

Thesis Proposal

Citizen Perceptions and Preferences on Hate Speech Declaration and Hate Speech Regulations

Joshita Pai

Sona Jose

Introduction

While its effect is disputably unquantifiable, targeted anti-Semitic attacks on Jewish journalists and the stirring of ethnic conflicts in Myanmar are attributable to sustained bouts of offensive speech circulated online.

From triggering polarization to creating riot-like situations (Arun et al., 2016), Governments across the world regard hate speech as outside the bounds of free speech and insist digital platform providers to evaluate their community standards and to prevent the spread of online hate speech. The larger emphasis has been on the policies by respective governments and platform regulations. However, this research project is an attempt to understand citizen preferences in terms of online hate speech regulation and also their perceptions regarding what essentially constitutes hate speech.

There is no single universally accepted definition of hate speech whether it is online or offline. Nevertheless, the most commonly used notion of hate speech is that it is bias-motivated, hostile, and malicious language targeted at a person or a group because of their actual or perceived innate characteristics (Cohen-Almagor 2011; Faris et al. 2016). Existing definitions of hate speech as well as the legal definition of what constitutes hate speech could be extremely broad or fairly narrow. Recently, the digital platforms themselves have come up with the definition of online hate speech. For example, Facebook defines hate speech as a direct attack on people based on what they call ‘protected characteristics’ — race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity, and serious disease or disability (Facebook, 2018).

There is no consensus in these definitions about hate speech and what constitutes hate speech could be subjective at times depending on individual contexts. A 2016 study by Wike observed that Americans are more tolerant of offensive speech (Wike, 2016). The relevance of this project emerges from the fact that neither government nor platform actions appear to be informed by public preferences. Public preferences of what speech may be regarded offensive

differ from context to context and these factors in the same vein, affect individual's preferences for regulation.

Research Question

The research question pertains to exploring and comparing the attributes that shape the perception of hate speech declarations and the preference for hate speech regulations in US and Germany. The second objective is to explore whether these public perceptions of hate speech declaration determine/shape the preferences for hate speech regulation in both the countries.

Scope of the Study

The overall scope of the project is confined to exploring the overlapping individual characteristics between people's perceptions of hate speech and their corresponding preference for the need for and form of regulation in US and Germany.

Research Design

To explore the influence of individual characteristics on their opinion of hate speech and preference for regulation, Principal Component Analysis can be conducted. This is done so as to not prematurely remove variables which may contribute to explaining the variance in the model, and instead reduce the dimensionality of the variables. The variables may have different effects on what constitutes hate speech and on preferences for hate speech regulation.

Methodology

The data on exploring the preferences for hate speech declaration and regulation is derived from the opt-in web panel data in US and Germany. The data entails multiple waves of survey conducted between 2018-2019 of 1300 respondents in each country.

Findings under 'Hate speech declaration' and 'hate speech responsibility' items will be the subject matter of the project. Each item contains specific options for the respondents to choose from to sharpen the findings. The survey presented the respondents with a variety of posts which entail different issues from Muslim immigrants to posts targeting women to posts on political leaders. The preferences for hate speech vary within the range of these posts. Similarly, the preference for hate speech regulation and the nature of regulation preferred also varies.

The data provides demographic and other necessary information of the respondents including political orientation and extent of use of social media etc., which offers insight into their background, against which their responses on the two items will be analyzed.

Variables under individual characteristics influence public opinion on classifying a certain speech as hate speech and the same characteristics can also influence the preference for hate speech regulation.

Few of the possibly overlapping variables to include:

Variable	Hate Speech Declaration	Hate Speech Responsibility
Gender	Are women more likely to find posts objectifying or targeting women, more offensive than men?	Women are more likely to opt for regulations (perhaps platform regulations and not necessarily in the form of government regulations).
Political orientation	Respondents identifying more with the left are very likely to declare speech targeting any community to be offensive.	People identifying more with the right may opt for regulations and the ones leaning to the left are likely to be shy of choosing government restrictions on online speech.
Level of Education	People with higher education may be more conscious of or have better means to express themselves so they are more likely to discern speech from offensive/hate.	The effect of education on the one hand may contribute to skepticism towards government regulation on speech but people with higher education are less likely to deem that victims deal with hate speech by themselves.
Media use and political information	A higher exposure to the internet platforms, especially for consumption of political information is likely to increase the acceptance of purportedly offensive speech.	If the threshold for offensive speech is lowered by increased exposure, the individual may be less likely to opt for a regulation.

Social media use	Use of Facebook, Twitter and now Whatsapp is more likely to evoke strong perceptions of hate speech declarations, may be by lowering the threshold again of offensive speech, compared to use of LinkedIn.	Preference for regulations, especially platform regulations is likely to be affected by regular users, especially of Twitter, Facebook etc.
General attitude towards politics		Individuals indifferent to politics are likely to accept regulating hate speech.
Parties and candidates	Individuals with affiliation to political parties on the right of the spectrum are more likely to find posts on Muslim immigrants, women etc. to be more acceptable than offensive.	This is an interesting intersection. Increasing platform regulations are viewed by right wing political party affiliates as a move by tech giants to suppress the voices on the left. However, support for government regulations in form of censorships is welcomed by those associated with political parties on the right.