Digital Trace Data in Computational Social Science Research

Ruben Bach, University of Mannheim
@DIGISURV, University of Manchester
Feb 14, 2025

Digital Trace Data

- Increasing importance for (computational) social science research (Bach et al., 2021)

- Many different sources

  - Research Apps (Kreuter et al., 2020)

  - Webscraping (Munzert et al., 2014)

  - Data donation (Keusch, Pankowska, Cernat & Bach, 2024)

  - Webtracking (Bach et al., 2022)

Webtracking

- Participants install small plugins in their web browsers / smartphones

- Keep track of websites visited

- May capture screenshots or HTML code of websites visited

- Ex-post scraping of URLs visited

- Increasing popularity in health research, media and communication research, and political communication research

Webtracking data in applied social research

Populist alternative media
Areal, Behnert, Lajic, Bach (2025)

- Anti-mainstream, right-wing populist agenda & communication style, anti-elitism (Müller & Schulz, 2021)

- Protection of the environment as part of a globalist & elite agenda (Lockwood, 2018)

- Environmental activism harms the economy and undermines national sovereignty (Huber et al., 2020; Freudenthaler & Wessler, 2022)

➡ **How do populist alternative media report on environmental topics compared to legacy media?**

Populist alternative media
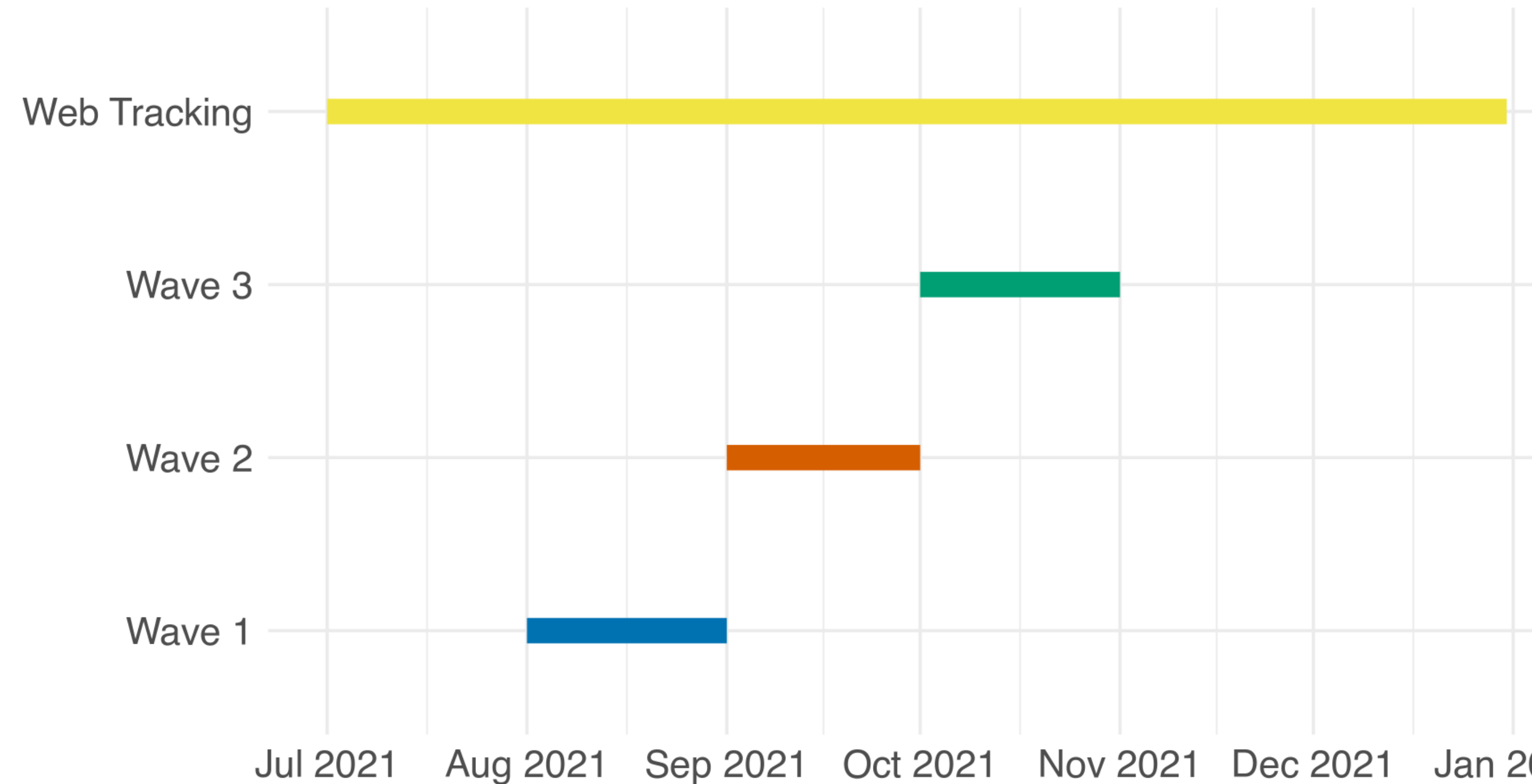## Areal, Behnert, Lajic, Bach (2025)

Media effects

- Limited media effects of legacy media (Valkenburg et al., 2016); hardly evidence for alternative media

- Alternative media: persuasive, opinionated, and ideologically coherent reporting (Müller &Schulz, 2021)

  - **Transportation Theory** (Green & Brock, 2000): Narrative involvement increases the likelihood of having attitudes and beliefs influenced by the story

  - **Elaboration Likelihood Model** (Petty & Cacioppo, 1986): Opinion-driven, ideologically biased content amplifies cognitive and emotional reactions.

  - **Stronger Media effects** (Dimitrova et al., 2014; Frischlich et al., 2022)

- **What are the short-term media effects of using alternative media on environmental attitudes?**

Research design
Bach et al. (2023)

- Six months of web tracking data (n=2,000)

  - July through December 2021

  - Web scraping of news websites visited

- Three-wave panel survey

  - August (n=1,603)

  - September (n=1,579)

  - October (n=1,546)

  - Three dependent variables: Economy before environment; Importance of protecting the environment; Sympathy with the Green Party

- Online Access Panel

Populist Alternative Media
Areal, Behnert, Lajic, Bach (2025)

Scraping of news media articles

- Ca. 96,000 articles (legacy media)

- Ca. 3,000 articles (alternative media)

Identification of articles on environmental topics via machine learning classifier

- Ca. 3,000 articles (legacy media)

- Ca. 300 articles (alternative media)

Thematic analysis with transformer-based language model

In what context do populist alternative media report on environmental issues?
Areal, Behnert, Lajic, Bach (2025)

- Identification of the narrative or context of an article

- Comparison between legacy and alternative media

➡️ **Populist alternative media emphasize the political aspects of environmental issues much more strongly than mainstream media**
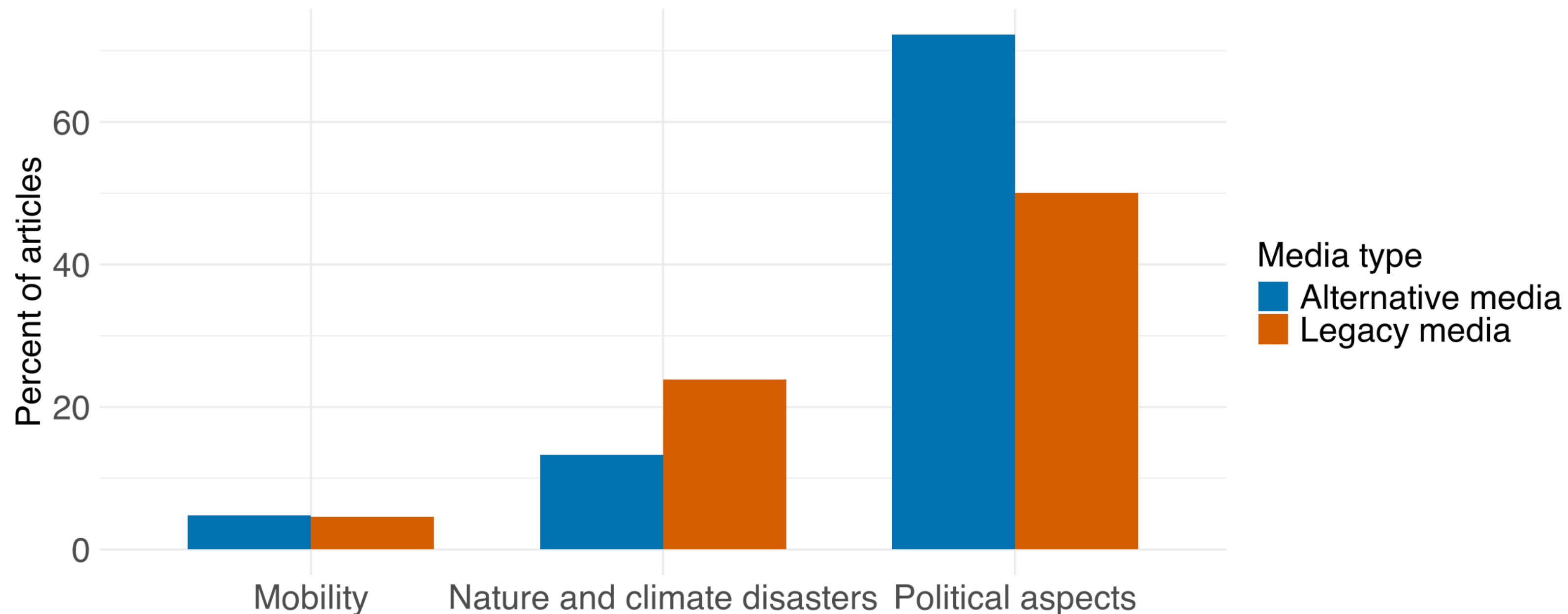
Media effects of populist alternative media

Areal, Behnert, Lajic, Bach (2025)

- Random-Effects Within-Between (Bell & Jones, 2015): Effects of alternative media consumption (web tracking data) on attitudes (survey data)

➡️ **Consumers of alternative media are generally less environmentally conscious**

➡️ **Consumption of alternative media reduces the perception of the importance of protecting the environment**
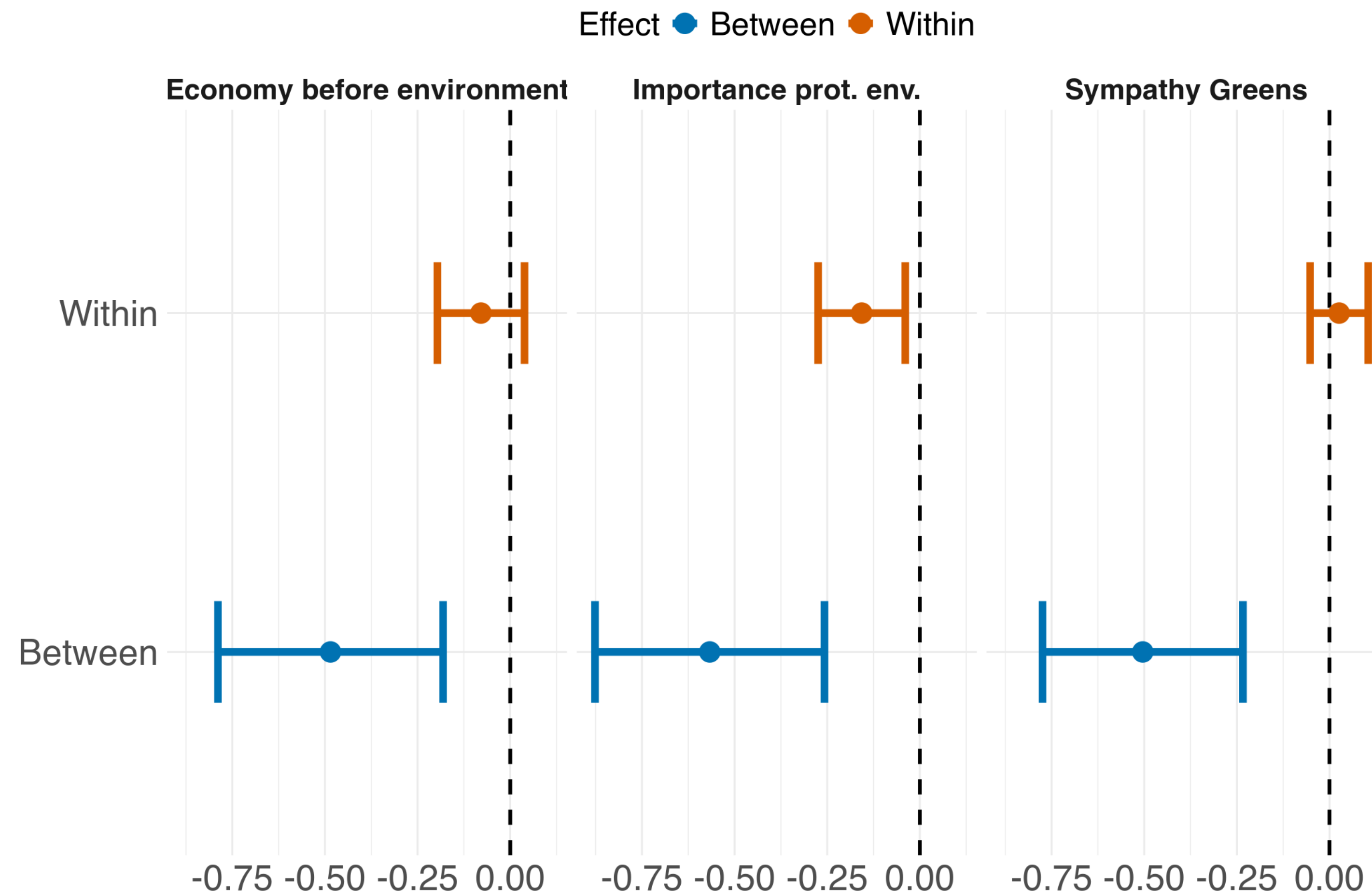
New measures of cognitive household labor — A Web Tracking Approach

New measures of cognitive household labor

- Gender differences in the division of cognitive household labor and mental load (z.B. Daminger, 2019; Haupt & Gelbgiser, 2024)

- Hardly useful survey instruments (Dean et al., 2021; Haupt & Gelbgiser, 2024)

- Measurement of cognitive household labor through activities and keywords (Daminger, 2019) in web tracking data

  - Parallel survey study with items from existing studies (Gender & Generations Survey)

  - GESIS Panel.dbd

Causal Effects of Social Media Use — A Data Donation Approach

Causal effects of social media use on body image

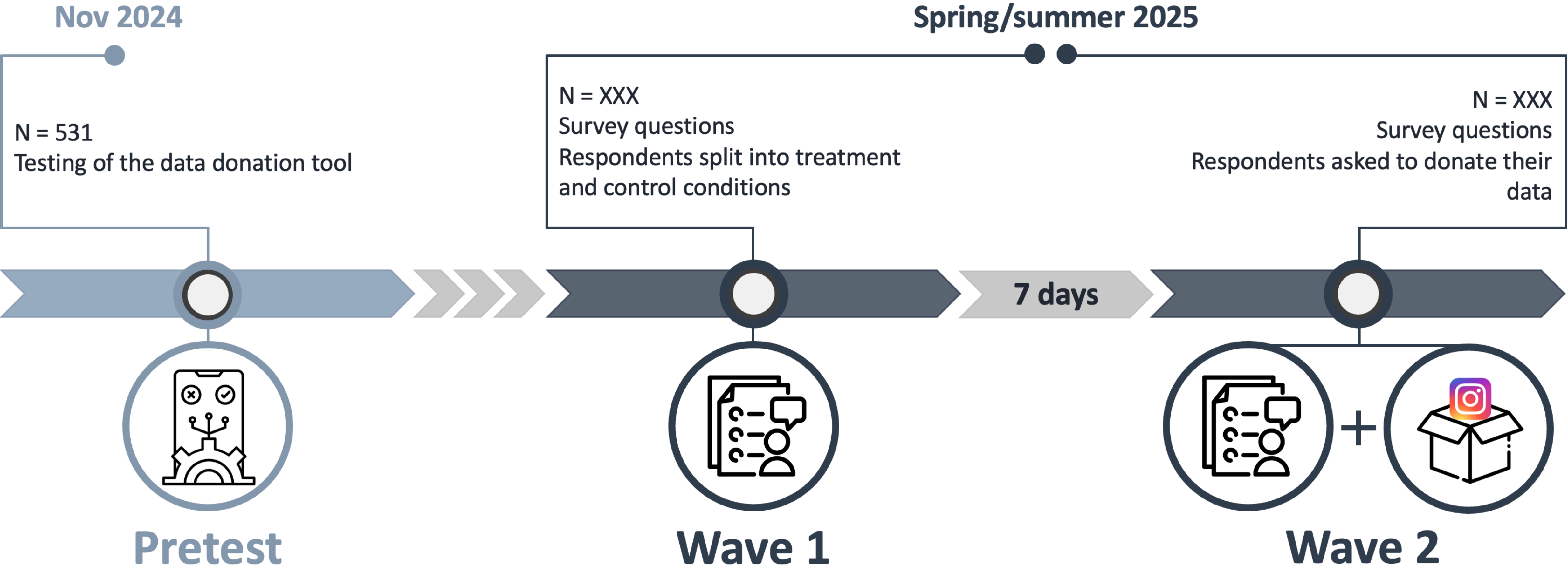## Instagram Makes Teen Girls Hate Themselves. Is That a Bug or a Feature?

7 MINUTE READ

The features within Instagram that cause teenage girls to develop negative feelings about their body image may be baked into the very core of the platform, researchers and former employees have said in the wake of new revelations that the company did not disclose what it knew about its impact on young users.

Facebook, which owns Instagram, has known for years that the platform is harmful to the mental health of many teenagers—particularly girls—but has kept internal research about the issue private, according to a *Wall Street Journal* report published Tuesday.

https://time.com/6098771/instagram-body-image-teen-girls/

Causal effects of social media use on body image

Nov 2024

Spring/summer 2025

N = 531
Testing of the data donation tool

N = XXX
Survey questions
Respondents split into treatment
and control conditions

N = XXX
Survey questions
Respondents asked to donate their
data

7 days

Pretest

Wave 1

Wave 2
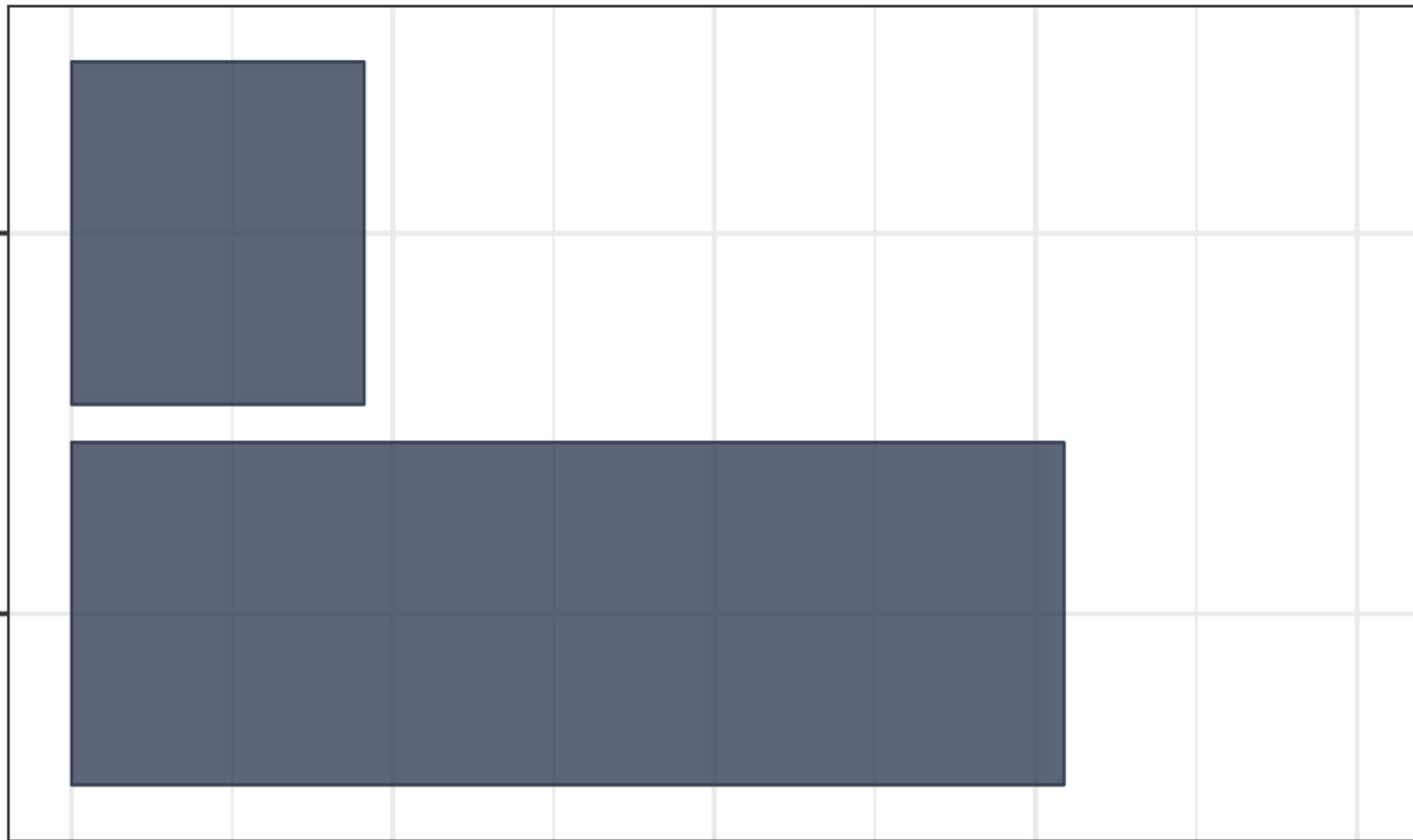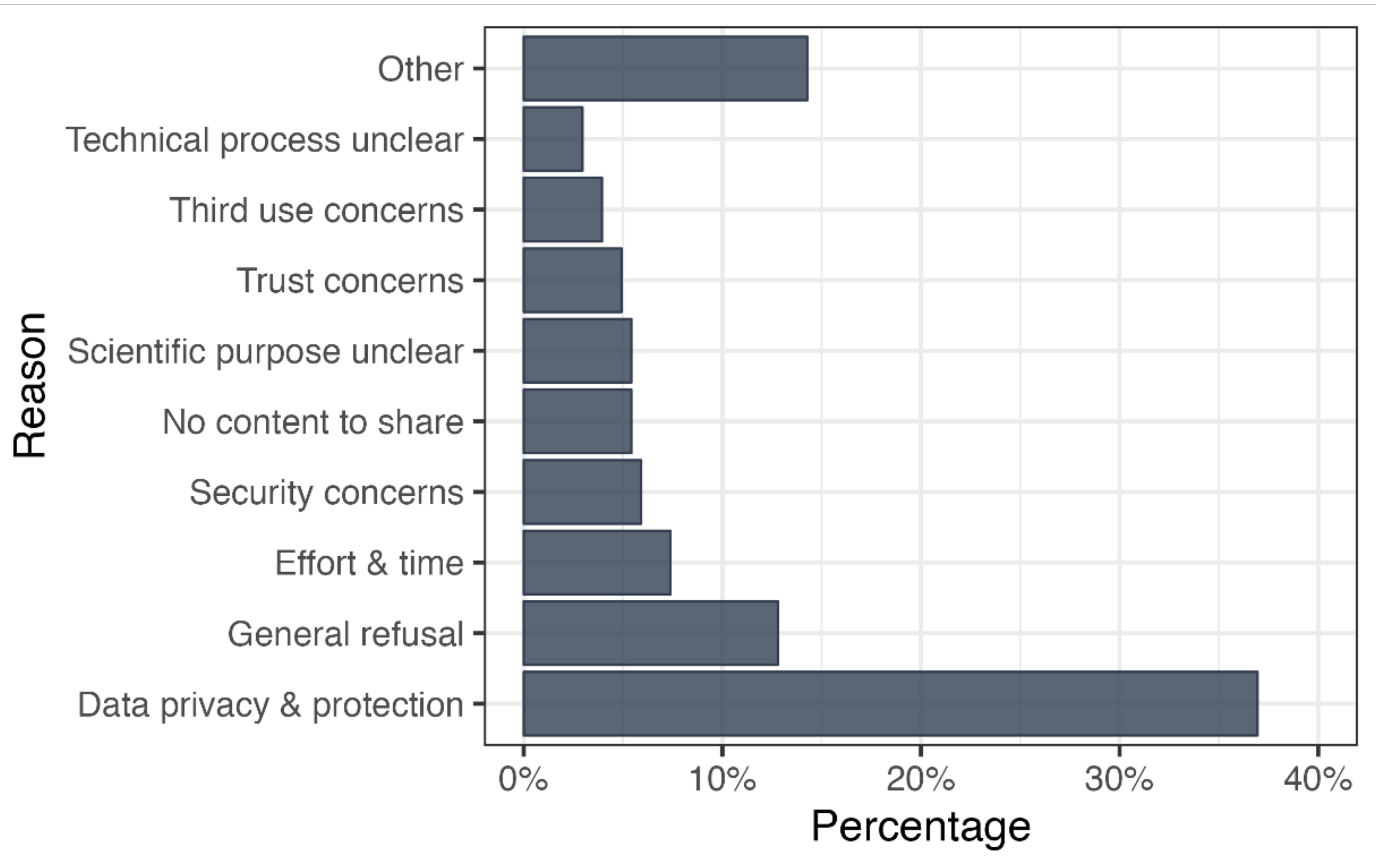
Thank you

Literature

- Araujo, T., Ausloos, J., van Atteveldt, W., Loecherbach, F., Moeller, J., Ohme, J., ... & Welbers, K. (2022). OSD2F: An open-source data donation framework. *Computational Communication Research*, 4(2), 372-387.

- Areal, J., Behnert, J., Lajic, D., & Bach, R. (2024). Environmental topics on the alternative agenda: content and effects of alternative media coverage. *International Journal of Public Opinion Research*.

- Bach, R., Kern, C., Amaya, A., Keusch, F., & Kreuter, F. (2021). Predicting Voting Behavior Using Digital Trace Data. *Social Science Computer Review*, 39(5), 862-883.

- Bach, R., Keusch, F., Areal, J., Pankowska, P., & Cernat, A. (2023). Political Identities and News Consumption in Election Times (PINCET). 2023, Version 2.0.0. (https://doi.org/10.7802/2585)

- Bach, R. L., Kern, C., Bonnay, D., & Kalaora, L. (2022). Understanding political news media consumption with digital trace data and natural language processing. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 185, S246-S269.

- Bauer, P. C., Landesvatter, C., & Behrens, L. (2022). APIs for social scientists: A collaborative review. https://paulcbauer.github.io/apis_for_social_scientists_a_review/

Literature

- Bosch, O. J., Sturgis, P., Kuha, J., & Revilla, M. (2024). Uncovering digital trace data biases: tracking undercoverage in web tracking data. *Communication Methods and Measures*, 1-21.

- Bosch, O. J., & Revilla, M. (2022). When survey science met web tracking: Presenting an error framework for metered data. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 185(S2), 408-436.

- Daminger, A. (2019). The cognitive dimension of household labor. *American Sociological Review*, 84, 609–33.

- Dean, L., Churchill, B., & Ruppanner, L. (2021). The mental load: building a deeper theoretical understanding of how cognitive and emotional labor overload women and mothers. *Community, Work & Family*, 25, 1–17.

- Dimitrova, D. V., Shehata, A., StrÅNombÅNack, J., & Nord, L. W. (2014). The effects of digital media on political knowledge and participation in election campaigns: Evidence from panel data. *Communication research*, 41(1), 95–118.

- Freudenthaler, R., &Wessler, H. (2022). How alternative are alternative media? Analyzing speaker and topic diversity in mainstream and alternative online outlets. *Digital Journalism*, online first, 1–21.

- Frischlich, L., Kuhfeldt, L., Schatto-Eckrodt, T., & Clever, L. (2022). Alternative counter-news use and fake news recall during the covid-19 crisis. *Digital Journalism*, online first, 1–23.

Literature

- Haupt, A. & Gelbgiser, D. (2024). The gendered division of cognitive household labor, mental load, and family–work conflict in European countries. *European Societies*, 26(3), 828-854.

- Gil-López, T., Christner, C., de León, E., Makhortykh, M., Urman, A., Maier, M., & Adam, S. (2023). Do (not!) track me: Relationship between willingness to participate and sample composition in online information behavior tracking research. *Social science computer review*, 41(6), 2274-2292.

- Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of personality and social psychology*, 79(5), 701.

- Huber, R. A., Fesenfeld, L., & Bernauer, T. (2020). Political populism, responsiveness, and public support for climate mitigation. *Climate Policy*, 20(3), 373–386.

- Keusch, F., Bach, R., & Cernat, A. (2023). Reactivity in measuring sensitive online behavior. *Internet Research*, 33(3), 1031-1052.

- Keusch, F., Pankowska, P. K., Cernat, A., & Bach, R. L. (2024). Do you have two minutes to talk about your data? Willingness to participate and nonparticipation bias in Facebook data donation. *Field Methods*, 36(4), 279-293.

- Kreuter, F., Haas, G. C., Keusch, F., Bähr, S., & Trappmann, M. (2020). Collecting survey and smartphone sensor data with an app: Opportunities and challenges around privacy and informed consent. *Social Science Computer Review*, 38(5), 533-549.

Literature

- Lockwood, M. (2018). Right-wing populism and the climate change agenda: Exploring the linkages. *Environmental Politics*, 27(4), 712–732.

- Müller, P., & Bach, R. L. (2023). Populist alternative news use and its role for elections: Web-tracking and survey evidence from two campaign periods. *New media & society*, 25(10), 2663-2683.

- Müller, P., & Schulz, A. (2021). Alternative media for a populist audience? Exploring political and media use predictors of exposure to breitbart, sputnik, and co. *Information, Communication & Society*, 24(2), 277–293.

- Munzert, S., Rubba, C., Meißner, P., & Nyhuis, D. (2014). *Automated data collection with R: A practical guide to web scraping and text mining*. John Wiley & Sons.

- Pankowska, P., Cernat, A., Bach, R., & Keusch, F. (R&R). Using hidden Markov models to assess and correct for measurement error in digital trace data. *Communication Methods and Measures*.

- Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer.

- Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). Integrating Survey and Digital Trace Data: Key Issues in an Emerging Field. *Social Science Computer Review*, 38(5), 503-516.