

Robertson, Claire E.; Pröllochs, Nicolas; Pärnamets, Philip; Van Bavel, Jay J.;
Feuerriegel, Stefan

Working Paper

Response to Replication Report of "Negativity Drives Online News Consumption"

I4R Discussion Paper Series, No. 200

Provided in Cooperation with:

The Institute for Replication (I4R)

Suggested Citation: Robertson, Claire E.; Pröllochs, Nicolas; Pärnamets, Philip; Van Bavel, Jay J.;
Feuerriegel, Stefan (2025) : Response to Replication Report of "Negativity Drives Online News
Consumption", I4R Discussion Paper Series, No. 200, Institute for Replication (I4R), s.l.

This Version is available at:

<https://hdl.handle.net/10419/311305>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen
Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle
Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich
machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen
(insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten,
gelten abweichend von diesen Nutzungsbedingungen die in der dort
genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

*Documents in EconStor may be saved and copied for your personal
and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to
exhibit the documents publicly, to make them publicly available on the
internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content
Licence (especially Creative Commons Licences), you may exercise
further usage rights as specified in the indicated licence.*

No. 200
DISCUSSION PAPER SERIES

Response to Replication Report of “Negativity Drives Online News Consumption”

Claire E. Robertson
Philip Pärnamets
Stefan Feuerriegel

Nicolas Pröllochs
Jay J. Van Bavel

This **paper responds to:**

Reiss, M.V., and H. Roggenkamp. 2025. A Comment on “Negativity Drives Online News Consumption”.
I4R Discussion Paper Series No. 199. Institute for Replication.

February 2025

I4R DISCUSSION PAPER SERIES

I4R DP No. 200

Response to Replication Report of “Negativity Drives Online News Consumption”

**Claire E. Robertson¹, Nicolas Pröllochs², Philip Pärnamets³,
Jay J. Van Bavel¹, Stefan Feuerriegel⁴**

¹*New York University, New York/USA*

²*University of Giessen/Germany*

³*Karolinska Institutet, Stockholm/Sweden*

⁴*Ludwig-Maximilians-University, Munich/Germany*

FEBRUARY 2025

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and meta-scientific work in the social sciences. Provided in cooperation with EconStor, a service of the [ZBW – Leibniz Information Centre for Economics](#), and [RWI – Leibniz Institute for Economic Research](#), I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Editors

Abel Brodeur
University of Ottawa

Anna Dreber
Stockholm School of Economics

Jörg Ankel-Peters
RWI – Leibniz Institute for Economic Research

Response to replication report of “*Negativity drives online news consumption*”

Claire E. Robertson ¹, Nicolas Pröllochs ², Philip Pärnamets ³, Jay J. Van Bavel ^{1,4}, Stefan Feuerriegel ⁵

¹ Department of Psychology, New York University, New York, NY, USA

² Department of Business and Economics, University of Giessen, Giessen, Germany

³ Division of Psychology & Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

⁴ Center for Neural Science, New York University, New York, NY, USA.

⁵ LMU Munich School of Management, LMU Munich, Munich, Germany.

We thank the authors for the extensive and thoughtful replication of our paper “*Negativity drives online news consumption*” (Robertson et al 2023). The authors conducted two types of direct replications and two types of conceptual replications of our analyses, and in all cases, found our main claim – that negative language drives news consumption – was highly robust. Indeed, they were able to exactly reproduce our key results and found they were robust to alternative model specifications and analytic decisions. We highly appreciate the depth of the replication, such as providing extensive documentation in their software repository as well as experimenting with additional methods based on large language models (LLMs). Overall, this gives us much greater confidence in the main conclusion from our original paper – that negative drives online news consumption in the Upworthy archive.

Interpretation of findings

We are pleased to see that the replication supports our main hypothesis. Let us briefly summarize the main takeaways: First, the authors repeated our exact analysis procedure with the code and data offered as supplementary materials (labeled as “unblind computational reproduction”; see Section 4). Here, the authors found numerically identical results. Second, the authors repeated the analysis with their own preprocessing pipeline (labeled as “blind computational reproduction; see Section 3). Here, the authors arrived at the same conclusive findings as our main hypothesis.

The replication also offers an additional conceptual replication analysis using a novel LLM-based approach (Section 5; which the replication refers to as “semantic” analysis in contrast to our “lexical” analysis), where the authors use the zero-shot functionality of LLMs to extract the perceived positivity of the overall headline. LLMs have recently garnered attention in behavioral science, especially after the launch of ChatGPT at the end of 2022 (e.g., Feuerriegel et al 2024). However, at the time of conducting our research, LLM-based approaches, especially those based on zero-shot functionality as used in the replication report, were not yet established for measuring psychological constructs. Hence, we are very excited to see the new analyses and hope that such analyses will help improve our theoretical understanding of the role of language in news consumption.

We were excited to see that, using this new methodology, the authors also found evidence of negativity bias in the news. However, they did not find robust evidence that positivity reduced news consumption. We had stronger theoretical reasons to expect negativity to increase consumption than we did for positivity to reduce consumption, and consider the former to be our main claim in the paper (this is why we made this finding the title of our paper).

Here, we highlight several important differences between the LLM-based approach in the replication report and our analysis, which is important for correctly interpreting the results and thus for generating new insights into the underlying psychological mechanisms. As such, the LLM-based approach should be seen as a complement to our analysis that asks a related but slightly different research question than our original manuscript.

First, we originally preregistered our hypotheses based on the frequency of negative and positive words as an important determinant of news consumption (see the Design Table in Table 1 of our manuscript). Throughout the manuscript, we were very careful to discuss our results in light of our hypothesis, namely that the presence of more negative (positive) words may increase (decrease) the click-through rate. Hence, it is important to recognize that the constructs in the LLM-based approach are different from ours, because these aim to capture the perceived negativity or perceived positivity. Such differences in the role of language were discussed earlier (e.g., Berger et al 2020; Feuerriegel et al 2025; Mohammad 2016), where language was theorized to have a dual role in that it can “reflect” the writer’s emotional state and “affect” the reader’s emotional state. Thus, we can consider these results as complementary rather than identical.

Second, there are differences in the definition of the statistical quantities. We measured the use of positive and negative words (which the replication refers to as “lexical” analysis) where we normalized for the length of the headline. In contrast, the LLM-based approaches used in the replication report computed probabilities as to whether the headline content was overall positive or negative (or neutral) but without normalizing for the length (which the replication refers to as “semantic” analysis). For example, a news headline that consists of a negative and a positive word would receive high scores for both variables in our frequency-based approach, but not necessarily in the LLM-based approach. Hence, it is not surprising that the estimated coefficients may vary, and the variations can be attributed to some extent to the differences in variable definitions.

In general, LLMs can offer several advantages in behavioral science. For example, many studies found that LLMs are highly accurate in capturing psychological constructs, even across different languages (e.g., Rathje et al 2024). At the same time, we recommend a careful approach that balances both the strengths and is also transparent about potential weaknesses. First, there is essentially no transparency of how inferences are made by LLMs (Grimes et al 2023). Second, LLMs are highly sensitive to the exact prompt (Demszky et al. 2023) and repeat human biases in assessing language (Caliskan et al 2017). Third, state-of-the-art LLMs such as GPT-4o are proprietary and subject to frequent changes or even discontinuation. Ironically, this could make down-the-line replication with these models difficult (although LLMs seem to improve with each iteration; see Rathje et al., 2024).

There are several other considerations one should think about when interpreting the LLM results. First, it is unclear how effective smaller LLMs like DistillBERT are in capturing semantics. Oftentimes, specialized LLMs such as Sentence-BERT are used instead for very short text expressions like tweets or sentences (Reimers et al 2019). Second, the DistillBERT approach is based on a custom HuggingFace model¹ that has not (yet) gone through peer review. Hence, it is unclear how well the model performs in capturing psychological constructs more broadly (e.g., in dedicated benchmarking studies because the model webpage is unclear for which corpus the model was fine-tuned). In particular, many of the smaller LLMs, such as those from the BERT family, are often not as good in zero-shot inferences as compared to large LLMs such as GPT-3.5 onward. Third, the prompting strategies may play an important role in order to generate accurate outputs. Some evidence suggests that LLMs perform more reliably when they are prompted for discrete outputs (e.g., “good” vs. “bad”) in comparison to numerical outputs (e.g., probabilities) (e.g., Atreja et al 2024). Fourth, it is very likely that GPT-4o has seen the *Upworthy* corpus during training. Here, it is unclear whether this may bias the results or even improve the accuracy (due to the pre-training).

Taken together, we think that more research is needed to compile rigorous evidence, and we appreciate the new insights that were generated through the replication. The extended analysis (labeled “semantic” analysis) asks theoretically distinct research questions about new constructs that are important and warrant further analysis.

The above LLM analysis suggested in the replication opens up several avenues for future research. We would like to suggest several directions for future research in the following, which we find particularly interesting:

- How robust are the LLM-based approaches to the exact prompting strategy? Would the use of in-context learning (i.e., adding some examples to the prompt) improve the agreement with human annotators? How robust is the analysis to specific parameter choices (e.g., different choices of the temperature parameter)?
- How can LLMs help in understanding the role of emotions in news consumption? We expect the LLMs to be especially effective in capturing basic emotions. Here, we originally found that the dictionary approaches had only relatively little overlap with human perceptions. Hence, a rewarding direction is to repeat our second research question from the Design Table by using an LLM-based approach for inferring perceived basic emotions.
- How does the role of language vary across cognitive vs. affective information processing? Evidence from fake news consumption finds that language can trigger information processing in different ways (Lutz et al 2023, Lutz et al 2024).

Lessons learned

We found some interesting implications from the replication effort, and we hope that this spurs further discussions to develop more concrete community guidelines.

¹ <https://huggingface.co/lxyuan/distilbert-base-multilingual-cased-sentiments-student>

First, routines for data preprocessing can become very long, and, even if described with verbal language, the descriptions are never as accurate as raw source code (e.g., different implementations of the same algorithm can even end up with different numerical results). Hence, there is an inherent tradeoff as to what extent preprocessing steps should be detailed in the main paper or additionally in the supplementary codes. In our case, we opted for the – arguably – more common way in computer science where some descriptions of the preprocessing steps are relegated to the software codes but where, therefore, the preprocessing steps are made publicly available as source code. In sum, one implication for the replication initiative is that it would be helpful to develop further guidelines for authors about how to balance the description of preprocessing steps between the main paper (with length limitations) and codes.

Second, this explains the minor differences between the blind and the unblind computational reproduction. Here, we would like to emphasize that our research was conducted at the highest level of reproducibility rigor, as it was a registered report, combined from two unique and interdisciplinary teams, and published after extensive and critical peer review both at the design and the results stage. Our analysis description was conditionally accepted for publication before we produced our results. Still, our in-text description of the analysis was not "sufficient" for perfect reproducibility. Quite frankly, if this rigorous process did not produce sufficient in-text descriptions of preprocessing and analysis pipelines, we can't imagine that any in-text description would. Hence, another takeaway is that *any* codes for analysis should be made available, including explicitly those for data processing and even simple statistics, as this replication effort is evidence that in-text description will almost certainly be deficient in some way.

Third, we have carefully reported the versions of the software packages that we used for the analysis, similar to many other papers. Upon reading the replication, we realized that packages in the data processing pipeline may have downstream implications for the analysis, and, hence, it would be great to document the versions of all used packages in the software environment, not just from the analysis but also for other processing steps. Here, standardized reporting forms such as the "Reporting Summary" as used by many *Nature* journals are a very important and helpful step but one implication may be that such reporting forms could be developed further as a result of the replication initiative (e.g., by encouraging the reporting of exact software versions or even exact software environment to be dumped to a metadata file in the repository).

Fourth, the replication sought to recreate our own software environment. This effort is commendable and also shows the limitations of many computational analyses. Software packages are updated frequently, and there are thus issues around compatibility, dependency issues, and limited support as packages can be discontinued very abruptly. Back when we conducted our analyses, there was limited support for package version control systems in *R*. For example, the *Microsoft R Application Network*, which is the backbone of packages like *groundhog* (as used by the replication), was just announced to be discontinued, and, hence, we decided against its use at the time when we conducted our analysis to avoid working with deprecated packages. As can be seen in our case, package version control systems may introduce further dependencies with unintended consequences. In sum, while some progress has been made to recreate software

environments, we hope that the replication initiative calls for more efforts to develop standardized package version control systems and ensure long-term maintenance.

Summary

We sincerely thank the authors for their thorough replication of our paper “*Negativity Drives Online News Consumption*” (Robertson et al., 2023). We are pleased to see that their findings support our main hypothesis. In sum, the authors (i) identically replicated the finding with our code, (ii) statistically replicated the finding without our code, and (iii) conceptually replicated our finding twice with completely new methods. We look for future research that extends our analysis in several ways, such as by leveraging recent advances from machine learning or by studying new constructs that are complementary to ours.

References

- Atreja, S., Ashkinaze, J., Li, L., Mendelsohn, J., & Hemphill, L. (2024). Prompt design matters for computational social science tasks but in unpredictable ways. *arXiv:2406.11980*.
- Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the tribes: Using text for marketing insight. *Journal of Marketing*, 84(1), 1–25.
- Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334), 183–186.
- Demszky, D., Yang, D., Yeager, D. S., Bryan, C. J., Clapper, M., Chandhok, S., ... & Pennebaker, J. W. (2023). Using large language models in psychology. *Nature Reviews Psychology*, 2(11), 688–701.
- Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. *Business & Information Systems Engineering*, 66(1), 111–126.
- Feuerriegel, S., Maarouf, A., Bär, D., Geissler, D., Schweisthal, J., Pröllochs, N., Robertson, C. E., Rathje, S., Hartmann, J., Mohammad, S. M., Netzer, O., Siegel, A. A., Plank, B., & Van Bavel, J. J. (2025). Using natural language processing to analyse text data in behavioural science. *Nature Reviews Psychology*.
- Grimes, M., Von Krogh, G., Feuerriegel, S., Rink, F., & Gruber, M. (2023). From scarcity to abundance: Scholars and scholarship in an age of generative artificial intelligence. *Academy of Management Journal*, 66(6), 1617–1624.
- Lutz, B., Adam, M. T., Feuerriegel, S., Pröllochs, N., & Neumann, D. (2023). Affective information processing of fake news: Evidence from NeurolS. *European Journal of Information Systems*, forthcoming.
- Lutz, B., Adam, M., Feuerriegel, S., Pröllochs, N., & Neumann, D. (2024). Which linguistic cues make people fall for fake news? A comparison of cognitive and affective processing. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW1), 1–22.
- Mohammad, S. M. (2016). Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In *Emotion Measurement* (pp. 201–237). Woodhead Publishing.
- Rathje, S., Mirea, D. M., Sucholutsky, I., Marjeh, R., Robertson, C. E., & Van Bavel, J. J. (2024). GPT is an effective tool for multilingual psychological text analysis. *Proceedings of the National Academy of Sciences*, 121(34), e2308950121.
- Reimers, N., and Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.

Robertson, C. E., Pröllochs, N., Schwarzenegger, K., Pärnamets, P., Van Bavel, J. J., & Feuerriegel, S. (2023). Negativity drives online news consumption. *Nature Human Behaviour*, 7(5), 812–822.