Does Interacting in Hate Groups Increase Hate?

Daniel Hickey[1] (hickeyda@oregonstate.edu), Matheus Schmitz[2], Daniel Fessler[3], Paul Smaldino[4], Goran Murić[5], Keith Burghardt[5]

[1]Oregon State University, [2]Qualcomm, [3]University of California, Los Angeles, [4]University of California, Merced, [5]University of Southern California Information Sciences Institute

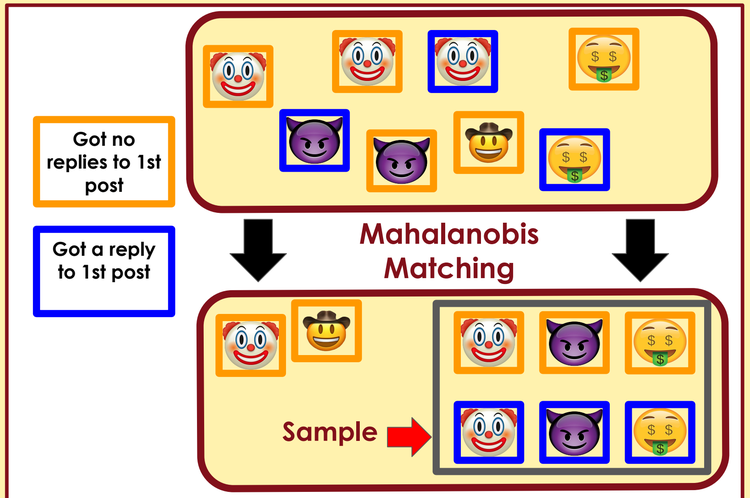
Work performed under REU Site program supported by NSF grant #2051101

Motivation:

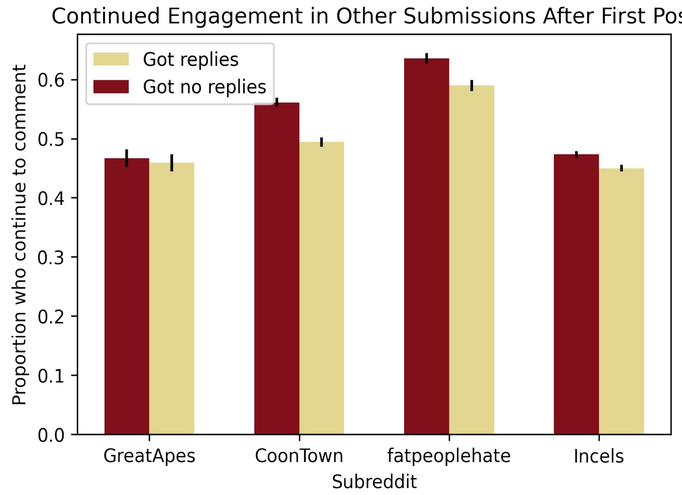
- What social interactions **cause** users of hateful subreddits to become more hateful?

Hate groups investigated

Subreddit	Type of Hate	Users Crawled
r/GreatApes	Anti-Black Racist	3140
r/CoonTown	Anti-Black Racist	9154
r/fatpeoplehate	Fat-shaming	10240
r/Incels	Misogynistic	20211



Schematic of quasi-experimental design for causal inference.

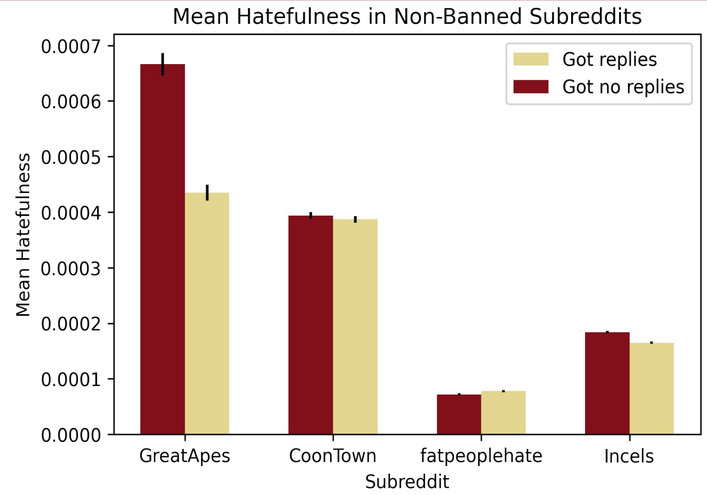


RQ1: Do users who get replies to their first post become more engaged in hate subreddits? - Users who got replies were less likely to continue posting.

Subreddit	Coefficient	P-value
r/GreatApes	0.29	**0.02**
r/CoonTown	0.17	**0.004**
r/fatpeoplehate	0.19	**$4 * 10^{-8}$**
r/Incels	0.027	0.4

Does sentiment of replies influence engagement?

Coefficients of logistic regression model. Response = engagement, predictor = sentiment of first reply.



RQ2: Do users who get replies to their first post become more antagonistic in non-hateful subreddits? - users who got replies were usually less antagonistic on average

Discussion/Conclusion:

- Getting reactions to **initial posts** in hateful subreddits does not seem to increase hate
- Matching is preliminary - will do more careful feature selection
- Sentiment analysis offers more clarity - some replies are encouraging, some discouraging.
- Will expand analysis to more hateful subreddits