Write-up: Facebook Field study with Local Politicians

# Introduction

Online political hostility, such as hate speech on the internet, is a widespread phenomenon that poses a threat to democracy by undermining political engagement. This unfortunate tendency possibly entails users to withdraw from the political debate. While political science research has extensively studied aggressors and the development of politically hostile behavior, there has been relatively little focus on bystanders and their potential pro-social reactions, and whether it is possible to encourage pro-social online behavior.

# Study objective

This project is part of the Standing by: Pro-social Bystander Reactions to Online Political Hostility (STANDBY) project, aiming to explain pro-social bystander reactions and their consequences, funded by the European Research Council (ERC).

The main objective of this study was to investigate if Facebook users respond more pro-socially to hateful content on the Facebook posts of Danish local politicians from Århus and Copenhagen, when an intervention comment encoring users to: ‘Sig Fra, Anmeld, støt op’ has been implemented. Furthermore, to investigate any more general tendencies in the commenting on politicians Facebook posts for the two municipalities.

Several research questions were formulated:

1. Is there a difference in the Facebook comments when the intervention is applied compared to when it is not?
2. How do people respond to negative comments? do negative comments also entail negative subcomments and vice versa?
3. Is there a difference across the two municipalities? Is the commenting different in Århus compared to Copenhagen?
4. Are there general tendencies in the commenting on the politicians’ posts? Do people generally comment negatively, aggressively, encouraging etc.?

# Study Design

Local politicians from Århus and Copenhagen was invited to participate in the study. 11 politicians, of which 5 were from Copenhagen, took part in the intervention part of the study. Politicians were instructed to comment a designed, pre-written text to their post, alongside with a short, animated video. In short terms, the text encourages users to speak up against hateful content by creating awareness of the hatefulness of certain comments, report the comment or user, and showing your support to the receiver of the hateful comment. The full text is available in the Appendix. The politicians were instructed not to change anything else about the post or the content of the post itself. The treatment weeks were randomly assigned to the politicians, hence the intervention was not implemented on all posts of a participating politician. The intervention schedule ran for 8 weeks in total.

Furthermore, Facebook thread data was collected for 66 additional local politicians from Århus and Copenhagen to expand the data pool, making comparison between treatment and control posts possible.

# Data

Data from 1601 Facebook threads were collected, of which 51 were intervention post. After data cleaning, a total of 13.102 comments remains, of which 489 are comments from intervention posts. Posts are collected from both the private and public accounts of politicians. Date of posting runs from 01/05/23 to 02/07/23 totalling 9 weeks. The data was collected and formatted to json files from 02/07/23 to 25/08/23.

To extract emotions and attitudes as a way of quantifying the comments, 5 types of sentiment analysis were conducted for every comment:

1. VADER: giving a negative, positive, neutral and compound score
2. ROBERTA: giving a negative, positive, neutral and compound score
3. Danish BERT: giving 1 of 9 different labels like tillid/accept, foragt/modvilje, glæde/sindsro etc.
4. Attack: giving 1 of 2 different labels: linguistic attack/no linguistic attack
5. Rec-nition: giving 1 of 2 different labels: appreciative language/no appreciative language

# Results

1. General tendencies for commenting

Plots

1. Sentiment of subcomments

Plots

1. Differences in municipality

Plots

1. Effect of the intervention

Plots

Models

# Discussion and implications

Why the backfire effect

Inconclusive!!

no certainty because of random assignments of treatment, cant be certain that the backfire effect are exclusively due to the intervention

# Conclusion?

# Appendix

NÅR DU OPLEVER FJENDTLIGE KOMMENTARER: SIG FRA, ANMELD, STØT OP!

Vi ønsker alle platforme, hvor vi kan diskutere åbent. Men fjendtlige kommentarer på sociale medier forekommer ofte, rammer alle og kommer i mange former. At opleve fjendtlige kommentarer giver folk lyst til at trække sig fra diskussioner. Når du oplever fjendtlighed på sociale medier, kan du vise din modstand på mange måder. Ved at vise din modstand bliver du en del af løsningen! Du kan gøre en stor forskel ved at følge tre enkle råd:

Sig fra: Gør høfligt opmærksom på indlæggets fjendtlige karakter, eller prøv at vække folks medfølelse. Reager aldrig på fjendtligt indhold med mere fjendtlighed.

Anmeld: Brug platformens anmeld-funktion til at gøre platformens moderatorer opmærksomme på den fjendtlige kommentar. Støt op: Vis din støtte til modtageren af det fjendtlige indhold i en kommentar eller i en privat besked.

Når du viser din modstand mod fjendtlige kommentarer, hjælper du offeret med at bearbejde det og få diskussionen tilbage på sporet. Målet er ikke at få afsenderen af den fjendtlige kommentar til at ændre mening, men at vise at du tager afstand fra det. Det, du synes er fjendtligt, er der sandsynligvis også andre, der synes er fjendtligt. Når du viser din modstand mod fjendtlige kommentarer er du ikke alene. Du er ikke alene, når du viser din modstand mod fjendtlige kommentarer.