Gender and Score Outcomes on Social Media: Evidence from a Reddit Field Experiment

John Andrus, Jared Dec, Weijia Li, Jamie Smith, Fengjiao Sun

Abstract

Social media and online communities continue to grow both in size and in importance to modern society. However, experimental study of online culture - particularly with regards to gender issues - has not been widely performed. In this paper we investigate whether gender signalling in Reddit usernames affects the scores that other users attribute to the content. Upon completion of the study, we did not observe significant differences in content scores based on username signalling in general.

1. Introduction

Online communities often resemble physical communities in the way that their members are subject to the same societal rules, expectations, and prejudices that affect them in person. Platforms such as Facebook, Instagram, TikTok, and many others allow individuals to represent themselves online in ways that closely align with their true appearance and identity.

However, online communities are unique as a result of the fact that they can connect extraordinarily large numbers of people together who may leave a quantitative value judgement on almost every interaction they have (likes, hearts, shares, upvotes, etc). These interactions can result in actual monetary value for creators who amass large followings. It is reasonable to suggest that, just as in real life, certain identities pay more than others. Likewise, the implications that immutable characteristics may have on online harassment and content removal can be severe.

The social media site Reddit is different from most other social networking platforms because its design promotes anonymity. New users do not need to input any demographic information or upload images or videos of themselves. Usernames may be anything a person chooses and the site rules explicitly ban taking actions that reveal a user's identity. This makes it a valuable testing ground for testing the relationship between identity and outcomes in virtual spaces.

In this study, we investigate the relationship between the perceived gender of Reddit users on the scores that their posts to the website receive. We have focused on gender in this study for two main reasons. First, similar observational studies have found relationships between gender and scoring outcomes on other platforms, but we could not find any examples of this question being addressed by a randomized controlled trial. Second, misogyny and sexism in online spaces have been a subject of considerable public and media interest in recent years with online cultural issues being tied to real-world impacts. The research question that we will attempt to answer through this experiment is: Does having a Reddit username that is readily identifiable as female influence the number of upvotes or comments that a post receives?

Our goal is to assess the difference in outcomes between posts made by Reddit accounts that signal a female user and posts made by accounts that signal no gender whatsoever. We will then identify if different categories of communities appear to show a greater effect than others

through covariate analysis. Finally, we will share lessons learned and areas for improvement in hopes of informing future research.

2. Literature Review

Before conducting a review of similar work, it is important to dedicate some time to explaining why one can make the assumption that there will be any difference in upvotes as a result of user gender. To explain this, we must exit the realm of data science and enter one of a seemingly unrelated field, gender studies.

The idea that the average person has preconceived notions about female or male areas of expertise and these reflect in their assumptions about what speech patterns males and females use is called the "genderlect" theory in the field of gender studies. The idea of the Genderlect theory was conceived by Deborah Tannen in 1990 and theorizes that communication between males focuses more on exchange of information, while female communication focuses more on feeling and supporting others.[1] The idea of whether or not this is true, is not relevant to our study but what is at hand is whether or not the average person, i.e. the average Reddit user who sees our Reddit posts, has this particular bias.

Numerous studies have been done to test if this theory applies to the average person, but we used as our primary reference "The Prevalence of Gender Communication in Social Media" from 2012.[2] This study showed that when a body of students were tasked with providing feedback on two anonymized Facebook profiles with differing levels of self-disclosure, the students worded their feedback using gendered pronouns despite there being no no gender information associated with either profile.

This study, although on a very small scale, shows that people tend to make assumptions about people based on their gender, i.e. the owner of Profile 1 was female because they cared about supporting their family and the owner of Profile 2 was male because they seemed disinterested in giving unnecessary details. Based on this study, there is some reason to believe that people perceive people differently based on gender. Based on these results and the general genderlect theory, we can expect that if this theory is correct and the general population of Reddit expresses it, we should expect a difference in upvotes.

The most similar study that the team could find to our conceived experimental design based on this hypothesis, was conducted in 2020 entitled "The Effects of Gender Signals and Performance in Online Product Reviews" which attempted to determine if gender signalling in usernames affected relative upvotes vs downvotes on Amazon product reviews with similar content.[3]

---

[1] Tannen, Deborah. "Rethinking power and solidarity in gender and dominance." *Annual Meeting of the Berkeley Linguistics Society*. Vol. 16. No. 1. 1990.
[2] Eckman, Alicia, Kelsey Fisher, and Talia Stifter. "The Prevalence of Gender Communication in Social Media." *Concordia Journal of Communication Research* 1.1 (2014): 1.
[3] Sikdar, Sandipan, et al. "The Effects of Gender Signals and Performance in Online Product Reviews." *arXiv preprint arXiv:2001.09955* (2020).

The key difference though is this study relied on using deep learning to find existing posts posted by actual users to find similar posts that only differed in terms of having a male or female signalling username. Such an approach could be difficult to implement as their study reviewed millions of posts with only a few thousand actually read through by humans to determine if the classification by the deep learning algorithm of "similar posts" was actually valid.

The study found that overall there was little difference in upvotes between male and female indicating usernames but there were large differences depending on the category. For instance, this study found there were differences as large as 40% favoring male posters in categories such as Electronics and differences as high as 20% favoring female posters in categories such as Beauty. Some categories such as Books or Toys were almost exactly even in upvotes for males vs female posters. The main conclusion to take away from this study is that bias towards a poster's perceived credibility depends heavily on what the subject of the post is. If this study is reproducible and if our study is not flawed in its design, we should expect to see differences in upvotes based on subreddit topics.

3. Experimental Design

3.1 Subjects

The subjects of this experiment are subreddits on Reddit. A subreddit is a forum dedicated to a specific topic on Reddit. Each subreddit has a group of subscribers and viewers that are interested in the posts regarding the specific topic the subreddit is dedicated to. The group of subscribers and viewers to each subreddit are able to see the posts on the subreddit and react to the post through upvotes, downvotes, and comments at their own discretion.

As of December 2021, there are over 2.8 million subreddits; however, not all the subreddits have suitable activity levels or content types for this experiment. For this reason, we established activity level and content type as selection criteria for the subjects in this experiment.

To ensure treatment effect can be measured during the time of the experiment, we selected subreddits with sufficiently high activity levels to provide the greatest opportunity for content engagement. We initially considered several different subreddit metrics such as number of subscribers, total subreddit upvotes and comments, and total post counts. Number of subscribers is not a good indicator of subreddit activity because we cannot determine the number of inactive subscribers per subreddits. Total numbers upvotes, comments and posts in each subreddit are also not good indicators of the subreddit activity because we cannot determine whether the upvotes, comments and posts are from recent subreddit activities.

We finally settled on daily comment counts as the acceptable metric of current activity. For this experiment, a subreddit is considered active if it has at least 100 comments a day[4]. There are 4350 subreddits meeting this criteria as of October 2021. Selecting active subreddit increases the potential treatment effect size and ultimately increases the statistical power of this experiment.

[4] The activity data is based on Subreddits Stats: https://subredditstats.com/

In order for the 4350 active subreddits to be eligible for this experiment, the subreddits need to allow users to post links of news articles to the subreddit. Each subreddit has a list of rules that restrict the type of content posted on it. Some of the restrictions are image only, video only, original content only, etc. Subreddits with the above restrictions are excluded in this experiment. There is no systematic way to identify subreddits with different rules. As a result, the restrictions are identified and excluded by the researchers executing this experiment. There are around 24% of the subreddits that have met the activity criteria excluded from the experiment due to rule based restrictions.

3.2 Treatment

To find out whether gender signaling impacts the response to social media posts, we will perform an experiment on the Reddit platform by tracking responses to posts generated from gender-neutral usernames (control) and usernames which signal that the poster is female (treatment).

Table 1: Control and Treatment usernames used throughout the experiment

Control Usernames (n=19)	Treatment Usernames (n=19)
Radiant-Fun5315, Alert-Change-6744, True_Basis5414, Professional_Bus8426, somber-groceries, PositiveCattle6485, Mammoth_Ad_6474, TastyTruth170, Objective_Comb_264, Eastern-Sandwich573, Pretend_Ad_55, ProfessionalTotal711, EfficientPie4436, Special-Jump3252, Impossible-Adagio-15, Asleep-Explorer3406, OddScarcity123, NumericalAlphabet123, Sudden_Report_3159	Hovercraft_empress, Impossible-Girl-0, Kooky-Princess, promising-woman, EverydayQueen2021, Old-Location-Queen, Rebel_Girl_564, LadyLovely123, Empress_3076, Queen-Journalist3833, Least_Figure_Gal, Fluffy_Start_Queen, QueenKong875, Independent_Lady798, Princess_Winter8286, DuchessoftheSeaside, MondaySenorita, TimelyDamsel, MoreThanaWoman135

***Notes**: This table shows list of usernames used for both the Control and Treatment groups throughout the experiment*

Furthermore, usernames in the control group were generated by Reddit, which provides suggested usernames when creating an account, which have unidentifiable demographic attributes. The usernames created in the treatment group were derivations of the randomly generated username, however, the researchers adjusted the usernames to add words associated with females (Girl, Queen, Lady, etc.). The reason that a total of 38 usernames were created was due to the fact that each of the experimenters worked independently to collect data, and Reddit times out usernames from posting too frequently to prevent spam. Having several usernames per experimenter allowed for continuous data collection. Furthermore, some of the selected subreddits were treated multiple times, and for subsequent posts on the same subreddit a different set of control and treatment usernames was employed to prevent any one username from becoming recognizable among the subreddit's users.

The subjects for this experiment were the subreddits of the Reddit platform, however, we administered the treatment, a post from a control username and a treatment username, within subreddits, to elicit upvotes and comments from individuals that follow those individual subreddit threads.

The administration of the treatment, a post by the female gendered username, occurred at approximately the same time as the post by our control usernames. It's important to note that the content of the posts are not identical. Subreddits do not allow duplication of content when individuals post to the Subreddit. The researchers took the approach of posting the top two most relevant news stories related to the subreddit thread via a Google search. The reason that only articles were posted and not other types of content such as images, videos, or other self-created content, was a desire to have content of more uniform quality. It would be difficult for five independent experimenters to create self-created content of consistent levels of relevance, quality, and witness The assignment of the news articles was randomly determined for each sample. We then waited 24 hours and recorded the number of upvotes and comments from the post generated by the treatment username and the control username.

Throughout the experiment, we found that accounts accumulated "karma," a Reddit feature that attributes valued content to users, which consequently allowed some of our usernames to post content with less moderation. This unintended result meant that certain accounts were able to post in subreddits without the content being reviewed by the moderators of the subreddit thread and thus had a higher likelihood of receiving upvotes and comments. Though this was initially a challenge, if our hypothesis is correct, then this would in turn mean that female gender usernames would have fewer upvotes and comments as a result of accumulating less karma. Though this was an unintended phenomenon, it is one that we were comfortable with not controlling for since it had the potential to further expose differences in the accumulated upvotes and comments generated in posts by our control and treatment usernames.

Figure 1: Example Treatment and Control Posts

Notes: *Control and Treatment example posts used in the experiment*

3.3 Random Assignment

This study used a within-subjects design which required each subject to receive both treatment and control as opposed to a between-subjects design that assigns each of the subreddits randomly to either treatment or control. However, of the 4350 subreddits that met the criteria as experimental subjects, 391 were randomly selected to be included in the experiment. Additionally, each subreddit was randomly assigned to receive either treatment first or control first as a way of mitigating any impact that post order might have on the results.

Additionally, The subreddits were not randomly assigned to each one of the people who conducted this experiment, and instead our experimenters self-selected for which of the valid subreddits they would post in, typically choosing subreddits for which they had some prior knowledge or relevant background to the subreddit's topic. This means that the subreddit

assignment was not truly random as the decision to "skip" an assigned subreddit for treatment was decided by the experimenters themselves based on observations of the subreddit's most recent top posts and the subreddit's listed moderation rules (which should be noted were not always complete, as posts that seemingly followed all listed rules were often removed without a reason given). Regardless, the lack of true random assignment to treatment should be considered a flaw in this experiment based on a within-subjects design.

We believe that our design meets the excludability assumption in that the only apparent causal effect is through receipt of a treatment or control post. We identified no evidence that our experiment resulted in actions from individuals outside the experiment that would impact outcomes and our team used the same estimands and ATE calculation for all subjects.

Of greater concern is the non-interference assumption. It is possible that, particularly in smaller communities, individual user accounts could gain within-subreddit or across-subreddit reputations that would result in different outcomes in a subreddit not because of how it was treated, but how others were treated as well. Additionally, certain user accounts might be judged partially on their total accumulated upvotes (or Karma) which is publicly available for all users to see. However, by only selecting large, active subreddits as described we feel that it is reasonable to assume that individual user reputations could only persist for the most established and prolific users within them. We do not feel that the accounts created for this study meet this criteria and thus this assumption is met.

3.4 Within Subjects Design

For this study the decision of whether to implement a between-subjects design or a within-subjects design had to be made. There are many factors to consider as each has significant potential benefits and drawbacks. A within subjects design requires fewer observations and does not require blocking by subreddit subject. However, it makes several key assumptions, no anticipation and no persistence, the former of which is easily met by our design due to no one really anticipating Reddit posts but the latter poses far more problems which will be discussed later. The between subjects design requires far more observations, but makes fewer assumptions, and blocking in of itself is easier to implement. To explain why a within subjects design was chosen, we need to discuss the exact number of subjects each one requires and which is more feasible for the social media platform of Reddit in the first place.[5] For this calculation the following assumptions are: the test family is t-test (difference in means between two groups), two tails were used, the effect size was set to small (0.2), significance level was set to 0.05, power was set to 0.8. Using two tails, a power level of 0.8, and a significance level of 0.05 are all based on these choices being semi-standard for experiments. Far more problematic is effect size, though we chose 0.2 because it is generally described as standard for detecting a "small" effect size. Based on these assumptions, a within subject design requires 199 subjects to determine statistical significance and a between subjects design requires 788 subjects (394 for treatment and 394 for control). If data collection was no issue, between subjects would be the preferable design as the assumption of no persistence for the within subjects design is quite problematic, we can only test after the fact if the posts we posted first in a subreddit have a statistically significant

[5] For these calculations, a statistical power calculator was used:
https://www.ai-therapy.com/psychology-statistics/sample-size-calculator

difference from those that were posted second. However there are a finite number of subreddits with enough activity that a random sample can be drawn from and herein lies the primary catalyst for the choice of a within subjects design.

The initial random sample of subreddits that the experiment was designed around was selected from the 4,350 “most active” subreddits on the site based on the criteria of having at least 100 comments per day. However initial tests that were run suggest that a large number of subreddits had strict auto-moderation in place that would automatically delete the type of posts that our experiment intended to post (links to articles that were related to the subreddit but had not been previously posted). The sample that was taken also contained many subreddits that were not conducive to posting articles, for various reasons. Some subreddits only allowed posts of memes, images, videos, or had other limiting factors that prevented them from being treated by our experiment design.Finally some subreddits specifically had very few articles on their subject matter that had not previously been posted (examples include entertainment products that were not released recently). It became apparent early on that it would be a challenge to find enough viable subreddits to post on and not have the post deleted to collect data. Given that the rough definition of an active subreddit only gave a subset of 4,350 subreddits to pull a sample from, and that a large portion of these subreddits did not have moderation rules or general post content that was conducive to the experimental posts being allowed, the between subjects design was highly problematic. It was theorized that due to the uncertainty of the auto-moderation in place on subreddits that allowed article posting of the type that was intended, at least twice as many subreddits would need to be treated as the statistical power calculator suggested. So for a between-subjects design, this would require 1,600 subreddits to receive posts. And given the fact that a large portion of the initial 4,350 subreddits that the sample was pulled from did not allow articles to be posted in the first place, it seemed unlikely that the reality of the active subreddits would allow the between subjects design to be successfully implemented. Consequently, a within subjects design was chosen.

To address the assumptions of a within subjects design once again, the two core assumptions are of no anticipation and no persistence. The no anticipation assumption is easily accounted for, it is extremely unlikely that any of the users who happened to access the subreddit and see the experimental posts linking articles anticipated those articles being posted. Reddit posts, especially those like the experimental posts that added no original content whatsoever, are not really ever anticipated, merely seen when the user happens to browse the subreddit.

No persistence is more of an issue. The no persistence assumption essentially means the first treatment option (treatment or control) received by the subject does not affect the second. For this experiment, almost all subreddits automatically removed posts linking to an article that had already been posted earlier. This meant that the articles posted for the treatment and control options had to differ. The result is that the posts themselves were different in a way that was not simply due to the username differing between being female or non-gender indicating, and the treatment outcome that was observed could be due to the article’s content and not simply because of the treatment itself.

Still, the assumption of no persistence has another application to this study which must be addressed as well. Some of the subreddits in this sample were treated twice. While different

accounts were used for the second round of posts (in order to ensure that no experimental user gained some form of following by the Reddit users who browsed the subreddit), a washout period of a minimum of five days was employed between the first and second rounds of treatment. To test if the assumption of no persistence was met, tests were run to see if there was a statistically significant difference between the first and second rounds of treatment and to test if the first post in any round of treatment had a statistically significant difference from the second (for each round of treatment and control, the choice to post the control post or the treatment post first was randomly decided). At first glance, despite the order of treatment and control being randomly assigned, it seems that treatment came first in more instances, 200 out of 391 total samples collected had treatment first while 191 had control first. Assuming a probability of 0.5 for control being first, the expected number of times control would be expected to be assigned to be posted first would be 195.5 with a standard deviation of 9.89 (found by taking the square root of 391*0.5*0.5).[6] This means that while treatment was assigned to be posted first more often than the expected mean, it is within 1 standard deviation of the expected number of assignments and therefore there is not a statistically significant difference in the number of times treatment was assigned to be posted first. Of the 206 treatment-control pairs in which neither post was deleted by moderators, the total number of upvotes for those which were posted first were 2,298 or an average of 11.87 upvotes while the posts which were posted second in the treatment-control pair, the total number of upvotes were 2,684 or an average of 14.13 upvotes per post. A paired, two-tailed t-test of these upvotes gives a p-value of 0.4611, ruling out the possibility of statistical significance in this difference.

In all, it appears that the no persistence assumption has been satisfied based on the t-test that was conducted. However, a significant flaw remains in this analysis which is the fact that the assignment of subreddits to be subject to the experimental posts is not truly random. While the experiment design generated a random number for all of the 4,000 subreddits which were judged to be the most active which was then sorted. However, as was mentioned previously, many subreddits that would have been treated based on random assignment could not be treated due to either subreddit rules, typical posts for that particular subreddit not being articles, or a lack of relevant articles to that particular subreddit's topic.

3.5 Estimand

There are two potential outcome metrics for this experiment: count of upvotes and count of comments within 24 hours of posting. Upvotes signal Reddit users' approval or support for a post. Comments allow users to express their opinion and discuss the topic with other Reddit users. Since users in each subreddit can upvote and comment on the post created, we would like to understand the impact of gender on both upvotes and comments. The treatment effect for this experiment is the difference in number of upvotes of the two posts created by the gender neutral username and the female indicated username in each subreddit under the within-subject design after the posts are created for 24 hours. Similarly, the treatment effect measured by the number of comments is analyzed as well.

---

6 https://stats.stackexchange.com/questions/21581/how-to-assess-whether-a-coin-tossed-900-times-and-comes-up-heads-490-times-is-bi

There is a difference between upvotes and comments. Each user in the subreddit can only upvote the posts once and can comment as many times as they want. There is a potential amplification of the potential outcome by measuring the comments alone. Therefore, measuring the number of upvotes of the post is a better indicator of subreddit's treatment effect with posts created by different gender-indicated usernames.

3.6 Non-Compliance

In this study, treatment and control posts were administered directly by the experimenters. This resulted in zero noncompliance.

3.7 Attrition

This study saw considerable attrition among its selected subject pool. We defined attrition as any subreddit that deleted the treatment post, control post, or both within the 24-hour observation period. Attrited subreddits were not included in the ATE calculation nor were they considered eligible for retreatment after our predetermined washout period. This experiment had only one treatment condition and observed a 47.6% attrition rate across all posts.

The missing data in this study is hard to justify as independent of potential outcomes. Posts on Reddit are deleted for three primary reasons:

1. Account age or karma restrictions
2. Violation of subreddit rules
3. Suspected bot activity

It is reasonable to assume that there may be a relationship between the types of subreddits that face these causes for attrition and the potential outcomes for these subreddits. For example, busier subreddits that would foreseeably result in larger outcomes may be more likely to strictly enforce rules and use automoderator features, thus resulting in a higher rate of attrition for larger subreddits. Subreddits that are more hostile to female users may also remove posts at a higher rate from these users, thus resulting in differential attrition that would bias the treatment effect toward zero.

4. Results

4.1 Data

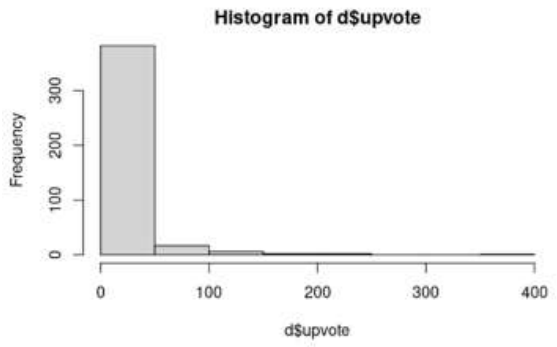
We started with 391 subreddits where we created posts by using both gender-neutral (control) and female (treatment) usernames. Of those 391 subreddits, 205 existed for at least 24 hours and were able to be analyzed. The rest 186 subreddits have at least one control or treatment post removed, which we define as attrition. We had 47.6% attrition in total.

The outcome that is being measured is whether the number of upvotes would vary by the gender of the username. First, we want to look at **the number of upvotes**. As seen below, while most posts received less than 20 upvotes, we also observed that a few posts were quite popular and received hundreds of upvotes. Considering that high-upvotes posts would attract more views and

potentially more upvotes, we decided to log transform the number of upvotes. Below histograms demonstrate the original upvotes data and log transformed data.

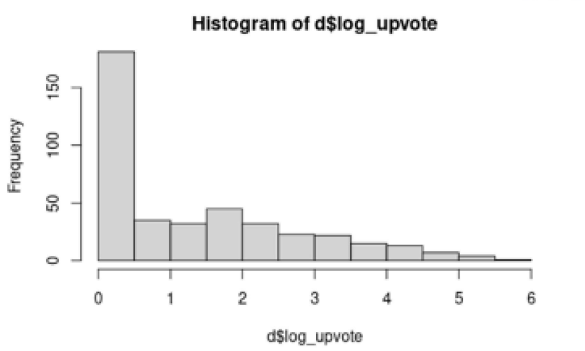
Figure 2: Original Upvotes Histogram vs Log Transformed Upvotes Histogram

Original **upvotes** (control & treatment)



***Notes**: Histogram of the number of upvotes for each of the subreddits*

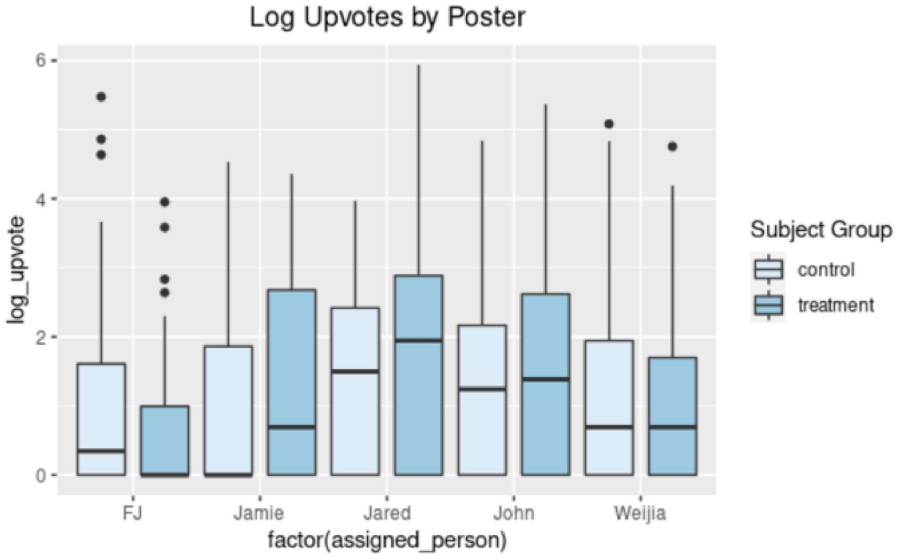
Log transformed **upvotes** (control & treatment)



***Notes:** Histogram of the log of upvotes for each of the subreddits*

Second, we are also concerned about the impact from the **poster**, or the posters. Because to make sure the sample size meets the statistical power, each of the team members selected the subreddits that are relatively familiar or interesting to him/her. This might inevitably result in a heterogeneous treatment effect. As seen below, when breaking down the log upvotes by poster, we can see some deviation between control and treatment groups by different posters. For the same reason mentioned above, we would not remove outliers.

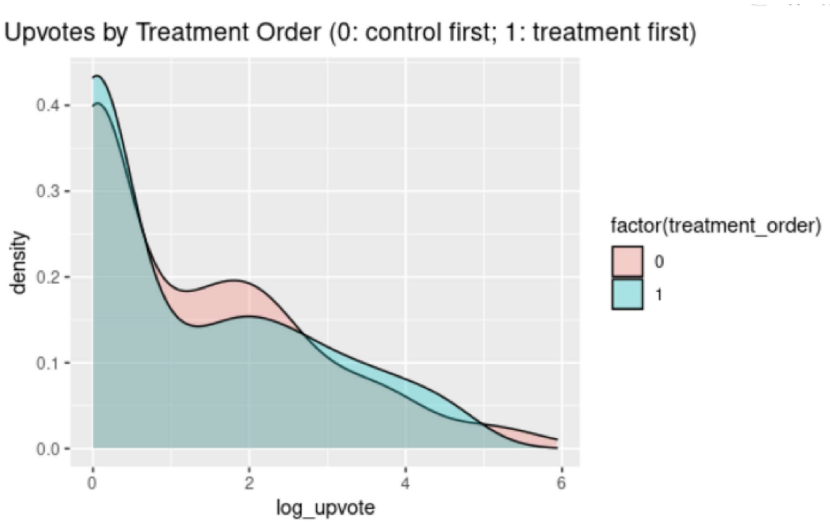
Figure 3: Boxplot of Log Upvotes by Poster



***Notes:** As shown on the boxplots, we can see some deviation between control and treatment groups by different posters when breaking down the log upvotes by poster.*

Lastly, though we randomized the sequence of posting and allowed wash-out period to minimize the impact from time, we want to make sure the **treatment order** would not cause any bias after we remove the attritted samples. As seen below, the effect from treatment order is not distinct – the treatment has a mixed effect on the log upvotes. Based on this finding, we will not include treatment order to the regression model.

Figure 4: Boxplot of Log Upvotes by Category



***Notes**: The effect from treatment order is not distinct – the treatment has a mixed effect on the log upvotes*

4.2 Regressions

As we adopted within-subject design, we apply the fixed effect of subreddit. Subreddit is unique for each sample and contains both the control post and treatment post . When using the linear regression model to estimate the mean difference in log upvotes between control (gender-neutral username) and treatment (female username) group, we structured models into 3 levels – the simplest model regressed log upvotes on the treatment (username gender) and the fixed effect of subreddits; the intermediate model by adding poster (`assigned_person`); and the improved model by adding interaction between poster and treatment.

$$Simplest\ Model\colon\ log\_upvote\ =\ \alpha\ +\beta_0 group\ +\gamma\cdot subreddits$$

From the simplest model, we found that there is no statistically significant effect of the username genders on the number of upvotes. In fact, the average treatment effect is only 0.038 with a robust standard error being 0.148 and R square being 0.672.

$$Intermediate\ Model\colon\ log\_upvote\ =\ \alpha\ +\ \beta_0 group\ +\beta_1 poster\ +\ \gamma\cdot subreddits$$

By adding poster (`assigned_person`) to the intermediate model, we found that there is still no statistically significant effect of the username genders on the number of upvotes, and the coefficient and standard error remains to be 0.038 and 0.148. The poster variable, however,

demonstrated a significant effect on the number of upvotes – posts that were created by Jamie and John had a statistically significant effect on the number of votes. A post that is created by Jamie or John would have a 4.149% and 3.473% increase in upvotes compared to posts that are created by FJ.

$$Improved\ Model\colon\ log\_upvote = \alpha + \beta_0 group + \beta_1 poster + \lambda\ group * poster + \gamma \cdot subreddits$$

We are also concerned with the potential heterogeneous treatment effects from the posters and created the improved model by adding an interaction of poster and treatment variable to the intermediate model. From the Table 2 below, we can see that the average treatment effect of username gender is not statistically significant, but the R square improved from 0.672 to 0.680. From the heterogeneous treatment effect perspective, there is no statistically significant evidence that username gender works differently for different posters in increasing the upvotes. Therefore, there is no heterogeneous effect.

Table 2: Regression Result - Username Gender Impact on Post Scores

Username Gender Impact on Post Scores

	Dependent variable:		
	log_upvote		
	simplest model	intermediate model	improved model
	(1)	(2)	(3)
grouptreatment	0.038	0.038	-0.398
	(0.148)	(0.148)	(0.465)
assigned_personJamie		4.149***	3.826***
		(0.086)	(0.407)
assigned_personJared		-0.000	-0.368
		(0.038)	(0.469)
assigned_personJohn		3.473***	3.245***
		(0.903)	(0.994)
assigned_personWeijia		0.000	-0.137
		(0.038)	(0.412)
grouptreatment:assigned_personJamie			0.646
			(0.571)
grouptreatment:assigned_personJared			0.735
			(0.553)
grouptreatment:assigned_personJohn			0.456
			(0.548)
grouptreatment:assigned_personWeijia			0.274
			(0.558)
Constant	4.130***	-0.019	0.199
	(0.111)	(0.079)	(0.371)
Subreddits Fixed Effects	Yes	Yes	Yes
Observations	410	410	410
R2	0.672	0.672	0.680
Adjusted R2	0.430	0.430	0.434
Residual Std. Error	1.098 (df = 235)	1.098 (df = 235)	1.094 (df = 231)
F Statistic	2.770*** (df = 174; 235)	2.770*** (df = 174; 235)	2.759*** (df = 178; 231)
Note:			*p<0.1; **p<0.05; ***p<0.01

Note: The table shows the post-level upvotes in gender and score outcomes study regressed on username gender indications (simplest model), poster (intermediate model) and interaction of poster and username (improved model). We applied fixed effects on "subreddits". "Group treatment" reflects whether the post was in control (gender-neutral username) or treatment (female username) group. "Assigned_person" shows the authors of the samples; "group treatment:assigned_person" reflects the interaction between posts and the authors. Standard errors use robust standard deviation and are shown in parentheses. Stars denote significance level of the difference: *, significant at 10 percent confidence level; **, significant at 5 percent confidence level; ***, significant at 1 percent confidence level.

For an easy read on the poster's impact on upvotes, we further broke down the improved model by each poster and the result is shown below (Table 3). There is no statistically significant impact of posters on the treatment effect. And the coefficient is in line with Table 2. For example, John's posts in the treatment (female username) group received an 0.058% higher upvotes than posts in his control (gender neutral) group. This is the same effect as calculated from Table 2: -0.398 + 0.456 = 0.058.

Table 3: Regression Result - Username Gender Impact on Post Scores by Posters

Username Gender Impact on Post Scores by Poster

	Dependent variable:				
	log_upvote				
	John	Jamie	Jared	Weijia	FJ
	(1)	(2)	(3)	(4)	(5)
grouptreatment	0.058	0.248	0.338	-0.124	-0.398
	(0.290)	(0.331)	(0.298)	(0.308)	(0.465)
Constant	-0.029	4.025***	4.055***	0.408*	0.199
	(0.145)	(0.166)	(0.160)	(0.247)	(0.269)
Subreddits Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	84	68	96	94	68
R2	0.514	0.811	0.751	0.673	0.562
Adjusted R2	0.349	0.648	0.537	0.366	0.136
Residual Std. Error	1.138 (df = 62)	0.912 (df = 36)	1.055 (df = 51)	1.025 (df = 48)	1.318 (df = 34)
F Statistic	3.118*** (df = 21; 62)	4.986*** (df = 31; 36)	3.502*** (df = 44; 51)	2.194*** (df = 45; 48)	1.320 (df = 33; 34)
Note:					*p<0.1; **p<0.05; ***p<0.01

Note: The table shows the post-level upvotes in gender and score outcomes study regressed on each individual poster. We applied fixed effects on "subreddits". "Group Treatment" measures the difference of upvotes between control and treatment group from the assigned person's posts. Standard errors use robust standard deviation and are shown in parentheses. Stars denote significance level of the difference: *, significant at 10 percent confidence level; **, significant at 5 percent confidence level; ***, significant at 1 percent confidence level.

We also measured the number of comments and the sum of comments and upvotes, but found there is no statistically significant impact from any of the treatment or covariate variables. For simplicity, we will keep the results with comments and the sum in the code report.

5. Omitted Variable Bias

There are a number of factors that could be impacting these results that are not directly captured in the controls that were presented above. First and foremost, as mentioned previously, the content of the posts differed within the same subreddit between control and treatment due to the requirements imposed by Reddit's auto-moderation, which automatically deletes posts of the same link within the same subreddit. This is a significant problem for this experiment because any differences that are detected within a treatment-control pair on the same subreddit could be entirely to the subreddit's users preferring the content of one of the articles over the other, rather than any actual difference in perception of a reddit poster due to the gender indicated by their username.

Another unmeasured variable which could significantly affect the interpretation of our results is that of the Reddit ranking algorithm. All posts on all subreddits of Reddit are subject to a ranking algorithm that sorts the posts on a given subreddit by a function that weighs the amount of time since they were posted and the net number of upvotes that the post has received. Reddit is open-source and the ranking algorithm is freely available online.[7] However, actually computing in real-time the average page rank of these posts on their respective subreddits is still beyond the scope of this experiment. Certainly posts that were posted at less active times for the target audience of that subreddit might have contributed to more time decay in this algorithm relative to the number of upvotes the post received, causing the post to be inherently lower on the page rankings for more of the 24-hour time period in which upvotes were measured. An example would be some of the posts made in subreddits dedicated to places in the United Kingdom. All of our experimenters were located in the United States and mostly made these posts in the evening hours in US time zones which are the very early morning hours in the United Kingdom. As most of the likely readers of these subreddits would not have seen these posts until much farther into the 24-hour measurement time period, these posts could have net lower upvotes compared to subreddits of other subjects, due to being ranked lower on the subreddit page.

A final potential omitted variable could have been the karma received by these accounts during the course of the experiments. Higher karma is almost universally recognized as a measure of the net quality of a user's posts across all subreddits. This is evidenced by the fact that many subreddits require a minimum amount of karma before allowing a user to post. However, even though almost none of the final subreddits included in this experiment had an overt karma requirement to post, some of the posts made by users may have gotten more upvotes because the account used to post had had more karma prior to posting. For example, a user may inherently be more willing to give more upvotes to a post made by a user just because they have a higher karma rating and therefore are considered a more trustworthy source of information. If there were no time constraint for this experiment, the way to truly prevent bias from this omitted variable would be to create a new account for every single post in the experiment.

6. Generalizability

The subreddits chosen for this experiment were randomly selected from a field of over 4,000. These 4,000 subreddits were identified through a selection method that ensured that they had a minimum daily comment threshold to gauge activity and also did not require that the posts within the subreddit were images, gifs, videos, or memes. Since the subreddits that we selected did not include certain subreddit posting types, we cannot conclude that our results would generalize to all subreddit threads. In particular, we excluded subreddits that did not allow links to be posted as well as those that required video, images, and memes. We also excluded NSFW subreddits, which allow explicit content, subreddits as part of this experiment.

[7] The Reddit ranking algorithm can be found here:
https://medium.com/hacking-and-gonzo/how-reddit-ranking-algorithms-work-ef111e33d0d9

7. Conclusions

This study contains a number of potential flaws with its given within-subjects framework. The largest and most glaring of which is the fact that the subreddit posts differ for reasons other than the gender indication of the username. This was necessary due to: a) a lack of viable subreddits that could lead to a between-subjects design having statistical power, and b) the universal auto-moderation rules for almost every subreddit that prevent identical article links from being posted in the same subreddit. However these circumstances leave us with the unenviable position of our results possibly being due to the article links posted and not the intended treatment effect. A further issue with the experimental design is the randomization not being truly random. Despite the experiment randomly sorting 4,350 "active" subreddits, deemed to be active by the definition set at the beginning of the experiment, there was a great deal of self-selection of subreddits by the experimenters due to the need to filter out subreddits which did not allow article links to be posted, and the experimenters could choose which subreddits they posted to based on familiarity with the subject of the subreddit.

Despite these flaws in the experimental design, the experiment gathered more observations than the initial statistical power estimated was needed based on a within-subjects design (206 observations were collected with neither the treatment nor control posts being deleted, and 199 observations was the projected total that would need to be collected to have statistical power). It should also be noted that due to the fact that the treatment effect was not found to be statistically significant, it is inappropriate to try and estimate a post-hoc statistical power.[8] There also appears to be no statistically detectable persistence effect with the posts nor is the random assignment of treatment or control to be posted first statistically unlikely to be truly random. It is also worth mentioning that the articles themselves were selected fairly consistently as the top trending news articles on Google for the subreddit's subject that had not been posted in the subreddit before. There is no reason to believe that either treatment or control was favored with superior article selection at the outset. Furthermore, there are no heterogeneous treatment effects at the 5% level when controlling for which member of the treatment group posted the articles. In the end, it seems that at least in aggregate, that there is no statistically detectable difference between treatment and control between upvotes received by posts from non-gender inciating usernames and female-indicating usernames, regardless of which controls are included.

7. Future Work and Areas for Improvement

There are some limitations to this experiment such as consistency between treatment and control post, confounding variables and attrition. In future iteration of this study, there are multiple opportunities of improvement. Due to limitations of resources and time, several assumptions are used in this experiment for treatment length and wash-out period. Additional analysis should be conducted prior to future experiments to validate assumptions used in this experiment. To prevent post deletion due to subreddit restriction on user account age and low karma, each Reddit user account going to be used for future studies should be created at least one month prior to the start of the experiment and have at least 60 karma. To achieve true randomization of subjects, future studies should remove the human factor in identifying which subreddits qualify for the

[8] See "The Dangers of Post-Hoc Analysis" at https://clincalc.com/stats/Power.aspx

experiment. Potential partnership with Reddit platform could help automate the identification of reddit restriction and reddit categories. To prevent inconsistency between treatment and control posts in each subreddit, the methodology of creating each post should be more systematic to reduce the variability between treatment and control posts such as content of the article, time of the post and the author of the post. To prevent karma and account specific metrics being a confounding variable causing user preference to the treatment and control post, the karma and account specific metric of the gender neutral account and the female indicated account treating each subreddit should be similar. To expand the generalizability of the experiment, other types of subreddits excluded in this experiment should be explored; such as images, videos and user created content. The types of treatment and control posts in the excluded subreddits should be appropriate according to the restrictions of the subreddits. Expanding the subreddit types also helps prevent manual interference of identifying subreddits that cannot be treated. Due to the execution time limit of the experiment, limited resources and number of active subreddits being treated, we carried out a within-subject design. This experiment can also be carried out by a between-subject design. With a between-subject design, we can eliminate the verification of anticipation and persistence assumption of the within-subject design. However, additional effort is required to prevent spill-over between treatment and control subject and to maintain sufficient power. Since subreddits do not allow duplicate articles shared within the subreddit, this experiment used top 2 popular articles based on google search. Therefore, the content of the treatment and control posts were different in each subreddit. Between-subject design can eliminate this downside of within-subject design. Future experiments should also include analysis on the posts deleted by the subreddit to examine whether there is a causal effect between the gender-indicated usernames and the posts deleted. With more detailed planning, design and resources that incorporate the above improvements, we can gain additional insights on the effect of gender-indicated username on the number of upvotes and/or comments that a post receives.

8. Appendix: Analysis performed in R is attached below

W241_final_project

```
library(data.table)

library(sandwich)
library(lmtest)
library(stargazer)
library(foreign)
library(AER)
library(ggplot2)
library(patchwork)

robust_se <- function(mod, type = 'HC3') {
  sqrt(diag(vcovHC(mod, type)))
  }
```

We started with 391 subreddits where we created posts by using both gender-neutral (control) and female (treatment) usernames. Of those 391 subreddits, 205 existed for at least 24 hours and were able to be analyzed. The rest 186 subreddits have at least one control or treatment post removed, which we define as attrition. We had 47.6% attrition in total.

```
d <- fread('./final_data_r.csv')

setnames(d, "1day_upvote","upvote")
setnames(d, "1day_comment", "comment")

head(d)
```

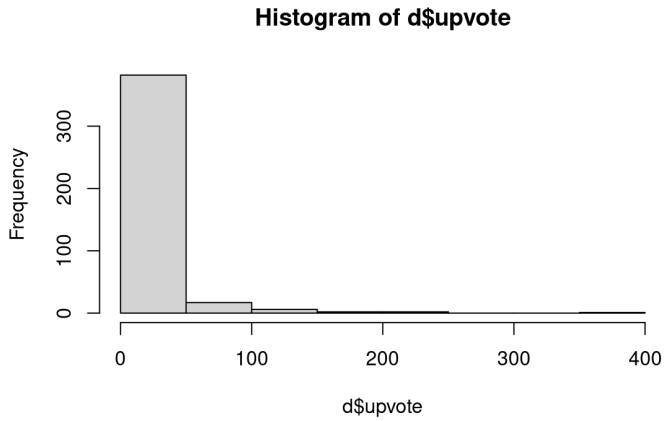
...	ID	subreddits	comments_per_day	random_assign_rank	NotUsedReason
<int>	<int>	<chr>	<int>	<int>	
1	3851	0xPolygon	110	478	
2	1027	49ers	542	745	
3	3193	ac_newhorizons	143	414	
4	654	Accounting	839	230	
5	654	Accounting - second time	839	230	
6	472	AirForce	1150	466	

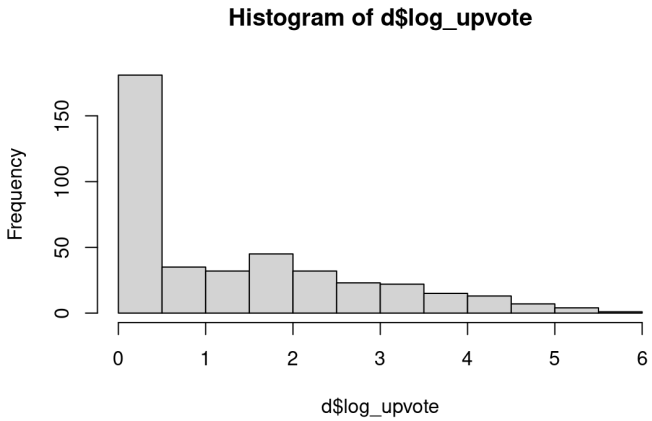
6 rows | 1-6 of 20 columns

Data exploration

The outcome that is being measured is whether the number of upvotes would vary by the account gender. First, we want to look at the number of `upvote`. As seen below, while most posts received less than 20 upvotes, we also observed that a few posts were quite popular and received hundreds of upvotes. Considering that high-

upvotes posts would attract more views and potentially more upvotes, we decided to log transform the number of upvotes. Below histograms demonstrate the original upvotes data and log transformed data.

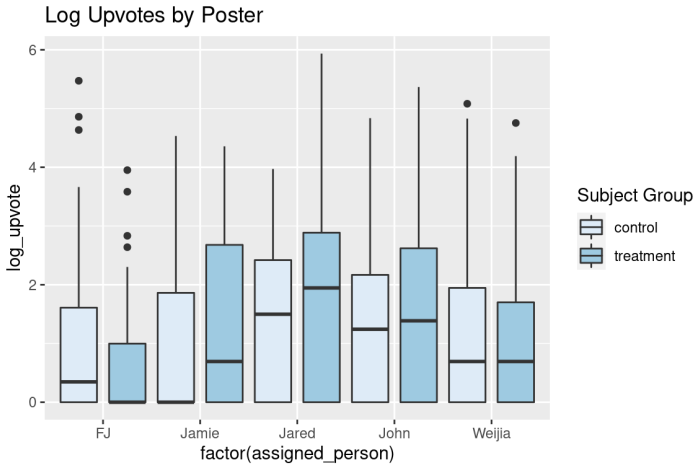




There are also a few important covariates that we want to explore before adding to the model:

`assigned_person` : we are also concerned with the potential heterogeneous treatment effects from the poster. To collect as many samples as possible, each of the team members selected the subreddits that are relatively familiar or interesting to him/her. This might inevitably result in a heterogeneous treatment effect. As seen below, when breaking down the log upvotes by poster, we can see some deviation between control and treatment groups by different posters. For the same reason as mentioned above, we would not remove outliers.

```
ggplot(data=d) +
    geom_boxplot( aes(x=factor(assigned_person), y=log_upvote, fill=factor(group)), posi
tion=position_dodge(1)) +
    theme_minimal() +
    scale_fill_brewer('Set3') +
    guides(fill=guide_legend(title="Subject Group")) +
    theme_update(plot.title = element_text(hjust = 0.5)) +
    ggtitle("Log Upvotes by Poster")
```



`treatment_order` : though we randomized the sequence of posting and allowed wash-out period to minimize the impact from time, we would still include this covariate to explore if any impact.As seen below, the effect from treatment order is not distinct – the treatment has a mixed effect on the log upvotes. Based on this finding, we will not include treatment order to the regression model.

```
ggplot(d, aes(x=log_upvote, fill=factor(treatment_order))) +
           geom_density(alpha=.3) +
            theme_update(plot.title = element_text(hjust = 0.5)) +
            ggtitle("Log Upvotes by Treatment Order (0: control first; 1: treatment firs
t)")
```

Regression

As we adopted within-subject design, we apply the fixed effect of subreddit, which has unique value for each sample. When using the linear regression model to estimate the mean difference in log upvotes between control (gender-neutral username) and treatment (female username) group, we structured models into 3 levels – the simplest model regressed log upvotes on the treatment (username gender) and the fixed effect of subreddits; the intermediate model by adding poster (`assigned_person`); and the improved model by adding interaction between poster and treatment.

```
mod_1 <- d[ , lm(log_upvote ~ group + subreddits)]
mod_2 <- d[ , lm(log_upvote ~ group + assigned_person + subreddits)]
mod_3 <- d[ , lm(log_upvote ~ group + assigned_person + group * assigned_person + subred
dits)]


stargazer(mod_1, mod_2, mod_3,
          type='text',
          se = list(
                      robust_se(mod_1),
                      robust_se(mod_2),
                      robust_se(mod_3)),
          column.labels = c('simplest model','intermediate model','improved model'),
          title = 'Username Gender Impact on Post Scores',
          omit = 'subreddits',
          add.lines = list(c("Subreddits Fixed Effects", "Yes", "Yes","Yes")),
          font.size = "tiny",
          column.sep.width = "1pt"
          )
```

```
##
## Username Gender Impact on Post Scores
## ================================================================================
==========================
##                                                              Dependent variable:
##                                        -----------------------------------------------
--------------------------
##                                                                   log_upvote
##                                             simplest model       intermediate model
improved model
##                                                  (1)                     (2)
(3)
## --------------------------------------------------------------------------------
--------------------------
## grouptreatment                                  0.038                   0.038
-0.398
##                                                (0.148)                 (0.148)
(0.465)
##
## assigned_personJamie                                                    4.149***
3.826***
##                                                                        (0.086)
(0.407)
##
## assigned_personJared                                                    -0.000
-0.368
##                                                                        (0.038)
(0.469)
##
## assigned_personJohn                                                     3.473***
3.245***
##                                                                        (0.903)
(0.994)
##
## assigned_personWeijia                                                    0.000
-0.137
##                                                                        (0.038)
(0.412)
##
## grouptreatment:assigned_personJamie
0.646
##
(0.571)
##
## grouptreatment:assigned_personJared
0.735
##
(0.553)
##
## grouptreatment:assigned_personJohn
0.456
##
```

```
(0.548)
##
## grouptreatment:assigned_personWeijia
0.274
##
(0.558)
##
## Constant                                    4.130***                  -0.019
0.199
##                                             (0.111)                  (0.079)
(0.371)
##
## ------------------------------------------------------------------------------------
-------------------------
## Subreddits Fixed Effects                      Yes                       Yes
Yes
## Observations                                  410                       410
410
## R2                                           0.672                     0.672
0.680
## Adjusted R2                                  0.430                     0.430
0.434
## Residual Std. Error                    1.098 (df = 235)          1.098 (df = 235)
1.094 (df = 231)
## F Statistic                         2.770*** (df = 174; 235) 2.770*** (df = 174; 23
5) 2.759*** (df = 178; 231)
## ====================================================================================
=========================
## Note:                                                                          *
p<0.1; **p<0.05; ***p<0.01
```

Breakdown by each poster

For an easy read on poster's impact on upvotes, we further broke down the improved model by each poster and the result is shown below. There is no statistically significant impact of poster on the treatment effect. And the coefficient is in line with Table 2. For example, John's posts in the treatment (female username) group received an 0.058% higher upvotes than posts in his control (gender neutral) group. This is the same effect as calculated from Table 2: -0.398 + 0.456 = 0.058.

```
mod_john <- d[assigned_person == 'John' , lm(log_upvote ~ group + subreddits )]
mod_jamie <- d[assigned_person == 'Jamie', lm(log_upvote ~ group  + subreddits)]
mod_jared <- d[assigned_person == 'Jared', lm(log_upvote ~ group  + subreddits)]
mod_weijia <- d[assigned_person == 'Weijia' , lm(log_upvote ~ group + subreddits )]
mod_fj <- d[assigned_person == 'FJ' , lm(log_upvote ~ group + subreddits)]

stargazer(mod_john, mod_jamie, mod_jared, mod_weijia, mod_fj,
          type='text',
          se = list(
                      robust_se(mod_john),
                      robust_se(mod_jamie),
                      robust_se(mod_jared),
                      robust_se(mod_weijia),
                      robust_se(mod_fj)),
          column.labels = c('John','Jamie','Jared','Weijia','FJ'),
          title = 'Username Gender Impact on Post Scores by Poster',
          omit = 'subreddits',
          add.lines = list(c("Subreddits Fixed Effects", "Yes", "Yes","Yes", "Yes","Yes"
)),
          font.size = "tiny",
          column.sep.width = "0.01pt"
          )
```

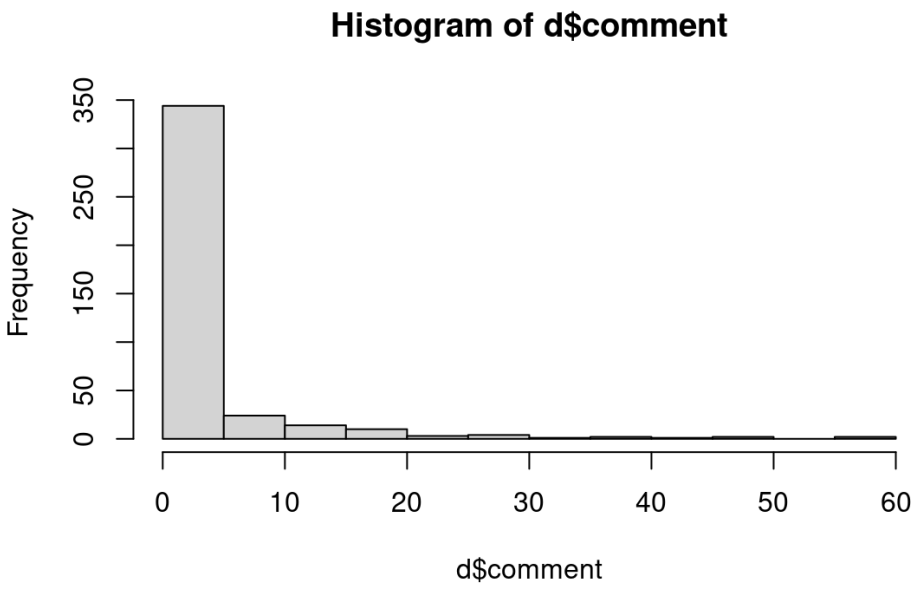
```
##
## Username Gender Impact on Post Scores by Poster
## ================================================================================================
================================================
##                                                                                 Dependent vari
able:
##                          ---------------------------------------------------------------------
-----------------------------------------------
##                                                                                        log_upvote
##                                    John                  Jamie                    Jared
Weijia                 FJ
##                                    (1)                    (2)                     (3)
(4)                   (5)
## ------------------------------------------------------------------------------------------------
-----------------------------------------------
## grouptreatment                    0.058                  0.248                   0.338
-0.124               -0.398
##                                  (0.290)                (0.331)                 (0.298)
(0.308)               (0.465)
##
## Constant                          -0.029                4.025***                4.055**
*                0.408*                0.199
##                                  (0.145)                (0.166)                 (0.160)
(0.247)               (0.269)
##
## ------------------------------------------------------------------------------------------------
-----------------------------------------------
## Subreddits Fixed Effects           Yes                    Yes                     Yes
Yes                   Yes
## Observations                       84                     68                      96
94                   68
## R2                                0.514                  0.811                   0.751
0.673                 0.562
## Adjusted R2                       0.349                  0.648                   0.537
0.366                 0.136
## Residual Std. Error         1.138 (df = 62)        0.912 (df = 36)        1.055 (df =
51)        1.025 (df = 48)       1.318 (df = 34)
## F Statistic              3.118*** (df = 21; 62) 4.986*** (df = 31; 36) 3.502*** (df =
44; 51) 2.194*** (df = 45; 48) 1.320 (df = 33; 34)
## ================================================================================================
================================================
## Note:
*p<0.1; **p<0.05; ***p<0.01
```

Regression on log_comments and comments+upvotes

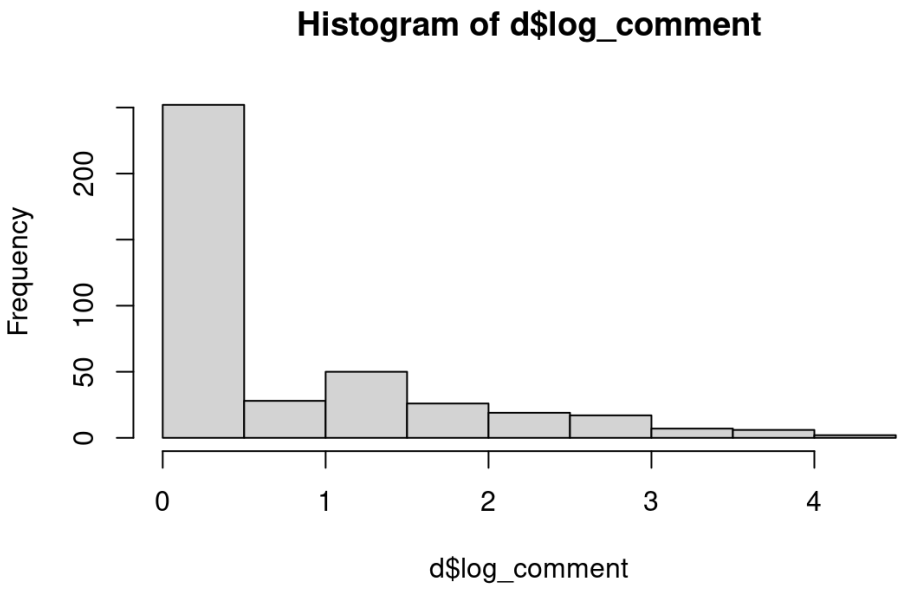
We also measured the number of comments and subset the model on each poster. The model that uses comments as the outcome variable was not informative, as there is no statistically significant impact from any of the treatment or covariate variables. For simplicity, we will keep the results with comments and the sum in the

code report.

```
hist(d$comment)
```




```
d[ , log_comment := ifelse(comment ==0, 0, log(comment))]
hist(d$log_comment)
```




```
d[ , sum_comm_upvote := upvote + comment]
d[ , log_sum_comm_upvote := ifelse(sum_comm_upvote ==0, 0, log(sum_comm_upvote))]
```

```
mod_4 <- d[ , lm(log_comment ~ group + subreddits)]
mod_5 <- d[ , lm(log_comment ~ group + assigned_person + subreddits)]
mod_6 <- d[ , lm(log_comment ~ group + assigned_person + group * assigned_person + subre
ddits)]


stargazer(mod_4, mod_5, mod_6,
          type='text',
          se = list(
                      robust_se(mod_4),
                      robust_se(mod_5),
                      robust_se(mod_6)),
          #column.labels = c('simplest model','intermediate model','improved model'),
          title = 'Username Gender Impact on Comments',
          omit = 'subreddits',
          add.lines = list(c("Subreddits Fixed Effects", "Yes", "Yes","Yes")),
          align = TRUE
          )
```

```
##
## Username Gender Impact on Comments
## ================================================================================================
==========================
##                                                                   Dependent variable:
##                                        ------------------------------------------------
--------------------------
##                                                                       log_comment
##                                                  (1)                        (2)
(3)
## ------------------------------------------------------------------------------------------------
--------------------------
## grouptreatment                                  0.006                      0.006
-0.263
##
##
## assigned_personJamie                                                       2.740
2.583
##
##
## assigned_personJared                                                       -0.000
-0.196
##
##
## assigned_personJohn                                                        2.454
2.262
##
##
## assigned_personWeijia                                                      -0.000
-0.102
##
##
## grouptreatment:assigned_personJamie
0.315
##
##
## grouptreatment:assigned_personJared
0.392
##
##
## grouptreatment:assigned_personJohn
0.386
##
##
## grouptreatment:assigned_personWeijia
0.203
##
##
## Constant                                        2.738                      -0.003
0.132
##
##
```

```
## ------------------------------------------------------------------------------------
--------------------------
## Subreddits Fixed Effects                          Yes                        Yes
Yes
## Observations                                      407                        407
407
## R2                                               0.606                      0.606
0.611
## Adjusted R2                                      0.310                      0.310
0.306
## Residual Std. Error                        0.835 (df = 232)          0.835 (df = 232)
0.837 (df = 228)
## F Statistic                              2.048*** (df = 174; 232) 2.048*** (df = 174; 23
2) 2.008*** (df = 178; 228)
## ====================================================================================
==========================
## Note:                                                                          *
p<0.1; **p<0.05; ***p<0.01
```

```
mod_7 <- d[ , lm(log_sum_comm_upvote ~ group + subreddits)]
mod_8 <- d[ , lm(log_sum_comm_upvote ~ group + assigned_person + subreddits)]
mod_9 <- d[ , lm(log_sum_comm_upvote ~ group + assigned_person + group * assigned_person
+ subreddits)]


stargazer(mod_7, mod_8, mod_9,
          type='text',
          se = list(
                      robust_se(mod_7),
                      robust_se(mod_8),
                      robust_se(mod_9)),
          #column.labels = c('simplest model','intermediate model','improved model'),
          title = 'Username Gender Impact on Comments & Upvotes',
          omit = 'subreddits',
          add.lines = list(c("Subreddits Fixed Effects", "Yes", "Yes","Yes"))
          )
```

```
##
## Username Gender Impact on Comments & Upvotes
## ================================================================================================
==========================
##                                                                 Dependent variable:
##                                     ---------------------------------------------------------
--------------------------
##                                                                 log_sum_comm_upvote
##                                              (1)                        (2)
(3)
## ------------------------------------------------------------------------------------------------
--------------------------
## grouptreatment                             0.003                      0.003
-0.555
##
##
## assigned_personJamie                                                  4.380
3.935
##
##
## assigned_personJared                                                  -0.000
-0.423
##
##
## assigned_personJohn                                                   3.784
3.483
##
##
## assigned_personWeijia                                                 0.347
0.145
##
##
## grouptreatment:assigned_personJamie
0.890
##
##
## grouptreatment:assigned_personJared
0.846
##
##
## grouptreatment:assigned_personJohn
0.602
##
##
## grouptreatment:assigned_personWeijia
0.402
##
##
## Constant                                   4.379                      -0.001
0.278
##
##
```

```
## ------------------------------------------------------------------------------------
-------------------------
## Subreddits Fixed Effects                          Yes                          Yes
Yes
## Observations                                      407                          407
407
## R2                                               0.682                        0.682
0.694
## Adjusted R2                                      0.444                        0.444
0.454
## Residual Std. Error                        1.081 (df = 232)          1.081 (df = 232)
1.071 (df = 228)
## F Statistic                              2.863*** (df = 174; 232) 2.863*** (df = 174; 23
2) 2.899*** (df = 178; 228)
## ====================================================================================
==========================
## Note:                                                                          *
p<0.1; **p<0.05; ***p<0.01
```