Challenges of comparative research on hate speech in media user comments: Comparing countries, platforms, and target groups

Schemer, Christian; Reiners, Liane

Erstveröffentlichung / Primary Publication
Sammelwerksbeitrag / collection article

Empfohlene Zitierung / Suggested Citation:
Schemer, C., & Reiners, L. (2023). Challenges of comparative research on hate speech in media user comments: Comparing countries, platforms, and target groups. In C. Strippel, S. Paasch-Colberg, M. Emmer, & J. Trebbe (Eds.), *Challenges and perspectives of hate speech research* (pp. 127-139). Berlin https://doi.org/10.48541/dcr.v12.8

Digital
Communication
Research.de

Recommended citation: Schemer, C., & Reiners, L. (2023). Challenges of comparative research on hate speech in media user comments: Comparing countries, platforms, and target groups. In C. Strippel, S. Paasch-Colberg, M. Emmer, & J. Trebbe (Eds.), *Challenges and perspectives of hate speech research* (pp. 127–139). Digital Communication Research. https://doi.org/10.48541/dcr.v12.8

Abstract: Hate speech is a phenomenon studied in numerous disciplines by many researchers. This research has produced a variety of findings, e.g., with regard to the prevalence of hate, common targets or differences between platforms or countries. However, previous research also comes with conceptual and methodological challenges, e.g., definitions or operationalizations of hate speech in empirical studies. The present chapter focuses on the issue of equivalence in previous hate speech research—a well-known problem of comparative research in general. To compare research findings relating to hate speech across different contexts scholars need to consider the equivalence with respect to definitions, methods, measurements, procedures, and also the sampling communication content. We provide an overview about potential pitfalls and biases that can be due to a lack of equivalence and point to strategies on how to address them.

DOI 10.48541/dcr.v12.8

Christian Schemer & Liane Reiners

Challenges of Comparative Research on Hate Speech in Media User Comments

Comparing countries, platforms, and target groups

1 Introduction

A vast body of research on hate speech in user comments is dispersed across disciplines, such as communication, political science, computer linguistics, and linguistics. From a comparative perspective, one major challenge exists: It is difficult to compare and make sense of results from different studies because they differ in terms of their definitions, sampling strategies and units, and measurements of hate speech (for a recent overview of comparative studies, see Pamungkas et al., 2021 and also Fortuna et al. in this volume). Therefore, it is often difficult to argue that hate speech prevalence is higher in one country compared to another one. This problem also arises when researchers compare platforms, the comment sections of different news outlets, and so on. A downstream consequence of biased estimates of hate speech is also that the prediction of hate speech across different contexts cannot be compared. For comparative researchers, this is a well-known problem.

Basically, a comparison or summary of results across contexts requires assumptions that relate to the equivalence of definitions, methods, and procedures

that are used in empirical research (for a more in-depth look at comparative research methodology, see Rössler, 2012; van de Vijver & Leung, 1997; Wirth & Kolb, 2012, 2014). This also holds true for single studies that annotate hate speech in user comments in different contexts, including platforms (Olteanu et al., 2018), media outlets (Paasch-Colberg et al., 2021; Zannettou et al., 2020), countries (Hanzelka & Schmidt, 2021; Ruiz et al., 2011), and targets or authors of hate speech (ElSherief et al., 2018). The problems related to the analysis of contexts, such as platforms or media outlets that host user comments, are often not easier to solve than those linked to cross-cultural analysis.

Most studies on hate speech are not explicitly comparative in nature, but may nevertheless be plagued by equivalence issues. This chapter aims to raise awareness among researchers of these methodological issues to encourage research that can be used for comparative purposes. To this end, this chapter emphasizes the role of equivalence at different levels and responds to some equivalence issues that occur in the first part of this edited volume. It demonstrates what the equivalence of definitions of key concepts, sampling, and measurements means and how violations of equivalence can bias the comparison of findings across contexts.

2 Equivalence of definitions, measurements, and procedures

Comparisons across contexts, such as actors, platforms, or cultures, require that a construct of interest, such as hate speech, can be considered as a single unitary construct that is manifest (i.e., located and observable) in user comments. If we start from an etic position and the existence of a universal phenomenon called "hate speech," which we can describe as a theoretical concept, the crucial question relates to whether this phenomenon can be assessed with measures that are specific to a context or not (Triandis & Marin, 1983). If we assume that manifestations of hate speech differ across contexts (e.g., users rely on different ethnic slurs for social groups in different contexts), an emic measurement strategy is required. Research that aims at comparisons of hate speech across such contexts would need to argue that different ethnic slurs of social groups are functional equivalents (for a detailed discussion of these issues, see Wirth & Kolb, 2012). Without this assumption we cannot know whether a particular group is more often the target of hate

speech, whether a particular event elicited more hate than another, or whether hate speech is more prevalent in some countries than others.

If we consider previous definitions of hate speech (see, for an overview, Paasch-Colberg et al., 2021; Reiners & Schemer, 2020; Siegel, 2020), it becomes clear that researchers frequently start with different conceptions of the construct of interest. This is sometimes guided by pragmatic considerations (e.g., the processing of large quantities of user-generated content). Additionally, ideographic aspects of an event or a culture motivate how researchers approach hate speech (e.g., user-generated content after Islamist terror attacks). The issue of the *equivalence of definitions* is complicated by the use of different labels when talking about hate speech. This can vary from abusive language to verbal aggression, toxic or dangerous speech, extremism, and many more (e.g., Schmidt & Wiegand, 2017; Siegel, 2020; see also the "Theoretical Perspectives" section in this volume).

Some researchers also include an effects dimension of hate speech (i.e., speech that incites hate or violence; see, e.g., Gagliardone et al., 2015). This complicates the assessment of hate speech even further because research then has to specify not only the content that is typical of hate speech but also the effects on users that may be difficult to observe. Thus, if definitions of central theoretical concepts differ across studies, then comparisons across these studies or contexts become difficult to interpret at best and meaningless at worst (Rössler, 2012). Therefore, a basic requirement of comparisons across contexts is that at least the functional equivalence of measures of hate speech exists.

Narrow theoretical conceptions of hate speech can simplify the task of achieving the *equivalence of measurements*. However, they are likely to underestimate the amount of hate that circulates on social media. Broad definitions are likely to result in overestimation. For instance, Silva et al. (2016) assume that hate speech is an expression of a user that describes a negative stance toward a social group (e.g., "I hate [or don't like or other expressions by users] some member of a social group"). This is a narrow conception of hate speech because other expressions, such as explicitly assigning negative attributes to social groups or using ethnic slurs, can frequently occur (Siegel, 2020). This definition also ignores subtle forms of hate speech (Schmidt & Wiegand, 2017). Implicit notions, such as humor or the use of specific metaphors as hate speech devices, are real challenges for equivalence (see Szczepańska & Marchlewska in this volume). Specifically, the authors demonstrate the diversity of slogans that a Polish protest movement uses against

the government, ranging from the outright derogation of the ruling party to subtle and humorous appeals, which are less explicit and negative but are meant to ridicule the governing elite.

Therefore, the amount of hate speech that researchers relying on a narrow definition of the same can find is likely an underestimation (i.e., 20,305 tweets out of 512 million, which is around 0.004 per cent; Silva et al., 2016). Burnap and Williams (2015) started with a broader definition of hate speech as offensive or antagonistic in terms of race, ethnicity, or religion. They found a prevalence of 11 per cent of hateful tweets. The broader definition of hate speech is likely to result in a higher prevalence estimate. In this study, hateful comments may include expressions that other authors would not consider hateful, but rather criticism or disagreement. Another study defines "hateful speech as discourse practiced by communities who self-identify as hateful towards a target group" (Saleem et al., 2016, p. 4). This means that every post in such a community is automatically classified as hate speech (for a similar approach, see Albuquerque & Alves in this volume). In this study, the authors focus on a pro-Bolsonaro network on Brazilian social media, which is labeled the "Office of Hate" and is considered a spreader of hate speech against social groups and established institutions. Although these studies on the structure of notorious hate nets offer important insights, ignoring heterogeneity in their communication is a limitation. Saleem et al. (2016) also acknowledge that some of this communication may be "non-hateful chatter." Thus, not all communication in the "Office of Hate" should be automatically categorized as hate speech if parts of the conversation there do not attack or derogate individuals or social groups.

This discussion on the heterogeneity of definitions and a quick look at operationalizations of hate speech in previous studies demonstrate that extant research is far from achieving functional equivalence, let alone the strict invariance of measures for the detection of hate speech. However, having unequivocal definitions of hate speech would produce truly valuable findings. For instance, research could provide evidence of which platforms, outlets, or sites are more likely to be plagued by hate speech. This can be helpful for practitioners and political authorities to tailor interventions or policies that aim to reduce hate speech. Research would also benefit from unequivocal and comparable definitions. So far, most research is concerned with the detection of hate speech and less so with the prediction of the same. If researchers can agree on definitions

of hate speech, predictive studies also become comparable, and we would learn more about the causes of hate speech at the levels of the technical infrastructure, the authors, and the specific situations and contexts within which this communication emerges.

The equivalence of definitions and (functionally) equivalent measures to assess hate speech in different contexts are a necessary condition of comparisons but not a sufficient one. *Procedural equivalence* is another issue that researchers need to be aware of. For instance, this refers to potential differences in how annotators apply a coding scheme to a given corpus. Ross et al. (2016) demonstrate that even providing detailed guidance for annotators can result in the low reliability of hate speech annotations. If the application of annotation guidelines varies across annotators or cultures, then comparisons across these contexts can be severely biased. There are also practical issues in multicultural studies that can emerge from common language guidelines and the use of translations for annotations in a given language (see Rössler, 2012 for a discussion of such procedures in content analysis). When it comes to translations, researchers need to be aware of instrument bias, which means that translations of measures and guidelines result in different interpretations by annotators or different applications of the instrument for a given corpus. Consequently, the assumption of (functional) equivalence is violated, and comparisons across these contexts are also biased. There are also means to quantify whether measurement invariance truly holds by comparing the reliability of annotations or accounting for differences in reliability when analyzing comparative data. However, in cross-cultural content analytic work, this is more complicated than in survey research (for an overview of this problem, see Rössler, 2012; Wirth & Kolb, 2012).

3 Sampling equivalence

Hate speech is frequently a moving target, and sampling strategies need to account for these dynamics. The comparison of studies is frequently hampered by differences in sampling. Similarly, single studies that compare user-generated content across outlets or platforms encountering issues, such as different publication and registration policies, moderation frequency and style, and many more, can threaten sampling equivalence (e.g., Ruiz et al., 2011). There are at least two

sources of bias that can occur and challenge comparability: first, bias that is due to the researchers' motivation and focus, and second, bias that is due to platform hosts or providers or community managers in comment sections.

Sampling bias due to the focus of a researcher refers to the sampling of user comments that are specifically tied to an event; a specific group; keywords, such as hashtags; or a specific time frame (e.g., Burnap & Williams, 2015; Chaudhry, 2015; see also Harb and Szczepańska & Marchlewska in this volume). For instance, Szczepańska and Marchlewska (this volume) study hate speech in the context of the "All-Poland Women's Strike" against the ruling government. It is unclear how the amount and quality of hate speech found in this context compare to other protests or other targets of hate within Poland. In a similar vein, Harb (this volume) focuses on hate speech by Lebanese journalists targeting the Shia community, among others. However, it is difficult to know how this compares to hate speech by other actors (e.g., ordinary users) or how the findings compare to less exceptional situations.

Prevalence estimates of hate speech based on these selected samples cannot be compared to each other nor to representative samples from platforms, websites, or comments sections without any further assumptions about the data generation process. Siegel et al. (2021) compared a representative sample of random tweets to samples related to Trump and Clinton from the election campaign and found considerable differences between daily occurrences of hate speech that were difficult to predict. Thus, research findings based on samples generated in the context of specific events or related to specific keywords or hashtags cannot be generalized to other contexts or routine communication situations. Other studies demonstrate that moderators and platforms behave differently in times of crises than in routine periods (Mladenović et al., 2020). These differences in moderation behavior are another issue that threatens the generalization of findings based on event-specific samples.

Bias due to providers or hosts of user comments result from different policies of countries, providers, platforms, or outlets that affect the deletion rate of hateful comments. Specifically, some platform providers filter hateful comments before they get published and before researchers can capture them. These policies may be platform-specific or result from legislation that is specific to a country (e.g., the liability of Holocaust denial in different countries; Kennedy et al., 2018). In addition to such interventions, comment moderators or lay community

managers can actively intervene in discussions. The potential interventions by all these actors are likely to reduce the amount of hate that researchers can obtain from comment sections on news websites or networking sites.

However, it is important to consider how actors from different platforms or news sites differ in terms of their intervention strategies. For instance, Facebook, YouTube, and Twitter differ in their policies with regard to dealing with hateful content (for an overview, see Fortuna & Nunes, 2018; Siegel, 2020). To complicate matters even more, the same platform can even differ in its treatment of hate speech attacking specific targets. On Facebook, hateful comments addressing protected groups, such as Muslims, violate community policies, while migrants do not qualify as a protected group (Fortuna & Nunes, 2018). Thus, researchers would consider "Fucking Muslims" and "Fucking Migrants" as instances of hate speech. However, given that Facebook automatically deletes the former, comparisons across such groups will return biased results.

The use of the same platforms for sampling user comments across different countries does not guarantee equivalence either. Comparisons can be biased by different legislations and the populations that use these platforms. For instance, Twitter is more widely used by the populations of the United States or the United Kingdom, but less so in Germany. In this case, not only the populations differ but maybe also the functions of such a service. Algorithmic treatment of user comments may also differ across countries when algorithms are tuned for a specific language but perform poorly in others. Any difference in the prevalence of hate speech on such platforms between countries can be due to different populations using the platform, different intervention policies, algorithms working differently, or true cultural differences. However, it is impossible to disentangle these sources of variance in observational data.

One option for avoiding this problem involves studying comment sections without any moderation or intervention. However, this is difficult to know beforehand despite some platforms having few restrictions (Strippel & Paasch-Colberg, 2020). Another option is to account for differences in moderation practices by observing moderation or checking for differences in moderation policies. However, Ahmad (in this volume) suggests that moderation practices can vary across moderators and within moderators over time. Similarly, researchers can take into account differences in populations that communicate on specific platforms. This informed approach can result in weighting procedures to reduce sampling bias.

If specific sampling strategies are chosen, it is important to discuss the findings against this background (Rössler, 2012). Otherwise, findings from comparative studies are difficult to interpret.

For instance, Ruiz et al. (2011) compared user comments on newspaper websites in five countries. They sampled posts from one single quality newspaper in each country, most of which had a liberal leaning. Obviously, a single outlet with a specific political leaning cannot represent a whole media system or culture. Nevertheless, the authors present their results as if this was the case and as if the cultural context can explain the findings. Specifically, Ruiz et al. (2011, p. 482) state that the "results of this study suggest that the cultural context is relevant to the democratic quality of the debates we analyzed." So, if research only looks at variation across countries without any variation across outlets within a country, inferences with respect to cultural differences are always confounded by differences across outlets. Other research that examined single cases across countries produced similarly problematic findings that are difficult to interpret (e.g., the comparison of the anti-Islam Facebook group Pegida in Germany and initiatives against Islam in the Czech Republic by Hanzelka & Schmidt, 2017). However, avoiding these pitfalls is important to secure sampling equivalence and to draw valid inferences with respect to differences across countries, platforms, sites, or targets of hate speech. At the very least, a thorough discussion on how the sampling strategies may have affected the given findings should be included in any research report (Rössler, 2012).

4 Equivalence of context

Securing equivalence is a prerequisite for comparisons. However, researchers also need to be aware of the broader context in which hate speech occurs. This context can be essential for understanding and interpreting research findings. From the perspective of public discourse in liberal democracy, where the freedom of expression is not an issue, hate speech is easily condemned when it is observed since it can be harmful to substantive debates. However, hate speech or elements of hate speech can also occur in other contexts. There are subcultures and minority groups, for example, that use offensive and sometimes hateful language in a positive sense to build and preserve a common ingroup identity without

devaluing their own or other social groups (see Davidson et al., 2017). The use of the n-word among the people of colored communities is one prominent example. On the other hand, incivility and hate speech are frequently an option for expressing one's opposition to corrupt or authoritarian regimes when offline opposition is impossible or dangerous. In this vein, hate speech is considered as a means of self-defense against oppressive actors (see Szczepańska & Marchlewska in this volume). For instance, according to Szczepańska and Marchlewska, protesters in the "All-Poland Women's Strike" relied on hate speech as a last resort to fight against the abortion policies of the ruling government. In the present chapter, we cannot discuss the legitimacy of hate speech as self-defense. However, it is important to distinguish hate speech that comes from oppressed minorities or from actors that aim at silencing oppositional forces.

Therefore, it is important to consider the sociopolitical context in which hateful speech is embedded (see Litvinenko in this volume). This context also matters for the normative evaluation of hate speech and policies designed to avoid, reduce, or moderate it. These differences in functions of hate speech within and across societies considerably complicate the regulation of the phenomenon at the national and global levels (see Litvinenko and Ilori in this volume). For instance, harsher restrictions to regulate hate speech on social network sites in Western democracies have inspired authoritarian regimes to copy more restrictive policies, but with the goal of banning or censoring any oppositional voices. Thus, as Ilori (this volume) points out, any regulation, be it legal or non-legal (e.g., by exerting social pressure on haters in social networks), needs to balance the civility of political discourse against the freedom of speech.

5 Agenda for future comparative research

Research on hate speech has increased in the past decade and has improved considerably with respect to the methods that are used and breadth of phenomena and outlets that are studied. Making sense of all these studies requires comparing the findings from different studies or the results across contexts within single studies. Otherwise, we end up with idiosyncratic explanations for the emergence and dynamics of hate speech. The present chapter demonstrates how a basic requirement for comparisons is equivalence with respect to definitions,

methods, measurements and procedures, and sampling. Equivalence with respect to context matters for the substantive interpretation of comparisons.

Research reviews in the field raise awareness of some of these issues by discussing problems of narrow versus broad definitions of hate speech (Schmidt & Wiegand, 2017; Siegel, 2020), issues of reliability (Ross et al., 2016), or the generalization of classification algorithms (Fortuna et al., 2021). However, most primary research rarely accounts for the problem that violations of equivalence assumptions invalidate comparisons across studies or across contexts within a given study. Therefore, future research needs to take issues of equivalence and potential bias more seriously. Specifically, reasoning about equivalence should inform the design of a study, the sampling and collection of data, the measurements of hate speech, and, finally, the analysis of data. Ideally, equivalence should be quantified and used in weighting procedures in the analysis of data to account for potential bias. At the very least, researchers need to show awareness of bias due to violations of equivalence and discuss their findings against this backdrop.

Christian Schemer is Professor for Communication at the Department of Communication at Johannes Gutenberg-Universität in Mainz, Germany. https://orcid.org/0000-0002-7808-2240

Liane Reiners is a research assistant at the Department of Communication at Johannes Gutenberg-Universität in Mainz, Germany.

References

Burnap, P., & Williams, M. L. (2015). Cyber hate speech on Twitter: An application of machine classification and statistical modeling for policy and decision making. *Policy & Internet*, *7*(2), 223–242. https://doi.org/10.1002/poi3.85

Chaudhry, I. (2015). #Hashtagging hate: Using Twitter to track racism online. *First Monday*, *20*(2). https://doi.org/10.5210/fm.v20i2.5450

Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. In O. Varol, E. Ferrara, C. A. Davis, F. Menczer, & A. Flammini (Eds.), *Proceedings of the 11th International AAAI Conference on Web and Social Media – ICWSM 2017* (pp. 512–515). AAAI. https://arxiv.org/pdf/1703.04009.pdf

ElSherief, M., Kulkarni, V., Nguyen, D., Wang, W. Y., & Belding, E. (2018). *Hate lingo: A target-based linguistic analysis of hate speech in social media.* https://arxiv.org/abs/1804.04257

Fortuna, P., & Nunes, S. (2018). A survey on automatic detection of hate speech in text. *ACM Computing Surveys*, *51*(4), article 85, 1–30. https://doi.org/10.1145/3232676

Fortuna, P., Soler-Company, J., & Wanner, L. (2021). How well do hate speech, toxicity, abusive and offensive language classification models generalize across datasets? *Information Processing & Management*, *58*(3), 102524. https://doi.org/10.1016/j.ipm.2021.102524

Gagliardone, I., Gal, D., Alves, T., & Martinez, G. (2015). *Countering online hate speech*. UNESCO Publishing.

Hanzelka, J., & Schmidt, I. (2017). Dynamics of cyber hate in social media: A comparative analysis of anti-Muslim movements in the Czech Republic and Germany. *International Journal of Cyber Criminology*, *11*(1), 143–160. https://doi.org/10.5281/zenodo.495778

Kennedy, B., Atari, M., Davani, A. M., Yeh, L., Omrani, A., Kim, Y., Koombs, K., Havaldar, S., Portillo-Wightman, G., Gonzalez, E., Hoover, J., Azatian, A., Hussain, A., Lara, A., Olmos, G., Omary, A., Park, C., Wang, C., Wang, X., Zhang, Y., & Dehghani, M. (2018). *Introducing the Gab Hate Corpus: Defining and applying hate-based rhetoric to social media posts at scale*. https://psyarxiv.com/hqjxn/

Mladenović, M., Ošmjanski, V., & Stanković, S. V. (2020). Cyber-aggression, cyberbullying, and cyber-grooming: A survey and research challenges. *ACM Computing Surveys*, *54*(1), article 1. https://doi.org/10.1145/3424246

Olteanu, A., Castillo, C., Boy, J., & Varshney, K. R. (2018). *The effect of extremist violence on hateful speech online*. https://arxiv.org/abs/1804.05704

Paasch-Colberg, S., Strippel, C., Trebbe, J., & Emmer, M. (2021). From insult to hate speech: Mapping offensive language in German user comments on immigration. *Media and Communication*, *9*(1), 171–180. https://doi.org/10.17645/mac.v9i1.3399

Panmungkas, E. W., Basile, V., & Patti, V. (2021). Towards multidomain and multilingual abusive language detection: A survey. *Personal and Ubiquitous Computing.* Advance online publication. https://doi.org/10.1007/s00779-021-01609-1

Reiners, L., & Schemer, C. (2020). A feature-based approach to assess hate speech in user comments. *Questions de Communication*, *38*, 529–548. https://doi.org/10.4000/questionsdecommunication.24808

Rössler, P. (2014). Comparative content analysis. In F. Esser & T. Hanitzsch (Eds.), *The handbook of comparative communication research* (pp. 459–468). Routledge.

Ross, B., Rist, M., Carbonell, G., Cabrera, B., Kurowsky, N., & Wojatzki, M. (2016). Measuring the reliability of hate speech annotations: The case of the European refugee crisis. *Proceedings of NLP4CMC*, *17*, 6–9. https://arxiv.org/abs/1701.08118

Ruiz, C., Domingo, D., Micó, J. L., Díaz-Noci, J., Meso, K., & Masip, P. (2011). Public sphere 2.0? The democratic qualities of citizen debates in online newspapers. *International Journal of Press/Politics*, *16*(4), 436–487. https://doi.org/10.1177/1940161211415849

Saleem, H. M., Dillon, K. P., Benesch, S., & Ruths, D. (2017). *A web of hate. Tackling hateful speech in online social spaces.* Text analytics for cybersecurity and online safety (TA-COS 2016). https://arxiv.org/pdf/1709.10159

Schmidt, A., & Wiegand, M. (2017). A survey on hate speech detection using natural language processing. *Proceedings of the 5th International Workshop on Natural Language Processing for Social Media*, 1–10. https://doi.org/10.18653/v1/W17-1101

Siegel, A. A. (2020). Online hate speech. In N. Persily & J. A. Tucker (Eds.), *Social media and democracy. The state of the field and prospects for reform* (pp. 56–88). Cambridge University Press.

Siegel, A. A., Nikitin, E., Barberá, P., Sterling, J., Pullen, B., Bonneau, R., … Tucker, J. A. (2021). Trumping hate on Twitter? Online hate speech in the 2016 U.S. Election campaign and its aftermath. *Quarterly Journal of Political Science*, *16*(1), 71–104. https://doi.org/10.1561/100.00019045

Silva, L., Mondal, M., Correa, D., Benevenuto, F., & Weber, I. (2016). Analyzing the targets of hate in online social media. *Proceedings of the 10th International AAAI Conference on Web and Social Media (ICWSM 2016).* https://arxiv.org/abs/1603.07709

Strippel, C., & Paasch-Colberg, S. (2020). Diskursarchitekturen deutscher Nachrichtenseiten [Discourse architectures of German news sites]. In V. Gehrau, A. Waldherr, & A. Scholl (Eds.), *Integration durch Kommunikation (in einer digitalen Gesellschaft): Jahrbuch der Deutschen Gesellschaft für Publizistik- und Kommunikationswissenschaft 2019* (S. 153–165). DGPuK. https://doi.org/10.21241/ssoar.68129

Triandis, H. C., & Marin, G. (1983). Etic plus emic versus pseudoetic: A test of the basic assumption of contemporary cross-cultural psychology. *Journal of Cross-Cultural Psychology*, *14*, 489–500. https://doi.org/10.1177/0022002183014004007

Van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Sage.

Wirth, W., & Kolb, S. (2012). Securing equivalence: Problems and solutions. In F. Esser & T. Hanitzsch (Eds.), *The handbook of comparative communication research* (pp. 469–485). Routledge.

Zannettou, S., Elsherief, M., Belding, E., Nilizadeh, S., & Stringhini, G. (2020). Measuring and characterizing hate speech on news websites. *12th ACM Conference on Web Science*, 125–134. https://doi.org/10.1145/3394231.3397902