9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Exposing Misinformation in Fringe Communities: Testing Source Exposure and Debunking of Anti-Ukrainian Misinformation among German Fringe Communities

Keywords: misinformation; intervention; source exposure; debunking; fringe communities

Extended Abstract

In the current digital age, misinformation is a threat for society, as it erodes trust in democracy, increases polarization, and taps into extremist movements ideology. As a result, both scholars and practitioners have been actively searching for ways to effectively stop misinformation in recent times [3, 5, 9]. Short of censorship, the standard counter-strategy against misinformation has become debunking [5], i.e. the exposure of factual and logical fallacies of misinformation.

The literature on debunking tends to be based on clinical experiments and focuses on general audience rather than members of fringe communities. Fringe communities are harder to reach with debunking approaches as they tend to avoid fact-checkers and traditional media [6]. Moreover, these communities have been shown to both be more susceptible to misinformation [2, 7] as well as have higher consumption levels [1] of misinformation than a general audience. As such, there is a need for more proactive counter-misinformation strategies among these high risk groups.

Another aspect that further complicates countering misinformation is the fact that misinformation spreaders have built an infrastructure of media, channels and bloggs to facilitate their efforts [8]. In addition to opportunistically exploiting new topics and claims and jumping between narratives to flood a platform with relatable content [4], misinformation sources also move from platform to platform which makes it impossible to shut them down.

Combining the fact that this misinformation infrastructure exists and these communities tend to fall outside of the scope of the existing misinformation treatment strategies, there is a need to investigate new ways with which this group can be reached and informed.

Exposing an outlet as a creator of misinformation within one of these fringe communities directly addresses both the aspects of this problem (i.e., combating the infrastructure and reaching the target audience). Consequently, such a strategy promises to be more effective than the existing techniques used by practitioners and researchers at this moment. Therefore, our study compares the use of source exposure directly within fringe communities as a technique to thwart consumption of misinformation as an alternative to existing approaches.

Candidate Facebook groups for the interventions were selected via an iterative process. First, we constructed an overview of fringe media in Germany ($N = 17$). Next, we identified fringe communities (i.e., public Facebook groups) who frequently consumed the fringe media content, by sharing and posting url's ($N = 157$). Finally, we identified groups who regularly consumed the two most popular German misinformation sources ($N = 156$), i.e. 'Reitschuster' and 'Deutschland Kurrier'. Only those communities who consumed content of these sources on a biweekly basis ($N = 35$) were included in our experiment.

In collaboration with the fact-checking organization VoxCheck, we then conduct an online field experiment with two interventions; by either 1) debunking anti-Ukrainian misinformation,

or 2) exposing the source of the message as a spreader of misinformation. Examples of these posts are shown in Figure 1.

Based on our experiment, we find that treated groups, who were given either debunking or source exposure posts, do have a statistically significant lower consumption two weeks after treatment, compared to the control group (i.e., no intervention). Looking at the long term effect of the interventions, we find that source exposure has a statistically significant long term effect of lowering misinformation consumption, whereas we do not observe this outcome for debunking, in light with previous work [5]. More surprisingly, however, is the fact that we saw a statistically significantly lower consumption of misinformation sources for the fringe communities for which the group admins rejected our treatment.

Despite the low number of observations, our results support our hypothesis that a more proactive counter-misinformation approach to reduce consumption of misinformation sources has an effect. Surprisingly we also find that notifying gatekeepers among fringe groups are also effective in lowering a groups consumption of misinformation, reducing the risk of backlash to such proactive approaches.

References

[1] A. Bor and M. B. Petersen. The psychology of online political hostility: A comprehensive, cross-national test of the mismatch hypothesis. *American Political Science Review*, 116(1):1–18, 2022.

[2] M. Bruder, P. Haffke, N. Neave, N. Nouripanah, and R. Imhoff. Measuring individual differences in generic beliefs in conspiracy theories across cultures: Conspiracy mentality questionnaire. *Frontiers in Psychology*, 4, 2013.

[3] M.-P. S. Chan, C. R. Jones, K. H. Jamieson, and D. Albarracín. Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, 28(11):1531–1546, 2017.

[4] C. Kriel and A. Pavliuc. Reverse engineering russian internet research agency tactics through network analysis. *Defence Strategic Communications*, 6:199–227, 2019.

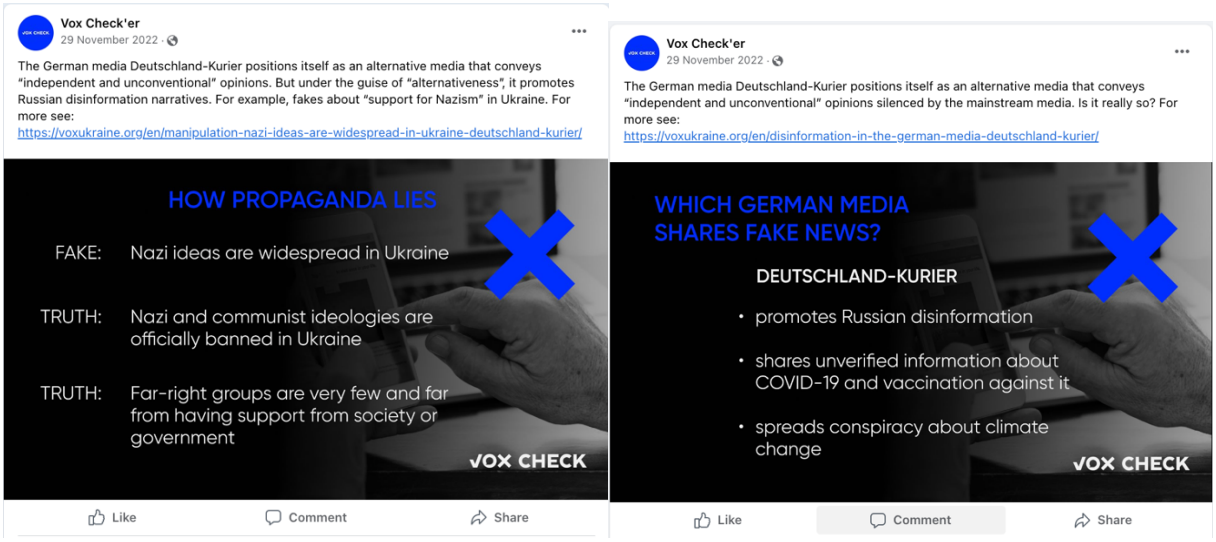
[5] D. P. McCabe and A. D. Smith. The effect of warnings on false memories in young and older adults. *Memory & Cognition*, 30(7):1065–1077, 2002.

[6] L. Nouri, N. Lorenzo-Dus, and A.-L. Watkin. Impacts of radical right groups' movements across social media platforms – a case study of changes to britain first's visual strategy in its removal from facebook to gab. *Studies in Conflict & Terrorism*, 0(0):1–27, 2021.

[7] M. B. Petersen, M. Osmundsen, and K. Arceneaux. The "Need for Chaos" and Motivations to Share Hostile Political Rumors. In *114th Annual Meeting of the American Political Science Association*, 2018.

[8] K. Starbird. Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1):230–239, May 2017.

[9] S. van der Linden, A. Leiserowitz, S. Rosenthal, and E. Maibach. Inoculating the public against misinformation about climate change. *Global Challenges*, 1(2):1600008, 2017.



(a) Debunking (b) Source Exposure

Figure 1: Examples of the intervention posts placed in target Facebook groups.