

# My title\*

Yichen Ji                      Xiaoxu Liu

February 11, 2024

In this report,.

## Table of contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Data</b>	<b>3</b>
<b>3</b>	<b>Results</b>	<b>3</b>
<b>4</b>	<b>Discussion</b>	<b>6</b>
4.1	First discussion point . . . . .	6
4.2	Second discussion point . . . . .	6
4.3	Third discussion point . . . . .	6
4.4	Weaknesses and next steps . . . . .	6
	<b>Appendix</b>	<b>7</b>
<b>A</b>	<b>References</b>	<b>8</b>

## 1 Introduction

The research has shown that roughly 4 in 10 Americans have experienced online harassment, including name-calling, physical threats, and sexual abuse (Pew Research Center 2021a). With the current widespread use of social media, the debation of regulating toxic content becomes even more pertinent. Cultivating civility within democratic discourse is strongly necessary,

---

\*Code and data are available at: [https://github.com/Selinayichenji/Toxic\\_speech\\_replication.git](https://github.com/Selinayichenji/Toxic_speech_replication.git).

such as emphasizing respect and social order when communicating online. However, the emergence of uncivil, intolerant content on social platforms raises concerns about its potential harm to public discourse and democracy. Our focus is on the crucial dilemma: should measures be implemented to moderate toxic content and uphold civility, or should we suggest allowing such speech on social media to remain unconstrained? The pursuit of addressing these challenges becomes significant in shaping the future of online democratic engagement. Although hate, harassment, and extremism significantly impact the country and online community negatively as toxic comments have saturated lots of common social media in the U.S., the application of strong and effective regulation over social media faces challenges due to multiple factors. Various factors, including technology companies, government, and NGOs, oppose the potential heavy regulation, which operates within a distinct legal framework in the U.S. Besides, Users, the ultimate recipients of online toxicity, are important in reporting objectionable content through flagging mechanisms. Therefore, we attempt to find the ideal platform standards from users' views to against toxic speech.

We aim to apply the initial analysis from the original paper "Toxic Speech and Limited Demand for Content Moderation on Social Media", which is from the American Political Science Review. The paper attempts to figure out the consistency of how users reply to toxic speech when using social media to find an appropriate solution to improve the harmony of online platforms, and it includes two pieces of research: 1). targeting social groups and 2). targeting partisans. In our reproduction, we use the original methods and the same dataset and shift the concentration from the level of toxic speech to the types of users while extending the research with more specific prevalent aspects of the groups of people's responses. Instead of talking about how people respond to toxic speech toward different labels of victims (LGBTQ, billionaire, and Christian) in study 1 of the original paper, we expand the study range to how people with different genders, education levels, and races react to toxic speech toward LGBTQ, billionaires, and so on. Beyond the changes, we hold all other perspectives to be the same as the original paper. For example, we set standards of toxic speech with different levels: incivility (disrespectful tone, lack of respect, rudeness, and inconsiderate language), intolerance (derogating, silencing, or undermining particular groups with specified labels), and violent threats(the tendency of physical harm).

We obtain the result of the reproduction work and find that women (gender), high school graduates (education), and Blacks / Hispanics (races) imply a tendency to be more sensitive to high-level toxic speech on social media; on the other side, men, high school graduates, and Blacks tend to be more sensitive to the billionaires target overall. The research result provides us with the information that 1). Different groups of people hold different attitudes and behaviors toward topic speech with different targets and 2). People usually want a more loose-controlled online environment and choose no heavy moderation toward heavy speech. These help us understand the user preference in social media, and stimulate the appropriate regulations of the speeches on online platforms. Generally, the reproduction talks about the summary of what we do based on the original paper and what we obtain, the data sources, detailed pictures and analysis through coding, and the discussion that concludes our results

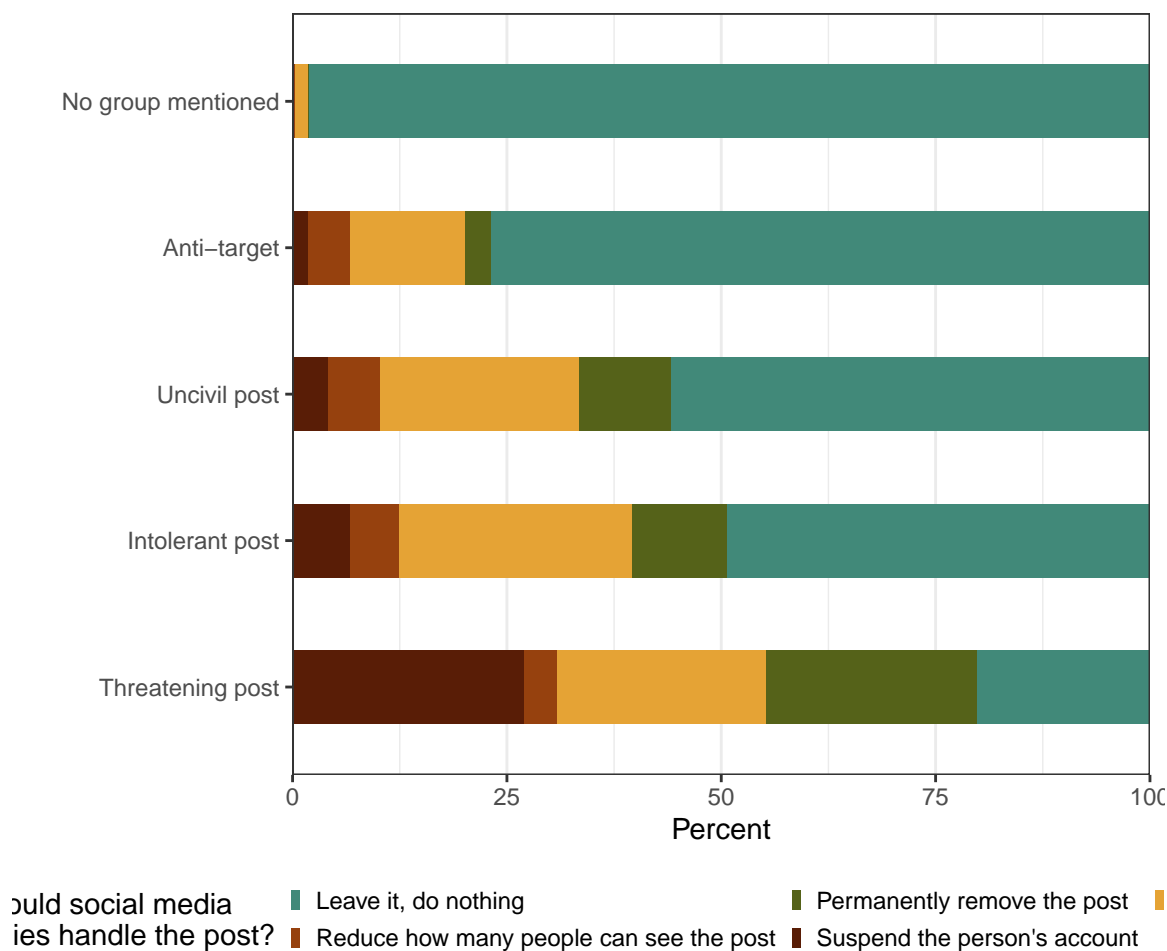
and lessons while discussing (potential) drawbacks and anticipated regulations/behaviors in the future. (can be improved)

## 2 Data

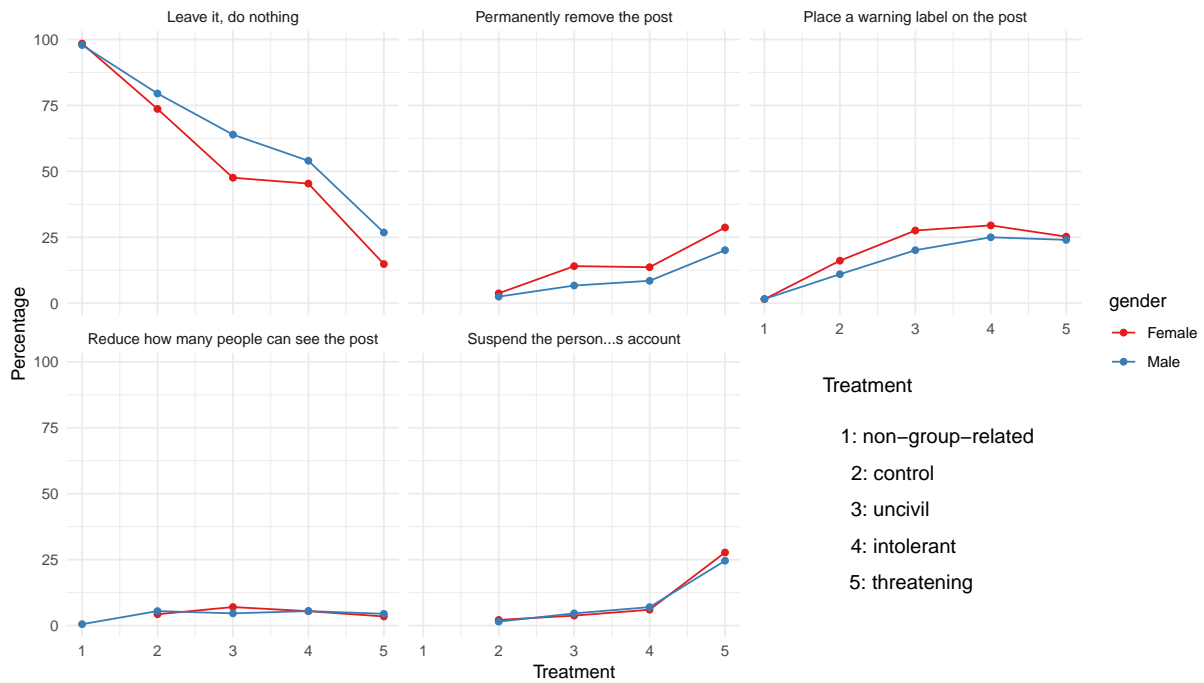
Our paper applies using the statistical programming language R (R Core Team 2020). Besides the programming tool, we also employ the following packages: readr (Wickham, Hester, and Bryan 2023), broom ( ), ggplot2 (Wickham H 2016), dplyr (Wickham et al. 2022), tidyverse (Wickham et al. 2019), MetBrewer ( ), and knitr (Xie 2014). (unsure about citation)

## 3 Results

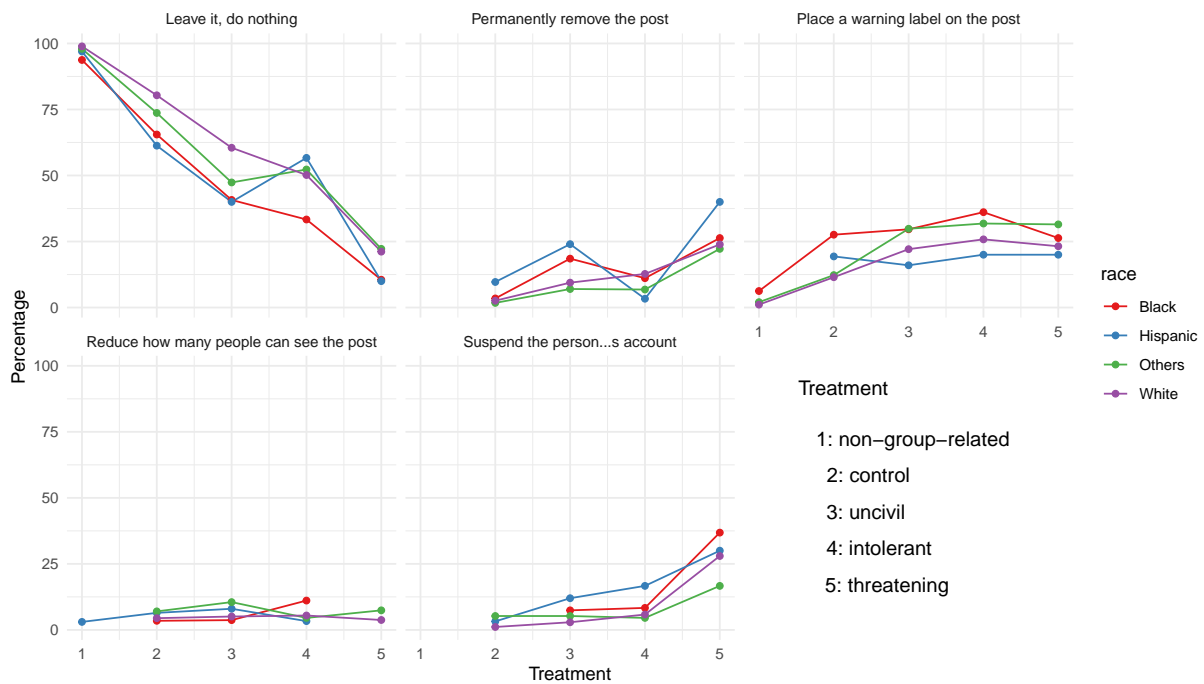
Our results are summarized in [?@tbl-modelresults](#).



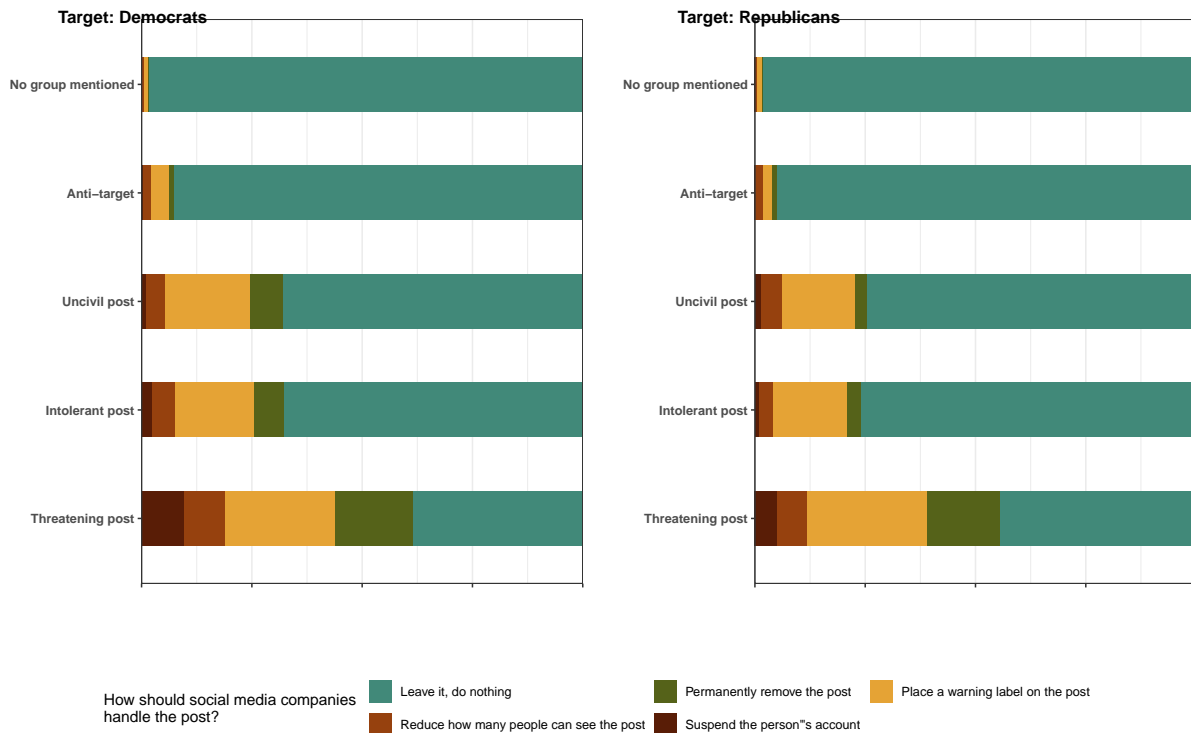
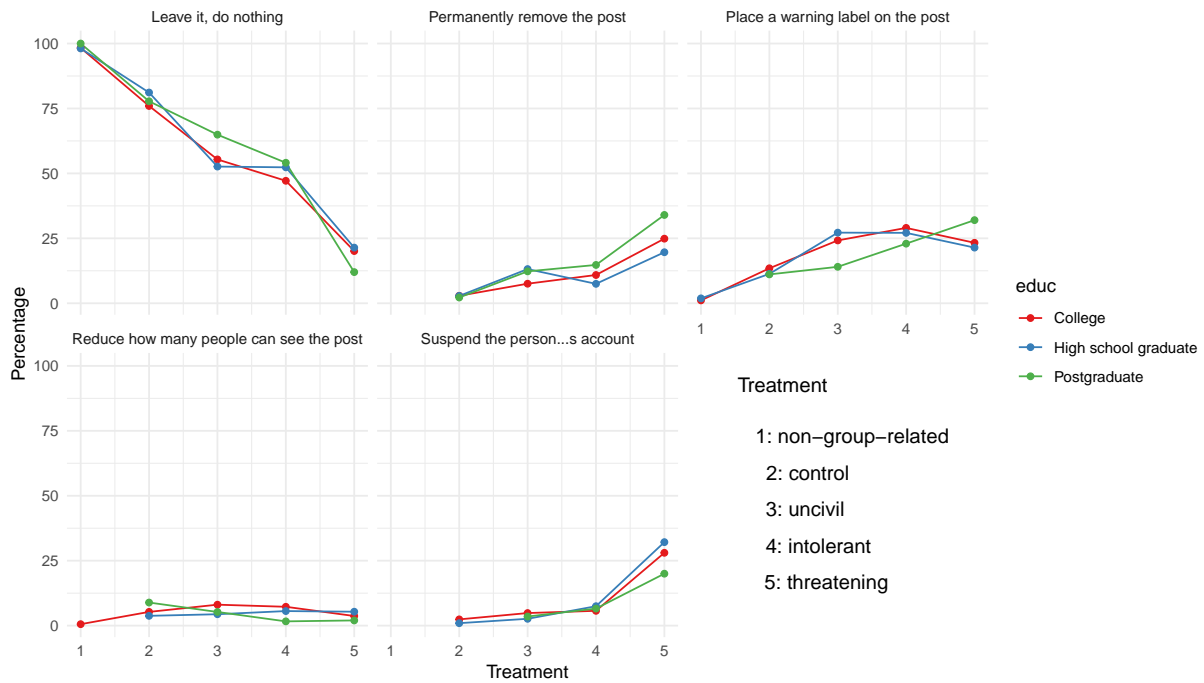
Percentage by Treatment and Gender in LGBTQ Topic



Percentage by Treatment and Race in LGBTQ Topic



## Percentage by Treatment and Education in LGBTQ Topic



## **4 Discussion**

### **4.1 First discussion point**

If my paper were 10 pages, then should be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

### **4.2 Second discussion point**

### **4.3 Third discussion point**

### **4.4 Weaknesses and next steps**

Weaknesses and next steps should also be included.

## Appendix

## A References