Comparing the Behavioral Effects of Different Interactions with Sources of Misinformation

Daniele Bellutta[0000−0002−6131−9846] and Kathleen M. Carley[0000−0002−6356−0238]

Carnegie Mellon University, Pittsburgh, PA 15213, USA
{dbellutt,kathleen.carley}@cs.cmu.edu

Abstract. Research into the behavioral effects of misinformation has been conflicting, with recent studies concluding that exposure to misinformation is limited to small, predisposed parts of a population. By investigating the impact of interacting with sources of misinformation on Twitter users' subsequent tweets, we developed a methodology that did not rely on self-reported survey responses and allowed us to compare the effects of two types of exposure to misinformation: sharing links to unreliable websites and replying to tweets sent by untrustworthy accounts. Using tweets from before and after a user was exposed to misinformation, we found little evidence of significant changes in factors such as users' hatefulness or choice of hashtags. However, we found that users replying to unreliable accounts tended to change whom they mentioned in their tweets, hinting that misinformation may sometimes influence a reader's social connections. Though this adds to recent evidence favoring a more limited view of the direct impact of misinformation, its worst repercussions may lie with its indirect effect on citizens' faith in democracy.

Keywords: Social Cybersecurity · Misinformation · Fake News

1 Introduction

The field of social cybersecurity has seen a massive growth in research as people have worked to quantify and understand the misinformation that spreads across social media. Many studies have analyzed online influence operations that sought to influence elections in multiple countries [5, 8, 10], and the misinformation that has proliferated during the COVID-19 pandemic has been serious enough to warrant being called an "infodemic" [19]. However, research into the actual effects of misinformation on people's behaviors and beliefs has been much rarer and more conflicting. While some work has concluded that false stories can affect voter behavior [7, 20], other studies have instead supported the notion that misinformation only has limited effects for small portions of the population [5, 6].

In this work, we introduce a methodology for studying the behavioral effects of misinformation without relying on surveys or lab studies, unlike past research in this area [5–7, 20]. By investigating changes in users' tweets from before they interacted with a source of misinformation to afterwards, we found little evidence

that users significantly changed their behavior, with one exception. We additionally compared the effects of two different types of interactions with sources of misinformation: sharing links to untrustworthy websites on Twitter and replying to tweets from untrustworthy accounts. This comparison revealed the only significant behavioral change found by our analysis: after replying to untrustworthy accounts, users tended to change who they were mentioning in their tweets.

These results support recent studies [5, 6] in concluding that online text-based misinformation may have limited direct effects on its readers. However, other media, such as video, may be more effective at changing beliefs [11, 1]. In addition, though our results call into question the potency of misinformation for directly affecting the behavior of social media users, new research has highlighted the significance of the secondary effects of misinformation, such as the potential for belief in misinformation's effectiveness to erode confidence in democracy [14].

2 Related Work

Some research has provided evidence for the harmful effects of video-based misinformation. At least two studies have tied belief in misinformation to changes in voter behavior. Gunther *et al.* analyzed data from a survey administered following the 2016 U.S. election [7]. After homing in on people who had voted for Obama in 2012, they found that belief in certain false stories about Hillary Clinton was strongly linked to not voting for Clinton in 2016. Though the authors evaluated alternative explanations for this defection and ultimately concluded that the effect of misinformation was significant, their study has been criticized for being correlational and relying on self-reporting of political preference and exposure to misinformation [6]. A survey carried out before and after the 2017 German parliamentary elections similarly led to the conclusion that greater belief in fake news increased the likelihood that a voter who had planned to vote for the governing party would instead shift to voting for a right-wing populist party [20]. However, this study also suffers from the issue of relying on self-reported data.

Other studies have instead concluded that the effects of online misinformation are likely limited to small groups of people with certain political predispositions. Grinberg *et al.* estimated the feeds of more than sixteen thousand Twitter accounts only to find that just one percent of those users accounted for eighty percent of misinformation consumption and that increased interest in politics was strongly linked to greater exposure to misinformation [5].

Using a pre-election survey and the recorded web traffic of its respondents, Guess *et al.* similarly found that participants who read more news overall also faced more exposure to untrustworthy content [6]. Those who visited untrustworthy websites also tended to have lower scores on cognitive reflection tests, suggesting that they may have been more likely to believe the unreliable information they saw. However, this study additionally concluded that untrustworthy sites constituted only a small portion of participants' information consumption. In conjunction with voter records, the authors also ruled out the possibility that misinformation had especially large effects on voter choice and turnout.

To more directly evaluate the behavioral effects of exposure to influence operations, Bail *et al.* surveyed a bipartisan set of Twitter users in 2017 and tracked whether these users interacted with accounts belonging to the Russian Internet Research Agency [2]. Their results did not reveal a significant relationship between these interactions and changes in a person's feelings towards the opposing political party, the number of political accounts a person followed, or the proportion of a person's followers who were of the same political party. Changes in a person's ideology were not significantly linked to interaction with the Russian accounts, independent of whether the interaction was direct (such as liking or retweeting a troll's messages) or indirect (such as simply following a troll and being exposed to its messages).

A different ramification was instead evaluated by Nisbet *et al.* [14]. They make the case that ubiquitous discussion of misinformation contributes to a person's belief in the influence of misinformation on other people, which in turn erodes confidence in democracy. Their survey provided some support for the notion that increased attention to news and politics was linked to an increase in the perceived influence of misinformation on others. The survey also supported a significant link between increased perceived influence of misinformation and greater dissatisfaction with democracy. Furthermore, a survey by Lyons *et al.* found that presenting participants with the aforementioned results on the limited effects of misinformation did not significantly impact participants' perceptions of the levels of fake news consumption [9]. The harmful secondary effects of misinformation in democratic countries may therefore be difficult to combat.

3 Methods

3.1 Data

Misinformation Sources In order to identify Twitter users who had interacted with sources of misinformation, we started with a set of misinformation sites compiled from several publicly available lists, such ones from Media Bias/Fact Check [12]. From this list, we isolated sixty-six sites that had associated Twitter accounts. Table 1 lists all of the websites and accounts used for this study.

Tweet Collection Using the Twitter API, we collected the users who shared links to an untrustworthy site or replied to an untrustworthy account between 17 March 2021 and 1 April 2021. For each of these two sets of users, we continued randomly selecting an account and attempting to download fifty of the user's tweets from before the exposure to misinformation and another fifty tweets from after the exposure. This continued until 2,500 link sharers and 2,500 tweet repliers had been successfully collected, each with 100 tweets downloaded.

Human Identification Since the goal of this work was to evaluate the behavioral effects of interacting with misinformation, we needed to ensure that we analyzed data from human users of Twitter. This meant filtering out any

Site(s)	Twitter Account(s)	Site(s)	Twitter Account(s)
activistpost.com	@activistpost.com	kingworldnews.com	@kingworldnews
jewsnews.co.il	@jews_news	occupydemocrats.com	@occupydemocrats
disclose.tv	@disclosetv	theduran.com	@theduran_com
indiaarising.com	@indiaarising	freedomoutpost.com	@freedomoutpost
firebrandleft.com	@firebrandleft	filmsforaction.org	@filmsforaction
eyeopening.info	@eyeopeninginfo	prntly.com	@prntly
therussophile.org	@therussophile	thedailysheeple.com	@thedailysheeple
madworldnews.com	@madworldnews	abovetopsecret.com	@abovetopsecret
dailybuzzlive.com	@dailybuzztv, @dailybuzzlive	barenakedislam.com	@barenakedislam
libertyunyielding.com	@libertyunyielding	24nyt.dk	@24nyt.dk
voltairenet.org	@voltairenetorg	topinfopost.com	@topinfopost
redflagnews.com	@redflagnews	beforeitsnews.com	@beforeitsnews
breitbart.com	@breitbartnews	yournewswire.com	@yournewswire
thenewsnerd.com	@thenewsnerd	endingthefed.com	@endingthefed
worldtruth.tv	@worldtruthtv	wakingtimes.com	@wakingtimes
newcenturytimes.com	@newcenturytimes	empireherald.com	@empireherald
coasttocoastam.com	@coasttocoastam	linkbeef.com	@linkbeef
lewrockwell.com	@lewrockwell	libertywriters.com	@liberty_writers
dailystar.com.lb, dailystar.co.uk, thedailystar.net	@dailystarnews, @dailystarleb	drudgereport.com, drudgereport.com.co	@drudge_report, @drudgereport
thecommonsenseshow.com	@thecommonseshow	truthfeed.com, truthfeednews.com	@truthfeednews
worldtribune.com	@worldtribune	infowars.com	@infowars
chinadaily.com.cn, chinadaily.net	@chinadailyusa	realfarmacy.com	@realfarmacy
amtvmedia.com	@amtvmedia	sgtreport.com	@sgtreport
presstv.com	@presstv	pravda.com.ua, pravdareport.com, pravda.sk	@pravdask
hangthebankers.com	@hangthebankers	govtslaves.com	@govtslaves
shiftfrequency.com	@shiftfrequency	journal-neo.org	@journalneo
ifyouonlynews.com	@ifyou0nlynews	mrctv.org	@mrctv
rt.com	@rt_com	thedcgazette.com	@thedcgazette
cnn-trending.com	@cnn_trending		

Table 1: The untrustworthy sites and Twitter accounts used in this study.

data collected from automated accounts, which we identified using the Tier-1 BotHunter model [3]. BotHunter is a random forest regressor that computes the probability that the account authoring a tweet is actually a bot. This machine learning model was trained on Twitter data labeled based on forensic analyses of events that were widely reported to have high bot activity, including the 2017 attack against the Atlantic Council Digital Forensic Labs. BotHunter makes use of multiple user-level attributes (such as screen length name and account age), several tweet-level features (such as content and timing), and various network-level features (including the numbers of friends and followers).

After generating bot probability scores for each account using its one hundred tweets, maximum probability thresholds were applied to only keep tweets that were likely authored by genuine users rather than by bots. We chose the thresholds of 40% and 50%, providing two levels of strictness with which bot tweets were filtered out of the data. Applying the 40% limit to the bot scores of the collected link sharers only left 437 human users, whereas the 50% threshold left 726 human users. For the users who replied to suspicious accounts, the 40% threshold yielded 805 human users, and the 50% threshold identified 1219 non-bot accounts. Notably, this meant that the link sharers were much less likely to be humans than the repliers in our data. Filtering out users with high bot scores also had the advantage of focusing our data on common people, since celebrity accounts often have bot-like characteristics [18].

3.2 Behavioral Features

Lexical Features To identify possible changes in user behavior, we computed the means of various features using tweets from before a user's interaction with an untrustworthy source and then calculated the corresponding average feature values for tweets sent after the interaction. We used the NetMapper software [13] to examine the presence of emotional cues within each user's tweets. The software computes various lexical features derived from literature on psycholinguistics [15, 16], such as the number of terms that refer to specific identity groups. These features were generated for English, Spanish, French, Arabic, and German, which together accounted for 90% of the tweets with a specified language. The features examined were a tweet's number of abusive terms, average sentence length, average word length, number of expletives, number of absolutist terms, number of sentences, reading difficulty, total number of identity terms, number of political identity terms, number of racial identity terms, number of gender identity terms, number of familial identity terms, number of religious identity terms, number of other identity terms, number of positive emoji, number of negative emoji, number of positive emoticons, and number of negative emoticons.

Hate Speech Scoring A significant concern as to the effect of misinformation is whether it may spur people to be more hateful towards others. We used a machine learning model for detecting hate speech developed by Uyheng and Carley [17] in order to quantify the hatefulness of each user's tweets before and after interacting with a source of misinformation. This hate speech model is a random forest classifier trained on labeled hate speech data [4] to categorize a tweet into one of three mutually exclusive categories: hate speech, offensive speech, and regular speech. Using the lexical features generated by NetMapper, the model gives each tweet three confidence values summing to one, with each value signifying the likelihood that the tweet belongs to that category of speech. The average confidence values across each user's sets of "before" and "after" tweets were used as features in our statistical analysis.

Topics of Discussion Another way in which users might change their behavior is by choosing to discuss different topics or mention different people. We therefore decided to compute the cosine similarity of users' choices of hashtags, mentions, and links between the period before a user interacted with a source of misinformation and the period afterwards. To determine whether this similarity was out of the ordinary for that user, we also calculated a baseline measurement using only tweets from before the interaction with an untrustworthy source.

Given that we only had fifty tweets per user from before the interaction being studied, we first calculated the number of times each hashtag, account, or website was mentioned in a user's fifteen tweets before the interaction. We then carried out the same calculation for the fifteen tweets after the interaction. This allowed us to compute the cosine similarity between these two usage patterns. To get the baseline measure of similarity, we calculated the usage numbers for tweets sixteen

through thirty from before a user's interaction with an untrustworthy source and then did the calculation again for tweets thirty-one through forty-five. In this way, we were able to compute the cosine similarity between another two sets of fifteen tweets, with both sets coming from before the interaction of interest without overlapping with the previous sets of tweets. Hence, the end result was a quantification of a user's change in discussion topics after interacting with a source of misinformation, along with a baseline measure of the user's variety of discussion topics before the interaction.

3.3 Statistical Analysis

Significance of Behavioral Changes We conducted three statistical experiments for determining whether there were significant changes in user behavior after interacting with an untrustworthy source of information. For each feature (except the cosine similarities), we computed the average value across a user's tweets before the interaction and the average value after the interaction. We therefore had two matched samples, with each pair representing a user. Wilcoxon signed-rank tests were used to determine the significance of the changes between these "before" and "after" samples. Except for the cosine similarity features, this process was carried out twice for each bot score threshold: once using all fifty tweets before and fifty tweets after the interaction and another time using only twenty-five tweets before and twenty-five tweets after the interaction. This improved the robustness of our analysis, since short-lived effects would be more likely to appear when examining fewer tweets.

Comparison of Link Sharing to Tweet Replying To compare the effects of sharing unreliable links with those of replying to unreliable tweets, we first used the "before" and "after" samples to compute the numerical change that occurred in each feature. We then compared these feature differences for the link sharers to those for the tweet repliers by running Mann-Whitney U tests. As before, this process was repeated for each combination of bot score threshold and number of tweets from before and after the interaction.

False Discovery Rate Correction Given the large number of features and parameters, our statistical analysis involved running a total of 318 statistical tests. We therefore applied the Benjamini-Hochberg procedure to adjust our p-values and control for a false discovery rate (FDR) of 5%.

4 Results

Most of the tested features did not show significant changes, including the hatefulness and offensiveness of tweets. Table 2 shows the results of running the Wilcoxon signed-rank tests on paired "before" and "after" features from users who shared links to unreliable sources of information. None of the tested features

Bot Score	Feature	[T-50, T-1] v. [T+1, T+50] Change	P-Value	Adjusted	[T-25, T-1] v. [T+1, T+25] Change	P-Value	Adjusted
$< 40\%$	# of expletives	0.002	0.201	0.616	0.005	0.049*	0.532
$< 50\%$	# of expletives	0.002	0.173	0.616	0.005	0.030*	0.515

Table 2: The p-values and changes in mean for significant features found by running Wilcoxon signed-rank tests on tweets from users sharing links to untrustworthy sites. FDR-adjusted p-values are also presented.

Bot Score	Feature	[T-50, T-1] v. [T+1, T+50] Change	P-Value	Adjusted	[T-25, T-1] v. [T+1, T+25] Change	P-Value	Adjusted
$< 40\%$	# of absolutist terms	0.003	0.010*	0.486	0.004	0.030*	0.515
	# of political identities	0.004	0.038*	0.531	0.003	0.141	0.614
	# of positive emoticons	-0.001	0.016*	0.486	-0.001	0.105	0.609
	% of capital letters	-0.001	0.062	0.534	-0.002	0.017*	0.486
$< 50\%$	# of expletives	0.002	0.078	0.561	0.002	0.043*	0.531
	# of positive emoticons	-0.001	0.033*	0.515	-0.000	0.329	0.705
	# of hashtags	-0.013	0.016*	0.486	-0.010	0.034*	0.515

Table 3: The p-values and changes in mean for significant features found by running Wilcoxon signed-rank tests on tweets from users replying to tweets from untrustworthy accounts. FDR-adjusted p-values are also presented.

showed significant changes when using fifty tweets before the link sharing and fifty tweets afterwards. The number of expletives in a tweet did show a slight increase after the link sharing, but only before the FDR correction.

Table 3 shows the results of running the Wilcoxon signed-rank tests for users who replied to untrustworthy Twitter accounts. Accounts with bot scores less than 40% showed a slight increase in the number of absolutist terms per tweet. They also showed a slight increase in the number of references to political identity groups and a slight decrease in the number of positive emoticons, but only when looking at fifty tweets on either side of the interaction. When looking at only twenty-five tweets before and after the reply, the proportion of capital letters in a user's tweets showed a very slight decrease. For the bot score threshold of 50%, the number of hashtags per tweet showed a slight decrease. The number of expletives per tweet also increased slightly in the short term, and the number of positive emoticons slightly decreased in the long term. However, none of these changes maintained significance after FDR correction.

Table 4 lists the results of running Wilcoxon signed-rank tests on the cosine similarities of hashtags and mentions. The similarity of shared links increased slightly for tweet repliers with bot scores lower than 40%, and the similarity of hashtag choices decreased slightly for tweet repliers with bot scores below 50%, but neither of these changes was significant after FDR correction. For both bot score thresholds, the similarity of user mentions showed a significant decrease across the time of interaction with a source of misinformation. This change remained significant even after FDR correction, but only for the tweet repliers.

Bot Score	Feature	Site Sharers			Tweet Repliers		
		Change	P-Value	Adjusted	Change	P-Value	Adjusted
$< 40\%$	Similarity of mentions	-0.030	0.023*	0.514	-0.034	0.000***	0.022*
	Similarity of URLs	Not significant			0.047	0.040*	0.531
$< 50\%$	Similarity of mentions	-0.027	0.011*	0.486	-0.041	0.000***	0.000***
	Similarity of hashtags	Not significant			-0.033	0.024*	0.514

Table 4: The changes and p-values found by running Wilcoxon signed-rank tests on the cosine similarity features. FDR-adjusted p-values are also presented.

Bot Score	Feature	[T-50, T-1] v. [T+1, T+50]			[T-25, T-1] v. [T+1, T+25]		
		Difference	P-Value	Adjusted	Difference	P-Value	Adjusted
$< 40\%$	# of absolutist terms	0.004	0.013*	0.322	0.003	0.188	0.551
	% of capital letters	-0.002	0.042*	0.442	-0.004	0.011*	0.321
	# of political identities	0.007	0.011*	0.321	0.006	0.056	0.442
$< 50\%$	# of absolutist terms	0.002	0.026*	0.401	0.002	0.172	0.551
	# of hashtags	-0.019	0.010*	0.321	-0.017	0.026*	0.401

Table 5: The mean feature value for tweet repliers minus the mean value for link sharers. Mann-Whitney U tests were used to compute the p-values, which were also adjusted.

Table 5 shows the results of the Mann-Whitney U tests for comparing link sharers to tweet repliers. For the bot score threshold of 40%, the change in the proportion of capital letters was slightly lower for tweet repliers than for link sharers, and the tweet repliers showed less similarity between the links they shared before and after interacting with an untrustworthy source of information. In addition, the tweet repliers showed a greater change in the number of political identities mentioned within fifty tweets of the interaction (but not within twenty-five tweets) and in the number of absolutist terms used within twenty-five tweets of the interaction. When looking at the 50% bot score threshold, tweet repliers showed a smaller change in the number of hashtags and in the similarity of users mentioned between before and after replying to an unreliable account. Tweet repliers also showed a slightly larger change in the number of absolutist terms per tweet, but only when looking within fifty tweets of the reply.

5 Discussion

The results of our analysis show little evidence that interacting with sources of misinformation inspired changes in Twitter users' behavior. Our analysis shows that the users in our particular data set did not become more hateful after being exposed to misinformation. Users replying to tweets from untrustworthy accounts did, however, show a significant change in the other users they were mentioning. This hints at replies to online misinformation being more indicative of misinformation having actually affected a person's state of mind. Another implication of this result is that misinformation may be more effective at influencing whom people communicate with rather than changing their beliefs.

Our study has a few important limitations that temper these conclusions. In particular, we cannot guarantee that the users whose tweets we analyzed in this work had not already interacted with misinformation before our data collection and thereafter changed their behavior. Our sample of Twitter users was also biased towards more active users because of the requirement to collect fifty tweets before and after the user's interaction with a source of misinformation. Additionally, our analysis only covered a nonrandom set of sources of textual misinformation. It may be that other sites or media (such as video) could be more effective at influencing beliefs or stimulating hate. Finally, we did not analyze all possible types of interactions that are possible on Twitter.

6 Conclusion & Future Work

After examining tweets from before and after a user's interaction with a source of misinformation, we have not seen evidence of significant behavioral changes that may have been inspired by those interactions, with one exception. People replying to a tweet from an untrustworthy account did tend to change their user mentions more than before they authored such a reply. This may perhaps mean that online misinformation is more effective at changing people's social connections than at modifying their beliefs. Notably, the users in our data did not show increasing hatefulness or offensiveness after interacting with the sources of misinformation we studied, though other sources or data sets may lead to different results.

Though our results question the potency of the direct effects of misinformation, the secondary effects of misinformation may present much more serious consequences. With recent research already showing that popular discussion of misinformation may be eroding the public's confidence in electoral democracy [14], future research could therefore focus on evaluating the strength and significance of the indirect effects of misinformation, which could turn out to be its most harmful repercussions for democratic societies.

Acknowledgements

This work was supported by the Center for Informed Democracy and Social Cybersecurity with funding from the Knight Foundation and Cognizant. Additional support was provided by the Center for Computational Analysis of Social and Organizational Systems. The views and conclusions contained herein are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Knight Foundation, Cognizant, or the U.S. government. The authors would further like to thank Joshua Uyheng for the use of his hate speech detection model as well as Dawn Robertson and Netanomics for having maintained the list of misinformation sources used in this work.

References

1. Albarracin, D., Romer, D., Jones, C., Hall Jamieson, K., Jamieson, P.: Misleading claims about tobacco products in youtube videos: Experimental effects of mis-

information on unhealthy attitudes. Journal of Medical Internet Research **20**(6) (2018)
2. Bail, C.A., Guay, B., Maloney, E., Combs, A., Hillygus, D.S., Merhout, F., Freelon, D., Volfovsky, A.: Assessing the Russian Internet Research Agency's impact on the political attitudes and behaviors of American Twitter users in late 2017 (2019)
3. Beskow, D., Carley, K.: Bot-hunter: A tiered approach to detecting characterizing automated activity on twitter. In: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation (July 2018)
4. Davidson, T., Warmsley, D., Macy, M., Weber, I.: Automated hate speech detection and the problem of offensive language. In: Eleventh International AAAI Conference on Web and Social Media (2017)
5. Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., Lazer, D.: Fake news on Twitter during the 2016 U.S. presidential election **363**(6425), 374–378 (2019)
6. Guess, A.M., Nyhan, B., Reifler, J.: Exposure to untrustworthy websites in the 2016 us election. Nature Human Behaviour **4**, 472–480 (2020)
7. Gunther, R., Beck, P.A., Nisbet, E.C.: "fake news" and the defection of 2012 Obama voters in the 2016 presidential election. Electoral Studies **61** (2019)
8. King, C., Bellutta, D., Carley, K.M.: Lying about lying on social media: A case study of the 2019 Canadian elections. In: Social, Cultural, and Behavioral Modeling. Lecture Notes in Computer Science, vol. 12268, pp. 75–85 (2020)
9. Lyons, B.A., Merola, V., Reifler, J.: How bad is the fake news problem? the role of baseline information in public perceptions. In: The Psychology of Fake News: Accepting, sharing, and correcting misinformation, pp. 11–26 (2020)
10. Machado, C., Kira, B., Narayanan, V., Kollanyi, B., Howard, P.: A study of misinformation in WhatsApp groups with a focus on the Brazilian presidential elections. In: Companion Proceedings of The 2019 World Wide Web Conference. p. 1013–1019 (2019)
11. Maurer, M., Reinemann, C.: Learning versus knowing: Effects of misinformation in televised debates. Communication Research **33**(6), 489–506 (2006)
12. Media Bias/Fact Check: Media bias/fact check, https://mediabiasfactcheck.com
13. Netanomics: NetMapper (2021), https://netanomics.com/netmapper/
14. Nisbet, E.C., Mortenson, C., Li, Q.: The presumed influence of election misinformation on others reduces our own satisfaction with democracy. Harvard Kennedy School Misinformation Review **1**(7) (2021)
15. Pennebaker, J.W., Mehl, M.R., Niederhoffer, K.G.: Psychological aspects of natural language use: Our words, our selves. Annual Review of Psychology **54**(1), 547–577 (2003)
16. Tausczik, Y.R., Pennebaker, J.W.: The psychological meaning of words: LIWC and computerized text analysis methods. Journal of Language and Social Psychology **29**(1), 24–54 (2010)
17. Uyheng, J., Carley, K.M.: Bots and online hate during the COVID-19 pandemic: Case studies in the United States and the Philippines. Journal of Computational Social Science pp. 1–24 (2020)
18. Zafar Gilani, Reza Farahbakhsh, G.T.L.W., Crowcroft, J.: Of bots and humans (on Twitter). In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. pp. 349–354 (2017)
19. Zarocostas, J.: How to fight an infodemic. The Lancet **395**, 676 (2020)
20. Zimmermann, F., Kohring, M.: Mistrust, disinforming news, and vote choice: A panel survey on the origins and consequences of believing disinformation in the 2017 german parliamentary election. Political Communication **37**(2), 215–237 (2020)