






Integrating Communication Science and Computational Methods to Study Content-Based Social Media Effects

J. Loes Pouwels ^a, Theo Araujo ^b, Wouter van Atteveldt ^c, Marko Bachl ^d,
and Patti M. Valkenburg ^b

^aBehavioural Science Institute, Radboud University, Nijmegen, The Netherlands; ^bAmsterdam School of Communication Research, University of Amsterdam, Amsterdam, The Netherlands; ^cDepartment of Communication Science, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; ^dDepartment of Political and Social Sciences, Institute for Media and Communication Studies, Freie Universität, Berlin, Germany



ABSTRACT

A pressing societal and scientific question is how social media use affects our cognitions, emotions, and behaviors. To answer this question, fine-grained insight into the content of individuals' social media use is needed. It is difficult to study content-based social media effects with traditional survey methods because such methods are incapable of capturing the extreme volume and variety of social media content that is shared and received. Therefore, this special issue aims to illustrate how content-based social media effects could be examined by integrating communication sciences and computational methods. We describe a three-step method to investigate content-based media effects, which involves (a) collecting digital trace data, (b) performing automated textual and visual content analysis, and (c) conducting linkage analysis. This Special Issue zooms in on these steps and describes the strengths and weaknesses of different computational methods. We conclude with some challenges that need to be addressed in future research.

Introduction

In today's society, the use of social media is an important part of our daily lives. As of 2022, adults worldwide use social media for about 2.5 hours per day on average (Statista, 2022), and this number is even larger among young cohorts (Pew Research Center, 2022). Social media are often seen as a double-edged sword with promises like near-constant access to support and information and concerns regarding the impact of algorithms, the spread of disinformation, and other harmful outcomes on well-being (Choukas-Bradley et al., 2023; European Commission, 2023). The European Union, therefore, aims to create a safer online environment through the Digital Service Act (European Commission, 2023), and the United States government introduced the Kids Online Safety Act to empower children and their parents to control children's online experiences and to protect their well-being (The Kids Online Safety Act, 2022). Therefore, an urgent societal and scientific question is how social media use affects our cognitions, emotions, and behaviors.

In the past decades, the effects of social media use have been investigated across a wide range of disciplines, particularly communication science. To date, most of the evidence on the effects of social media use is based on variables that focus on the quantity of people's social media use, such as their time spent using social media or the frequency with which specific platforms are used (Parry et al., 2021). Although time-based measures of social media use could provide valuable insights into social

CONTACT J. Loes Pouwels  loes.pouwels@ru.nl  Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

media effects on outcomes like distraction or social displacement, they may be too rough to investigate media effects on other outcomes (Valkenburg et al., 2022). For example, to find out whether and how social media use contributes to polarization in today's society, we should be able to investigate what specific social media content people encounter, share, or create (Kubin & von Sikorski, 2021). Content-based measures of social media use are also needed to reveal how social media use relates to well-being because well-being fluctuations may be more dependent on the valence or type of social media interactions than on their frequency or duration (Valkenburg et al., 2022). Therefore, urgent societal questions require more fine-grained insights into the effect of certain types of content shared or encountered on social media.

This special issue aims to illustrate how content-based social media effects could be examined by integrating communication science and computational methods. Building on previous research (De Vreese et al., 2017; Kleinnijenhuis et al., 2019; Scharkow & Bachl, 2017), we aim to demonstrate how computational methods could be implemented to study content-based media effects using a three-step approach. As a first step, we should be able to collect digital trace data containing the textual and audiovisual content (i.e., multimodal data) of a person's social media use. As a second step, (automatic) textual (Step 2a) and visual (Step 2b) content analyses are required to extract meaningful dimensions, topics, or emotional expressions from this content. In the third and final step, the content should be linked to meaningful outcome variables, and its effects should be analyzed in methodologically sound research designs.

Each of the four review articles within this special issue zooms in on one of these steps. These articles identify not only the advantages and disadvantages of different state-of-the-art methods relevant for each step but also their best practices, including how they could be combined in a way that helps communication and social science researchers to rigorously answer pressing research questions in the realm of social media effects research. These articles also showcase the crucial need for interdisciplinary collaborations because investigating content-based media effects using computational methods requires sophisticated methodological expertise and solid theoretical knowledge of both social media use and the outcomes of interest.

This special issue is relevant for two types of scholars. First, we aim to provide media effects researchers with more information about computational methods' promises, challenges, and limitations. These insights will enable them to critically review communication science articles that have implemented computational methods and to consider how these methods can help them answer their social media effects research questions. Researchers interested in acquiring more hands-on expertise in implementing the computational methods themselves are referred to more practical articles and book chapters, as extensive tutorials are beyond the scope of this special issue. Second, we provide computational social scientists, computer scientists, and data scientists without a background in communication science with more insights into the study of media effects and the current challenges we experience in the field. This special issue may help them better understand how their knowledge could be implemented into the broader social sciences to solve urgent methodological and societal questions. We hope to foster effective communications and collaborations between media effects researchers, computational social scientists, computer scientists, and data scientists by targeting these two audiences.

In the remainder of this editorial, we will define social media use and computational methods. Subsequently, we will describe the three-step method to study content-based media effects. Specifically, we will discuss current challenges in the field and zoom in on each special issue article to illustrate how computational methods could overcome these challenges and bring the field forward. We end with some overarching conclusions and suggestions for future research.

Definitions of social media use and computational methods

In this special issue, we define social media use as “computer-mediated communication channels that allow users to engage in social interaction with broad and narrow audiences in real-time or

asynchronously” (Bayer et al., 2020, p. 316). This broad definition of social media use refers to a wide variety of more general (e.g., Facebook, Instagram, TikTok, YouTube) or specialized (e.g., Tinder or LinkedIn) social media apps, as well as messaging applications (e.g., WhatsApp, Telegram, Signal). Changes in cognitions, emotions, or behavior that occur within persons as a consequence of their social media use are defined as social media effects (Valkenburg et al., 2016). A distinction could be made between reception effects (how individuals are affected by social media content produced by others) and self-effects (i.e., how individuals are affected by their own produced social media content) (Valkenburg, 2017). Social media effects may occur both in the short term (e.g., within seconds, hours, or days) and in the long term (e.g., within weeks, months, or years) (Pouwels et al., 2021). Furthermore, they may be group- and person-specific, indicating that different groups and persons may be affected in different ways by using social media (Pouwels et al., 2021; Valkenburg et al., 2021).

We define computational methods as collecting and analyzing large amounts of data (Lazer et al., 2009). Computational communication science studies generally involve: “(1) large and complex data sets; (2) consisting of digital traces and other naturally occurring data; (3) requiring algorithmic solutions to analyze; and (4) allowing the study of human communication by applying and testing communication theory” (van Atteveldt & Peng, 2018, p. 82). In the context of this special issue, we specifically focus on digital trace data collection of social media use, which could be defined as all records of social media activity and content (trace data) that are digitally undertaken (Howison et al., 2011; Ohme et al., 2023). We also discuss computational methods used to analyze content (e.g., transformer-based models and image analysis methods) and linkage designs to investigate media effects.

A three-step method to study content-based social media effects

Step 1. Digital trace data collection

The first challenge in investigating the effects of social media use is to collect digital trace data of the textual and audiovisual content of a person’s social media use. Various methods enable researchers to collect digital trace data regarding the content of people’s social media use (Ohme et al., 2023). These data consist of a wide variety of textual and audiovisual content that is (a) *produced by* social media users (e.g., posts, stories, private messages, likes, tags), (b) *selected by* social media users (e.g., searches), (c) *selected for* social media users (e.g., algorithmic recommendations/filtering), and (d) *received by* social media users (e.g., scrolling through posts of others, reading private messages). To better understand these data, it is essential to collect not only the produced, selected, and received content itself but also the context in which the activity took place (e.g., does the content originate from an original message or repost, a targeted or untargeted message, and is the message produced by a profile with many or a few followers?). However, although promising and insightful, each digital trace data collection method has its strengths and weaknesses, and each may not cover all types of data relevant to a research question.

To help researchers make an informed choice about which digital trace methods most closely align with their research questions, Ohme et al. (2023) compare three digital trace data collection methods: API data, data donation, and tracking. They introduce each of these methods and describe how they differ in terms of platform and user dependency, timeframe for data collection, the required data and content types, data quality, and privacy risks. They conclude that APIs might outperform data donation and tracking regarding predictability and unobtrusiveness. However, at the time of this writing, many APIs made available by social media platforms are being severely restricted or even discontinued. In addition, API data may be less appropriate for media effects research as it often cannot be linked to predictors and (or self-reported) outcomes of social media use, such as perceptions, emotions, or offline behavior – which is often measured via surveys. Data donation and tracking are user-centric approaches that enable researchers to link user-centric data with self-reported outcomes

more directly. However, these methods may be subject to sampling bias and require more participant effort. For all three methods, researchers should take extra measures to verify to what extent the collected digital traces are complete and fully reflect the concepts of interest.

Step 2. Automated content analysis

After the digital trace social media data have been collected, the second challenge is to analyze the textual and visual social media trace data in a theoretically meaningful manner. Given the extreme *volume* (i.e., a large amount of data) and *variety* (i.e., different formats/styles, languages, and modalities) of textual social media data that could be obtained from one single individual (Kroon et al., 2023), it is very hard and labor intensive to perform a (manual) content analysis on visual or textual data. As an alternative, automatic, top-down classifications could be used by teaching a computer how input features (e.g., text and audiovisual information) relate to predefined dimensions or topics (van Atteveldt et al., 2021).

Two commonly used strategies for automated textual content analyses that are discussed by Kroon et al. (2023) are dictionary analyses and bag-of-word (BoW) approaches, which identify and classify predefined theoretical concepts in a social media text. Kroon et al. argue that these methods often fall short of capturing semantic meaning because they ignore the contextualized meaning of language. Indeed, previous research has shown that implementing BoW methods in social media research falls short compared to trained human coding (van Atteveldt et al., 2021). As a promising alternative, Kroon et al. (2023) argue for using transformed-based models based on deep-language analyses (i.e., large language models; LLMs). By bringing the contextual meaning back into the model, these LLMs could help us to understand and classify the large variety of digital trace data. Especially for categories that are *meaningful* for answering relevant research questions, such as “Why does social media make some people feel happy while leaving others feeling blue?” LLMs may outperform BoW models in terms of meaningful categorization of social media content.

Concerning the automated content analysis of visual data, Peng et al. (2023) describe four different methods, along with their strengths and weaknesses: (a) Commercial APIs, (b) open-source models and commercial libraries, (c) customized supervised machine learning, and (d) customized unsupervised machine learning. They explain how these methods vary in purpose, flexibility, technical expertise, resources, replication, and ethics. Commercial APIs and open-source libraries could perform a predetermined selection of tasks that are part of existing computer vision tools that link visual data to relevant concepts of interest (e.g., face, object, or text detection). Supervised learning methods are more flexible because they train models to predict specific visual attributes of interest (e.g., fitness images). Unsupervised machine learning methods could be implemented if researchers want to explore potential visual categories, topics, or themes in their digital trace data without predefined categories or attributes.

The quality of automated textual and visual content analysis methods depends on the dataset on which the methods are trained. Textual and visual training data can be biased because they often originate from developed countries and inherit structural and social biases (Kroon et al., 2023; Peng et al., 2023). As such, all language and visual models are sensitive to learning stereotypical associations and capture racial, political, and gender biases. Both Kroon et al. and Peng et al., therefore, warn that automated textual and visual content analysis methods are prone to social and cultural biases. Peng et al. (2023) mention that computer vision programs contain biases that can lead to inaccurate predictions of minority groups. Therefore, researchers who implement transfer learning to measure social media effects must critically judge the quality and validity of the dataset that they use to fine-tune their model in terms of diversity and representativeness of the target population (Kroon et al., 2023). In addition, researchers should be aware that their findings could potentially only be generalized to a limited sample.

Step 3. Linkage analysis

The third challenge relates to the broader research design being implemented – and how this design connects the data collection methods and analyses discussed so far. Answering social media effects questions will often require the combination of self-reports – measuring, for example, the dependent variables of interest (e.g., attitudes, perceptions, intentions, offline behavior) and potentially individual-level predictors (e.g., socio-demographics) – with social media digital trace data collected and analyzed with the methods discussed above. In addition, this challenge also relates to how the digital trace data will be linked to the self-reports and which analytical techniques will be used to answer the research question.

Combining self-reports with digital trace data via linkage analysis opens several avenues for research on social media effects (Otto et al., 2023). For example, using user-centric digital trace data collection methods measuring content and exposure longitudinally, social media content and exposure could easily be linked to ESM or longitudinal survey data to examine within-person and person-specific changes in cognitions, emotions, and behavior. Considering the granularity of digital trace data, researchers can focus on almost immediate social media effects across seconds or hours by linking this exposure to self-reports measured with ESM data. Alternatively, they can focus on long-term effects by aggregating this exposure and linking it to longitudinal panel surveys. In addition, digital trace data collection methods also enable the examination of bi-directional associations, like reinforcing spirals (Slater, 2007).

Although these opportunities match our advanced and more detailed thinking of linkage analyses, Otto et al. (2023) also mention some challenges of so-called modern linkage designs. For example, while it is possible to measure social media content and exposure almost continuously using digital trace methods, self-reported surveys often measure only a snapshot of the outcome of interest because they are often part of cross-sectional designs or longitudinal designs with a large timespan between waves (e.g., months or years). Even ESM studies cannot continuously measure outcomes like well-being throughout the day, as the number of questionnaires that can be administered per day is limited. Another challenge they mention is that linkage analyses often require advanced data pre-processing, which requires transparency, given that pre-processing decisions may affect the outcome of a linkage analysis. Preregistration is therefore recommended.

Ethical and privacy challenges and potential solutions

The articles in this special issue show that collecting and analyzing social media content poses three new ethical and privacy challenges. The first challenge is that meaningful consent is warranted for collecting digital trace data (Ohme et al., 2023) and linking digital traces with surveys (Otto et al., 2023). Individuals must be well-informed about the detail and magnitude of the data they share (Otto et al., 2023). Informed consent directly provided to researchers may only be possible in user-centric digital trace data collection methods – as these are the methods in which researchers have direct access to participants' timelines (Ohme et al., 2023). Even so, some data collection methods (e.g., data donation) would allow participants to provide consent about the specific existing content they would share with researchers because they could know this content at the beginning of the study. However, other data collection methods (e.g., screen tracking) may need to rely on broader informed consent because participants and researchers do not know at the beginning of the study what type of content participants will share or will be exposed to.

A second privacy challenge of collecting and analyzing social media content is that digital traces of such content often contain privacy sensitive and personal information, such as profile information (e.g., names, birthdate, contact information), status updates (e.g., relationship status, political and religious beliefs), location data, personal interests, and posts from families and friends. According to the GDPR, the European privacy law, researchers should try to minimize the amount and detail of collected, stored, and shared data (Otto et al., 2023) and be mindful of different levels of privacy risks

posed by the data collection methods (Ohme et al., 2023). Researchers should, therefore, critically reflect upon their use of digital trace data collection methods, which data is required for answering their research question, and carefully consider how data minimization and anonymization will be performed. These considerations are crucial and need to be established before the data collection is considered – already envisioning the complete research design for the study, including the specific linkage analysis that will be performed.

A third and related privacy challenge, mentioned by Peng et al. (2023), is that third parties may get unwanted access to highly sensitive social media data (e.g., profile pictures) by using specific visual and textual data-analysis methods. Some APIs keep the uploaded data to improve their algorithms. As such, researchers may violate privacy protection rules if they use these algorithms without informing the participants that all their pictures will be uploaded to the database. Researchers should, therefore, carefully review the terms and services of the APIs they use to analyze (visual) data. Preferably, researchers should use methods that do not store the participants' data.

Avenues for future research

Open science

The articles in this special issue led to several avenues for future research. The first avenue for future research is to be open and transparent about the representativeness of the samples, data collection methods, and data analysis techniques. Such openness is essential because participants may not want to participate in digital trace data collection studies due to privacy concerns (Ohme et al., 2023), and participants may drop out due to technical errors or their inability to access and donate their data. As such, research samples may be biased, limiting the generalizability of studies that use social media digital trace data methods. Therefore, researchers should provide information on their sample's representativeness (Ohme et al., 2023; Otto et al., 2023).

In addition, researchers should be open and transparent about the data collection and analysis techniques they have used to enable other researchers to replicate their findings and judge the quality of their methods. Although open methods are essential, computational data collection methods and textual and visual content analysis methods are not always transparent. For example, it is not always clear what data are included in a data donation package, and the structure of data donation packages may change over time (Otto et al., 2023). Furthermore, many computer vision and API methods are developed by commercial entities that do not provide insight into training datasets, algorithms, and procedures to researchers (Peng et al., 2023). In addition, given that many digital trace data of social media use contain privacy-sensitive information, researchers cannot share all raw data after publication. Even when they anonymize their data, participants could be identified based on a unique combination of variables (Otto et al., 2023). Therefore, new guidelines should be developed that allow researchers to make data and computational algorithms available for reproducibility while guaranteeing participants' privacy (Ohme et al., 2023; Otto et al., 2023; Peng et al., 2023).

Interdisciplinary collaborations

The second avenue for future research is the development of interdisciplinary collaborations. As the implementation of digital trace methods is costly in terms of resources like time, money, skills, and expertise, the articles in this special issue highlight the importance of interdisciplinary collaborations. For example, Ohme et al. (2023) promote the development of interdisciplinary digital trace data consortiums. These consortiums may contribute to openly available high-quality data sources. Furthermore, they may increase synchronization between researchers regarding methodological standards and data quality criteria.

Although interdisciplinary collaborations could move the field forward more quickly and forcefully, communication scientists should be aware that many engineering tasks within the field of

communication science are not interesting from a computer science perspective. Communication scientists are often not looking for scientific collaboration but rather for an engineer or programmer to do some work for them. These programmers often have limited job prospects in universities. If communication scholars want to obtain long-lasting collaborations with research engineers and computational scientists, they should try to make such collaborations more attractive for them by actively contributing to their field. This could, for example, be accomplished by identifying the types of research questions they are most interested in answering and publishing in statistical software journals.

Finally, researchers from different disciplines differ in background knowledge. For example, scholars may have difficulties understanding each other as they use their own terminology. To have fruitful collaborations with computational scientists, Ohme et al. indicated that it is essential that we educate a greater number of communication scholars and students- in computational language, analyses, and methods.

Refining communication methods and measures

A third and final avenue for future research is redefining our communication methods and measures, thereby refining how communication theories are applied to understand media effects in today's environment. Until recently, most studies have focused on quantitative measures of social media use. The findings of these studies are often non-significant or inconsistent. An important step toward content-based social media research is to refine our measures and methods by considering the content of individuals' social media use. To improve our investigation of content-based social media effects, Peng et al. (2023) suggest that developing benchmark datasets that reflect theoretically meaningful categories related to communication domains would be valuable. Communication theories should give us an idea of the categories that should be included in such benchmarks.

Last but not least, theories and models may need to better address the diversity of individuals' social media use. Kroon et al. (2023) mentioned that social media diets are highly diverse in terms of format and style, language, and modality (e.g., text/audio/vision). Furthermore, Otto et al. (2023) highlighted the personalized, fragmented, and fast-lived social media environment, which challenges traditional linkage analysis. Individuals differ in the social media content they produce, select, or receive and how they are affected by this content (Otto et al., 2023). Furthermore, individuals exposed to similar content may respond differently (Valkenburg, 2022). Computational methods enable the collection and analyses of fine-grained diverse and fragmented digital trace data of individuals, which fosters person-specific content-based media effects analyses and theory building.

Conclusion

This special issue presents a three-step approach to studying content-based social effects using computation methods by (1) collecting digital trace data, (2) performing automated textual and visual content analysis, and (3) performing linkage analysis. Several challenges must be addressed in future research to foster ethically responsible, open, interdisciplinary, and theory-driven content-based social media effects research. We hope that this special issue helps researchers weigh the pros and cons of different computational methods to better understand how our social media use affects our cognitions, emotions, and behaviors.

Acknowledgments

A KNAW early career partnership grant awarded to J.L. Pouwels by the Royal Netherlands Academy of Arts and Sciences was used to organize an interdisciplinary expert meeting on the special issue topic. The special issue builds on insights that were obtained during this meeting.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the Koninklijke Nederlandse Akademie van Wetenschappen [Early career partnership grant].

Notes on contributors

J. Loes Pouwels is an assistant professor at the Behavioral Science Institute at Radboud University, Nijmegen, The Netherlands. Her research interests include adolescents, social media use, psychosocial functioning, and peer relations.

Theo Araujo is Scientific Director of the Amsterdam School of Communication Research (ASCoR) and Associate Professor in the Department of Communication Science at University of Amsterdam. His research focuses on the increasing adoption of artificial intelligence and related technologies within our communication environment, including conversational agents and automated decision-making.

Wouter van Atteveldt is professor of Computational Communication Science and Political Communication at the Vrije Universiteit Amsterdam. He focuses on automatic analysis of (political) communication, including both traditional and social media, and the methods and data required for studying this.

Marko Bachl (Ph.D., University of Hohenheim) is an assistant professor in the Institute for Media and Communication Studies at Freie Universität Berlin. He studies digital research methods, political communication, and health communication. He was an associate editor at *Communication Methods and Measures* from 2021 to 2023.

Patti M. Valkenburg is a University Distinguished Professor at the University of Amsterdam. Her research focuses on the cognitive, emotional, and social effects of (social) media on youth and adults. She is particularly interested in theorizing and studying how adolescents and adults differ in their susceptibility to the effects of (social) media.

ORCID

J. Loes Pouwels  <http://orcid.org/0000-0002-9586-392X>

Theo Araujo  <http://orcid.org/0000-0002-4633-9339>

Wouter van Atteveldt  <http://orcid.org/0000-0003-1237-538X>

Marko Bachl  <http://orcid.org/0000-0001-5852-4948>

Patti M. Valkenburg  <http://orcid.org/0000-0003-0477-8429>

References

- Bayer, J. B., Triệu, P., & Ellison, N. B. (2020). Social media elements, ecologies, and effects. *Annual Review of Psychology*, 71(1), 471–497. <https://doi.org/10.1146/annurev-psych-010419-050944>
- Choukas-Bradley, S., Kilic, Z., Stout, C. D., & Roberts, S. R. (2023). Perfect storms and double-edged swords: Recent advances in research on adolescent social media use and mental health. *Advances in Psychiatry and Behavioral Health*, 3(1), 149–157. <https://doi.org/10.1016/j.ypsc.2023.03.007>
- De Vreese, C. H., Boukes, M., Schuck, A., Vliegthart, R., Bos, L., & Lelkes, Y. (2017). Linking survey and media content data: Opportunities, considerations, and pitfalls. *Communication Methods and Measures*, 11(4), 221–244. <https://doi.org/10.1080/19312458.2017.1380175>
- European Commission. (2023). *The Digital Services Act Package*. <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package>
- Howison, J., Wiggins, A., & Crowston, K. (2011). Validity issues in the use of social network analysis with digital trace data. *Journal of the Association for Information Systems*, 12(12), 767–797. <https://doi.org/10.17705/1jais.00282>
- The Kids Online Safety Act, S.3663, 118th Congress. (2022). <https://www.blackburn.senate.gov/services/files/D89FC49B-0714-4124-B8B1-4F35A85F5E02>.
- Kleinnijenhuis, J., van Hoof, A. M. J., & van Atteveldt, W. (2019). The combined effects of mass media and social media on political perceptions and preferences. *Journal of Communication*, 69(6), 650–673. <https://doi.org/10.1093/joc/jqz038>

- Kroon, A. C., Welbers, K., Trilling, D., & van Atteveldt, W. (2023). Using transfer-learning for measuring media effects: Challenges in automated analysis of multiple formats, languages, and modalities. *Communication Methods and Measures*. <https://doi.org/10.1080/19312458.2023.2167197>
- Kubin, E., & von Sikorski, C. (2021). The role of (social) media in political polarization: A systematic review. *Annals of the International Communication Association*, 45(3), 188–206. <https://doi.org/10.1080/23808985.2021.1976070>
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009). Computational Social Science: Obstacles and opportunities. *Science*, 323(5915), 721–723. <https://doi.org/10.1126/science.1167742>
- Ohme, J., Araujo, T., Boeschoten, L., Freelon, D., Ram, N., Reeves, B. B., & Robinson, T. N. (2023). Digital trace data collection for social media effects research: APIs, data donation, and (screen) tracking. *Communication Methods and Measures*, 1–18. <https://doi.org/10.1080/19312458.2023.2181319>
- Otto, L. P., Loecherbach, F., & Vliegenthart, R. (2023). Linkage analysis revised – linking digital traces and survey data. *Communication Methods and Measures*, 1–19. <https://doi.org/10.1080/19312458.2023.2257595>
- Parry, D. A., Davidson, B. I., Sewall, C. J. R., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour*, 5(11), 1535–1547. <https://doi.org/10.1038/s41562-021-01117-5>
- Peng, Y., Lock, I., & Sallah, A. A. (2023). Automated visual analysis for the study of social media effects: Opportunities, approaches, and challenges. *Communication Methods and Measures*. <https://doi.org/10.1080/19312458.2023.2167197>
- Pew Research Center. (2022). *Teens, social media and technology 2022*. Pew Research Center. <https://www.pewresearch.org/internet/2022/08/10/teens-social-media-and-technology-2022/>
- Pouwels, J. L., Valkenburg, P. M., Beyens, I., van Driel, I. I., & Keijsers, L. (2021). Some socially poor but also some socially rich adolescents feel closer to their friends after using social media. *Scientific Reports*, 11(1), 21176. <https://doi.org/10.1038/s41598-021-99034-0>. Article 21176.
- Scharkow, M., & Bachl, M. (2017). How measurement error in content analysis and self-reported media use leads to minimal media effect findings in linkage analyses: A simulation study. *Political Communication*, 34(3), 323–343. <https://doi.org/10.1080/10584609.2016.1235640>
- Slater, M. D. (2007). Reinforcing spirals: The mutual influence of media selectivity and media effects and their impact on individual behavior and social identity. *Communication Theory*, 17(3), 281–303. <https://doi.org/10.1111/j.1468-2885.2007.00296.x>
- Statista. (2022). *Daily Time Spent on Social Networking by Internet Users Worldwide from 2012 to 2022 (In Minutes)*. Statista. <https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/#:~:text=How%20much%20time%20do%20people,minutes%20in%20the%20previous%20year.>
- Valkenburg, P. M. (2017). Understanding self-effects in social media. *Human Communication Research*, 43(4), 477–490. <https://doi.org/10.1111/hcre.12113>
- Valkenburg, P. M. (2022). Social media use and well-being: What we know and what we need to know. *Current Opinion in Psychology*, 45, 101294. <https://doi.org/10.1016/j.copsyc.2021.12.006>
- Valkenburg, P. M., Beyens, I., Meier, A., & Vanden Abeele, M. M. P. (2022). Advancing our understanding of the associations between social media use and well-being. *Current Opinion in Psychology* 47, 101357. Article 101357. <https://doi.org/10.1016/j.copsyc.2022.101357>
- Valkenburg, P. M., Beyens, I., Pouwels, J. L., van Driel, I. I., & Keijsers, L. (2021). Social media and adolescents' self-esteem: Heading for a person-specific media effects paradigm. *Journal of Communication*, 71(1), 56–78. <https://doi.org/10.1093/joc/jqaa039>
- Valkenburg, P. M., Peter, J., & Walther, J. B. (2016). Media effects: Theory and research. *Annual Review of Psychology*, 67(1), 315–338. <https://doi.org/10.1146/annurev-psych-122414-033608>
- van Atteveldt, W., & Peng, T.-Q. (2018). When communication meets computation: Opportunities, challenges, and pitfalls in computational communication science. *Communication Methods and Measures*, 12(2–3), 81–92. <https://doi.org/10.1080/19312458.2018.1458084>
- van Atteveldt, W., van der Velden, M. A. C. G., & Boukes, M. (2021). The validity of sentiment analysis: Comparing manual annotation, crowd-coding, dictionary approaches, and machine learning algorithms. *Communication Methods and Measures*, 15(2), 121–140. <https://doi.org/10.1080/19312458.2020.1869198>