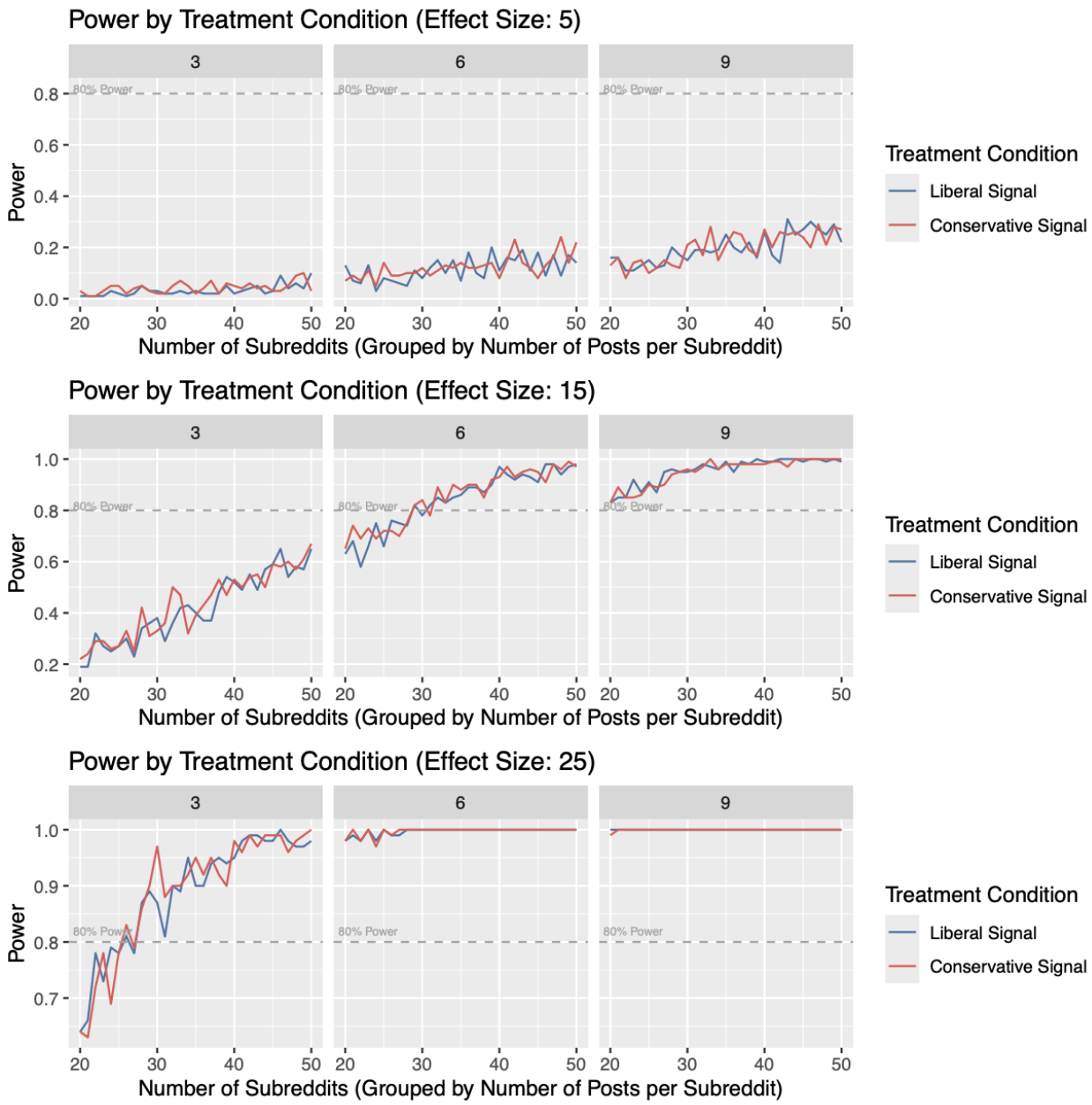



**Figure 2.** Initial power curves for detecting effects on raw net score. The experiment targeted 30 subreddits with an effect size of 15 to reach the 80% baseline confidence.

We updated our simulation-based power analysis after conducting the experiment to reflect the outcome variable log transformations and normalization and understand how these manipulations affected our anticipated power. Because we operationalized the outcome variable as logged net score, normalized over logged view count, we used the same pre-experiment scores distribution to simulate scores and logged them, simulating the outcome variable's numerator. We then used the experiment's raw (unimputed) view counts to generate a beta-distribution with mean 29,169, standard deviation 108,041, and range from 0 to 898,346. We simulated the outcome variable's denominator by taking random draws from this beta

distribution and logging them. This simulated view count distribution mimicked the view count values we saw in our experiment. We retained the same effects sizes of 5, 15, and 25 more votes for treatment conditions than control, transforming the values accordingly within the power simulation function.

The updated analysis found that detecting a small effect (an average difference of 5 votes, or about 0.16 logged and normalized votes), medium (an average difference of 15 votes, or about 0.26 logged and normalized votes), or large effect (an average difference of 25 votes, or about 0.31 logged and normalized votes) at least 80% of the time would not be possible with up to 50 subreddits and 9 posts each. Figure 3 plots the power curves for 20 to 50 subreddits, each with 3, 6, or 9 posts, under small, medium, and large normalized effect size conditions. Unfortunately, our planned 30 subreddits with 6 posts each would only achieve an estimated power of below 10%, without attrition, for the large effect size. As a result, our experiment was severely underpowered. (For more information on both the unnormalized and normalized analyses, see the power analysis folder in the project repository.)

Underpowered studies carry the risk of failing to reject the null hypothesis when it is false (Type II error), generating false-negative results because there are not enough data points to surpass the threshold for significance and thus detect a true effect. Statistically significant results in an underpowered study can result in an exaggerated estimate of a true effect (Type M error) because greater sampling variability widens the distribution around the population parameter, increasing the probability of observing a higher proportion of extreme values. This is particularly true when Type I errors occur, where the effect size will tend to be larger in underpowered studies.

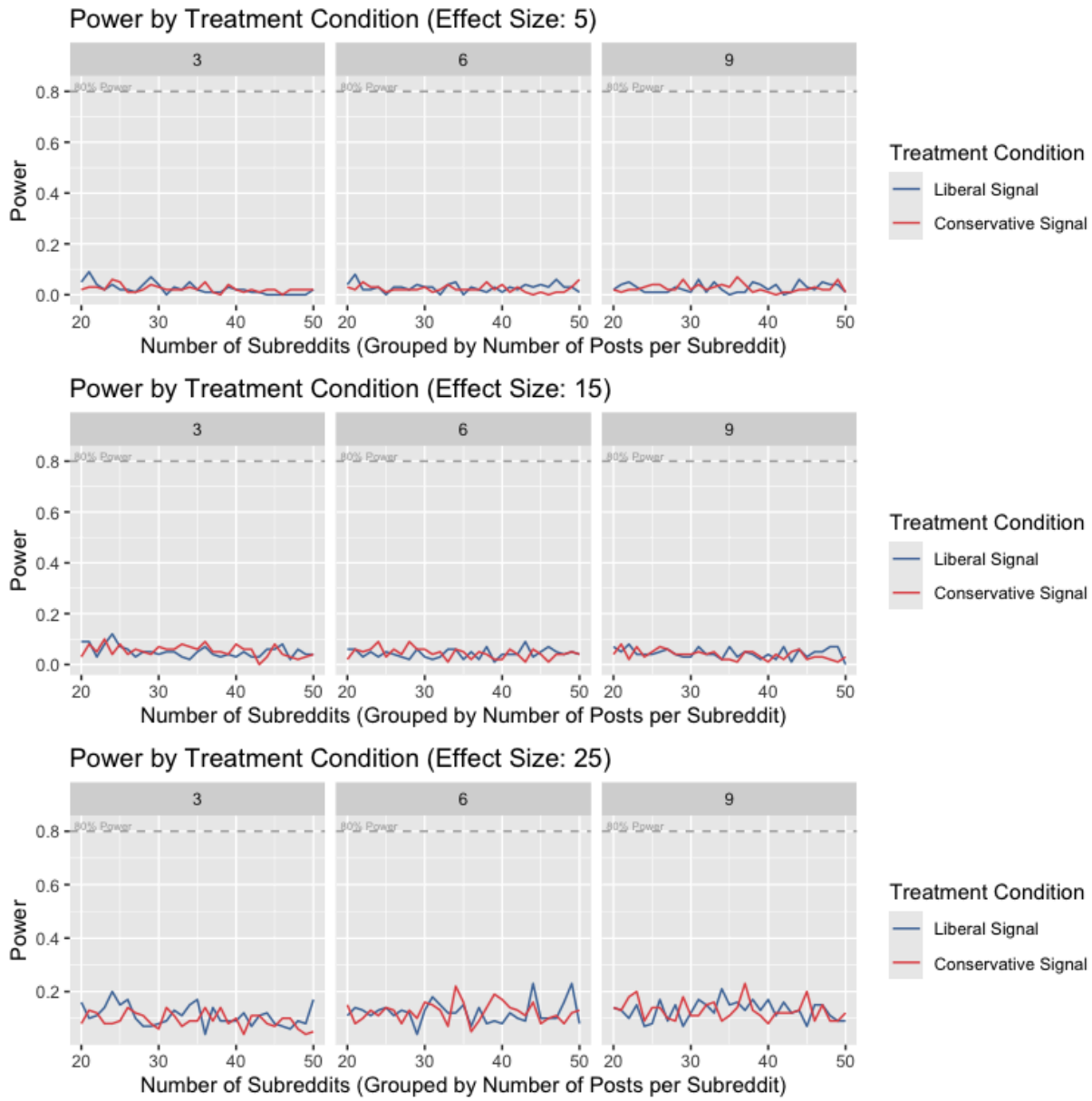



**Figure 3.** Power curves for detecting effects on normalized scores. This shows that our experiment was ultimately underpowered.

# Analysis

## Data

As stated previously, individual posts constitute our unit of analysis. Posts were randomly assigned to neutral or politically signaled usernames, representing control and treatment conditions, respectively. Posts assigned to treatment were either assigned to a liberal (“no_kings”) or conservative (“MAGA”) signaled usernames. The two treatment conditions are indicator variables in our regression models, with the control condition as the referential

baseline. Because our experimental design blocks by subreddit, each subreddit becomes a covariate in the regression equation.

Reddit generates a suite of metrics for each post—including view counts, net score, upvote ratios, and comment counts—which we gathered 12, 24, and 48 hours after publication using an automated scraper (Appendix C details the scraping automation process). View counts represent the number of Redditors who scrolled over or through a post. Net score is the total number of upvotes minus downvotes, and upvote ratio is the proportion of the net score that constituted upvotes. Comment counts are the number of unique comments on each post. From these Reddit-provided metrics, we calculated the downvotes for each post and the total interactions, defined as the sum of upvotes, downvotes, and comment counts.

Our outcomes of interest are net score and total interactions, both of which we log and normalize over the post's logged view count, after 24 hours of publication. This results in a ratio of logs for both outcomes (for details on the decision to log-transform net score, total interactions, and view counts, as well as the decision to have ratio of logs instead of log ratio outcomes, see Appendix E). We normalize by view count to account for algorithmic bias that may push more popular posts higher up on Redditors' feeds, making such posts more visible and more likely to generate interactions.

### Attrition

Attrition in this experiment occurred when posts were not published as intended and thus were unable to produce measurable results. Examples of attrited posts include those that were unable to be published due to insufficient karma requirements, posts that were automatically or manually removed by subreddit moderators, and a post that was accidentally published in the wrong (non-experimental) subreddit.

Of the 180 experimental posts, 52 (28.9%) attrited. Of these, 8 were never published (due to insufficient karma requirements or because they were posted to the incorrect subreddit) and 16 were immediately removed (autoremoved). These posts therefore generated no usable data from which to impute their missing values. Additionally, 28 of the 52 attrited posts were manually removed. While we can still access vote metrics for manually removed posts, Reddit obscures their view counts, which we must impute.

One subreddit, r/SelfImprovement, had karma requirements that prevented all six posts from being published, rendering the subreddit unobservable. Another subreddit, r/TravelHacks, consisted of 6 unpublished and 2 manually removed posts; subreddit-based imputations on the view counts for the manually removed posts was therefore impossible, since there were no successfully published posts with view counts on which to base this imputation. This left 28 subreddits with observable values for analysis and a total of 26 manually removed posts with missing view counts, recoverable through imputation.

Figure 4 maps out the attrition by treatment group. Chi square tests and logistic regression models found no evidence of significant difference in missingness by experimental condition. No

username group was more likely—at a statistically significant level—to have missing values than others. A separate chi square test found no evidence of significantly different missingness among experimental groups by removal status. No one username group was differentially likely to be manually removed, automatically removed, or not removed. For more detail on how we handled attrited posts and imputations, see Appendix D.

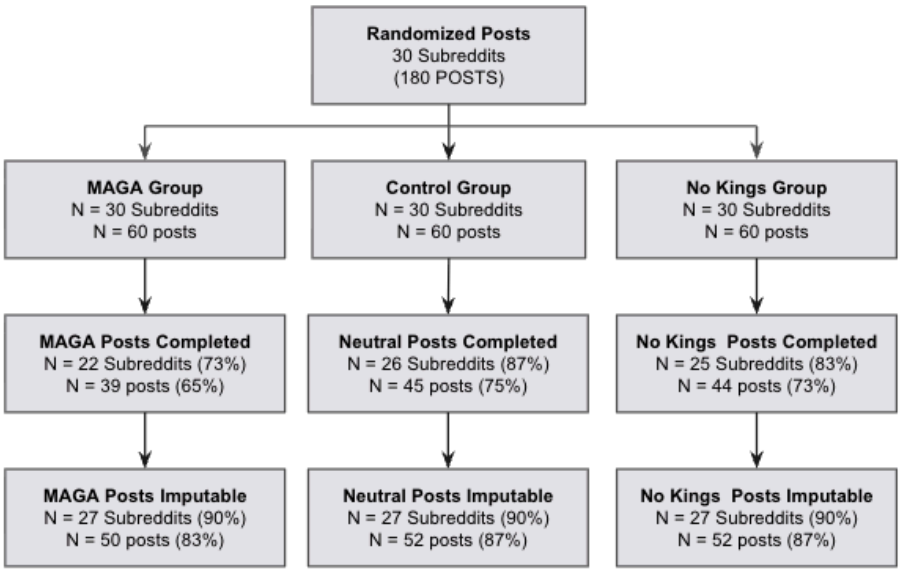



**Figure 4.** Diagram of attrited posts.

## Imputation

Unrecoverable attrition left us with at least partial data for 154 posts spanning 28 subreddits. Of the 462 rows of data pertaining to the 12-, 24-, and 48-hour metrics of the remaining 154 posts, 78 rows were missing view count values. Additionally, two conservative posts were missing their 24-hour upvote ratio due to an auto-scraper error, resulting in missing upvote and downvote data. View count and score metrics for both posts were successfully collected.

Although our analysis looks only at post metrics 24 hours after publication, the 12- and 48-hour metrics collected for each post provided essential information about the underlying metrics' distributions, allowing for imputation.

We followed three parallel imputation methods: one modeled the underlying distribution from the existing data to infer the missing values; another was a lower-bounds approach that sought to generate the minimum possible outcome measurement for each missing value; and the last was an upper-bounds approach to generate the maximum possible outcome measurement for each missing value. This conservative method of using both upper and lower bound imputed values in our analysis helps us assess how large of a difference in treatment effect our results could survive in the presence of attrition, by looking at what would happen to our overarching results if the missing values happened to all be as contrary as possible to our hypotheses.

In each case, we imputed the two missing upvote ratios as well as the 24-hour and removal-time view counts. Table 2 displays the summary statistics for each post metric under each imputation type.

**Imputation Comparisons**

| | Metric | Mean | Median | Min | Max | StDev | NAs |
|---|---|---|---|---|---|---|---|
| 1 | Raw Upvotes | 25 | 4 | 0 | 956 | 103 | 2 |
| 2 | Distribution-Based Upvotes | 25 | 4 | 0 | 956 | 102 | 0 |
| 3 | Lower Bounds Upvotes | 26 | 4 | 0 | 956 | 102 | 0 |
| 4 | Upper Bounds Upvotes | 25 | 4 | 0 | 956 | 102 | 0 |
| 5 | Raw Downvotes | 3 | 0 | 0 | 61 | 8 | 2 |
| 6 | Distribution-Based Downvotes | 3 | 0 | 0 | 61 | 8 | 0 |
| 7 | Lower Bounds Downvotes | 3 | 0 | 0 | 61 | 8 | 0 |
| 8 | Upper Bounds Downvotes | 3 | 0 | 0 | 61 | 8 | 0 |
| 9 | Raw Upvote Ratio | 1 | 1 | 0 | 1 | 0 | 2 |
| 10 | Distribution-Based Upvote Ratio | 1 | 1 | 0 | 1 | 0 | 0 |
| 11 | Lower Bounds Upvote Ratio | 1 | 1 | 0 | 1 | 0 | 0 |
| 12 | Upper Bounds Upvote Ratio | 1 | 1 | 0 | 1 | 0 | 0 |
| 13 | Raw 24-Hour Views | 29,169 | 1,868 | 0 | 898,346 | 108,041 | 26 |
| 14 | Distribution-Based 24-Hour Views | 41,529 | 3,300 | 0 | 898,346 | 132,389 | 0 |
| 15 | Lower Bounds 24-Hour Views | 48,807 | 3,539 | 0 | 898,346 | 157,224 | 0 |
| 16 | Upper Bounds 24-Hour Views | 35,012 | 3,025 | 0 | 898,346 | 112,852 | 0 |
| 17 | Distribution-Based Removal-Time Views | 35,459 | 3,144 | 0 | 898,346 | 115,103 | 0 |
| 18 | Lower Bounds Removal-Time Views | 43,858 | 3,539 | 0 | 921,390 | 141,641 | 0 |
| 19 | Upper Bounds Removal-Time Views | 30,738 | 2,310 | 0 | 898,346 | 103,856 | 0 |

**Table 2.** Post metric summary statistics by imputation type.

For each outcome of interest, we run seven regression models (for a total of 14), each based on a different imputation method, to examine whether the treatment effect withstands these varied approaches:

- Models labeled “Measured” are based on the raw, non-attrited data.
- “Hourly (D)” models use 24-hour view counts, where available, or view counts at time of post removal, imputed using a linear regression model of the corresponding subreddit’s distribution of 12-, 24-, and 48-hour view counts.
- “Weighted (D)” models use 24-hour view counts scaled by the proportion of the time period for which the post remained published. Posts with observed 24-hour view counts were scaled by a factor of 1 (no change). We used the corresponding subreddit’s average 24-hour view count to impute unobserved 24-hour view counts, then scaled them by the removal hour as a fraction of 24.
- “Hourly (LB)” models follow the same imputation process as “Hourly (D)” models, adding the corresponding subreddit’s maximum residual to the imputed view count to generate the minimum possible outcome measurement (by maximizing the denominator engagement-to-view count ratio) for each missing value.

- "Weighted (LB)" models follow the same imputation process as "Weighted (D)" models, except instead of taking the corresponding subreddit's average 24-hour view count, it takes the maximum.
- "Hourly (UB)" models follow the same imputation process as "Hourly (D)" models, subtracting the corresponding subreddit's minimum residual from the imputed view count to generate the maximum possible outcome measurement (by minimizing the denominator in the engagement-to-view count ratio) for each missing value.
- "Weighted (UB)" models follow the same imputation process as "Weighted (D)" models, except instead of taking the corresponding subreddit's average 24-hour view count, it takes the minimum.

## Hypotheses

Because political affiliations can draw in-group/out-group boundaries, we expect treated usernames to generate a different amount and type of engagement than neutral ones:

> $H_{01}$: There is no difference in net scores between posts published by political usernames and posts published by neutral usernames.
>
> $H_{A1}$: There is a significant difference in net scores between posts published by political usernames and posts published by neutral usernames.

An upvote signifies approval, interest, support, or promotion. We expect posts to generate more upvotes when Redditors perceive the poster favorably and as part of, or more closely aligned with, their political "in-group." (We discuss in-group/out-group behavior in the Justification.) While downvotes signify disapproval, disinterest, a lack of support, or intention to suppress, we expect disagreement to materialize as apathy more often than as active downvoting. Because Reddit generally leans left (Godfrey, 2024; see also Ramakko, 2024) as do a lot of online platforms, we expect net score to be higher among the liberal treatment group than the conservative treatment group:

> $H_{02}$: There is no difference in net scores between liberal and conservative usernames.
>
> $H_{A2}$: Liberal username net scores are significantly higher than conservative username net scores.

We expect total interactions to differ between the control and treatment conditions:

> $H_{03}$: There is no difference in the number of total interactions between posts published by political usernames and posts published by neutral usernames.
>
> $H_{A3}$: There is a significant difference in the number of total interactions between posts published by political usernames and posts published by neutral usernames.

As with net score, we expect total interactions to be higher among the liberal treatment condition than its conservative counterpart due to Reddit's left lean. Users may feel more comfortable interacting with liberal usernames than conservative ones, because they perceive these individuals to be part of, or more closely aligned with, their in-group:

> $H_{04}$: There is no difference in the number of total interactions between liberal and conservative usernames.
>
> $H_{A4}$: Liberal username number of total interactions are significantly higher than conservative username number of total interactions.

## Models

To evaluate the relationship between political usernames and our score and interaction outcome variables, we conducted the following two linear regressions:

$$\frac{log(Net\ Score)}{log(Views)} = \beta_0 + \beta_1(Treatment) + \beta_2(Subreddit) + \varepsilon$$

**Equation 1.** Linear regression model for logged net score, standardized by logged view counts.

$$\frac{log(Total\ Interactions)}{log(Views)} = \beta_0 + \beta_1(Treatment) + \beta_2(Subreddit) + \varepsilon$$

**Equation 2.** Linear regression model for logged total interactions, standardized by logged view counts.

We evaluated both regressions using the raw, un-attrited data as well as the range of outcomes calculated from the different imputation strategies, with the goal of understanding the robustness of our treatment effects in the presence of post attrition. In other words, to what extent do treatment effect estimates vary when using imputed attrited values from the different methods described above? Because we blocked by subreddit during treatment assignment at the experiment design level, we added subreddit into our model as a covariate to account for the variation removed in the blocking process. Given the high variability in both outcomes of interest as well as view counts in our data, we used robust standard errors in all models.

## Results

Table 3 shows the results of our regressions for the log-normalized net score outcome. We observe only a statistically significant relationship in the model that uses non-attrited data, which predicts a decrease in liberal usernames post scores relative to neutral username post scores when viewed by the same number of users. We do not, however, see this effect across the other models, which could signify that this relationship is a result of attrition bias in our sampled posts and is not an effect that exists across the larger target population.

**Score Regression Results (Omitting Subreddit Covariates)**

| | Dependent Variable: Log Normalized Score | | | | | | |
|---|---|---|---|---|---|---|---|
| | Measured | Hourly (D) | 24hr Weighted (D) | Hourly (LB) | 24hr Weighted (LB) | Hourly (UB) | 24hr Weighted (UB) |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Conservative | -0.036 | -0.040 | -0.033 | -0.040 | -0.034 | -0.038 | -0.029 |
| | (0.036) | (0.034) | (0.035) | (0.033) | (0.035) | (0.034) | (0.037) |
| Liberal | -0.073** | -0.038 | -0.036 | -0.039 | -0.037 | -0.036 | -0.033 |
| | (0.036) | (0.034) | (0.035) | (0.034) | (0.035) | (0.034) | (0.036) |
| Constant | 0.278*** | 0.125 | 0.114 | 0.126 | 0.115 | 0.125 | 0.112 |
| | (0.024) | (0.085) | (0.098) | (0.083) | (0.095) | (0.087) | (0.101) |
| Observations | 128 | 154 | 154 | 154 | 154 | 154 | 154 |
| $R^2$ | 0.392 | 0.300 | 0.304 | 0.302 | 0.306 | 0.298 | 0.296 |
| Adjusted $R^2$ | 0.212 | 0.137 | 0.141 | 0.138 | 0.143 | 0.134 | 0.131 |
| Residual Std. Error | 0.167 (df = 98) | 0.174 (df = 124) | 0.179 (df = 124) | 0.172 (df = 124) | 0.177 (df = 124) | 0.176 (df = 124) | 0.186 (df = 124) |
| F Statistic | 2.182*** (df = 29; 98) | 1.834** (df = 29; 124) | 1.865** (df = 29; 124) | 1.847** (df = 29; 124) | 1.884*** (df = 29; 124) | 1.816** (df = 29; 124) | 1.797** (df = 29; 124) |

*Note:* *p<0.1; **p<0.05; ***p<0.01

Parenthetical codes denote which imputed value types were used: D - distribution-based, LB - lower-bounds, UB - upper-bounds.

**Table 3.** Regression results, omitting subreddit covariates, of the effect of treatment conditions on log-normalized score. Full results can be found in Appendix F.

Table 4 shows the results of our regressions for the log-normalized interaction outcome. Here we observe a statistically significant relationship between the conservative username treatment and the measured outcome in all seven regression models. This suggests that even with the worst case imputation methods, which use the highest and lowest bounds of imputed outcomes for missing data, the significant relationship holds. The presence of this effect across all models reduces the likelihood that the significant results in the measured data are due to attrition bias, increasing confidence that this effect actually exists across the target population.

**Total Interactions Regression Results (Omitting Subreddit Covariates)**

| | Dependent Variable: Log Normalized Total Interactions | | | | | | |
|---|---|---|---|---|---|---|---|
| | Measured | Hourly (D) | 24hr Weighted (D) | Hourly (LB) | 24hr Weighted (LB) | Hourly (UB) | 24hr Weighted (UB) |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Conservative | -0.049** | -0.050*** | -0.042** | -0.054*** | -0.046** | -0.047** | -0.030** |
| | (0.019) | (0.019) | (0.020) | (0.019) | (0.020) | (0.019) | (0.015) |
| Liberal | -0.048*** | -0.027 | -0.029 | -0.027 | -0.030 | -0.022 | -0.021 |
| | (0.017) | (0.018) | (0.018) | (0.019) | (0.018) | (0.018) | (0.014) |
| Constant | 0.407*** | 0.333*** | 0.363*** | 0.326*** | 0.357*** | 0.338*** | 0.278*** |
| | (0.024) | (0.040) | (0.037) | (0.040) | (0.036) | (0.039) | (0.029) |
| Observations | 128 | 154 | 154 | 154 | 154 | 154 | 154 |
| $R^2$ | 0.632 | 0.460 | 0.487 | 0.448 | 0.486 | 0.472 | 0.529 |
| Adjusted $R^2$ | 0.523 | 0.334 | 0.367 | 0.319 | 0.366 | 0.348 | 0.418 |
| Residual Std. Error | 0.081 (df = 98) | 0.097 (df = 124) | 0.099 (df = 124) | 0.098 (df = 124) | 0.098 (df = 124) | 0.097 (df = 124) | 0.075 (df = 124) |
| F Statistic | 5.806*** (df = 29; 98) | 3.647*** (df = 29; 124) | 4.064*** (df = 29; 124) | 3.474*** (df = 29; 124) | 4.050*** (df = 29; 124) | 3.820*** (df = 29; 124) | 4.793*** (df = 29; 124) |

*Note:* *p<0.1; **p<0.05; ***p<0.01

Parenthetical codes denote which imputed value types were used: D - distribution-based, LB - lower-bounds, UB - upper-bounds.

**Table 4.** Regression results, omitting subreddit covariates, of the effect of treatment conditions on log-normalized total interactions. Full results can be found in Appendix F.

Figure 5 shows a swarm plot for the distribution-based, hourly imputed total interaction outcome presented in column 2 of these regression sequences. In this plot, posts are split into treatment groups, with each point representing a single post's outcome measurement. Notice the slight shift down of the red conservative points relative to the grey neutral points.

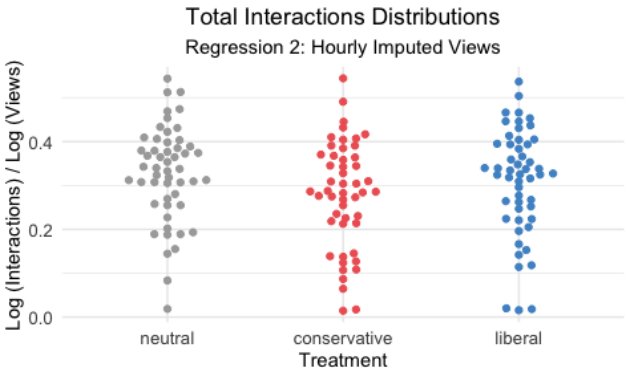



**Figure 5.** Swarm plot comparing total interactions (based on hourly imputed views) by experimental group.

How do we interpret the estimates in our regression table, given that outcome variables are calculated on a log/log scale? For a treatment group regression coefficient of $\beta_1$, the proportion of predicted interactions on that treatment group can be calculated using Equation 3 (see Appendix G for the algebraic transformations used to derive this and the following equations).

$$\frac{Interactions_{treatment}}{Interactions_{neutral}} = Views^{\beta_1}$$

**Equation 3.** Interpretation of positive regression coefficients by view count.

As Equation 3 implies, a negative treatment coefficient $\beta_1$ indicates that as view count increases, the ratio of interactions with the treatment username versus interactions with the neutral username decreases. When $\beta_1$ is negative:

$$Views^{\beta_1} = \frac{1}{Views^{|\beta_1|}}$$

**Equation 4.** Interpretation of negative regression coefficients by view count.

This means that for a given negative $\beta_1$, the right side of Equation 3 declines as views increase. The left side of Equation 3 implies that as views increase (and the right side of the equation decreases) treatment usernames accumulate interactions more slowly than neutral username posts. In our case, conservative treatment coefficients in the regression models for interaction outcomes were consistently negative, meaning that as views increased, conservative username posts were predicted to accumulate interactions at a lower rate relative to neutral username posts.

The results from our hourly imputed regression, where $\beta_1$ = -0.05, and a view count of 30,000 (average number of views for measured posts at 24 hours), correspond to a ratio of treatment interactions over neutral interactions of 0.60, or a 40% decrease in the predicted number of interactions on the conservative username posts.

Equation 5 shows how we can use the regression results to predict the number of interactions for a post given its view count, treatment group, and subreddit page.

$$Predicted\ Interactions = Views^{\beta_0 + \beta_1 X_1 + \beta_2 X_2}$$

**Equation 5.** Calculation for Predicted interactions using View count, treatment group ($X_1$), and subreddit ($X_2$)

Since views enter Equation 5 as the base of a power function, for a given set of coefficients, the marginal effect of each additional view decreases. In other words, as a post receives additional views, each additional view contributes progressively fewer new interactions. Using Equation 5, Table 5 shows how the resulting coefficients from the hourly imputed total interactions regression (column 2 of table 4) map to predicted total interactions for a range of view counts:

| Views | Predicted Total Interactions | | | | |
|---|---|---|---|---|---|
| | **Neutral Username ( $\beta_0$ = 0.33)** | **Conservative Username ($\beta_{1,1}$ = -0.050)** | | **Liberal Username ($\beta_{1,2}$ = -0.027)** | |
| | # Predicted Interactions | # Predicted Interactions | % Decrease from Neutral | # Predicted Interactions | % Decrease from Neutral |
| 10,000 | 20.9 | 13.2 | 36.9% | 16.3 | 22.0% |
| 30,000 | 30.0 | 17.9 | 40.3% | 22.7 | 24.3% |
| 50,000 | 35.5 | 20.7 | 41.8% | 26.5 | 25.3% |

**Table 5.** Predicted total interactions relative to the neutral baseline, by view count and treatment condition.

From this table we see that at 30,000 views, a post from a neutral username is expected to see ~30 interactions, whereas our treatment posts are expected to see ~18 interactions (conservative) and ~23 interactions (liberal) respectively. These correspond to a 40% decrease in interactions for conservative posts and a 24% decrease in interactions for liberal posts, compared to the neutral post interactions. These example calculations are based on the subreddit variable baseline category r/AskCullinary so exclude a specific subreddit coefficient ($\beta_2$).

While these values are predictions based on a specific regression model's results, and the exact magnitude of interactions on any specific post may vary, the statistical significance of the conservative coefficient across all of our total interaction regression models indicates that the observed difference in total interactions on conservative posts is unlikely to be a false positive and instead reflects a true underlying effect.

## Sentiment Analysis

Beyond the outcome measures that look at quantitative interaction numbers, we also wanted to take an initial look at the language being used to engage with our accounts. We ran a state-of-the-art sentiment analysis model, twitter-roberta-sentiment-base-latest (Loureiro et al., 2022), over the replies to each username and compared the results across treatment groups.

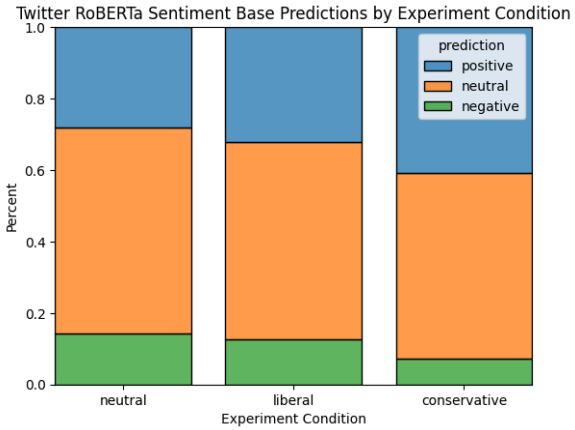



**Figure 6.** Sentiment of direct replies based on poster username signal

While upon first look this result appears to contradict our theory that liberal posts will receive more positive interactions and conservative posts will receive less, it is important to note that we did not calibrate our posts for eliciting neutral replies. As a result, some posts will inherently elicit positive or negative comments regardless of treatment condition, and a few may get many comments, skewing results. We can see this in the results on a per-post basis:

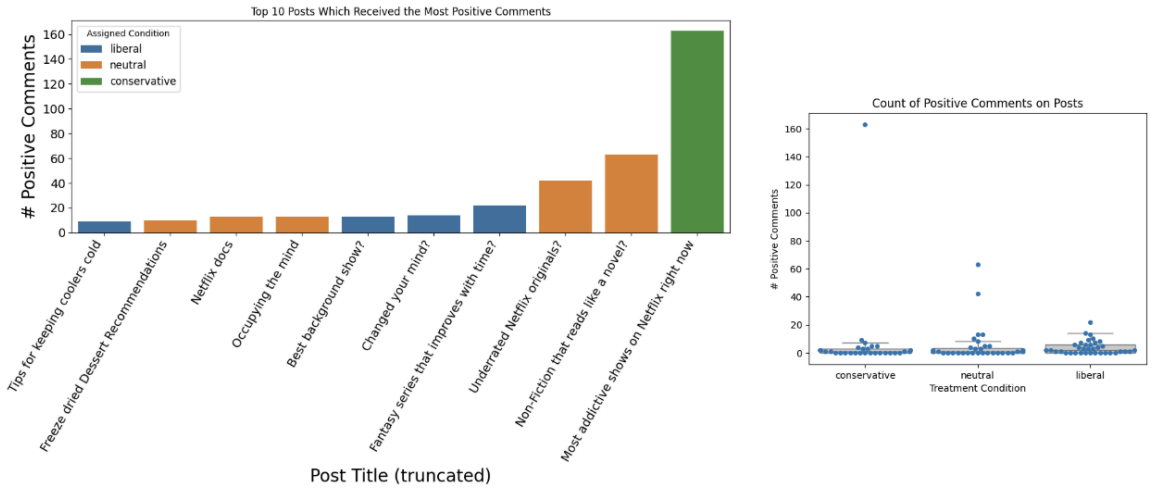



**Figure 7.** Distribution and top 10 positive comment counts on posts

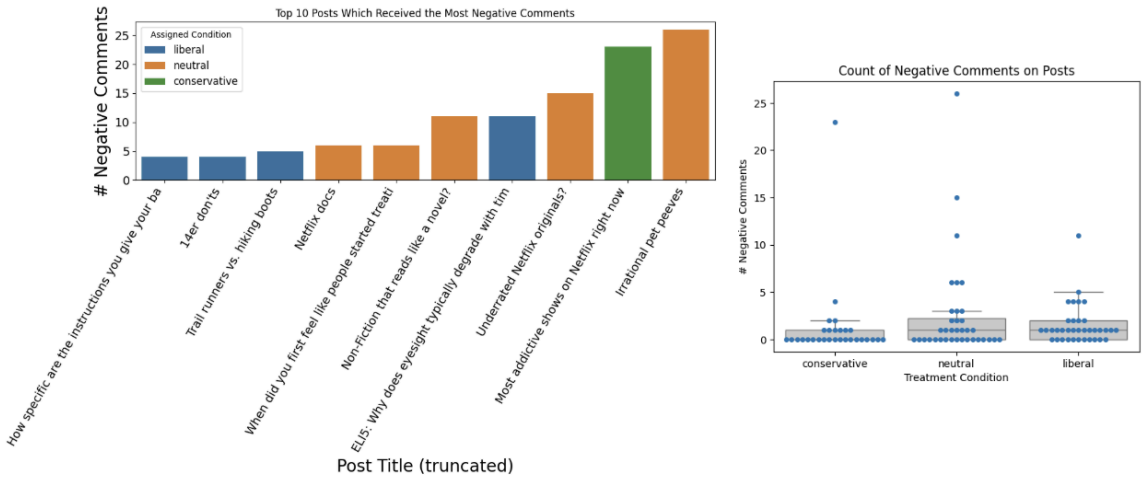



**Figure 8.** Distribution and top 10 positive comment counts on posts

There is a single post, “Most Addictive Shows on Netflix Right Now”, which received 350 total direct replies. That is double the number of direct replies on the next most popular post “Non-Fiction that reads like a novel?”. The nature of the post meant that nearly half of those replies were classified positive, with only 22 being classified negative. This contrasts with a post like “Irrational Pet Peeves” which was in the control condition and generated mostly negative responses. Removing the single outlier post with the most positive replies results in a more muted difference between treatment groups:

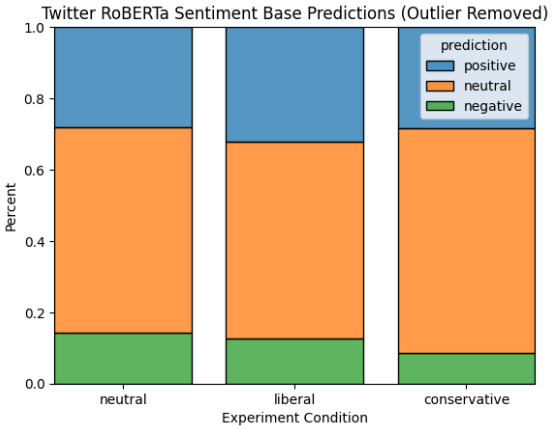



**Figure 9.** Sentiment of direct replies with the “Most Addictive Shows…” outlier post removed

Regression results for the sentiment classification are included in Appendix I.

# Discussion

## Practical Significance

We see statistically significant results for the conservative treatment in our regression models for the total interactions outcome, indicating fewer interactions for conservative usernames compared to neutral usernames. Regression coefficients compare treated groups to the neutral baseline. The lack of robust significant results among liberal usernames could mean that Reddit users, who lean left, might not differentiate between neutral and “no_kings” usernames. Meanwhile, this liberal-leaning population may be quicker to notice “MAGA”-branded usernames and have a discernibly different reaction to them. In many cases, disagreement may result in apathy—that is, lack of engagement—instead of explicit rejection or overt opposition in the form of a downvote or a negative comment. Reddit users may be less willing to acknowledge and contribute to (whether positively or negatively) conversations that were started by conservative usernames. This may indicate some propensity to prefer to avoid contact with these usernames, and avoidance may come from a place of disapproval or discomfort. Why put time into engaging with someone that you dislike? Such mechanisms could be responsible for the statistically significant lower total interactions among conservative usernames relative to the neutral control group.

The inherent complexity of the “total interactions” outcome makes other interpretations of our results possible. This outcome is a summation of upvotes and downvotes—which encapsulates both positive and negative reactions to the post—along with a tally of total comments. This aggregate measure could therefore encode many different emotions and meanings simultaneously, without differentiating between types of interactions. Instead, it tracks the number of contributions to a discussion, acknowledgements of communication, or acts of contact, without accessing any information about the interaction’s intent or emotional valence.

We also found that the liberal treatment had a statistically significant negative impact on the net score and total interactions outcomes for the non-attrited data. However, since these effects did not persist across regression models for the datasets that included imputed data, we believe this result could be due to attrition bias. Further studies, ideally with fewer attrited results, could confirm this supposition.

## Geopolitical Context and Considerations

In addition to considering the significance of negative reactions and comments in our total interactions outcome, it is also important to consider the broader domestic and international political climate that could have been impacting reception of our treatment conditions.While this experiment was conducted over the course of 22 days (March 9th - 31st, 2026) with the goal of minimizing the influence of geopolitical changes on the observed treatment effect, two potentially confounding events are worth noting. First, the Iran war, which began on February 28th, has had deleterious effects on President Trump’s approval ratings, which could have knock-on effects on the US public’s perception of the “MAGA” symbol. The war has also caused

a rift among MAGA subscribers, some of whom believe foreign entanglements do not align with an “America First” agenda. Prominent figures such as Congresswoman Marjorie Taylor Greene and journalists and talk show hosts Tucker Carlson and Megyn Kelly, previously ardent supporters of the MAGA movement, have come out vocally against the war, splintering the movement (Moore. et. al., 2026). As such, the MAGA movement may have experienced its weakest point during the experimental period, exacerbating negative treatment effects for the conservative treatment group.

In addition to the Iran war, a No Kings protest occurred on March 28th, toward the end of our experiment period. Increased awareness of the meaning of “No Kings” in the days leading up to this event could potentially have impacted the observed treatment effect, as well as a perception that the protests were large and widespread; organizers estimated that over 8 million people participated in 3,300 cities worldwide (CBS News, 2026). Redditors, who were likely more supportive of the protests, may have felt more comfortable interacting with “no_kings” usernames and less so with MAGA usernames, since the protests were a reminder of all of the wrongdoing that Trump’s critics accuse him of.

## Other Areas of Interest and Future Research

The following proposed changes to this experiment could improve and deepen future analysis. First, while we did undergo a 3.5 week warmup period with our experiment accounts, that proved to be insufficient for gathering enough karma to post in all experiment subreddits. Extending that warmup likely would have improved our post attrition rate. Increasing the number of subreddits and posts per subreddit could also enhance a recreated experiment’s power, even if we saw the same levels of attrition.

As discussed, while the 180 posts were crafted to minimize variation in political and interest levels, variation in positivity of engagement was not specifically designed for. Formulating a future experiment to address the question of quality and nature of interaction would be an interesting addition. The language content of replies may provide more information as to the ways that users react differently towards political-signaling usernames. Future experiments may look to probe this more deeply by designing posts which do not inherently elicit positive or negative replies. This would help isolate the effects of reacting to the username from other post-related effects. Ultimately, the choice to provide an unhelpful or hurtful comment, even through subtle language cues, is more profound and impactful than the plain decision of whether or not to interact.

Lastly, exploring more nuanced political signals in usernames, beyond the binary “MAGA” and “no_kings” dichotomy, could speak to the culture wars afflicting U.S. politics more broadly. Similarly, testing engagement with more overt political signalling, such as in the post itself instead of in the username, could generate more pointed and explicit engagement and speak more directly to how affective polarization—the dislike of people with opposing viewpoints—materializes online.

## Conclusion

We find ourselves in a moment in national history where political polarization is at an all-time high. Donald Trump's entry into the political scene, first in the lead-up to the 2016 presidential election and recently again with his 2024 win, has both reflected and built upon a decades-old rift between conservative and liberal America. Our experiment explores the consequences and manifestations of this rift in the social media context, asking whether and how political identity cues impact online interactions. Conversation between citizens is a bedrock of democratic politics, and our experiment probes whether political polarization has done any damage.

In particular, we explore how the political identity that a user "wears" in the Reddit-sphere through a username handle impacts how Redditors engage with apolitical posts. We find that usernames that are associated with the MAGA movement attract fewer total interactions, at a statistically significant level, than neutral usernames. In other words, people are less willing to upvote, downvote, or comment on posts made by the conservative treatment group; they are less willing to acknowledge and engage with MAGA users. This result is robust across our 7 model specifications; statistical significance was found when modeling both the measured data, which excluded attrited posts, and the 6 datasets derived from different imputation methods (distribution-based and upper/lower bound most extreme value imputations). However, because our experiment was underpowered, we can only cautiously suggest that the treatment effect of the conservative username on total interactions had the magnitudes we report. While our robust experimental design indicates the effect is likely negative, the magnitude may be inflated.

For our net score outcome, which represents the degree of support, approval, or promotion that a post attracted, we found that liberal "no_kings" treatment had a lower net score, at a statistically significant level, than the neutral non-political signaling username. In other words, liberal posts seemed to be less positively received. However, because this result was confined to measured data and not robust across our 7 datasets, we believe it could be an artifact of attrition rather than a substantive effect of the liberal treatment.

In short, our experiment finds that people were less willing to interact with posts published by MAGA-affiliated usernames, even to discuss minimally controversial topics related to everyday life, and not politics. Future studies can build on these findings by examining not only frequency of interaction but also its quality and nature, as well as by expanding types of political signaling by introducing further variation in usernames. Human communication is remarkably complex, and social media provides us with a unique opportunity to study recorded acts of communication at scale. These data can help us to understand how people calibrate their interactions, often in ways that are subtle and difficult to catch in real time, when they encounter people they perceive to be very different from themselves.

# References


Abramowitz, A. I., & McCoy, J. (2018). United States: Racial resentment, negative partisanship, and polarization in Trump's America. *The ANNALS of the American Academy of Political and Social Science, 681*(1), 137–156. https://doi.org/10.1177/0002716218811309

Associated Press. (2026, March 28). *Millions turn out for "No Kings" rallies held worldwide to protest against Trump*. CBS News. https://www.cbsnews.com/news/no-kings-rallies-protest-trump-millions/

Campaign Legal Center. (2026, February 24). *Tracking the Trump administration's most corrupt transactions*. https://campaignlegal.org/document/tracking-trump-administrations-most-corrupt-transactions

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*(5), 752–766. https://doi.org/10.1037/0022-3514.39.5.752

Finkelstein, S. (2025, October 5). *Trump vs. party affiliation*. *The Hill*. https://thehill.com/opinion/campaign/5818072-trump-vs-party-affiliation/

Godfrey, JT. (2024, October 18). *New study on Reddit explores how political bias in content moderation feeds echo chambers*. University of Michigan Ross School of Business. https://michiganross.umich.edu/news/new-study-reddit-explores-how-political-bias-content-moderation-feeds-echo-chambers

Haddock, K., & Roy, D. (2026, February 27). *ICE and deportations: How Trump is reshaping immigration enforcement*. Council on Foreign Relations. https://www.cfr.org/articles/ice-and-deportations-how-trump-reshaping-immigration-enforcement

Hutchens, M. J., Silva, D. E., Hmielowski, J. D., & Cicchirillo, V. J. (2019). What's in a username? Civility, group identification, and norms. *Journal of Information Technology & Politics, 16*(3), 203–218. https://doi.org/10.1080/19331681.2019.1633983

Kleinfeld, R. (2023, September 5). *Polarization, democracy, and political violence in the United States: What the research says*. Carnegie Endowment for International Peace. https://carnegieendowment.org/research/2023/09/polarization-democracy-and-political-violence-in-the-united-states-what-the-research-says

Lonsdorf, K. (2025, May 21). *Trump accepts Qatar plane for potential use as Air Force One*. NPR. https://www.npr.org/2025/05/21/nx-s1-5406420/trump-accepts-qatar-plane-air-force-one

Moore, E., & Inskeep, S. (2026, April 9). *Some of Trump's biggest supporters become his loudest critics over Iran war*. NPR. https://www.npr.org/2026/04/09/nx-s1-5778195/some-of-trumps-biggest-supporters-become-his-loudest-critics-over-iran-war

Ramakko, B. (2024, September). *Reddit's American political left-wing bias: A study of the top 100 posts from September 12–21, 2024*. *International Journal of Research and Innovation in Social Science, 8*(9), 2510–2517. https://rsisinternational.org/journals/ijriss/articles/reddits-american-political-left-wing-bias-a-study-of-the-top-100-posts-from-september-12-21-2024/

wa2b. (2022, June 9). *Does "total views" on post means that the post was opened?* Reddit. https://www.reddit.com/r/help/comments/v8inp4/does_total_views_on_post_means_that_the_post_was/

Segal, J. (2025). *Donald Trump: Source and symptom of polarization*. *The Social Contract, 16*(1). https://doi.org/10.5206/sc.v16i1.22577

Taylor, S. J., Muchnik, L., Kumar, M., & Aral, S. (2023). Identity effects in social media. Nature Human Behaviour, 7(1), 27–37. https://doi.org/10.1038/s41562-022-01459-8

Wagner, U., & Christ, O. (2007). Intergroup aggression and emotions: A framework and first data. In G. Steffgen & M. Gollwitzer (Eds.), *Emotions and aggressive behavior* (pp. 133–148). Hogrefe & Huber Publishers. https://psycnet.apa.org/record/2007-00066-008

Loureiro, Daniel, et al. (2022). TimeLMs: Diachronic language models from Twitter. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 251–260). https://doi.org/10.18653/v1/2022.acl-demo.25

# Appendix A: Username Balance Check

We deployed a survey to assess the strength of political signaling across experimental groups and their usernames. The survey asked respondents to score each username's political leaning on a 5-point ordinal scale with the options "Far Left," "Left," "Neutral," "Right," and "Far Right". We recorded 57 responses from 27 February 2026 to 7 March 2026.

Across all three username groups, the 57 responses overwhelmingly coincided with our intended political signaling. Table A1 summarizes the results by username, and Table A2 summarizes the aggregate results by experimental group. Each of the "no_kings" usernames was identified as liberal, as intended, over 91% of the time; each of the "MAGA" usernames was identified as conservative over 94% of the time; and each of the autogenerated usernames was identified as neutral over 87% of the time. Respondents labeled the liberal usernames as leaning "left" around 40% of the time and "far left" around 54% of the time. The conservative usernames' "right"-"far right" labeling had greater variation, with "right" labels ranging 12 percentage points from 16% to 28% (average: 23%) and "far right" labels ranging 16 percentage points from 67% to 83%, compared to 7-percentage point ranges among liberal usernames labeled both "left" and "far-left." Additionally, while liberal usernames had an average difference of 14 percentage points between "left" and "far left" labels, conservative usernames had an average difference of 50 percentage points between "right" and "far right."

| | Far left | Left | Neutral | Right | Far right | Intended group |
|---|---|---|---|---|---|---|
| **no_kings_victory24** | 50.9% | 43.9% | 0.0% | 1.8% | 3.5% | **94.7%** |
| **no_kings_movement1** | 54.4% | 38.6% | 1.8% | 1.8% | 3.5% | **93.0%** |
| **no_kings_now1** | 57.9% | 36.8% | 1.8% | 0.0% | 3.5% | **94.7%** |
| **no_kings_rising2** | 52.6% | 38.6% | 3.5% | 1.8% | 3.5% | **91.2%** |
| **no_kings_united2** | 54.4% | 40.4% | 3.5% | 0.0% | 1.8% | **94.7%** |
| **CharacterDig2229** | 1.8% | 7.0% | 87.7% | 3.5% | 0.0% | **87.7%** |
| **Wooden-Night-4696** | 1.8% | 3.5% | 91.2% | 3.5% | 0.0% | **91.2%** |
| **Fresh_Window_6484** | 1.8% | 7.0% | 87.7% | 3.5% | 0.0% | **87.7%** |
| **Interesting-Nerve663** | 1.8% | 3.5% | 93.0% | 1.8% | 0.0% | **93.0%** |
| **Purple-Rooster4613** | 3.5% | 5.3% | 87.7% | 3.5% | 0.0% | **87.7%** |
| **Evening_Issue_8448** | 0.0% | 3.5% | 91.2% | 3.5% | 1.8% | **91.2%** |
| **MAGA_victory24** | 0.0% | 0.0% | 1.8% | 24.6% | 73.7% | **98.2%** |
| **MAGA_rising2** | 0.0% | 0.0% | 3.0% | 15.8% | 82.5% | **98.2%** |
| **MAGA_movement1** | 0.0% | 1.8% | 3.5% | 26.3% | 68.4% | **94.7%** |
| **MAGA_united2** | 0.0% | 0.0% | 5.3% | 28.1% | 66.7% | **94.7%** |

| | | | | | | |
|---|---|---|---|---|---|---|
| **MAGA_now1** | 0.0% | 0.0% | 3.5% | 21.1% | 75.4% | **96.5%** |

**Table A1.** Proportion of responses categorizing usernames into each level of political signaling. “Intended group” is the proportion of responses categorizing the username as “left” or “far left” for “no_kings” usernames and as “right” or “far right” for “maga” usernames.

Although there is imbalance in the extremity of the political signal between the conservative and liberal handles, perceived political signaling overwhelmingly aligned with the usernames’ intended treatment group. On average, 94% of respondents categorized the liberal usernames as leaning “left” or “far left”; 96% categorized the conservative usernames as leaning “right” or “far-right”; and 89% categorized the neutral usernames as neutral.

| | "no_kings" usernames | | "maga" usernames | | Autogenerated usernames | |
|---|---|---|---|---|---|---|
| **Far left** | 54.04% | 93.68% | 0.00% | 0.35% | 2.11% | 7.37% |
| **Left** | 39.65% | | 0.35% | | 5.26% | |
| **Neutral** | 2.11% | 2.11% | 3.41% | 3.41% | **89.47%** | **89.47%** |
| **Right** | 1.05% | 4.21% | 23.16% | 96.49% | 3.16% | 3.16% |
| **Far right** | 3.16% | | 73.33% | | 0.00% | |

**Table A2.** Summary of responses categorizing usernames into each level of political signaling, aggregated at the username group level. Highlighted cells represent the intended signal category.

## Appendix B: Post Balance Check

To ensure posts were equally interesting and apolitical, we deployed a second survey that asked respondents to determine how interesting and how political each post was in relation to the other posts assigned to the same subreddit, on an ordinal scale from 1 to 5, with 3 representing “neutral” or no imbalance. We received 24 responses, each of which reviewed 186 posts, grouped into 31 prospective subreddits.

We expected each post to obtain a neutral interest-level and political-lean score of 3 on average and set a tolerance of ±0.5 to identify potentially imbalanced posts. We identified 16 posts with values outside of the 2.5 to 3.5 tolerance range: 7 posts were scored as more interesting, 2 were scored as less interesting, 1 was scored as more right-leaning, and 6 were scored as more left-leaning. Tables B1 and B2 summarize the post-level average scores and their average deviations from the baseline neutral score of 3. Aggregating by subreddit, no subreddit has a mean political lean or interest score that falls outside the 2.5-3.5 range (min: 2.56, max: 3.28). Based on this, we drop r/Meditation, which has most of the posts viewed as skewing liberal, and proceed with the remaining 30 subreddits.

| Subreddit | Post 1 | Post 2 | Post 3 | Post 4 | Post 5 | Post 6 | Avg | Avg dev from 3 |
|---|---|---|---|---|---|---|---|---|
| **Baking** | 3.773 | 2.273 | 2.591 | 3.182 | 3.000 | 3.409 | 3.038 | 0.417 |
| **Bodyweight Fitness** | 2.818 | 2.955 | 2.955 | 3.182 | 2.955 | 2.545 | 2.902 | 0.159 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Explain Like I'm Five** | 3.136 | 3.636 | 3.273 | 3.591 | 3.091 | 2.955 | 3.280 | 0.295 |
| **NoStupidQuestions** | 2.864 | 3.318 | 3.500 | 2.773 | 2.818 | 3.455 | 3.121 | 0.303 |
| **AskDocs** | 2.727 | 3.091 | 2.727 | 2.909 | 3.591 | 2.818 | 2.977 | 0.250 |
| **Cooking** | 3.318 | 3.182 | 2.727 | 2.864 | 3.045 | 3.091 | 3.038 | 0.174 |
| **Travel Hacks** | 2.636 | 2.818 | 3.136 | 3.545 | 3.500 | 3.136 | 3.129 | 0.311 |
| **Netflix** | 3.364 | 2.636 | 3.364 | 3.045 | 2.818 | 3.045 | 3.045 | 0.227 |
| **Ask Culinary** | 3.273 | 3.500 | 3.091 | 2.909 | 3.091 | 3.136 | 3.167 | 0.197 |
| **Learn Programming** | 3.273 | 3.045 | 2.955 | 2.864 | 2.909 | 3.136 | 3.030 | 0.121 |
| **Backpacking** | 3.045 | 3.182 | 3.000 | 2.864 | 3.182 | 3.000 | 3.045 | 0.091 |
| **Camping** | 2.636 | 2.955 | 2.818 | 2.955 | 2.955 | 3.182 | 2.917 | 0.144 |
| **Social Skills** | 3.182 | 3.409 | 3.136 | 3.409 | 3.182 | 3.091 | 3.235 | 0.235 |
| **Book Suggestions** | 2.682 | 2.636 | 3.136 | 3.227 | 3.545 | 3.273 | 3.083 | 0.311 |
| **Fantasy** | 2.773 | 2.727 | 3.318 | 3.273 | 3.000 | 2.500 | 2.932 | 0.265 |
| **Cycling** | 2.818 | 3.000 | 2.955 | 2.955 | 2.727 | 2.591 | 2.841 | 0.159 |
| **Sleep** | 3.227 | 3.000 | 3.273 | 3.409 | 2.864 | 2.773 | 3.091 | 0.212 |
| **Casual Conversation** | 3.455 | 3.273 | 3.136 | 3.591 | 3.227 | 2.864 | 3.258 | 0.303 |
| **Ask Men** | 2.773 | 2.955 | 2.955 | 3.182 | 3.273 | 2.773 | 2.985 | 0.167 |
| **Hiking** | 2.909 | 3.182 | 2.955 | 2.818 | 3.000 | 2.591 | 2.909 | 0.152 |
| **House Plants** | 2.727 | 2.909 | 2.955 | 2.591 | 2.864 | 2.773 | 2.803 | 0.197 |
| **Language Learning** | 3.455 | 2.818 | 3.273 | 2.909 | 3.045 | 3.273 | 3.129 | 0.220 |
| **Photography** | 3.000 | 3.091 | 2.773 | 3.045 | 3.136 | 2.773 | 2.970 | 0.121 |
| **Frugal** | 2.909 | 3.136 | 3.136 | 3.091 | 3.227 | 3.364 | 3.144 | 0.174 |
| **Self Improvement** | 3.364 | 2.909 | 3.318 | 3.000 | 2.955 | 3.364 | 3.152 | 0.197 |
| **Pets** | 2.455 | 2.864 | 3.136 | 2.727 | 2.773 | 2.909 | 2.811 | 0.235 |
| **Skiing** | 2.818 | 2.727 | 2.500 | 2.955 | 2.909 | 2.864 | 2.795 | 0.205 |
| **I Want to Learn** | 2.955 | 3.045 | 3.227 | 2.909 | 2.955 | 3.136 | 3.038 | 0.098 |
| **Beauty** | 3.045 | 2.955 | 2.545 | 3.000 | 2.773 | 3.136 | 2.909 | 0.152 |
| **Gift Ideas** | 2.545 | 2.727 | 3.227 | 3.045 | 3.273 | 2.864 | 2.947 | 0.235 |
| **Meditation** | 3.045 | 2.864 | 2.818 | 3.045 | 3.182 | 2.909 | 2.977 | 0.114 |

**Table B1.** Average post interest scores by subreddit, including average absolute deviation from the neutral baseline score of 3.

| Subreddit | Post 1 | Post 2 | Post 3 | Post 4 | Post 5 | Post 6 | Avg | Avg dev from 3 |
|---|---|---|---|---|---|---|---|---|
| **Baking** | 2.864 | 2.636 | 2.955 | 2.818 | 3.136 | 3.273 | 2.947 | 0.189 |
| **Bodyweight Fitness** | 3.091 | 3.045 | 2.864 | 3.045 | 3.045 | 3.045 | 3.023 | 0.068 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Explain Like I'm Five** | 2.955 | 2.909 | 2.864 | 2.909 | 3.136 | 2.864 | 2.939 | 0.106 |
| **NoStupidQuestions** | 3.045 | 3.091 | 2.909 | 3.182 | 3.091 | 3.091 | 3.068 | 0.098 |
| **AskDocs** | 2.818 | 2.955 | 2.955 | 3.045 | 2.909 | 3.045 | 2.955 | 0.076 |
| **Cooking** | 3.091 | 2.909 | 2.091 | 3.000 | 3.227 | 3.227 | 2.924 | 0.258 |
| **Travel Hacks** | 3.000 | 2.864 | 3.045 | 3.045 | 2.955 | 2.909 | 2.970 | 0.061 |
| **Netflix** | 3.000 | 3.000 | 2.909 | 3.045 | 3.091 | 2.682 | 2.955 | 0.091 |
| **Ask Culinary** | 3.136 | 3.000 | 3.000 | 2.955 | 2.955 | 3.091 | 3.023 | 0.053 |
| **Learn Programming** | 2.227 | 3.000 | 2.955 | 2.909 | 2.955 | 3.000 | 2.841 | 0.159 |
| **Backpacking** | 2.955 | 2.591 | 2.909 | 3.045 | 2.545 | 3.045 | 2.848 | 0.182 |
| **Camping** | 3.000 | 2.773 | 2.727 | 2.909 | 3.091 | 2.864 | 2.894 | 0.136 |
| **Social Skills** | 2.455 | 2.591 | 2.773 | 2.955 | 2.682 | 2.773 | 2.705 | 0.295 |
| **Book Suggestions** | 2.682 | 3.000 | 2.955 | 2.773 | 2.636 | 2.909 | 2.826 | 0.174 |
| **Fantasy** | 3.000 | 2.864 | 2.727 | 2.955 | 3.000 | 3.000 | 2.924 | 0.076 |
| **Cycling** | 2.500 | 2.727 | 2.727 | 2.818 | 2.682 | 2.864 | 2.720 | 0.280 |
| **Sleep** | 2.909 | 2.727 | 2.909 | 2.864 | 2.773 | 2.682 | 2.811 | 0.189 |
| **Casual Conversation** | 2.909 | 2.955 | 3.000 | 3.000 | 3.045 | 2.818 | 2.955 | 0.061 |
| **Ask Men** | 3.045 | 2.864 | 3.091 | 3.182 | 3.182 | 3.545 | 3.152 | 0.197 |
| **Hiking** | 2.818 | 2.727 | 2.864 | 3.000 | 2.773 | 2.773 | 2.826 | 0.174 |
| **House Plants** | 2.818 | 2.727 | 2.773 | 2.818 | 2.818 | 2.864 | 2.803 | 0.197 |
| **Language Learning** | 2.591 | 2.455 | 2.591 | 2.864 | 2.727 | 2.773 | 2.667 | 0.333 |
| **Photography** | 2.727 | 2.955 | 2.727 | 2.818 | 2.864 | 2.864 | 2.826 | 0.174 |
| **Frugal** | 3.000 | 2.955 | 2.909 | 2.955 | 3.136 | 3.045 | 3.000 | 0.061 |
| **Self Improvement** | 2.909 | 2.909 | 2.636 | 3.227 | 2.909 | 2.864 | 2.909 | 0.167 |
| **Pets** | 2.955 | 2.955 | 2.864 | 2.818 | 2.636 | 2.727 | 2.826 | 0.174 |
| **Skiing** | 3.091 | 3.045 | 3.091 | 3.045 | 3.136 | 3.000 | 3.068 | 0.068 |
| **I Want to Learn** | 3.000 | 3.091 | 3.000 | 2.955 | 3.045 | 3.000 | 3.015 | 0.030 |
| **Beauty** | 2.864 | 3.045 | 3.182 | 2.955 | 2.909 | 3.136 | 3.015 | 0.106 |
| **Gift Ideas** | 2.818 | 3.045 | 3.045 | 2.909 | 2.864 | 2.864 | 2.924 | 0.106 |
| **Meditation** | 2.455 | 2.727 | 2.682 | 2.500 | 2.455 | 2.545 | 2.561 | 0.439 |

**Table B2.** Average post political lean scores by subreddit, including average absolute deviation from the neutral baseline score of 3.

As a randomization check, we looked at the distribution of post political and interests scores assigned to each treatment group.

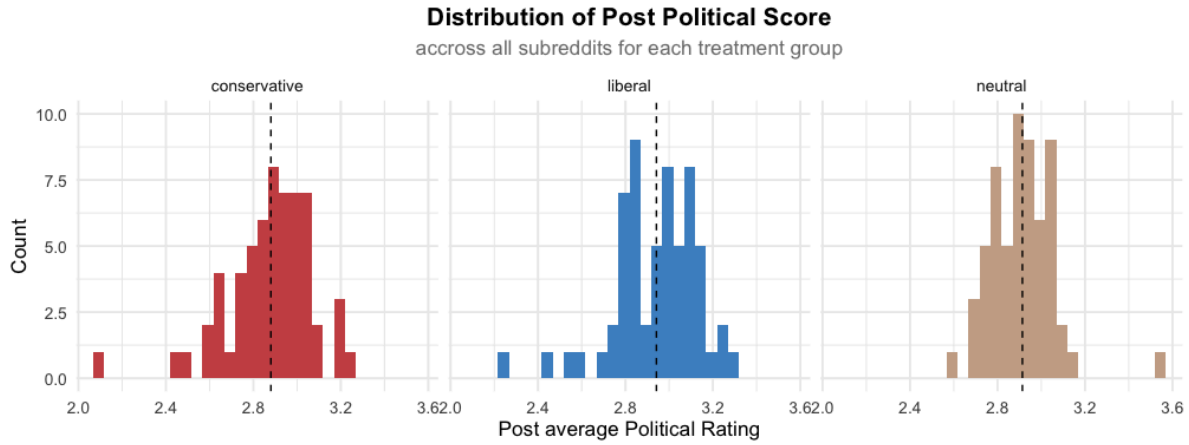



**Figure B1.** Distribution of post political scores across treatment groups

Figure B1 shows the distribution of the average political scores for the 60 posts randomly assigned to each treatment group.

```
t test of coefficients:

                          Estimate Std. Error  t value Pr(>|t|)
(Intercept)               2.913636   0.019451 149.7973   <2e-16 ***
treatmentconservative -0.033333   0.032388  -1.0292   0.3048
treatmentliberal       0.028030   0.031593   0.8872   0.3762
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The above t-test results look at whether the difference in the distributions is statistically significant, and finds that there is not a statistically significant difference in the political rating distributions of the three treatment groups.

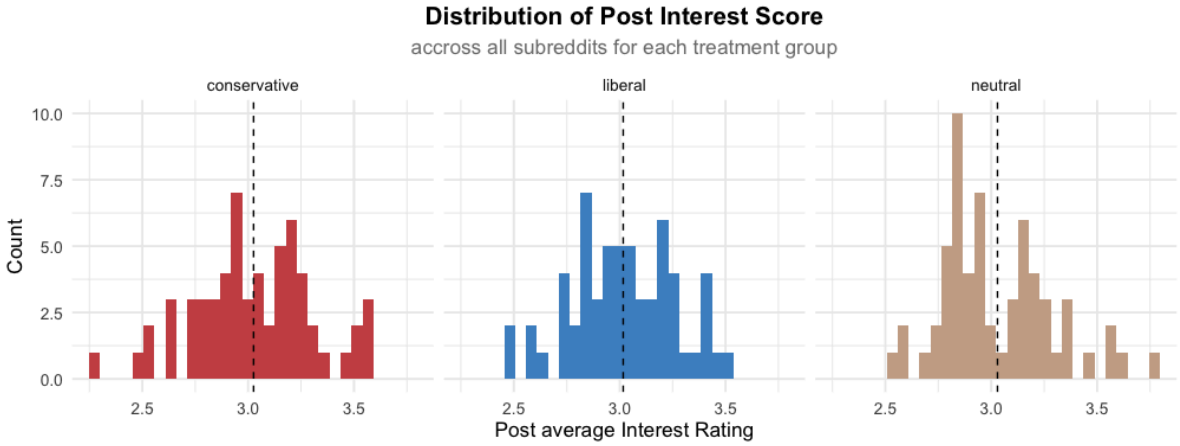



**Figure B2.** Distribution of post interest scores across treatment groups

Figure B2 shows the distribution of the average interest scores for the 60 posts randomly assigned to each treatment group.

```
t test of coefficients:

                           Estimate Std. Error t value Pr(>|t|)
(Intercept)              3.0295455  0.0345027 87.8061   <2e-16 ***
treatmentconservative -0.0037879  0.0504866 -0.0750   0.9403
treatmentliberal      -0.0128788  0.0473263 -0.2721   0.7858
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The above t-test results look at whether the difference in the distributions is statistically significant, and finds that there is not a statistically significant difference in the interest rating distributions of the three treatment groups.

# Appendix C: Post Timing

In addition to posts and usernames, posting times were also randomized. This randomization assigned only a date and a suggested posting order for each experimenter on that date. Experimenters were given latitude to post at any time on that date (in the Berkeley, California timezone), so long as they followed the posting order. This decision was backed by an analysis which scraped the most recent thousand posts from a list of potential experimental subreddits and looked at vote score at the time of collection. The results showed no clear difference in long-run vote score based on either time of day or day of week.

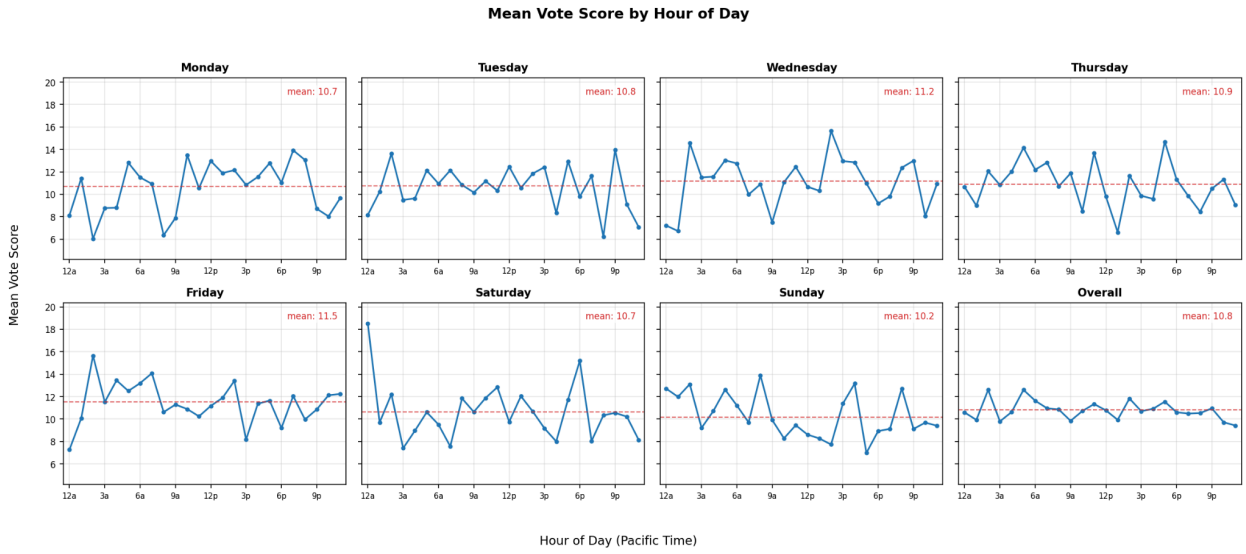



**Figure C1.** Mean vote score for a large sample of posts by posting time and day of week

# Appendix D: Automated Metrics Scraping

Post metrics were collected using a custom-built software tool which scraped Reddit for all available information on the posts from our 15 experiment accounts. In order to avoid various restrictions on querying Reddit, we needed to run these collections from a local IP address, rather than one associated with a cloud provider such as AWS, and including both query

throttling and robust retry mechanisms. We ultimately arrived at a solution that would run on a Raspberry Pi Zero 2w connected to the local network of one of the experimenters. For the duration, a cron job would run every 6 hours to discover new posts, then schedule at-jobs to run precisely when each post hit an age of 12 hours, 24 hours, and beyond in 12 hour increments up to 96 hours. These cron jobs would collect all of the available information for the post and upload to Google Drive.

Due to the need for precisely timed data collection, numerous mechanisms were implemented to detect and mitigate failures, such as sweeps every 2 hours to ensure all collections succeeded and an email notification system which monitored all activity. Despite this, we saw two significant failures over the course of the experiment. The first, about 1 week into the experiment, was a result of a known issue with the Raspberry Pi Zero 2W where the Wi-Fi chip enters a power-saving sleep mode, causing full disconnection from the internet. This was resolved by disabling the mode. A second issue occurred about a week later where Reddit started rejecting all queries from the automation script, which was resolved by restarting the device. These failures, along with other undiscovered minor issues, resulted in some collections being either delayed or missed entirely.

For the 24 hour collections, which we used in our analysis, we saw the vast majority successfully occur between a post age of 23 hours 59 minutes and 24 hours 0 minutes. There are 5 exceptions, 2 which occur approximately 7 minutes late while the other 3 were entirely missed. By contrast, no 48 hour collections were missed entirely but multiple occurred over 6 hours late.

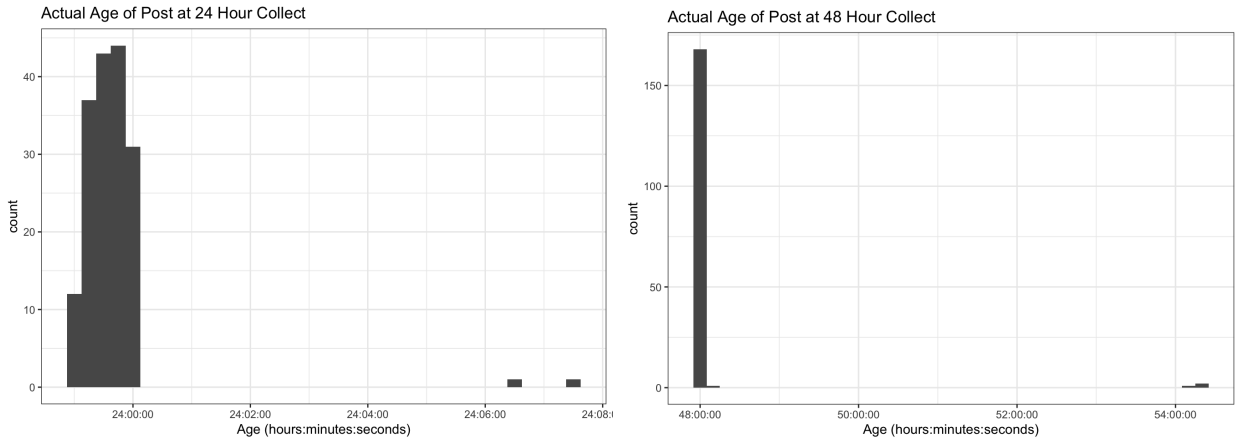



**Figure D1.** Distribution of post age at time of data collection

# Appendix E: Attrition Analysis

Attrition in this experiment occurred when posts were not published as intended and thus were unable to produce measurable results. Examples of attrited posts include those that were unable to publish due to insufficient karma requirements, posts that were automatically or manually removed by subreddit moderators, and one post that was accidentally published in the

wrong (non-experimental) subreddit. Of the 180 experimental posts, 52 (28.9%) attrited. Figure 4 maps out the attrition by treatment group. Chi square tests and logistic regression models found no evidence of significant difference in missingness by experimental condition nor by removal status (i.e., automatic removal, manual removal, or no removal).

We used a chi square test to assess whether there was a significant difference in missingness among the three experimental groups and found no evidence of differential attrition (p=0.4324). A multiple logistic regression model intended to test for the effect size and direction of missingness among experimental groups, followed by pairwise comparisons between each group, also found no significant differences in missingness. A separate chi square test also found no significant differences in missingness between treatment (posts assigned to either treatment condition) and control (p=0.5225). Figure E1 shows the distribution of removed posts by treatment group.

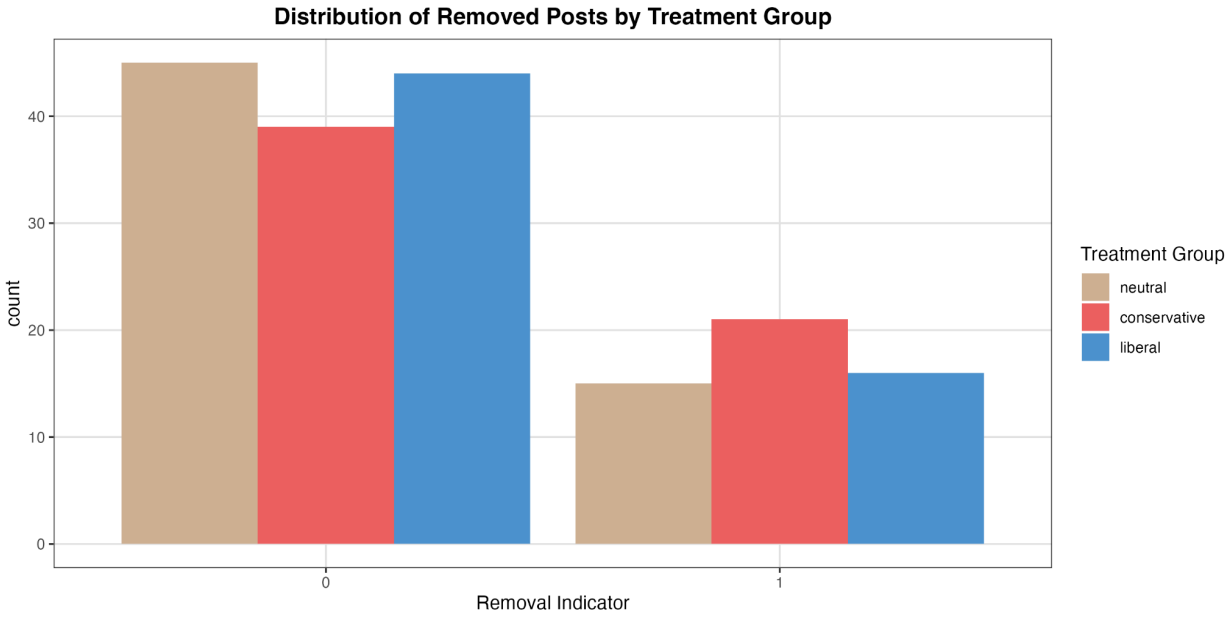



**Figure E1.** Distribution of removed posts by treatment group.

Further, we tested for significant differences in missingness by post removal type to identify whether attrited posts from any group were automatically removed, manually removed, or not removed at systematically different rates. A chi square test found no significant association in missingness by type of removal and treatment group (p=0.8825). Figure E2 summarizes the distribution of posts by removal status and treatment group.

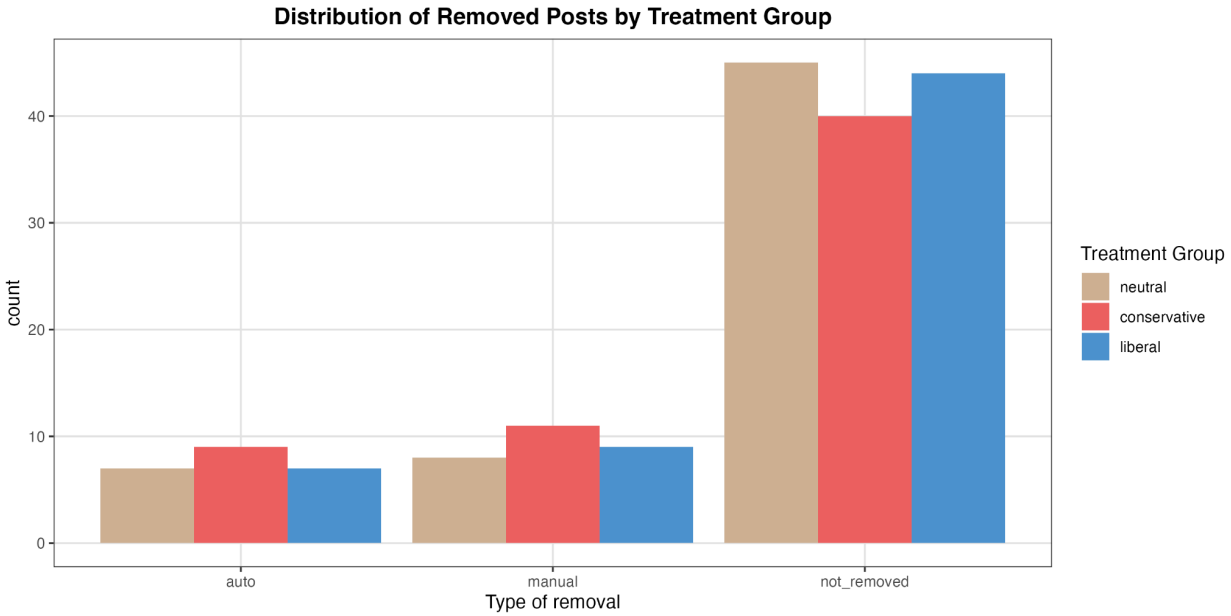



**Figure E2.** Distribution of removed posts by type of removal, by treatment group.

## Imputation

### Distribution-Based Imputation

We impute the missing 24-hour upvote ratios from each post's 12- and 48-hour records:

$$Upvote\ Ratio_{24\ Hours} = Upvote\ Ratio_{12\ Hours} + \frac{1}{3}(Upvote\ Ratio_{48\ Hours} - Upvote\ Ratio_{12\ Hours})$$

We use 1/3 of the 48-hour recorded upvote ratio because $\frac{24-12}{48-12} = \frac{1}{3}$.

With complete vote metrics, we proceed to impute the missing 24-hour view counts by taking the corresponding subreddit's average 24-hour view count.

We also impute the missing view counts at the time of post removal by regressing views on collection hour for each subreddit. We log transform view counts due to the strong right skew in the variable's distribution, then build a model that regresses views on collect hour, controlling for subreddit and post time, based on existing view data. This gives us the linear relationship between view counts and time (particularly at the 12-, 24-, and 48- hour marks). We use this regression model to predict removed posts' unobserved view counts at the time of removal.

### Lower Bounds Imputation

We additionally conduct a lower-bounds imputation to model the minimum possible outcome measurement for each missing view count value, where we impute missing view counts from the lowest view count result in each subreddit. Because we expect to see an engagement treatment

effect, our lower-bounds analysis should minimize vote scores and maximize view counts so that the score-to-views ratio reflects lower engagement.

While the two conservative posts with missing vote metrics' net scores are available, we still aim to minimize their upvotes by imputing the minimum upvote ratio from the post's 12 and 48 hour data.

Higher view counts will mean a smaller score-to-view count ratio, making it harder to show a treatment effect, so we maximize the imputed 24-hour view counts by taking the maximum view count value from posts in the same subreddit at the 24-hour mark.

To impute hour-removal view counts under the lower-bounds scenario, we use the same regression model as above. Instead of taking the direct prediction, we use the prediction plus corresponding subreddit's maximum residual from the model, so that we predict based on observed extreme view counts. This models a scenario in which the attrited posts behave like our most extreme high-view count posts.

### Upper Bounds Imputation

We additionally conduct an upper-bounds imputation to model the maximum possible outcome measurement for each missing view count value, where we impute missing view counts from the highest view count result in each subreddit. Opposite to the lower-bounds scenario, our upper-bounds analysis should maximize vote scores and minimize view counts so that the score-to-views ratio reflects higher engagement.

We maximized the two conservative posts' missing vote metrics by imputing the maximum upvote ratio from the post's 12 and 48 hour data.

Lower view counts will mean a higher score-to-view count ratio, making it easier to show a treatment effect, so we minimize the imputed view counts by taking the minimum view count value from posts in the same subreddit at the 24-hour mark.

To impute hour-removal view counts under the upper-bounds scenario, we use the same regression model as above. Instead of taking the direct prediction, we use the prediction plus the corresponding subreddit's minimum residual from the model, so that we predict based on observed minimum extreme view counts. This models a scenario in which the attrited posts behave like our most extreme low-view count posts.

## Appendix F: Variable Transformations

To achieve a more normal distribution of our outcome variables, with the goal of minimizing the effect of outliers in our relatively small pre-attrition sample size of 180 posts, we applied scaled transformations to our outcome variables. The following plots show transformations for our total interactions outcome. The net score outcome had similar variable distributions, resulting in the same transformations as outlined below.

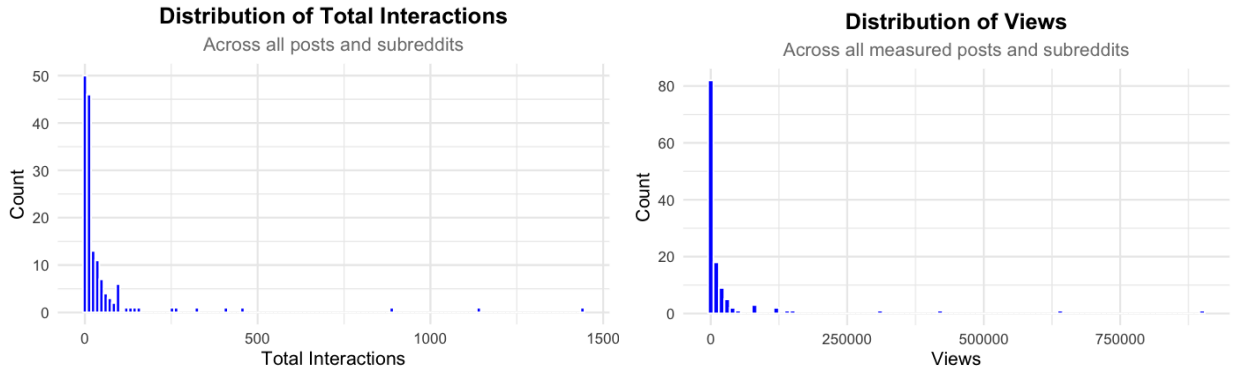



**Figure F1.** Distribution of measured total interactions and views

The figures above show the untransformed distributions of Total Interactions (figure F1 left) and Views (figure F1 right) for the non-attrited posts in our experiment. Both distributions show an extreme right skew of the data.

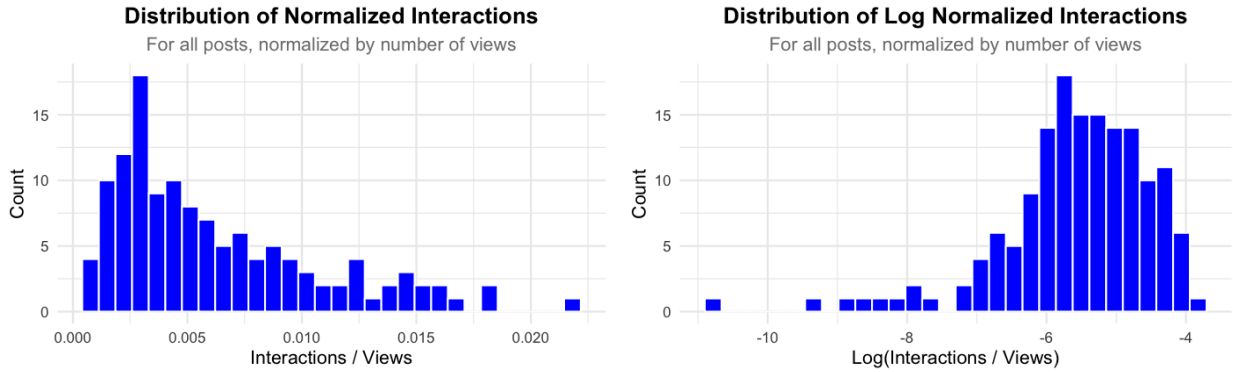



**Figure F2.** Distribution of log transformed total interactions and views

Normalizing interactions on the number of views by dividing each post's interaction measurements at 24 hours by the number of views at 24 hours still results in a right skewed distribution of our outcome measurement (figure F2 left). Because the magnitude of skew in Total Interactions and Views is different, applying a log transform to the normalized measurement results in a left skewed distribution (figure F2 right).

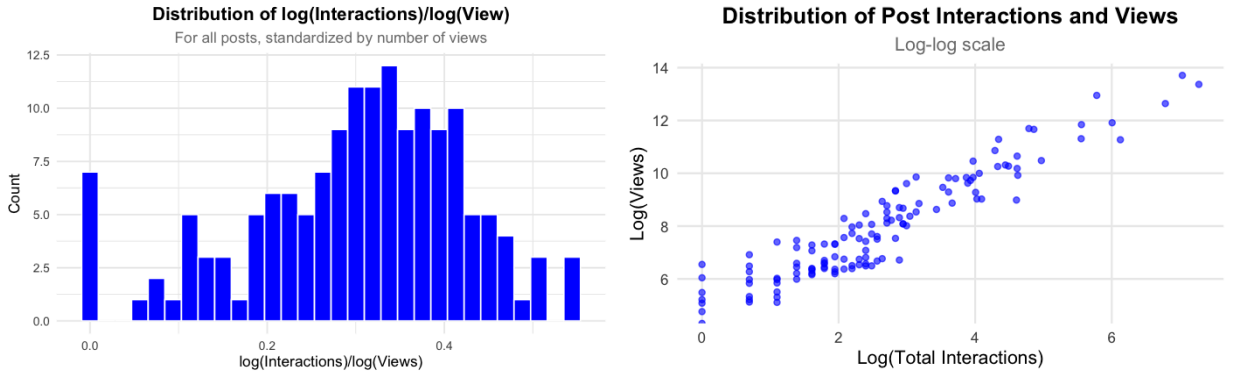



**Figure F3.** Final transformation of outcome variables

Therefore, to achieve a closer to normal distribution that reduces the effect of outliers on our models, we logged the variables independently before normalizing, resulting in a normalized distribution of our outcome measurement. Figure F3 shows the distribution of this log(Total Interactions) / log (Views) measurement for each post, as well as the resulting linear relationship between the two logged values.

# Appendix G: Model Results

**Score Regression Results**

| | Dependent Variable: Log Normalized Score | | | | | | |
|---|---|---|---|---|---|---|---|
| | Measured | Hourly (D) | 24hr Weighted (D) | Hourly (LB) | 24hr Weighted (LB) | Hourly (UB) | 24hr Weighted (UB) |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Conservative | -0.036 | -0.040 | -0.033 | -0.040 | -0.034 | -0.038 | -0.029 |
| | (0.036) | (0.034) | (0.035) | (0.033) | (0.035) | (0.034) | (0.037) |
| Liberal | -0.073** | -0.038 | -0.036 | -0.039 | -0.037 | -0.036 | -0.033 |
| | (0.036) | (0.034) | (0.035) | (0.034) | (0.035) | (0.034) | (0.036) |
| r/AskDocs | -0.181*** | -0.039 | -0.031 | -0.038 | -0.031 | -0.039 | -0.030 |
| | (0.034) | (0.084) | (0.095) | (0.083) | (0.093) | (0.086) | (0.098) |
| r/AskMen | 0.015 | 0.168** | 0.185** | 0.166** | 0.183** | 0.169** | 0.186* |
| | (0.032) | (0.080) | (0.093) | (0.078) | (0.090) | (0.082) | (0.096) |
| r/Backpacking | -0.162* | -0.019 | -0.011 | -0.019 | -0.011 | -0.020 | -0.011 |
| | (0.085) | (0.111) | (0.119) | (0.109) | (0.117) | (0.112) | (0.121) |
| r/Baking | -0.037 | 0.106 | 0.114 | 0.106 | 0.114 | 0.105 | 0.114 |
| | (0.060) | (0.095) | (0.105) | (0.094) | (0.104) | (0.097) | (0.108) |
| r/Beauty | -0.200* | -0.053 | -0.044 | -0.055 | -0.046 | -0.051 | -0.041 |
| | (0.105) | (0.113) | (0.122) | (0.112) | (0.120) | (0.115) | (0.125) |
| r/BodyWeightFitness | -0.026 | -0.018 | -0.021 | -0.013 | -0.018 | -0.026 | -0.035 |
| | (0.067) | (0.148) | (0.164) | (0.144) | (0.161) | (0.155) | (0.177) |
| r/BookSuggestions | -0.078 | 0.065 | 0.073 | 0.065 | 0.073 | 0.064 | 0.073 |
| | (0.073) | (0.107) | (0.116) | (0.106) | (0.115) | (0.109) | (0.119) |
| r/Camping | -0.246** | -0.106 | -0.097 | -0.106 | -0.098 | -0.107 | -0.097 |
| | (0.120) | (0.141) | (0.147) | (0.140) | (0.146) | (0.141) | (0.149) |
| r/CasualConversation | -0.049 | 0.081 | 0.129 | 0.077 | 0.122 | 0.091 | 0.158 |
| | (0.104) | (0.107) | (0.131) | (0.105) | (0.126) | (0.111) | (0.149) |
| r/Cooking | -0.470*** | -0.315** | -0.307** | -0.315** | -0.308** | -0.315** | -0.307** |
| | (0.125) | (0.147) | (0.153) | (0.146) | (0.151) | (0.147) | (0.154) |
| r/Cycling | -0.138 | 0.011 | 0.023 | 0.006 | 0.018 | 0.015 | 0.035 |
| | (0.101) | (0.118) | (0.126) | (0.116) | (0.124) | (0.119) | (0.130) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| r/ExplainLikeImFive | 0.211*** | 0.226** | 0.247** | 0.221** | 0.243** | 0.232** | 0.255** |
| | (0.025) | (0.103) | (0.112) | (0.103) | (0.111) | (0.104) | (0.114) |
| r/Fantasy | -0.028 | 0.088 | 0.137 | 0.087 | 0.137 | 0.090 | 0.138 |
| | (0.036) | (0.089) | (0.097) | (0.088) | (0.095) | (0.091) | (0.100) |
| r/Frugal | 0.205*** | 0.233* | 0.240* | 0.234* | 0.240* | 0.232* | 0.239* |
| | (0.046) | (0.128) | (0.135) | (0.127) | (0.134) | (0.129) | (0.138) |
| r/GiftIdeas | -0.077** | 0.065 | 0.073 | 0.066 | 0.073 | 0.065 | 0.074 |
| | (0.030) | (0.082) | (0.093) | (0.080) | (0.091) | (0.083) | (0.096) |
| r/Hiking | -0.122 | 0.045 | 0.069 | 0.038 | 0.061 | 0.052 | 0.085 |
| | (0.122) | (0.115) | (0.126) | (0.113) | (0.123) | (0.117) | (0.131) |
| r/HousePlants | -0.244*** | -0.101 | -0.093 | -0.101 | -0.093 | -0.102 | -0.093 |
| | (0.081) | (0.108) | (0.117) | (0.107) | (0.115) | (0.109) | (0.120) |
| r/IWantToLearn | -0.070 | 0.073 | 0.081 | 0.073 | 0.080 | 0.072 | 0.081 |
| | (0.046) | (0.090) | (0.100) | (0.088) | (0.098) | (0.091) | (0.103) |
| r/LanguageLearning | -0.259* | -0.096 | -0.089 | -0.096 | -0.089 | -0.092 | -0.088 |
| | (0.134) | (0.125) | (0.133) | (0.124) | (0.132) | (0.127) | (0.136) |
| r/LearnProgramming | -0.142*** | 0.001 | 0.009 | 0.001 | 0.009 | 0.0003 | 0.009 |
| | (0.031) | (0.083) | (0.094) | (0.082) | (0.092) | (0.085) | (0.098) |
| r/Netflix | -0.023 | 0.120 | 0.128 | 0.121 | 0.128 | 0.119 | 0.129 |
| | (0.102) | (0.134) | (0.140) | (0.132) | (0.139) | (0.135) | (0.143) |
| r/NoStupidQuestions | -0.198** | -0.067 | -0.059 | -0.067 | -0.059 | -0.067 | -0.058 |
| | (0.099) | (0.116) | (0.124) | (0.115) | (0.123) | (0.117) | (0.127) |
| r/Pets | -0.165*** | -0.023 | -0.015 | -0.022 | -0.015 | -0.023 | -0.014 |
| | (0.027) | (0.082) | (0.094) | (0.081) | (0.092) | (0.084) | (0.097) |
| r/Photography | -0.028 | -0.054 | -0.050 | -0.043 | -0.043 | -0.062 | -0.066 |
| | (0.060) | (0.128) | (0.141) | (0.120) | (0.134) | (0.135) | (0.158) |
| r/Skiing | -0.156 | -0.008 | -0.0001 | -0.008 | -0.0004 | -0.008 | 0.0004 |
| | (0.099) | (0.121) | (0.129) | (0.119) | (0.128) | (0.122) | (0.132) |
| r/Sleep | -0.201*** | -0.059 | -0.051 | -0.058 | -0.051 | -0.059 | -0.050 |
| | (0.038) | (0.084) | (0.095) | (0.082) | (0.093) | (0.085) | (0.098) |
| r/SocialSkills | -0.037 | 0.105 | 0.113 | 0.106 | 0.113 | 0.105 | 0.114 |
| | (0.054) | (0.092) | (0.102) | (0.090) | (0.100) | (0.093) | (0.105) |
| Constant | 0.278*** | 0.125 | 0.114 | 0.126 | 0.115 | 0.125 | 0.112 |
| | (0.024) | (0.085) | (0.098) | (0.083) | (0.095) | (0.087) | (0.101) |
| Observations | 128 | 154 | 154 | 154 | 154 | 154 | 154 |
| $R^2$ | 0.392 | 0.300 | 0.304 | 0.302 | 0.306 | 0.298 | 0.296 |
| Adjusted $R^2$ | 0.212 | 0.137 | 0.141 | 0.138 | 0.143 | 0.134 | 0.131 |
| Residual Std. Error | 0.167 (df = 98) | 0.174 (df = 124) | 0.179 (df = 124) | 0.172 (df = 124) | 0.177 (df = 124) | 0.176 (df = 124) | 0.186 (df = 124) |
| F Statistic | 2.182*** (df = 29; 98) | 1.834** (df = 29; 124) | 1.865** (df = 29; 124) | 1.847** (df = 29; 124) | 1.884*** (df = 29; 124) | 1.816** (df = 29; 124) | 1.797** (df = 29; 124) |

*Note:* *p<0.1; **p<0.05; ***p<0.01

Parenthetical codes denote which imputed value types were used: D - distribution-based, LB - lower-bounds, UB - upper-bounds.

**Table G1.** Model results for net score outcome

**Total Interactions Regression Results**

| | Dependent Variable: Log Normalized Total Interactions | | | | | | |
|---|---|---|---|---|---|---|---|
| | Measured | Hourly (D) | 24hr Weighted (D) | Hourly (LB) | 24hr Weighted (LB) | Hourly (UB) | 24hr Weighted (UB) |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Conservative | -0.049** | -0.050*** | -0.042** | -0.054*** | -0.046** | -0.047** | -0.030** |
| | (0.019) | (0.019) | (0.020) | (0.019) | (0.020) | (0.019) | (0.015) |
| Liberal | -0.048*** | -0.027 | -0.029 | -0.027 | -0.030 | -0.022 | -0.021 |
| | (0.017) | (0.018) | (0.018) | (0.019) | (0.018) | (0.018) | (0.014) |
| r/AskDocs | -0.194*** | -0.126*** | -0.159*** | -0.118** | -0.151*** | -0.134*** | -0.144*** |
| | (0.037) | (0.049) | (0.046) | (0.049) | (0.046) | (0.048) | (0.035) |
| r/AskMen | 0.070*** | 0.128*** | 0.109*** | 0.132*** | 0.114*** | 0.122*** | 0.068** |
| | (0.025) | (0.041) | (0.037) | (0.042) | (0.036) | (0.040) | (0.029) |
| r/Backpacking | -0.024 | 0.043 | 0.010 | 0.051 | 0.018 | 0.035 | -0.004 |
| | (0.029) | (0.043) | (0.040) | (0.043) | (0.040) | (0.042) | (0.032) |
| r/Baking | -0.134** | -0.066 | -0.099 | -0.058 | -0.091 | -0.074 | -0.101** |
| | (0.066) | (0.071) | (0.070) | (0.071) | (0.070) | (0.071) | (0.050) |
| r/Beauty | -0.095** | -0.054 | -0.085 | -0.049 | -0.080 | -0.057 | -0.080** |
| | (0.048) | (0.057) | (0.055) | (0.058) | (0.055) | (0.054) | (0.040) |
| r/BodyWeightFitness | -0.037 | -0.045 | -0.070 | -0.039 | -0.065 | -0.050 | -0.058 |
| | (0.045) | (0.081) | (0.077) | (0.082) | (0.078) | (0.079) | (0.058) |
| r/BookSuggestions | -0.073 | -0.006 | -0.038 | 0.002 | -0.030 | -0.013 | -0.049 |
| | (0.061) | (0.068) | (0.066) | (0.068) | (0.066) | (0.068) | (0.054) |
| r/Camping | -0.039 | 0.026 | -0.005 | 0.034 | 0.002 | 0.019 | -0.014 |
| | (0.058) | (0.063) | (0.061) | (0.064) | (0.061) | (0.063) | (0.049) |
| r/CasualConversation | 0.044 | 0.075 | 0.106 | 0.073 | 0.102 | 0.090 | 0.062 |
| | (0.044) | (0.059) | (0.082) | (0.061) | (0.077) | (0.059) | (0.057) |
| r/Cooking | -0.130** | -0.056 | -0.090 | -0.047 | -0.082 | -0.062 | -0.075 |
| | (0.056) | (0.065) | (0.064) | (0.065) | (0.063) | (0.065) | (0.048) |
| r/Cycling | -0.042 | 0.047 | 0.027 | 0.038 | 0.020 | 0.054 | 0.017 |
| | (0.039) | (0.054) | (0.058) | (0.050) | (0.050) | (0.062) | (0.052) |
| r/ExplainLikeImFive | 0.097*** | 0.090* | 0.078 | 0.089* | 0.080 | 0.091* | 0.078** |
| | (0.025) | (0.052) | (0.049) | (0.054) | (0.049) | (0.050) | (0.039) |
| r/Fantasy | 0.095*** | 0.015 | 0.051 | 0.021 | 0.059 | 0.011 | 0.018 |
| | (0.027) | (0.068) | (0.056) | (0.068) | (0.055) | (0.067) | (0.046) |
| r/Frugal | 0.152*** | 0.105 | 0.074 | 0.113 | 0.081 | 0.097 | 0.065 |
| | (0.027) | (0.112) | (0.110) | (0.112) | (0.110) | (0.110) | (0.087) |
| r/GiftIdeas | -0.001 | 0.067 | 0.034 | 0.074* | 0.042 | 0.059 | -0.009 |
| | (0.027) | (0.042) | (0.039) | (0.042) | (0.039) | (0.042) | (0.031) |
| r/Hiking | -0.002 | 0.080* | 0.080 | 0.073* | 0.070* | 0.089* | 0.060 |
| | (0.028) | (0.042) | (0.049) | (0.040) | (0.043) | (0.046) | (0.040) |
| r/HousePlants | -0.232*** | -0.164** | -0.197*** | -0.156** | -0.189*** | -0.172** | -0.166*** |
| | (0.056) | (0.067) | (0.065) | (0.067) | (0.064) | (0.067) | (0.046) |
| r/IWantToLearn | -0.115*** | -0.048 | -0.080 | -0.040 | -0.073 | -0.056 | -0.079** |
| | (0.038) | (0.051) | (0.049) | (0.052) | (0.049) | (0.051) | (0.038) |
| r/LanguageLearning | -0.062*** | -0.025 | -0.061 | -0.022 | -0.055 | 0.025 | -0.063* |
| | (0.023) | (0.055) | (0.053) | (0.060) | (0.054) | (0.044) | (0.038) |
| r/LearnProgramming | -0.090*** | -0.023 | -0.055 | -0.015 | -0.047 | -0.030 | -0.057 |
| | (0.034) | (0.048) | (0.046) | (0.049) | (0.045) | (0.048) | (0.035) |
| r/Netflix | 0.056 | 0.123** | 0.090* | 0.131** | 0.098** | 0.115** | 0.075* |
| | (0.040) | (0.053) | (0.050) | (0.053) | (0.050) | (0.053) | (0.042) |
| r/NoStupidQuestions | -0.065** | -0.012 | -0.051 | -0.016 | -0.049 | -0.011 | -0.055* |
| | (0.027) | (0.043) | (0.043) | (0.048) | (0.045) | (0.040) | (0.030) |
| r/Pets | -0.115*** | -0.048 | -0.080** | -0.040 | -0.073* | -0.056 | -0.090*** |
| | (0.026) | (0.042) | (0.039) | (0.043) | (0.039) | (0.042) | (0.031) |
| r/Photography | -0.002 | -0.006 | -0.025 | -0.015 | -0.030 | -0.001 | -0.018 |
| | (0.040) | (0.061) | (0.061) | (0.063) | (0.060) | (0.060) | (0.047) |
| r/Skiing | 0.003 | 0.073 | 0.040 | 0.082* | 0.048 | 0.066 | 0.026 |
| | (0.037) | (0.047) | (0.046) | (0.047) | (0.045) | (0.047) | (0.037) |
| r/Sleep | -0.241*** | -0.174*** | -0.206*** | -0.166*** | -0.199*** | -0.182*** | -0.174*** |
| | (0.048) | (0.058) | (0.055) | (0.058) | (0.055) | (0.058) | (0.040) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| r/SocialSkills | -0.059* | 0.009 | -0.024 | 0.017 | -0.016 | 0.001 | -0.039 |
| | (0.032) | (0.045) | (0.043) | (0.046) | (0.042) | (0.045) | (0.034) |
| Constant | 0.407*** | 0.333*** | 0.363*** | 0.326*** | 0.357*** | 0.338*** | 0.278*** |
| | (0.024) | (0.040) | (0.037) | (0.040) | (0.036) | (0.039) | (0.029) |
| Observations | 128 | 154 | 154 | 154 | 154 | 154 | 154 |
| $R^2$ | 0.632 | 0.460 | 0.487 | 0.448 | 0.486 | 0.472 | 0.529 |
| Adjusted $R^2$ | 0.523 | 0.334 | 0.367 | 0.319 | 0.366 | 0.348 | 0.418 |
| Residual Std. Error | 0.081 (df = 98) | 0.097 (df = 124) | 0.099 (df = 124) | 0.098 (df = 124) | 0.098 (df = 124) | 0.097 (df = 124) | 0.075 (df = 124) |
| F Statistic | 5.806*** (df = 29; 98) | 3.647*** (df = 29; 124) | 4.064*** (df = 29; 124) | 3.474*** (df = 29; 124) | 4.050*** (df = 29; 124) | 3.820*** (df = 29; 124) | 4.793*** (df = 29; 124) |

*Note:* *p<0.1; **p<0.05; ***p<0.01

Parenthetical codes denote which imputed value types were used: D - distribution-based, LB - lower-bounds, UB - upper-bounds.

**Table G2.** Model results for total interactions outcome

# Appendix H: Outcome Interpretation Equations

This section details the algebraic steps used for easier interpretation of our outcome variables. While these equations use total interactions as the variable of interest, the same equations hold true for net score.

**Step 0:** Regression Model:

$$\frac{log(Total\ Interactions)}{log(Views)} = \beta_0 + \beta_1(Treatment) + \beta_2(Subreddit) + \varepsilon$$

**Step 1:** Rearranging to solve for Total Interactions:

$$log(total\ interactions) = [\beta_0 + \beta_1(Treatment) + \beta_2(Subreddit)] \times log(Views)$$

$$total\ interactions = e^{[\beta_0 + \beta_1(Treatment) + \beta_2(Subreddit)] \times log(Views)}$$

**Step 2:** Using $e^{a\,log(b)} = b^a$, we derive the equation that can be used to find the predicted total interaction for a number of views, given a treatment group and subreddit

$$total\ interactions = Views^{[\beta_0 + \beta_1(Treatment) + \beta_2(Subreddit)]}$$

**Step 3:** Using $\frac{a^m}{a^n} = a^{m-n}$, we find the ratio of interactions between a neutral ($X_1$=0) and treatment post ($X_1$=1) in the baseline category subreddit ($X_2$=0)

$$\frac{Interactions_{treatment}}{Interactions_{neutral}} = \frac{Views^{\beta_0 + \beta_1 X_1}}{Views^{\beta_0}} = Views^{\beta_1}$$

# Appendix I: Sentiment Analysis Regression

A regression on the results of running a sentiment classifier on the replies to posts in each treatment condition reveals some small but significant effects. Specifically, replies to posts in the liberal condition were 9.3% more likely to be positive compared to control, while replies to posts in the conservative condition were 6.3% less likely to be negative compared to control, both significant at the 5% level. The sentiment analysis portion of this study was not considered in the original experimental design, so questions were not crafted to elicit neutral language in users' responses. Because of this, it is difficult to account for the cause of any differences in sentiment. Therefore, more work is needed to fully understand the interplay between comment language and political signaling. As such, these results are left in an appendix as a suggestion to spur further research.

**Sentiment Regression Results**

| | Dependent Variable: Likelihood of Reply in Given Sentiment Class | | |
|---|---|---|---|
| | Positive Sentiment | Neutral Sentiment | Negative Sentiment |
| | (1) | (2) | (3) |
| Conservative | 0.073* | -0.010 | -0.063** |
| | (0.038) | (0.036) | (0.026) |
| Liberal | 0.093** | -0.061 | -0.032 |
| | (0.047) | (0.047) | (0.035) |
| Constant | 0.059* | 0.740*** | 0.202*** |
| | (0.033) | (0.047) | (0.048) |
| Observations | 1,591 | 1,591 | 1,591 |
| $R^2$ | 0.099 | 0.067 | 0.088 |
| Adjusted $R^2$ | 0.088 | 0.056 | 0.077 |
| Residual Std. Error (df = 1571) | 0.450 | 0.484 | 0.307 |
| F Statistic (df = 19; 1571) | 9.112*** | 5.925*** | 7.963*** |
| *Note:* | | | *p<0.1; **p<0.05; ***p<0.01 |

**Sentiment Regression Results**

Subreddit factors in regression omitted from table. Robust Standard Errors clustered by post.

**Table I1.** Model results for sentiment analysis outcome