

Experiment Plan: Promoting Inclusion and Participation in an Online Feminism Community

Emily Hedlund, J. Nathan Matias, Marianne Reddan, Tyler Simko*

June 22, 2018

Contents

Introduction	1
Steps in this Experiment	2
Scope of this Pre-Analysis Plan	2
Outcomes	2
Main Outcome: Number of Comments	3
Main Outcome: Inclusion	3
Estimation Procedures and Assumptions	3
Covariates for Use in Regression Adjustment and Assignment	3
Banned	3
Variables To Report In Publication	3
Ingroup	3
Code for Estimation of Treatment Effect	4
Main Outcome: ATE on Number of Comments Made by a Newcomer	4
Main Outcome: ATE on a Newcomer's Reported Inclusion	4
Acknowledgments	4

Introduction

Online communities provide a unique space for discussion in the twenty-first century. Making all participants feel welcome is of the utmost importance to those who moderate these communities, as prior research has found that forums that “create and support the development of active and interactive communities” boost users’ well-being and foster a strong commitment to forum issues as well as a desire to address them in an offline setting. [1]. Pendry and Salvatore find that “this is especially true for forums covering more stigmatised issues where high levels of trust can develop, leading to ‘a virtuous circle of caring’ which can greatly enhance users’ feelings of social cohesion and emotional intelligence.” Accordingly, the issue of participant inclusion is especially important in communities such as r/feminism.

r/feminism is a community dedicated to “discussing and promoting awareness of issues related to equality for women.” Many users come to the subreddit to discuss current events as well as their personal experiences. However, users sometimes feel unwelcome in r/feminism for two reasons. First of all, they may be interested in participating in constructive discussions, but their opinions differ from the subreddit’s vocal majority. Accordingly, they feel uncomfortable voicing these opinions due to fear of an overwhelmingly negative reaction. On the other hand, some users

*ordered alphabetically

don't support the existence of the subreddit, and aren't interested in learning from its participants. In a survey of commenters, we discovered that many people report participating in conversations on r/feminism with the intent of annoying, harassing, and "trolling" people who participate in good faith.

Harassment may also discourage newcomers from participating even if they are not recipients of the harassment, since newcomers who witness harassing behavior may come to believe that the community overall is unwelcoming toward discussion of feminism. In this experiment, moderators are interested to learn if they can retain more participants by helping newcomers hold more accurate social norms (newcomers' beliefs about the community) about the community. To do so, moderators propose sending welcome messages to first-time commenters, messages that explain the values of the community, tell newcomers that any harassment is a minority view, and offer links to resources if they do experience some kind of harassment. By doing so, moderators hope to increase newcomers' sense of closeness to the community and increase the rate of participation in the community. To summarize their hypotheses:

Does sending encouraging welcome messages to first-time r/feminism commenters:

- increase their rate of participation on average?
- increase the closeness to the community that they report to on average?

Steps in this Experiment

The experiment has the following steps:

- We observe the accounts of newcomers to the community within several minutes of their comment through the following procedure
 - Use the reddit API to observe new comments to the community
 - If the account has not made any previous comments to the community within the past six months, we consider the account to be a first-time commenter
- We randomly assign newcomer accounts to one of two conditions
 - Control: No message
 - Intervention: Send the newcomer a private message
 - * Welcoming them to the community
 - * Providing them with information about managing online harassment
- We observe the activity of all enrolled participants over the following two weeks
- We send participants a survey that asks questions about the perceived closeness of their relationship the subreddit community.¹
- We conclude the study after receiving 1666 survey respondents

Scope of this Pre-Analysis Plan

This document describes the outcomes, adjustment variables, and estimation procedures.

Outcomes

The following variables will be used to estimate the effect promoting inclusion and participation in the subreddit. The unit of observation is a newcomer account enrolled in the study.

¹Example at <https://goo.gl/forms/zHi60GxaibKs4HEf2>

Main Outcome: Number of Comments

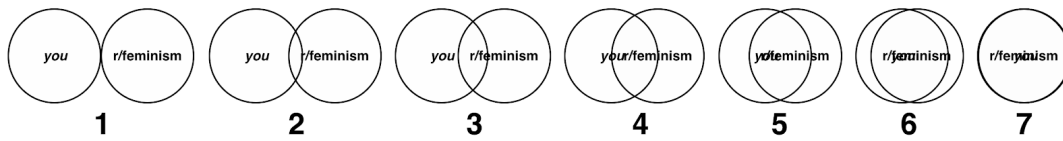
This behavioral measure is the number of comments posted in the subreddit in the two weeks following the assignment of a condition to an account.

`newcomers$num.comments`

Main Outcome: Inclusion

This survey measure is the level of inclusion in the subreddit community reported by a participant. The survey item is a version of the “inclusion of community in self” scale [2]:

In this image, imagine that the circle on the left represents you and the circle on the right represents r/feminism. Select which number, corresponding to the images above, best describes your relationship with r/feminism.



`newcomers$inclusion`

Estimation Procedures and Assumptions

We will use the procedures and assumptions below to calculate the average treatment effect estimate, standard errors / confidence intervals, and p-values.

Covariates for Use in Regression Adjustment and Assignment

We will use the following variables from accounts and account days as covariates to improve the precision of experiment estimates, predict the outcome and increase power, or to determine assignment.

Banned

This count variable represents the number of 24 hour periods that the account was forbidden from participating in the community by moderators or by the reddit platform operators during the 2 week period after enrollment, rounded up to the next integer value. We adjust for this variable because accounts that have been banned by a subreddit or by the platform are not able to comment in the subreddit.

`newcomers$banned.days`

Variables To Report In Publication

We plan to report the following variables in publications. They may also be used for exploratory analyses not covered by this pre-analysis plan.

Ingroup

An ordinal variable from a newcomer’s answer to an in-group question in the survey:

Did you identify as a feminist at the time you made your first comment in r/feminism?

In this measure, 1 represents those who self-report as feminists, -1 those who do not, and 0 those who are undecided.

```
newcomers$ingroup
```

Code for Estimation of Treatment Effect

These are the main analyses. The first main analysis tests the average treatment effect (ATE) on the incidence rate of comments made by a newcomer. The second main analysis tests the complier average treatment effect (CATE) on level of reported closeness to the community. The complier-average treatment effect is the effect within the group of newcomers in the treatment and control groups who complied with the survey[3].

We use negative binomial models to estimate incidence rates because the distribution of our count variables violate the assumptions of normality in linear regression, and because our questions are fundamentally about incidence rates [4]. The decision rule for all analyses will be $\alpha = 0.05$. The two main analyses will be adjusted for multiple comparisons using the Bonferroni method.

Main Outcome: ATE on Number of Comments Made by a Newcomer

We adjust this model for the number of days that an account was banned by moderators, since accounts that have been banned are not able to comment in the subreddit.

```
comments.model <- glm.nb(num.comments ~ TREAT + banned.days, data=newcomers)
```

Main Outcome: ATE on a Newcomer's Reported Inclusion

```
inclusion.model <- lm(inclusion ~ TREAT, data=subset(newcomers, complied==1))
```

Acknowledgments

We are grateful to the r/feminism moderators who facilitated this research, and providing feedback on this experiment plan.

References

- [1] Louise F. Pendry and Jessica Salvatore. Individual and social benefits of online discussion forums. *Computers in Human Behavior*, 50:211–220, 2015.
- [2] Debra Mashek, Lisa W. Cannaday, and June P. Tangney. Inclusion of community in self scale: A single-item pictorial measure of community connectedness. *Journal of Community Psychology*, 35(2):257–275, 2007.
- [3] Alan S. Gerber and Donald P. Green. *Field experiments: Design, analysis, and interpretation*. WW Norton, 2012.
- [4] J. Scott Long. Regression models for categorical and limited dependent variables. *Advanced Quantitative Techniques in the Social Sciences*, 7, 1997.