“How would you define hate speech?”

An Analysis on How Formal Education Shapes Individual Conceptions of Hate Speech on Social Media

MASTER THESIS BY JOHANNA MEHLER
j.mehler@students.hertie-school.org

CODE AND SUPPLEMENTARY MATERIAL
https://github.com/j-mehler/Hertie-Thesis-Mehler

BACKGROUND & RESEARCH QUESTION

“The definition of hate speech in a study or regulatory environment may be the most important part of the project’s design.” (Sellars, 2016, p. 32)

- **How inclusive** is the discourse on the trade-off between freedom of speech and the protection of vulnerable societal groups?
- Participation in political discourse is also a **question of education**.
- Literature suggests that **higher education** correlates with **more sophisticated and ideologically defined** political knowledge.

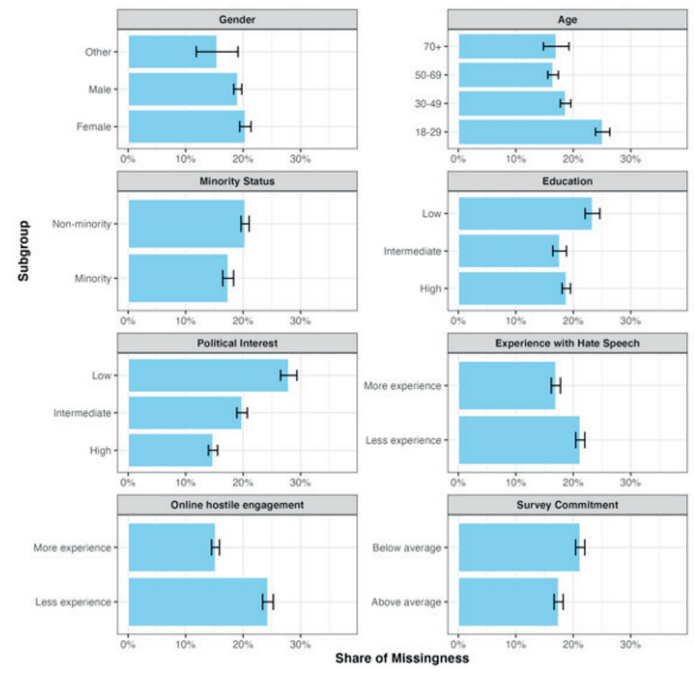
{How does formal education shape conceptions of hate speech?}

{H1} An academic level of education leads to a longer and better readable personal definition of hate speech than lower educational levels.

{H2} Academic respondents’ definitions of hate speech are more politically defined than those of non-academics.

Who is missing?

The 19.4% (n=19,000+) of respondents who did not answer the question had relatively **low education, low political interest, less prior experience with hate speech**, and were most frequently in the **18-29 age group.**



DATA & METHODS

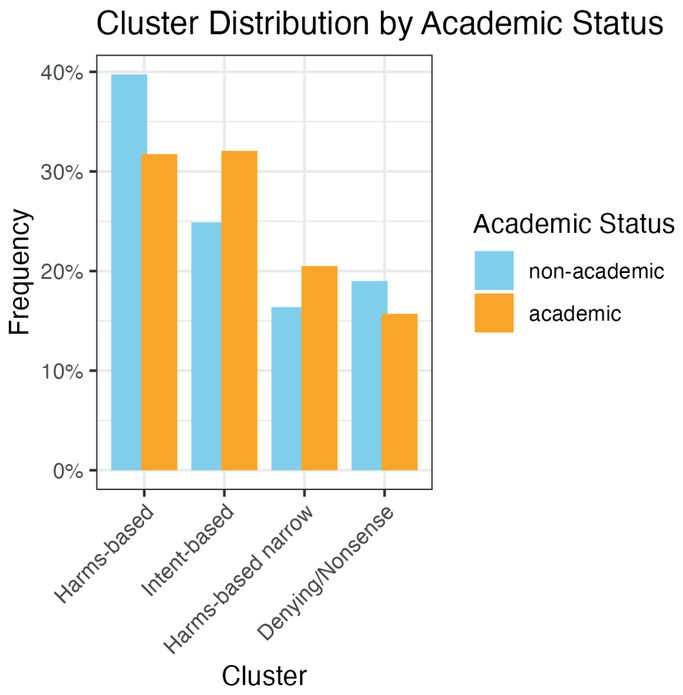
- Analytical Sample: **1,049 annotated hate speech definitions** from **19,000+ participants across eleven countries**
- **T-tests and chi-squared tests** compare indicators across academic status groups
- **Linear and multinomial regression models** explore the impact of academic status on each indicator, also in interaction with political interest, political ideology and hate speech experience.

2 Form Indicators

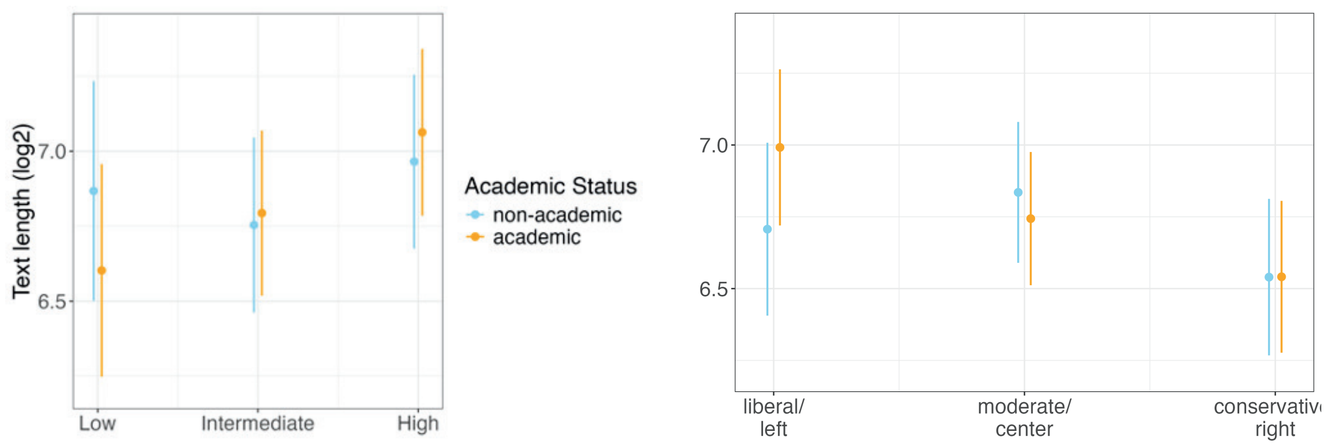
Text length	Readability	Content Type	Prediction of Political Orientation
R function nchar(), logarithmically transformed with a base of 2 due to a high frequency of lower values	Average of the five readability scores Flesch Kincaid Flesch (1948), Gunning Fog Index Gunning (1968), Coleman Liau Coleman and Liau (1975), SMOG (Harry and Laughlin, 1969), and the Automated Readability Index (Smith and Senter, 1967)	Partitioning Around Medoids (PAM) • Unsupervised clustering using the binary annotation dataset (indicating certain content, target, sender, scope features of the definitions)	Linear model with outcome variable “leftright” (1-11) to predict the political spectrum • Training data: binary annotation dataset • Calculation of the absolute error of predicted leftright value (1-11) = “Leftright prediction error”

2 Content Indicators

RESULTS



Findings show that the **definitions of academics** in the sample are generally **not longer or remarkably more readable**. However, they tend to define hate speech **more intent-based** or **narrowly harms-based** than non-academics. **Academics with higher political interest and left-leaning ideologies** provide **longer and more readable definitions** as those with lower interest and further to the right, while there is **no clear pattern for non-academics** in either context.



Marginal effects for text length by academic status in interaction with political interest (l.) and political ideology (r.)

CONCLUSION & POLICY RECOMMENDATIONS

- The results support prior findings that **most people possess an intuitive understanding** of hate speech
- Findings suggest that **political attitudes are crucial** in defining hate speech, but **even greater for academics** than non-academics.
- Policymakers should focus on measures that **address normative aspects next to knowledge** about hate speech through e.g. democratic citizenship education, promotion of internet awareness and counter-speech among students, or public awareness campaigns.

KEY REFERENCES

Heijden, E. v. d., & Verkuyten, M. (2020). Educational Attainment, Political Sophistication and Anti-Immigrant Attitudes. Journal of Social and Political Psychology, 8 (2), 600–616.

Kansok-Dusche, J., Ballaschk, C., Krause, N., Zeißig, A., Seemann-Herz, L., Wachs, S., & Bilz, L. (2023). A Systematic Review on Hate Speech among Children and Adolescents: Definitions, Prevalence, and Overlap with Related Phenomena. Trauma, Violence, & Abuse, 24 (4), 2598–2615.

Keen, E., Georgescu, M., & Gomes, R. (2020, May). Bookmarks (2020 Revised ed): A manual for combating hate speech online through human rights education. Council of Europe.

Munzert, S., Traunmüller, R., Barber´a, P., Guess, A., & Yang, J. (2023). Citizen Preferences for Online Hate Speech Regulation (Working Paper).

Sellars, A. (2016, December). Defining Hate Speech.