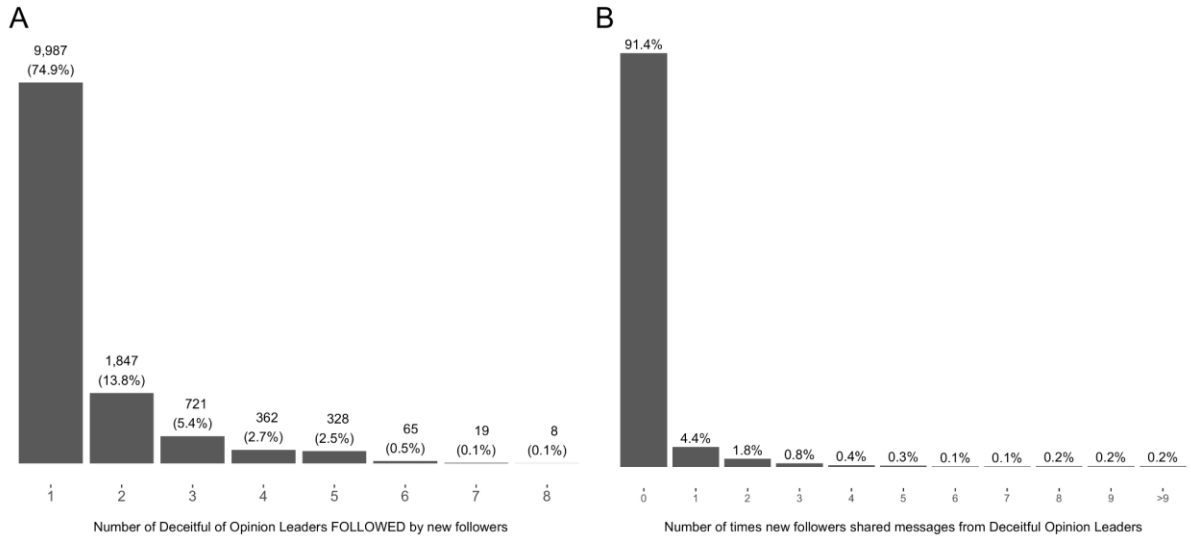
Appendix A. Number of DOLs followed and number of DOL tweets shared.

To provide some further descriptives on the sample of followers we study, in Figure A.1 we show the number of DOLs followed by all new followers (N = 13,337) and the number of times these followers shared messages of DOLs. Figure A.1.A displays how many DOLs the new followers started to follow during data collection. It indicates that 74,9% followed one DOL, and 25,1% followed two or more DOLs. In other words, during a period of 2 months, almost one-third of new followers started to follow more than one DOL. This finding supports existing literature, as it indicates that DOLs and their followers form homogeneous social networks online (Barbéra 2015; Shu et al., 2017).

Figure A.1.B displays the percentage of followers that shared messages from a DOL. The Figure shows that the majority (91,4%) of users in this dataset shared 0 tweets of DOLs. The other 8,6% that did share messages of a DOL, mostly only shared 1 message (4,4%). These findings align with other work studying the dissemination of deceitful content. For example, Guess et al. (2019) found that 8,5% of the respondents in their study shared fake news articles with their Facebook friends. This finding illustrates that, in line with previous literature (Guess et al., 2019; Vosoughi et al., 2018), only a small proportion of SM users contribute to disseminating misleading content online.

Figure A.1. Number of DOLs followed by new followers (A), as well as the number of times they shared a message from a DOL (B).



Appendix B. Codebook for Types of Disinformation.

Fake news

- Has a journalistic format, but is low in facticity (Egelhofer & Lecheler, 2019)
- Everything that looks like a reel news article, but comes from sources that aren't (always) factual. Such as the DOLs themselves.
- Open the links in tweets to see the articles.
- When a tweet is fake news, it cannot be disinformation as well.

Disinformation

- False information that is purposely spread to deceive people, seeking to amplify social divisions and distrust (Bennett & Livingston, 2018; McKay & Tenove 2020)
- Similar to fake news, but does not always have a link to an article with a journalistic format.
- Sometimes an article of a legitimate news source is linked but spun or interpreted in a way that makes the information not factual.

Conspiracy

- Efforts to explain events, practices, or secret plots that consist of two or more powerful actors, acting in secret for their benefit and working towards a malevolent or unlawful goal against the common good (Douglas et al., 2019; Sunstein & Vermeule, 2009).
- Everything that creates an us against them feeling, or hints at the elite being corrupt or misleading, etc.

Rumor

- Circulating information whose veracity status is yet to be verified at the time of spreading (DiFonzo & Bordia, 2007; Friggeri et al., 2014)
- Whenever tweets contain phrases such as: if they will do X then …, if X is true then …, they are saying that X will happen …, etc.

Appendix C. Codebook for Political, Uncivil, and Affective Polarization Content.

Political variable

- Binary variable (0 = not politics | 1 = politics
- Use a broad understanding of what's political: not only when political parties/institutions are mentioned, but also when topics that may have political implications, be picked up by media/politicians, eventually be part of a policy change, etc.
- If the text field is empty, add NAs to all coding columns/variables
- If a tweet is from a political leader but doesn't say anything about the political party or politics in general, code it as 0. For instance, if Jesse Klaver tweets what he had for breakfast this morning. If the party is mentioned, code it as political. So, for instance, if Jesse Klaver tweets: *This morning I had a lot of veggies for breakfast #GroenLinks*
- Code foreign languages via translate

Uncivil language

- Binary variable (0 = no | 1 = yes)
- Insulting, harassing, cursing words, racist, misogynist, very dismissive towards others, against a (minority) group: immigrants, LGBTIQ community, women
- "Vote them out" → not uncivil, but is negative
- Uncivil → Think about if you are Twitter or Facebook would you ban this tweet?
- Opinions that are factual/neutral aren't uncivil for example: *The U.S. Capitol Was Attacked by a Nation of Islam Follow and "GOP Leader" Kevin McCarthy Hasn't Said a Word About It. It's Time for Him to Resign. https://t.co/wSCWUcAsj9*

Affectively polarized

- It has to show dislike towards the opposing group, it has to be negative feelings towards the outgroup (affectively polarized)
- Check the nicknames for politicians, ideologic groups, and newspapers GeenStijl uses for political groups
- Needs to mention a political party or person or group or left/right (ideology)
- When tweets are about politicians in general, but it's not clearly directed at a political person, a political group etc, we code it as NA. For example: *Aan @Tjeenk Willink wat zijn wij Nederlanders zielig om te denken aan Rutte.wat nu als ze hem van zijn fiets rijden in Den Haag en als gevolg hiervan overlijd hij ,heeft Nederland een fucking probleem????gewoon zielig.*
- When a tweet is directed at a political person and uncivil we code it as AP for example: *@hugodejonge @GGD_RR Stumpert, je houdt geen 1.5m afstand. Leugenaar.*
- When groups are posed against each other and politicians in general or by name, or political parties are mentioned

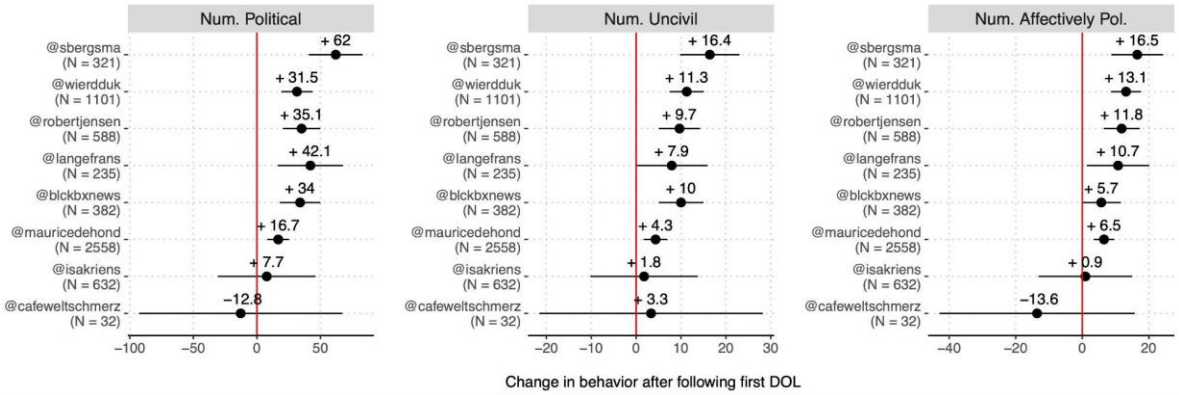
Misleading content

- Binary variable (but one tweet can be multiple types of misleading content) (0 = no | 1 = yes)
- We do not code videos, therefore we use NA when a tweet links to a video. Or we code it based on the text.
- If a tweet has been deleted, we code it as NA

Appendix D. Models with DOL-fixed effects and heterogenous effects by DOL.

In this Appendix, we provide further details about the change in behavior of new DOL followers after following the first DOL in our sample. In particular, we are interested in disentangling whether the effects observed in the paper are mainly driven by following a particular DOL, or whether these are patterns consistent among the new followers of all the eight DOLs in our sample.

Figure D.1. Coefficients (+95% confidence intervals) from linear models estimating a change in behavior after following the first Deceitful Opinion Leader (DOL). Same model specification as the models reported in Figure 3, but with the inclusion of DOL-fixed effects.



In Figure D.1 we report the results for models similar to those in Figure 3, but with the difference that in these new models we include DOL-fixed effects. In the figure we report the results for the fixed effects, which indicate the difference between the messages sent in the 30 days after (vs. before) following each particular DOL.

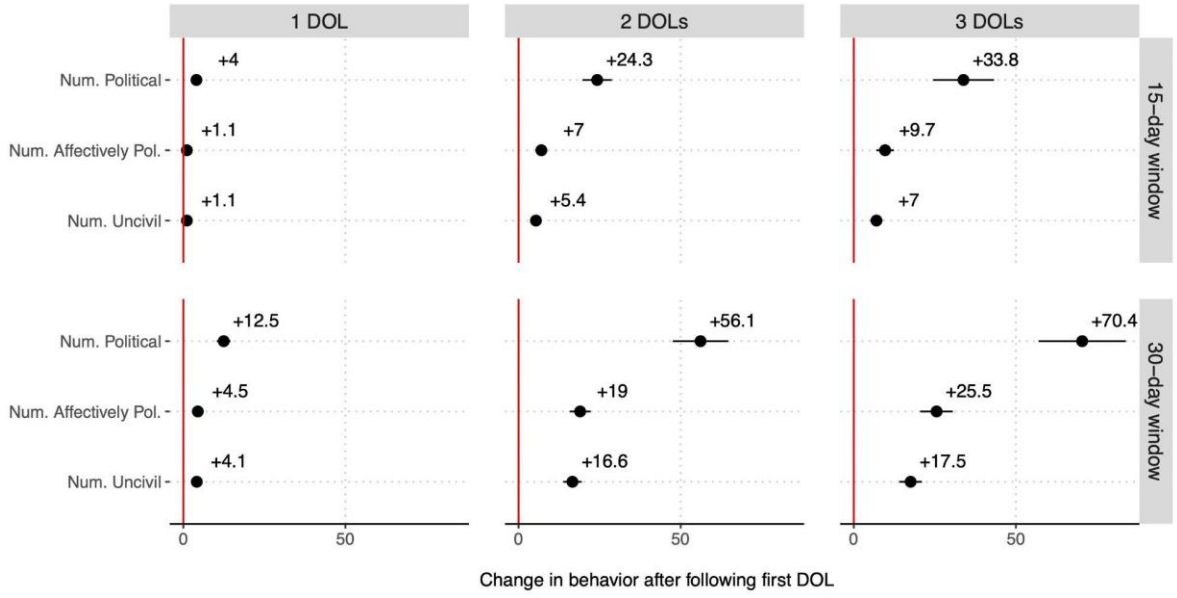
We observe positive, significant, and substantive effects for the users in our sample who decided to follow most of these DOLs, with two minor exceptions: we find positive but not significant results for the 632 users who started following @isakriens, and we find clear null results for the 32 users in our sample who started following @cafeweltschmerz. The followers of the remaining 6 DOLs clearly increased the volume of tweets in general during the 30 days after they started following them, as well as the number of political, uncivil, and affectively polarizing tweets. Although we observe some mild variations (for example, the 321 that started following @sbergsma are the ones who changed their behavior more substantially), overall we find very similar results. In sum, these findings indicate that the results reported in the paper are not driven by those who started following a particular DOL but are a reflection of a general pattern that happened across those who followed the different DOLs in our sample.

Appendix E. Analysis after excluding outliers.

In this Appendix, we replicate Figures 3 and 4 of the paper, where we explore changes in behavior among new DOL followers, but in this case, we assess the robustness of the original findings by excluding potential outliers in the data.

In Figure E.1, which replicates Figure 3, we used Cooks Distance (Cook, 1977) to remove any potential outlier from the regression models. For each of the models, we first calculated the same regression reported in Figure 3, then we calculated the Cooks Distance for each of the regression observations, and we finally estimated a new regression model after excluding the observations with a Cooks Distance that was four times larger than the average distance. In Figure E.1 we observe the size of the effects to be smaller once we remove these potential outliers. For example, when using a 15-day window to explore changes in behavior after following the first DOL (top-left panel), we find that users tweeted 4 more political tweets, 1.1 more uncivil messages, and 1.1 more affectively polarized tweets during the 15 days after following the DOL (compared to the previous 15 days), rather than 9.6, 2.3, and 2.2 respectively, as reported in Figure 3. Nevertheless, we still observe effects of a substantive magnitude, particularly if we look at the effects for longer periods (e.g. 30-day window) and when following more than one DOL; and all the statistical findings reported in Figure 3 hold when excluding these potential outliers.

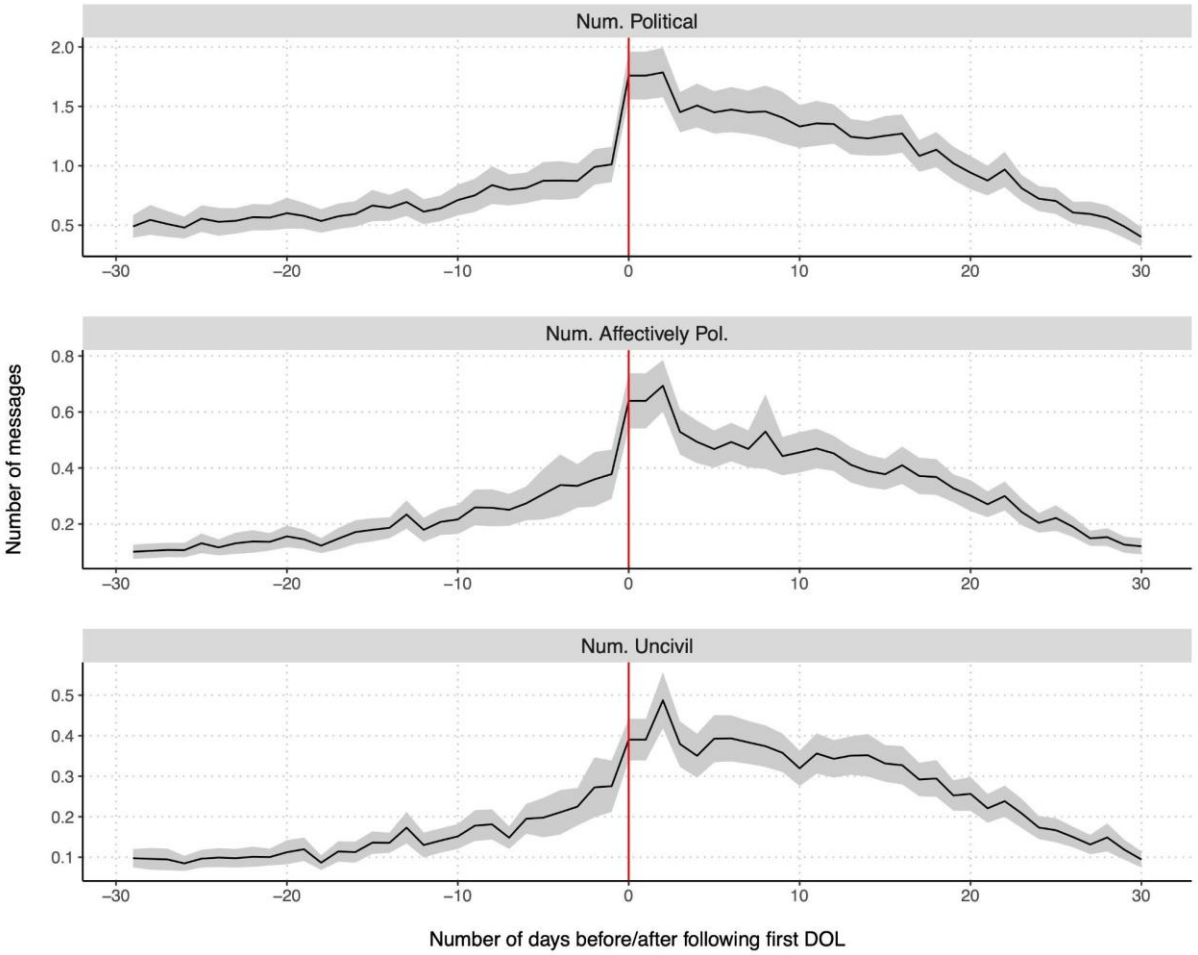
Figure E.1. Replication of the model results reported in Figure 3, after excluding regression outliers (those with a Cooks Distance 4 times larger than the average).



In Figure E.2 we replicate Figure 4 of the paper after removing from the dataset those users considered as outliers when estimating the four regressions reported in the bottom-left panel in Figure E.1 (so those considered as outliers when estimating a change in the political, uncivil, and affectively polarized messages; 30 days after following the first DOL). Note that a given user may be considered an outlier for one of the models (e.g. change in the number of political tweets) but not the others, and so information from a given user may be excluded when calculating the moving averages from some panels in Figure E.2, but not the others. Similar to the findings in Figure E.1, the overall volume

of tweets reported in Figure E.2 is slightly lower than those reported in Figure 4. Nevertheless, we observe the same patterns once these potential outliers are removed.

Figure E.2. Average number of political, uncivil, and affectively polarized tweets sent by followers of DOLs, during the 30 days before and after following the first DOL. Potential outliers have been removed from the data.



References

Barberá, P. (2015). Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data. *Political Analysis*, *23*(1), 76–91. https://doi.org/10.1093/pan/mpu011

Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, *33*(2), 122–139. https://doi.org/10.1177/0267323118760317

Cook, R. Dennis (February 1977). "Detection of Influential Observations in Linear Regression". Technometrics (American Statistical Association)

DiFonzo, N., & Bordia, P. (2007). Rumor, Gossip, and Urban Legends. *Diogenes*, *54*(1), 19–35. https://doi.org/10.1177/0392192107073433

Douglas, K. M., Uscinski, J. E., Sutton, R. M., Cichocka, A., Nefes, T., Ang, C. S., & Deravi, F. (2019). Understanding Conspiracy Theories. *Political Psychology*, *40*(S1), 3–35. https://doi.org/10.1111/pops.12568

Egelhofer, J. L., & Lecheler, S. (2019). Fake news as a two-dimensional phenomenon: A framework and research agenda. *Annals of the International Communication Association*, *43*(2), 97–116. https://doi.org/10.1080/23808985.2019.1602782

Friggeri, A., Adamic, L., Eckles, D., & Cheng, J. (2014, May). Rumor cascades. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 8, No. 1).

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, *5*(1), eaau4586. https://doi.org/10.1126/sciadv.aau4586

McKay, S., & Tenove, C. (2021). Disinformation as a Threat to Deliberative Democracy. *Political Research Quarterly*, *74*(3), 703–717. https://doi.org/10.1177/1065912920938143

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter*, *19*(1), 22–36. https://doi.org/10.1145/3137597.3137600

Sunstein, C. R., & Vermeule, A. (2009). Conspiracy Theories: Causes and Cures*. *Journal of Political Philosophy*, *17*(2), 202–227. https://doi.org/10.1111/j.1467-9760.2008.00325.x

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559