

Reducing Misinformation Online through Corrections

A Twitter Field Study

Shaye-Ann Hopkins, Duke's University's Center for Advanced Hindsight

Introduction

- Election-based misinformation continues to be amplified in the US, with social media users more likely to engage with 'fake vs. true' news stories¹.
- 'Inoculation theory' proposes several strategies (e.g. providing alternative narratives) to effectively correct misinformation^{2,3}.
- Our primary goals were to determine
 1. The effectiveness of corrections for reducing the spread of online misinformation.
 2. If effectiveness varies based on the type of corrections (i.e. *factual, reputational, or empathetic*) provided.
- There is limited field work on the utility of direct corrections on social media, so we aim to add to the body of experimental literature available on the topic.

Method

- Field experiment on Twitter with a between and within-subjects experimental design from Oct-Nov '21 (post-Cyber Ninjas audit).
- Tweets from misinformation 'super-spreaders' (N = 323) in Maricopa County, Arizona were tracked & tagged as misinformation or not for 14 weeks (pre-exp = 5 weeks, corrections = 5 weeks, post-exp = 4 weeks).
- Sample was randomly assigned to a treatment or control group. Our messenger (@AmericansFEI) responded to election misinformation with a correction (either **fact-checking, reputational, or empathetic**) or didn't (**control**).

Results

- **Corrections worked** for active spreaders relative to the control. They decreased relative election misinformation volumes by **1.80 ** tweets/week** during the experimental period and misinformation rates by 2.25% (p > 0.1).
- The **fact-checking** condition was most effective, with a similar reduction of **2.04 ** inaccurate tweets/week**. All feedback types had a significant impact.

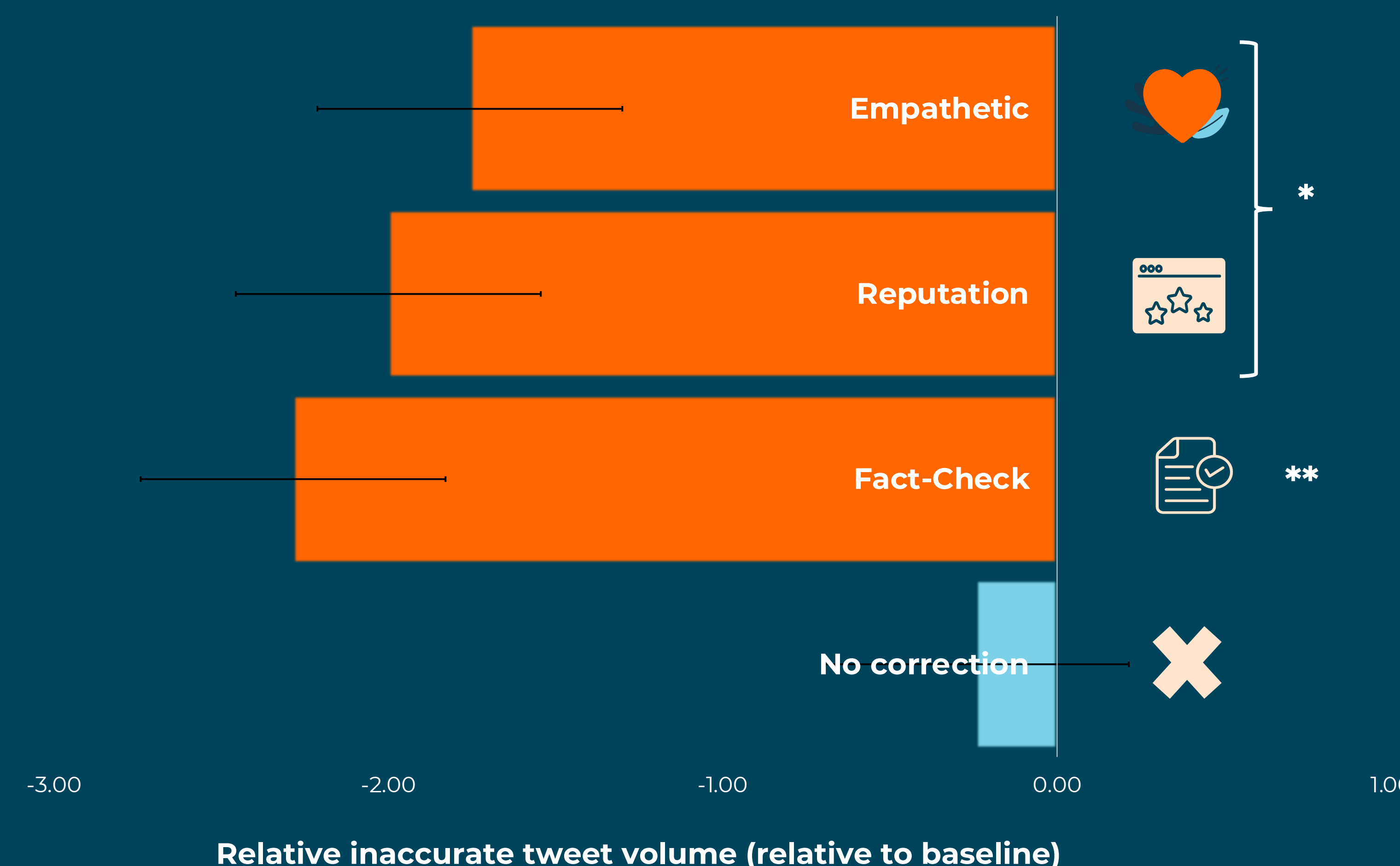
Conclusion

- Much bigger sample is needed to determine the impact.
- Timing (e.g. right before/after an election) is also essential.
- Consistent, direct corrections from trusted sources could be a useful option to reduce misinformation spread.
- At scale, lighter touch corrections (e.g. canned responses, bots) could be an option on social media.



Corrections reduce inaccurate tweet volume by **2 false tweets per week** for misinformation spreaders.

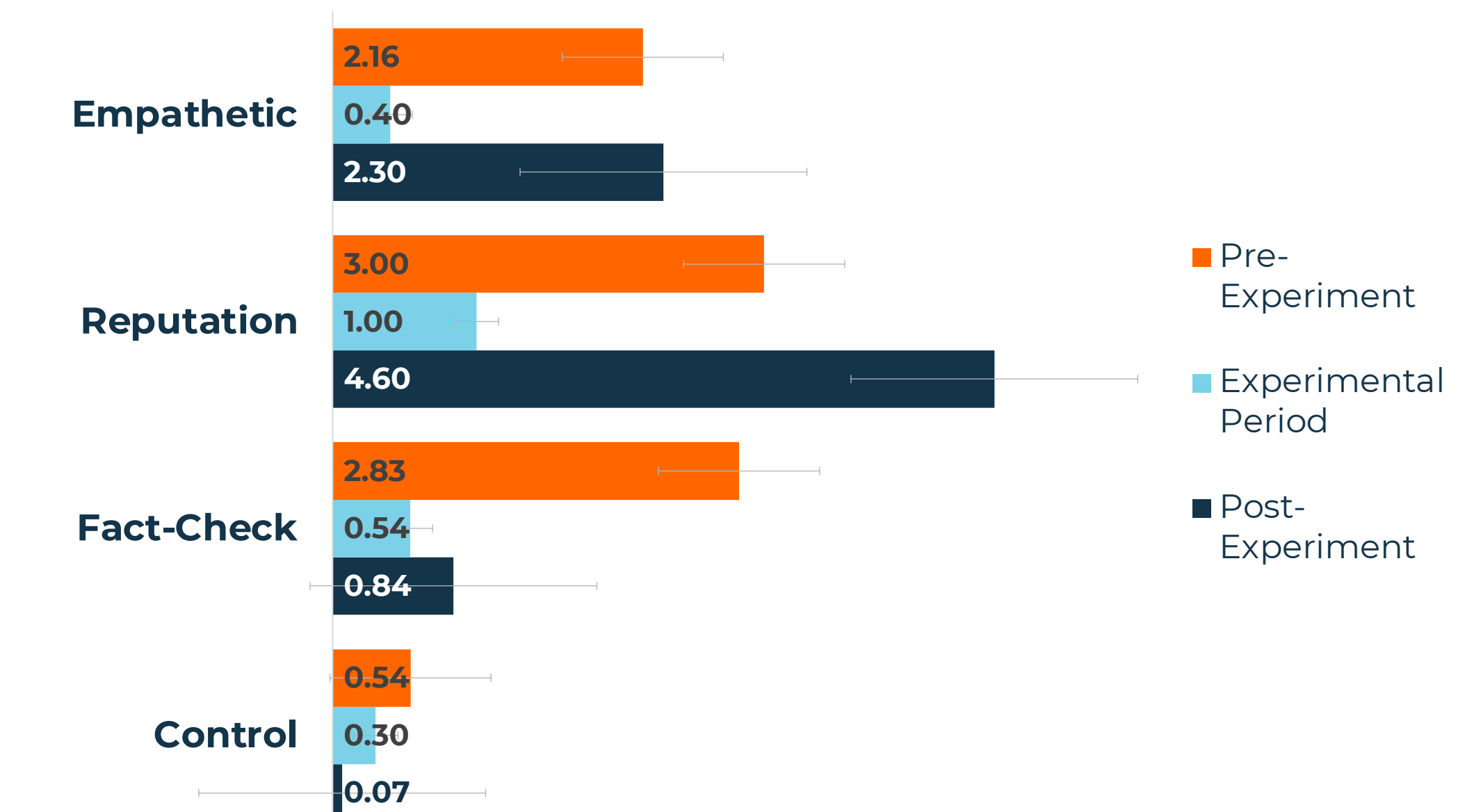
Relative Inaccurate Tweet Volume by Condition



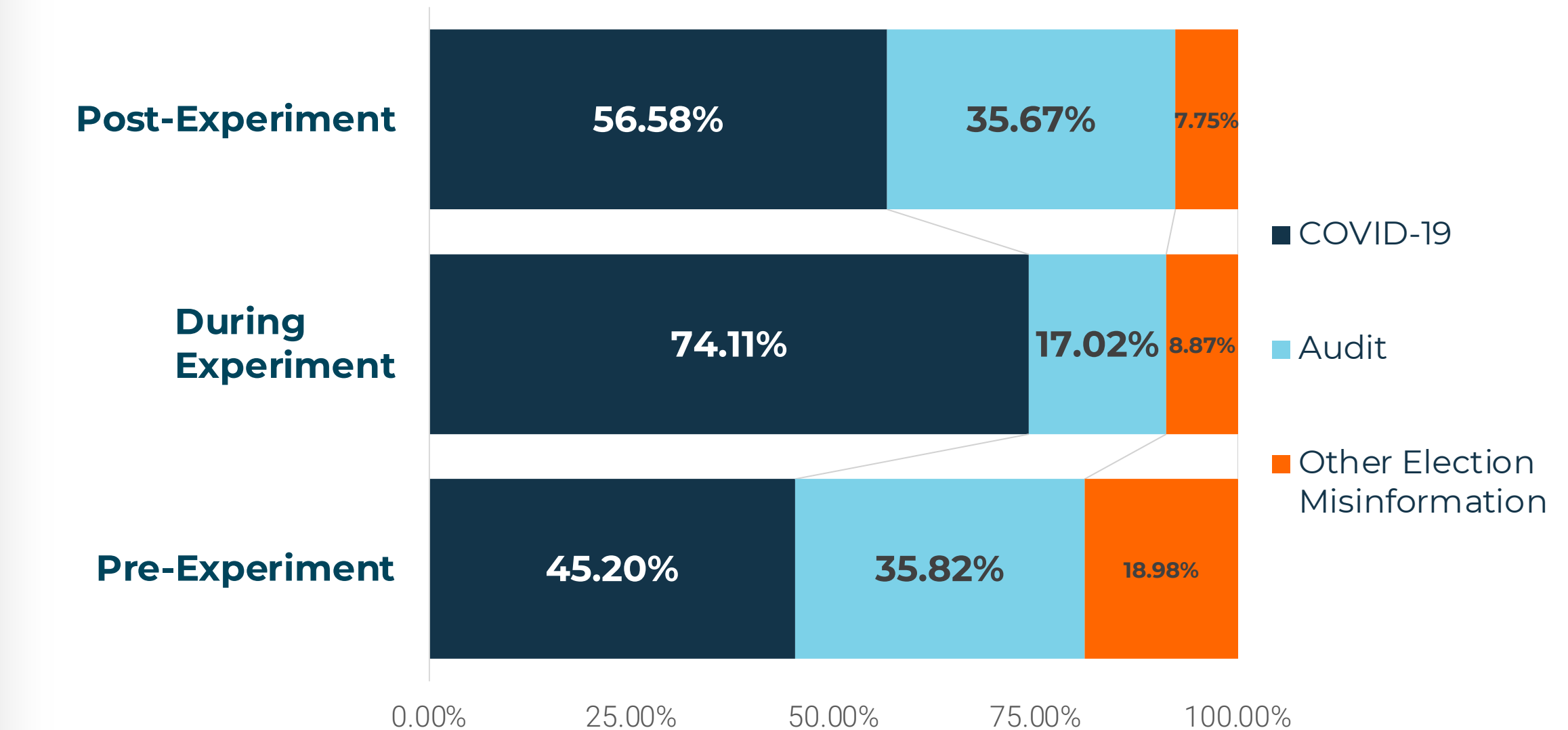
Additional Findings

- Misinformation was generally low (only ~1-2% of direct tweets), even for super-spreaders (~10%). This resulted in a much smaller active sample (N = 29).
- Throughout the experimental period, election-based misinformation reduced with feedback with COVID-19 misinformation increased and remained consistently high.

Misinformation Volume Before, During & After for the Active Sample



Misinformation Type Distribution for the Active Sample



Analysis Notes

- Panel regressions & ANOVAs used to determine impact & stat sig significance.
- Post-intervention period wasn't significant for the sample with an extremely small final sample.
- Control also had a much lower baseline (tweets were retroactively coded after the experiment).

Related literature

1. Silverman, C. (2016, November 16). *This analysis shows how viral fake election news stories outperformed Real News on facebook*. BuzzFeed News. Retrieved November 1, 2022, from <https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook#.qqE7PoA2QI>
2. Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological science in the public interest*, 13(3), 106-131.
3. Sangalang, A., Ophir, Y., & Cappella, J. N. (2019). The potential for narrative correctives to combat misinformation. *Journal of communication*, 69(3), 298-319.

Authors Shaye-Ann Hopkins ^o, Joseph Sherlock ^a
^aDuke University's Center for Advanced Hindsight in collaboration with the Brennan Center for Justice, Irrational Labs, & Precision Strategies.