Incivility as a Violation of Communication Norms in Public Online Discussions: Systematization, Perceptions, and Reactions

Inaugural dissertation
to obtain a doctoral degree in philosophy (Dr. phil.)
at the Faculty of Arts and Humanities of
Heinrich Heine University Düsseldorf

submitted by
Marike Bormann

First supervisor:
Prof. Dr. Gerhard Vowe
Heinrich Heine University Düsseldorf

Second supervisor:
Prof. Dr. Nicole Krämer
University of Duisburg-Essen

Düsseldorf, September 2022

D61

This document contains the synopsis of my cumulative dissertation, which was submitted in a slightly different version at the Faculty of Arts and Humanities of Heinrich Heine University Düsseldorf in February 2022. The synopsis provides an overview of the thematic focus of the individual research articles and their scientific connections. In addition to the synopsis, the dissertation consists of the following five research articles:

Article I:	Bormann, M., Tranow, U., Vowe, G., & Ziegele, M. (2022). Incivility as a Violation of Communication Norms – A Typology Based on Normative Expectations toward Political Communication. *Communication Theory, 32*(3), 332–362. https://doi.org/10.1093/ct/qtab018
Article II:	Bormann, M., & Ziegele, M. (in press). Incivility. In C. Strippel, S. Paasch-Colberg, M. Emmer, & J. Trebbe (Eds.), *Challenges and Perspectives of Hate Speech Analysis. An Interdisciplinary Anthology*. Digital Communication Research.
Article III:	Bormann, M. (2022). Perceptions and Evaluations of Incivility in Public Online Discussions – Insights from Focus Groups with Different Online Actors. *Frontiers in Political Science*, *4*(812145). https://doi.org/10.3389/fpos.2022.812145
Article IV:	Bormann, M., Heinbach, D., & Kluck, J. P. (2022). *Perceptions of and Reactions to Different Types of Incivility in Public Online Discussions – Results of an Online Experiment*. Manuscript submitted for publication.
Article V:	Ziegele, M., Jost, P., Bormann, M., & Heinbach, D. (2018). Journalistic Counter-Voices in Comment Sections: Patterns, Determinants, and Potential Consequences of Interactive Moderation of Uncivil User Comments. *SCM - Studies in Communication and Media*, *7*(4), 525–54. https://doi.org/10.5771/2192-4007-2018-4-525

Article II and *Article IV* are currently under review or in press and are included in full length in the Appendix.

This dissertation was supported by the Digital Society research program funded by the Ministry of Culture and Science of the German State of North Rhine-Westphalia.

Danksagung

Ohne die Unterstützung von vielen besonderen Menschen aus meinem beruflichen und privaten Umfeld wäre die Entstehung dieser Dissertation nicht möglich gewesen. Ihnen und euch möchte ich herzlich dafür danken.

An erster Stelle gilt mein Dank meinem Erstbetreuer Prof. Dr. Gerhard Vowe. Vielen Dank für die Möglichkeit, im Rahmen des Tandemprojekts kumulativ zu promovieren, und für die fachliche Begleitung, unermüdliche Motivation und großartige Unterstützung im gesamten Entstehungsprozess dieser Arbeit. Prof. Dr. Nicole Krämer danke ich für die Zweitbetreuung, für inspirierende Diskussionen und zahlreiche hilfreiche Ratschläge zu meiner Dissertation und wissenschaftlichen Karriere. Ein weiterer, besonderer Dank gilt meinem Mentor und „Zweitchef" Prof. Dr. Marc Ziegele. Von seiner fachlichen Expertise, der Aufnahme in die Nachwuchsforschungsgruppe DEDIS und der Zusammenarbeit haben meine Dissertation und wissenschaftliche Entwicklung maßgeblich profitiert.

Bei den Mitgliedern des Graduiertenkollegs „Digitale Gesellschaft NRW" bedanke ich mich für produktive und lehrreiche Research Retreats, Workshops und Summer Schools. Besonders meinem Tandempartner Jan Philipp Kluck danke ich für die hervorragende Zusammenarbeit und „zivile Kommunikation" in allen Phasen des Projekts. Bei den Ko-Autorinnen und Ko-Autoren bedanke ich mich für die allzeit konstruktive gemeinsame Arbeit an den verschiedenen Aufsätzen meines Kumulus.

Den Kolleginnen und Kollegen der Abteilung Kommunikations- und Medienwissenschaft an der Heinrich-Heine-Universität Düsseldorf danke ich herzlich für eine angenehme Arbeitsatmosphäre und wertvolle Gespräche zu Dissertation und Disputation. Ein ganz besonderer Dank gilt meinen Teamkolleginnen und Freundinnen Inga Brentel, Katharina Frehmann, Dominique Heinbach, Anke Stoll und Lena Wilms. Ihr habt mich in allen Höhen und Tiefen der Entstehung dieser Arbeit unterstützt und empowert, mit mir über Inzivilitätsfragen diskutiert und jederzeit mit Rat und Tat zur Seite gestanden. Vielen Dank auch für gemeinsame Konferenzen all over Europe und zahlreiche unvergessliche Events außerhalb des Arbeitskontexts.

Schließlich geht ein riesiges Dankeschön an meine Familie und meine Freundinnen und Freunde. Danke, dass ihr für die nötige Ablenkung neben der Dissertation gesorgt habt und für euer Verständnis, wenn ich mal wieder einige Wochen wegen „Diss-Tunnel" nicht erreichbar war. Mein wohl größter Dank gilt meinem Mann Julian – ohne deine bedingungslose Unterstützung, Geduld und Motivation wäre diese Dissertation schlicht nicht möglich gewesen.

Abstract

Scholars, politicians, journalists, and the general public are worried about an increase of so-called *incivility* in public online discussions. In recent years, a growing body of online incivility research has emerged. However, three main shortcomings can be identified: First, scholars usually conceptualize incivility as a violation of norms, while approaching different norms. Thus, research lacks a unified systematization of incivility, which makes it difficult to measure its prevalence, causes, and effects in a reliable and valid manner. Second, scholars largely agree that incivility is a perceptual construct but most studies conceptualize incivility based on approaches that prescribe norms, and studies on incivility perceptions of participants involved in public online discussions are scarce. Finally, it is largely unclear how different participants in public online discussions react to distinct types of incivility. Therefore, this dissertation aims at (1) providing a theoretically well-founded systematization of incivility in public online discussions, (2) empirically examining incivility perceptions of participants involved in public online discussions, thereby refining and validating the systematization of incivility, and (3) investigating participants' reactions to incivility. Drawing on analytical theories on cooperation, communication, and norms, a new theoretical framework of incivility in public online discussions was developed that is based on five communication norms and the disapproval of the participants involved in online discussions (*Article I*, *Article II*). Afterwards, five heterogeneous focus groups with three types of participants in public online discussions were conducted, namely lay participants (i.e., ordinary users), semi-professional participants (i.e., online activists collectively combating incivility), and professional participants (i.e., community managers), and they discussed what they perceive as (mildly and severely) uncivil. The results suggest that incivility encompasses violations of all five communication norms, that different types of norm violations are not assessed as equally severe, and that several criteria shape the processing of norm violations (*Article III*). To empirically test and validate the systematization of incivility, an online experiment was conducted. Lay participants were confronted with norm violations in a mock-up online discussion forum. The results validate the concept of incivility though revealing varying severity levels among different types of incivility, and indicate that distinct types of incivility elicit different responses (*Article IV*). Lastly, within a quantitative content analysis of the comments on the Facebook sites of 15 German news outlets, professional participants' reactions to incivility were examined. The results showed that professional participants react more often to one specific type of incivility, but that their response styles do not differ between different types of incivility (*Article V*).

Table of Contents

List of Figures and Tables

1. Introduction

1.1 Background and Research Question

Democracies thrive on the political participation of their citizens. At the very heart of political participation is often seen public debate, in which citizens discuss various political issues, freely express their opinions and learn about other views (e.g., Gastil, 2008, p. 8; Habermas, 1996, ch. 8; Kim et al., 1999, pp. 361-362). In this sense, the emergence of online discussions on social media, on news websites and on other online platforms has raised several democratic hopes. Public online discussions are principally accessible for everyone, can connect citizens of different places, cultures, milieus and political views, and enable exchange between journalists, politicians, and citizens (e.g., Rowe, 2015, pp. 121-122; Ruiz et al., 2011, p. 466). Indeed, online media have contributed to (1) simplifying participation in public debate, (2) disrupting the traditional communicator and recipient role, and (3) enabling ordinary citizens to comment on journalistic content and political debates with potentially high reach and impact (Neuberger, 2017, p. 102).

Over the years, online discussions have become popular elements of political participation. In different European countries, between 14% and 28% of the online users write comments on news websites at least once a week (Newman et al., 2016, p. 99). German online users write comments less frequently, 10% post own comments on a weekly basis while 30% comment at least once a month (Ziegele, 2019, p. 3; Ziegele et al., 2017). In the U.S., 24% of the citizens write comments on news sites at least once a week (Newman et al., 2017, p. 44). Reading comments is more widespread, with 40% of the German online users reading comments at least once a week as do half of the U.S. citizens (Ziegele, 2019, p. 3).

Besides the potential of the Internet and online discussions in particular as well as their popularity, the changed communication conditions unquestionably entail some negative consequences (Kümpel & Rieger, 2019, p. 5). For a fruitful online discussion to take place, several conditions must be met. One of the most important ones is civility. Scholars largely agree that civility "has always been considered a requirement for democratic discourse" (Papacharissi, 2004, p. 260), pertains to the "fundamental tone and practice of democracy" (Herbst, 2010, p. 3), and that "a free flowing, but mainly civil, online discourse is crucial to the public deliberation necessary in a vigorous democracy" (Chen, 2017, p. 5). Recently, however, scholars, politicians, journalists and the general public have expressed concerns about a decline of civility and an increase of *incivility* (Boatright, 2019, pp. 1-2; Muddiman, 2019, p. 31). Such concerns are also reflected in current survey data: 93% of the U.S. citizens perceive incivility in public discourse to be a serious problem and 74% believe it is getting worse compared to previous years (Weber Shandwick et al., 2019, pp. 2-3). Moreover, the vast majority reported to have experienced online incivility and 80% assume that it poses dangerously high risks to society (Weber Shandwick et al., 2019, p. 3). Surveys among German online users demonstrate a similar development: While 65% of the online users reported personal encounters with hateful comments in 2016, the number raised to 76% in 2021 (LfM, 2021, p. 2). Participating in online discussions, however, not only increases the likelihood of encountering incivility, but also of becoming a victim of such forms of communication (e.g., Costello et al., 2017; Ybarra et al., 2006). For those directly affected, incivility can have serious consequences. Victims of hate comments report emotional stress, fear and anxiety, problems with self-image, and even depression (Geschke et al., 2019, p. 27).

Moreover, the public debate about incivility in online contexts is increasingly shaped by fears of a violent spillover into the analog world. Certain forms of incivility, such as hate speech, calls for violence against specific social groups, and the spread of fake news and conspiracy theories, are assumed to have an effect in real life: In Germany, the rise in violent hate crime against marginalized groups and against politicians is often associated with hate speech and incitement to such crimes in online contexts (e.g., Jansen, 2021; Neuerer, 2022). In the U.S., potential effects of different forms of online incivility have been discussed not only since the violent riots on Capitol Hill (e.g., Silk & Connor, 2021). These developments and presumed effects have led to calls for stronger legal regulation of online platforms. In Germany, the "Netzwerkdurchsetzungsgesetz" (Engl.: Network Enforcement Act) was enacted in response to online incivility, however, critics consider the law to be insufficiently restrictive. And the European Commission has recently proposed the "Digital Service Act," which is supposed to regulate platforms and create a safe digital space (Herwartz, 2022; Schleif & Kettemann, 2021).

In this light, it is hardly surprising that scholarly attention towards the phenomenon of incivility in online discussions has increased in recent years, producing a valuable body of research on the prevalence of incivility across platforms (e.g., Coe et al., 2014; Rowe, 2015), on perceptions of incivility (e.g., Stryker et al., 2016, 2021), on causes and effects of incivility (e.g., Kluck & Krämer, 2021; Gervais, 2015), on automatic detection of incivility (e.g., Stoll et al., 2020; Su et al., 2018), and on interventions against incivility (e.g., Friess et al., 2021; Stroud et al., 2015). Several empirical findings suggest that the concerns mentioned above are valid: Content analyses indicated a share of uncivil comments in online discussions between 22% and 53% (Coe et al., 2014, pp. 667-668; Santana, 2014, p. 27; Su et al., 2018, p. 3690). Experimental research demonstrated that uncivil comments can, for example, lead to decreased openmindedness towards other opinions (Hwang et al., 2018), to negative emotions (Gervais, 2015, 2017), and to attitude polarization (Anderson et al., 2014). On the other hand, some studies revealed positive effects of incivility, such as an increased willingness to engage online in a civic manner (Borah, 2014) and to get politically active (Chen & Lu, 2017). These diverging results already point to a major problem in incivility research. Different studies can only be compared to a limited extent and thus valid statements about the prevalence and effects of online incivility are difficult to make because incivility research is lacking a uniform definition and operationalization of the construct.

One common denominator between different studies can be found, namely that incivility is mostly approached as a violation of norms (e.g., Coe et al., 2014, p. 660; Muddiman, 2017, pp. 3183-3184; Su et al., 2018, p. 3681). Moreover, scholars largely agree that incivility is highly subjective and thus depends on what the participants involved in a discussion perceive as uncivil (e.g., Chen et al., 2019, p. 2; Herbst, 2010, p. 3; Stryker et al., 2016, p. 540; Stryker et al., 2021, p. 2). The results of the few existing studies on incivility perceptions further suggest that incivility is a multidimensional construct that includes violations of several norms (e.g., Muddiman, 2017; Stryker et al., 2021). Such a multidimensional concept of incivility that is theoretically well-founded has not yet been provided. It is still largely unclear what exactly incivility is, what discussion participants perceive as (mildly and severely) uncivil, and how they react to different types of incivility. Only few studies have examined incivility perceptions and most of them have focused on citizens' perceptions of incivility between political elites (e.g., Muddiman, 2017, 2019; Stryker et al., 2016, 2021) instead of what participants involved in an online discussion perceive as uncivil. Moreover, there is little research on reactions to distinct types of incivility by different participants in online discussions. Prior research has usually focused on a particular type of incivility and responses to it (e.g., Kalch & Naab, 2017), or on a particular type of response by a certain group of participants (e.g., Wilhelm et al., 2020).

This dissertation addressed these shortcomings within a research program that resulted in five research articles. The dissertation thus aimed at (1) providing a theoretically well-founded concept of incivility that includes both a new definition and a comprehensive typology of incivility, (2) empirically refining and validating the concept of incivility by examining incivility perceptions of participants involved in online discussions, and (3) analyzing reactions to distinct types of incivility by different discussion participants. The overarching research question is therefore as follows:

How can incivility in public online discussions be systematized, how do communication participants perceive it, and how do they react to it?

1.2 Research Program and Relevance

To answer the overarching research question, a specific research program was conducted within this PhD project. The research program employed methodological triangulation, applying qualitative and quantitative research methods to obtain multi-layered scientific knowledge of the research subject. The research program resulted in five articles that are published, in press, or have been submitted to a journal. Figure 1 provides an overview of the overarching research question of this dissertation in relation to the sub-research questions addressed in the five articles.

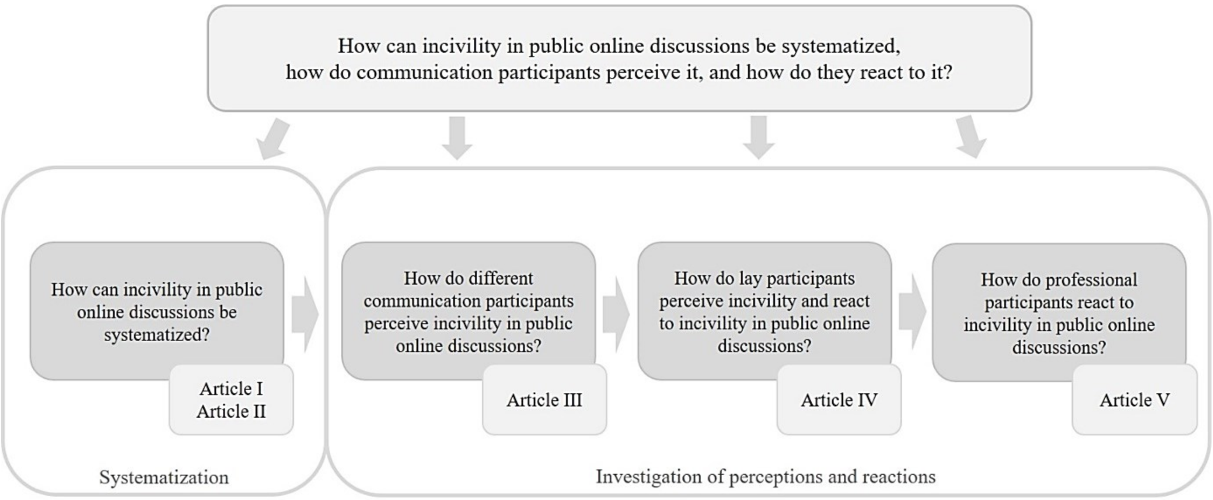
In a first step, we *developed a novel theoretical approach* to the concept of incivility, which is presented in *Article I*. Based on different theories on cooperation, communication and norms (e.g., Grice, 1975; Tomasello, 2009, 2019), we argued that incivility is a disapproved violation of communication norms. Five communication norms build the basis for our new definition and typology of incivility. *Article II* provides a literature review of incivility concepts, contextualizes our own concept within the literature, discusses challenges and perspectives of incivility research, and outlines normative implications.

Afterwards, a *qualitative focus group study* was employed with different participants in public online discussions, namely lay participants (i.e., ordinary users), semi-professional participants (i.e., members of online activist groups), and professional participants (i.e., community managers of online discussion platforms). In five heterogeneous focus groups, these online actors discussed what they perceive as (mildly and severely) uncivil in online discussion, where they agree and differ in their perceptions, and which criteria they apply to evaluate the severity of different types of norm violations. Based on the results, the theoretically developed typology of incivility was complemented and refined. The results are presented in *Article III*.

Following the qualitative study, a *quantitative experimental study* was conducted to empirically validate the concept of incivility. In a fully functional mock-up online discussion forum, lay participants were exposed to different types of norm violations. It was examined how they perceive the distinct norm violations, how they evaluate them in terms of severity, and how they react to them. The results are presented in *Article IV*.

Finally, data of a *quantitative manual content analysis* were used to analyze reactions to online incivility by professional participants. To investigate how community managers engage with different types of incivility, a content analysis of the Facebook sites of 15 different news media outlets was conducted. The results are presented in the final *Article V*.

Figure 1. Overarching Research Question of the Dissertation and Research Questions of the Five Articles.


How can incivility in public online discussions be systematized,
how do communication participants perceive it, and how do they react to it?
How can incivility in public online discussions be systematized?
Article I
Article II
Systematization
How do different communication participants perceive incivility in public online discussions?
Article III
How do lay participants perceive incivility and react to incivility in public online discussions?
Article IV
How do professional participants react to incivility in public online discussions?
Article V
Investigation of perceptions and reactions

Answering the overarching research question is highly relevant from a theoretical, empirical, practical, and socio-political perspective. Incivility research is in dire need of a theoretically well-grounded concept of incivility that integrates previous concepts into a comprehensive framework and considers the perspective of the participants involved in a discussion. By providing a novel theoretical approach to incivility based on theories on cooperation, communication, and norms, this dissertation strengthens the theoretical foundation of incivility research. Further, a methodological shift is pursued by approaching incivility analytically from the perspective of communication participants rather than prescriptively.

For future empirical studies on different aspects of incivility, the doctoral thesis provides an empirically validated typology of incivility that can be used as a research model to measure incivility in different political contexts. Such a uniform model of incivility can contribute to better comparable empirical findings of future studies. Moreover, a uniform multidimensional model helps to ensure that incivility is no longer approached as monolithic but in a much more differentiated way in that distinct types of incivility are treated as such empirically. This can enable, for example, a more nuanced determination of the effects of different forms of uncivil content, and thus lead to a more fine-grain understanding of the consequences of uncivil discourse for democratic life in a digital age. In addition, researchers can more accurately examine the prevalence of incivility and develop explanations based on solid conceptual and operational grounds.

From a practical perspective, insights into what incivility in online discussions is, how different types are evaluated and reacted to, are also highly relevant. Such insights are necessary, for example, to develop and test tailored interventions against uncivil behavior on various online platforms. The results show media companies and platform providers what their users evaluate as most problematic behavior, how they react to which types of deviant communication, and when community managers are expected to intervene. In addition, *Article V* provides insights into what forms of intervention work best for which types of uncivil comments. Lastly, a comprehensive systematization of incivility can be used by platform providers, media companies, and research to train algorithms and develop software to automatically detect uncivil comments in online discussions.

Finally, the overarching research question addressed in this dissertation is also relevant from a political and societal perspective. As mentioned above (see chapter 1.1), there is currently an intense public debate about the potential consequences of online incivility, the limits of freedom of speech, and the legal regulation of online platforms. More scientific knowledge on what incivility actually encompasses and which types are perceived as most harmful can contribute to the regulatory debate and public opinion-forming.

1.3 Structure of Synopsis

The dissertation consists of this synopsis and the five research articles. The structure of this synopsis is aligned with the five articles and embeds them in the respective state of research and theory:

- *Systematization:* The synopsis starts with an overview of different approaches to incivility in political (online) discussions in the extant literature. This is followed by a summary of *Article I* (chapter 2.3.1) and *Article II* (chapter 2.3.2).
- *Investigation of perceptions and reactions:* Given that incivility is approached as a perceptual construct, the next section reviews studies on perceptions of incivility in public online discussions in general and by different participants in particular. Furthermore, research on reactions to online incivility by lay participants and professional participants is discussed. Afterwards, the results of the qualitative study in *Article III* (chapter 3.2.1) and of the experiment in *Article IV* (chapter 3.2.2) are outlined. The chapter 3.2.3 presents the results of the content analysis published in *Article V.*
- *Discussion and conclusion:* Lastly, in chapter 4, the theoretical and empirical results of all individual articles are comprehensively discussed, limitations are outlined, and theoretical and practical implications are presented.

2. Systematization of Incivility

In this chapter, the existing literature on incivility in public online discussions will be reviewed in terms of theoretical approaches to, definitions and operationalizations of incivility. Against this backdrop, our new theoretical approach to as well as the definition and typology of incivility will be outlined (*Article I*, chapter 2.3.1) and contextualized within the literature (*Article II*, chapter 2.3.2). Before delving into theoretical approaches, however, the research subject of this dissertation will be specified more precisely and situated in the wide-ranging field of incivility research.

2.1 Research Subject

Previous research has examined incivility in various environments, including political contexts (e.g., Coe et al., 2014; Jamieson, 2000; Mutz, 2007) on which this dissertation focuses, and non-political contexts such as incivility in workplaces and classrooms (e.g., Schilpzand et al., 2014; Bjorklund & Rehling, 2009). Because of its Latin root "civis" (i.e., citizen) and "civitas" (i.e., citizenship) and associated historical meanings that refer to the civic role, civil society, and the order of the polity (Jamieson et al., 2018, p. 207; Simpson, 1960, p. 109), much of the research on incivility explicitly refers to (communication and debates in) the political public sphere[1]. Incivility has been studied in debates in parliaments (e.g., Jamieson, 2000; Jamieson & Falk, 2000), in talkshows, interviews and news on TV, radio or in print magazines and newspapers (e.g., Ben-Porath, 2010; Mutz, 2007; Sobieraj & Berry, 2011), as well as in political campaigns and advertising (e.g., Brooks & Geer, 2007). Studies analyzed both the prevalence of incivility in journalistic and political content, and in debates between politicians and journalists, as well as its effects on citizens[2].

While earlier studies have examined offline contexts and primarily focused on incivility in interactions between elites, the advent of Web 2.0 technology allowed research to extend the analysis of incivility to interactions between ordinary citizens and between citizens, journalists and politicians. In recent years, research has investigated incivility in online discussions on blogs (e.g., Anderson et al., 2014; Borah, 2013), Usenet news groups (e.g., Papacharissi, 2004), websites of news media (e.g., Coe et al., 2014), and on social networking sites such as Facebook

[1] For an overview on the evolution of the incivility concept, definitions of civility and incivility, and functions of civility and incivility, see also the chapter "The political uses and abuses of civility and incivility" by Jamieson et al. (2018) in *The Oxford Handbook of Political Communication.*

[2] For a meta-analysis of experimental research on the effects of political incivility on citizens, see Van 't Riet and Van Stekelenburg (2021). For a comprehensive review of the research on effects of televised political incivility on citizens, see Mutz (2015).

(e.g., Su et al., 2018), the microblogging platform Twitter (e.g., Oz et al., 2018; Theocharis et al., 2020) or the video sharing platform YouTube (e.g., Yun et al., 2020).

The wide-ranging research across platforms already reveals that the Internet is not a monolithic public sphere or put differently, one online discourse, but a network of diverse communication spaces in which various (semi-)public online discussions take place (Esau et al., 2017, p. 322). When referring to *public online discussions,* this dissertation includes public and semi-public discussions on matters of public interest on Web 2.0 platforms that are visible and accessible to all online users. Semi-public online discussions are also principally visible and accessible to the disperse audience of online users but take place on platforms where prior registration is required to actively participate, that is, to contribute one's own content to the discussion (Ziegele, 2016, pp. 30-31). Such semi-public online discussions take place, for example, in various discussion forums or on social networking sites such as Facebook, Instagram, or Twitter.

In public online discussions, a wide variety of people can participate and is potentially behaving uncivil or exposed to incivility. Scholars have studied occurrences, causes and effects of online uncivil behavior by ordinary citizens (e.g., Coe et al., 2014; Gervais, 2015; Kluck & Krämer, 2021), by politicians (e.g., Otto et al., 2020; Zompetti, 2019), and by journalists (e.g., Ziegele & Jost, 2020)[3]. As a reaction to the increase of uncivil comments on their news websites and social media pages, media outlets have additionally created a new journalistic role, that of community managers. Community managers monitor comment sections, delete severely uncivil comments and engage with their news audience (e.g. Friess et al., 2021, pp. 627-628; Frischlich et al., 2019, pp. 2016-2017). Besides community managers combating incivility, several activist groups have emerged and collectively engage against uncivil behavior in online discussions (e.g., Porten-Cheé et al., 2020, p. 515; Ziegele et al., 2020a, p. 732). Thus, incivility research has identified and studied diverse lay and (semi-)professional participants in online discussions, all of whom are addressed as (potential) *communication participants* in the dissertation. Communication participants do not necessarily have to be actively involved in a discussion, for example by writing comments and visibly engaging with other participants; the term also encompasses passively involved persons who read comments, so-called "lurkers" (Blanchard & Markus, 2004, p. 70; Springer et al., 2015, p. 799).

Lastly, uncivil behavior can be expressed through various channels (e.g., Kümpel & Rieger, 2019, p. 9). Although the vast majority of incivility studies addressed text-based

[3] For a literature review of research on causes and consequences of incivility in public online discussions, see Kümpel and Rieger (2019).

communication, i.e., written comments, posts or tweets (e.g., Coe et al., 2014; Santana, 2014; Su et al., 2018), some studies also analyzed (audio-)visual uncivil communication expressed in, for example, memes, images or short video-clips (e.g., Khedkar et al., 2021; Lobinger et al., 2020; McSwiney et al., 2021). The dissertation deals with principally *all forms of communication* in online discussions, which includes both text-based and (audio-)visual communication.

In sum, the dissertation focuses on incivility (a) in (semi-)public political online discussions between (b) various communication participants, ranging from ordinary citizens to semi-professional activists, professional community managers or journalists and politicians, that (c) can be text-based or audio(-visual). In the next section, the concepts of incivility in the extant literature will be discussed.

2.2 Concepts of Incivility

The concepts of civility and incivility are subject to various academic disciplines. Historians have focused on, for example, what (in)civility meant in earlier eras and how civility norms have changed over time (e.g., Bullock, 2019; Schwerhoff, 2020). Philosophers are interested in the moral implications of (in)civility (e.g., Mower, 2019). And social scientists, including communication scholars, political scientists and psychologists, have studied, for example, the causes of (un)civil behavior, the effects of (in)civility on political debates and communication participants, and correlates of (in)civility in communication (for an overview of different disciplines' approaches to civility and incivility, see e.g., Boatright et al., 2019). This dissertation focuses on incivility research in social sciences, drawing primarily on incivility concepts in communication studies, but incivility concepts of related disciplines are also considered in the following sections[4].

Due to multidisciplinary research on the phenomenon incivility, its widespread academic popularity and a fragmented research landscape, a diverse array of approaches to incivility has been developed. Further, the concept itself poses major challenges on incivility scholars, making it difficult to define. As Herbst (2010) stated in her influential book on "rude

[4] Moreover, it should be mentioned that there are several other concepts of deviant forms of online communication, on which a vast amount of research and literature exists that cannot be considered in the following literature review. These concepts include, for example, "flaming" (for an overview on the concept, see e.g., O'Sullivan & Flanagin, 2003), "hate speech" (for an overview on the concept and interdisciplinary approaches to the concept, see e.g., Wachs, Koch-Priewe, & Zick, 2021), "trolling" (for an overview on the concept, see e.g., Buckels et al., 2014; Rieger et al., 2020), "dark participation" (Quandt, 2018), "fake news" (for an overview on the concept, see Zimmermann & Kohring, 2018), and "offensive language" (for an overview on the concept, see e.g., Davidson et al., 2017; Risch et al., 2020).

democracy", the decision of what is civil and uncivil lies "very much in the eye of the beholder" (p. 3). Indeed, the vast majority of incivility scholars share the notion that incivility is highly subjective (e.g., Chen, 2017, p. 5; Chen et al., 2019, p. 2; Coe et al., 2014, p. 660; Jamieson et al., 2018, p. 206; Kalch & Naab, 2017, p. 400; Kenski et al., 2020, p. 797; Kluck & Krämer, 2021, p. 3; Muddiman, 2019, pp. 33-34; Stryker et al., 2016, p. 540; Stryker et al., 2021, p. 2; Sydnor, 2018, p. 97). Besides its subjective nature, scholars additionally emphasize that incivility is dependent on the context. More specifically, depending on the cultural context, platform, and discussion topic, among others, the same phrases and words can be defined as civil in one context and as uncivil in another one (e.g., Benson, 2011, p. 26; Chen et al., 2019, p. 2; Coe et al., 2014, pp. 673-674; Wang & Silva, 2018, p. 73; Sydnor, 2018, p. 99). Consequently, incivility is elusive and can be considered as "a notoriously difficult term to define" (Coe et al., 2014, p. 660).

Given that incivility is addressed by several academic disciplines and that its determination is subjective and context dependent, it is not surprising that research lacks an agreed-upon definition and systematization of the concept. Across the range of approaches to incivility, however, one common denominator can be identified, namely that most studies conceptualize incivility in the broadest sense as a violation of norms (e.g., Chen et al., 2019, p. 3; Coe et al., 2014, p. 660; Jamieson et al., 2018, pp. 205-206; Kluck & Krämer, 2021, p. 3; Hopp, 2019, p. 206; Muddiman, 2017, pp. 3183-3184, 2019, p. 32; Papacharissi, 2004, p. 271; Rossini, 2020, p. 2; Su et al., 2018, p. 3681; Sydnor, 2018, p. 99). The studies can be categorized into four categories: (1) Studies that approach incivility as a violation of politeness norms based on politeness theories, (2) studies that conceptualize incivility as a violation of democratic norms based on normative theories of democracy, (3) studies that define incivility as a violation of deliberative norms against the backdrop of deliberative democracy, and (4) studies that classify incivility as a violation of multiple norms.

The following section outlines these approaches to incivility and elaborates on similarities and differences between the concepts. Against the backdrop of the literature review, the integrative and multidimensional incivility concept of this dissertation is then presented, as well as the new theoretical approach to cooperative communication on which it is based (*Article I*).

2.2.1 Incivility as a Violation of Politeness Norms

Studies conceptualizing incivility as a violation of politeness norms often refer explicitly or implicitly to face and politeness theories (Brown & Levinson, 1987; B. Fraser, 1990; Goffman, 1955, 1967). Against this backdrop, incivility is defined as a face-threat (e.g., Chen

& Lu, 2017, p. 110; Chen & Ng, 2017, p. 182) or as violating the social norms of politeness of a particular culture (e.g., Su et al., 2018, p. 3681; Sydnor, 2018, p. 99).

Following the seminal work of Goffman (1955) on human interaction, "face" was originally defined as "the positive social value a person effectively claims for himself by the line others assume he has taken during a particular contact" (p. 213). According to Goffman (1955) people seek to maintain their face during interactions by various verbal and non-verbal efforts which are called "face-work" (p. 216). Goffman's theorizing on face was later adapted by Brown and Levinson (1987) to explain social interaction that revolves around being polite, which resulted in a widely recognized politeness theory. Defining face as the constructed public self-image that each participant involved in a communication or interaction seeks to claim for herself/himself, face theory was expanded by arguing that we have two faces, a positive and a negative one (Brown & Levinson, 1987, p. 13). While positive face refers to an individual's desire for approval and acceptance, negative face expresses the desire to be unimpeded and autonomous. During interactions, the participants aim to protect their own and the other persons' faces and the strategy to maintain face or to restore face after "face-threatening acts" (Brown & Levinson, 1987, p. 60) is politeness. Threats against the positive face can include, for example, harsh criticism of or disagreeing with the communication partner, ridiculing or insulting the person. The negative face is mainly threatened when the participant's freedom is restricted, for example, by being forced to do or refrain from doing something.

Originating in face-to-face communication, various research indicates that people expect the same face maintaining and restoring politeness rules in computer-mediated-communication (CMC) as they do face-to-face (e.g., Chen & Lu, 2017; Graham, 2007; Reeves & Nass, 1996). Regardless of whether participants in an online discussion know each other, they would expect each other to consider politeness norms, and face-threatening acts would be a violation of these norms (Chen, 2015, p. 821). Against this backdrop, incivility studies have approached uncivil behavior as a threat to positive face and thus as a challenge to communication participants' desire for approval and acceptance. Incivility is then defined as specific use of words or phrases in online comments that are impolite and pose a danger on the communication partner to lose her/his face, such as insults, name-calling, profanity, and using capital letters to indicate yelling in CMC (e.g., Chen & Lu, 2017, p. 110, 114; Chen & Ng, 2017, p. 182).

Several other studies do either implicitly refer to face-theory or build more broadly on a "social-norm view" (B. Fraser, 1990, p. 220). The social-norm view roots in a historical understanding of politeness that is generally anchored in Western societies. It states that each society has a certain set of social norms prescribing, for example, specific forms of communicative behavior in a given context (B. Fraser, 1990, pp. 220-221). Behavior that is compliant with the norms is evaluated as being polite. The social-norm view reflects, among others, what is known as etiquette, i.e., rules that are defined to govern polite discourse, and what constitutes good manners in human interactions (for manuals on etiquette, see e.g., Tuckerman & Dunnan, 1995; Vanderbilt & Baldridge, 1978). In terms of communication, the approach associates politeness explicitly with speech style and therefore more with the tone than the substance (B. Fraser, 1990, p. 221).

Particularly earlier studies on incivility in the media, in political advertising, in political debates and its effects on the public conceptualized incivility based on the social-norm view. As such, incivility was equated with impoliteness and rudeness, and operationalized, for example, as name-calling, eye-rolling, yelling, emotional language, sharp criticism, and interruptions (e.g., Brooks & Geer, 2007, pp. 4-5; Mutz, 2007, p. 622, 625; Mutz, 2015, pp. 1-16; Mutz & Reeves, 2005, pp. 3-5). Since then, a considerable number of studies on incivility in CMC have adapted this definition of uncivil behavior and applied it to online contexts (e.g., Borah, 2013, p. 459; Rossini, 2020, pp. 4-7; Sobieraj & Berry, 2011, p. 20; Su et al., 2018, pp. 3680-3681; Syndor, 2018, pp. 98-99). These studies' operationalizations of incivility are similar to but extend those already mentioned of other studies following the politeness approach. Their operationalization of uncivil online behavior can be grouped into the following four categories: (1) insulting language such as name-calling, mockery, and derogatory, condescending remarks, (2) ominous language including threats and curses, (3) foul language such as vulgarity, profanity, obscenity, and crudeness, and (4) norm violations pertaining to the style of a message's delivery, namely sarcasm, irony, emotional language and yelling (Borah, 2013, p. 463; Rossini, 2020, p. 13; Sobieraj & Berry, 2011, p. 26; Su et al., 2018, p. 3687; Syndor, 2018, p. 99).

Furthermore, three additional aspects of the definition of incivility are noteworthy which have been elaborated in different studies in this field:

First, some scholars have argued that different *degrees of severity* of incivility should be considered in its definition (e.g., Su et al., 2018, p. 3681; Sydnor, 2018, p. 99). While Sydnor (2018, p. 99) approached incivility as a continuum ranging from mildly to moderately and highly uncivil behavior, Su and colleagues (2018) proposed to distinguish between two severity levels, namely minor violations of politeness norms, which they call "rudeness" (p. 3687), and more serious impoliteness, defined as "extreme incivility" (p. 3687).

Second, several scholars suggested to consider the *target* of an uncivil act when conceptualizing incivility (Su et al., 2018, p. 3681; see also Coe et al., 2014; Papacharissi, 2004;

Rowe, 2015 in the next chapters 2.2.2, 2.2.3) and distinguished between interpersonal and other-directed incivility (or “personal” vs. “impersonal” in terms of Su et al., 2018, p. 3687), with the former type targeting participants involved in the discussion and the latter type pertaining to messages directed at individuals or groups that are not conversationally present or at non-human objects.

Third, various scholars differentiated between the *tone* and *substance* of a message when defining incivility (e.g., Mutz, 2007, p. 625; Sydnor, 2018, p. 98; see also Hopp, 2019; Papacharissi, 2004; Rowe, 2015 in the next chapters 2.2.2, 2.2.3). While the tone of a message refers to the speech-style and word choice, the substance of a message pertains to its content and information value. Studies defining incivility as impoliteness tend to link the construct to the tone of a message as Sydnor (2018), for example, pointed out “incivility is a function of the tone of communication [not substance], identified by the use of vulgarity, obscenity, mockery, name-calling and insults, among other categories of speech” (p. 98). However, several other scholars disagree with this definition and link incivility to the substance of the message. Papacharissi (2004), for example, argues that serious norm violations can be polite and well-mannered in tone at first glance, but contain substantial, “impeccable incivility” (p. 279) like covert racism or threats to individual rights. These approaches are addressed in the next chapter, which presents studies that conceptualized incivility as a violation of democratic norms.

2.2.2 Incivility as a Violation of Democratic Norms

Several scholars who approach incivility as a violation of democratic norms sharply distinguish incivility from impoliteness (e.g., Kalch & Naab, 2017, pp. 399-400; Oz et al., 2018, p. 3403; Papacharissi, 2004, pp. 266-267; Rowe, 2015, p. 128; Santana, 2014, p. 21; Stoll et al., 2020, pp. 111-113). Papacharissi (2004) pioneered this approach in the context of CMC, stressing that incivility cannot be confined to impoliteness and should be understood more broadly and politically as “disrespect for the collective traditions of democracy” (p. 267). Since politeness is determined by maintaining face and by our understanding of etiquette and formality, Papacharissi (2004) argues that political discussions that follow politeness norms are “reserved, tepid, less spontaneous” (p. 260), and “limit the extent and diversity of discussion” (p. 262), thereby compromising democratic plurality and open, democratic exchange. Political discussions should instead allow for passionate and heated exchanges that promote and pursue overarching democratic goals. Civility, then, is understood as a means of ensuring “that the conversation is guided by democratic principles” (Papacharissi, 2004, p. 260), instead of adherence to etiquette and polite speech-style.

Based on a historical review of the concept of civility, also in the context of normative theories of democracy, Papacharissi expands her argument for a separation of (in)civility and (im)politeness. Thereby, she clearly distinguishes her (in)civility concept from that referring to deliberative theory of democracy and Habermas' (1989, 1991) notion of deliberation in the public sphere, according to which participants debate civic matters rationally, reciprocally, and well-mannered. Papacharissi (2004, pp. 265-266) criticizes that the civility concept underlying deliberation is too narrow as it focuses above all on norms that are prescribed by a powerful elite. As a consequence, mainly privileged groups would be able to participate in public discourse and social groups that have a different speech-style that might not be considered norm-compliant are either excluded or lose their uniqueness.

This argument derives primarily from a tradition of democratic theory that can be classified as "constructionist theory" (Ferree et al., 2002, p. 306). The starting point of the leading exponents of this tradition (e.g., Benhabib, 1992; N. Fraser, 1990; Young, 1996) is what they describe as the ongoing reproduction of power and inequality in the political process in the public sphere. Theorists criticize that marginalized groups are systematically excluded from political participation and decision-making since those privileged groups in power set the rules and norms for it. In contrast to the Habermasian public sphere, constructionist theories create a notion of a diverse public sphere in which the powerful and their assumptions of what is norm-compliant are decentered because these norms limit who participates and silence or devalue social groups wo habitually communicate in alternative modes (N. Fraser, 1990; Young, 1996; for an overview of the arguments of the different theorists, see Ferree et al., 2002, pp. 306-315). Thus, they take a critical stance on the concept of civility as proposed by deliberative theorists, considering it a means of impeding empowerment and inclusion of marginalized groups.

In line with constructionist theories' notions, Papacharissi (2004) sketches a concept of civility that does not focus on speech-style but rather "promote[s] respect for the other, enhance[s] democracy, but also allow[s] human uniqueness and unpredictability" (p. 266). She argues that the decision of whether something is civil or uncivil should be assessed by its implications for democratic society and thus whether basic liberal democratic principles are considered or violated (Papacharissi, 2004, p. 267). Accordingly, behavior that is classified uncivil includes, for example, denying or attacking individual rights, stereotyping social groups, and posing threats on democracy such as proposing to overthrow a democratically elected government by force (Papacharissi, 2004, p. 274).

The seminal concept of incivility as a violation of liberal democratic norms has been adapted by various scholars examining incivility in online discussions (e.g., Chen et al., 2019;

Friess et al., 2021; Kalch & Naab, 2017; Naab et al., 2021; Oz et al., 2018; Rowe, 2015; Santana, 2014; Stoll et al., 2020; Ziegele et al., 2020a). The underlying arguments have also been discussed outside empirical incivility research and are, for example, reflected in Garton Ash's (2016) call for a culture of "robust civility" (p. 316), where problematic or marginalized opinions are not silenced, where even the most difficult and contradictory political issues can be openly discussed, and where heated arguments can take place. The thesis that political discussions need to be more robust in order to provide democratic value to the polity, and a related broader understanding of civility and incivility, however, has not emerged with the advent of CMC, in fact it has already been discussed in numerous earlier works (e.g., Bejan, 2017; Cahoon, 2000; Lyotard, 1984; Schudson, 1997).

Finally, three aspects are again noteworthy regarding this approach to incivility, with two aspects already known from the (im)politeness approach. First, this concept also distinguishes between the *tone* and *substance* of a message. Contrary to scholars defining incivility as impoliteness and as already indicated above, this approach explicitly links incivility to the substance of a message. Second, the *target* of an uncivil act is differentiated into interpersonal, if the uncivil comment is directed at discussion participants, and other-directed, if the comment is aimed at a non-present person. Third, incivility is linked to the *consequences of a behavior* and is therefore evaluated by its outcome. Once a behavior has negative consequences for democracy or, in other words, endangers the common good, it is considered uncivil. In this regard, a distinction is made, for example, between insulting an individual person and attacking an entire social group. According to Papacharissi (2004, p. 267), the first is merely impolite as it has no lasting implications for democracy, while the latter poses worse consequences on democratic society and is thus considered uncivil.

Whereas this approach is distinct from (in)civility concepts in the context of deliberative theories of democracy, a large number of other incivility studies can be located there. These studies will be outlined in the next chapter.

2.2.3 Incivility as a Violation of Deliberative Norms

A large body of empirical incivility studies can be located in the context of deliberation. These studies implicitly or explicitly refer to deliberative theories of democracy and define incivility as norm violations deleterious to deliberative debate (e.g., Anderson et al., 2014, p. 375; Gervais, 2015, p. 169) or as a violation of (deliberative) respect norms (e.g., Coe et al., 2014, p. 660; Muddiman & Stroud, 2017, p. 588; Stroud et al., 2015, p. 190, 194).

Deliberative theories of democracy refer to an ideal form of participatory democracy in which deliberation among equal citizens about matters of public interest is essential to decision-making (e.g., Barber, 1984; Dryzek, 2000; Gutman & Thompson, 2004; Habermas, 1996). The core of deliberative democracy is the process of deliberation in the public sphere. Ideally, the public sphere should be accessible to everyone, and equal citizens should exchange arguments about matters of public interest in a rational, reciprocal, and respectful manner (e.g., Dahlberg, 2001, p. 616; Habermas, 1996; Ruiz et al., 2011, p. 466). The deliberation process is supposed to bridge social differences and legitimize political decisions (e.g., Dryzek et al., 2019, p. 1145; Habermas, 1996).

With the advent of Web 2.0 technology, several deliberative advocates have argued that the technology creates ideal conditions for deliberation by providing the infrastructure for a deliberative public sphere. In principle, public online discussions are accessible to all citizens and provide a forum for fair and respectful exchange about social and political issues (e.g., Dahlberg, 2001; Graham & Witschge, 2003; Ruiz et al., 2011; Wright & Street, 2007). Accordingly, deliberative democracy has become an influential theoretical concept widely applied in research on political online communication, which has led to a large body of online deliberation research in recent years (for an overview, see Friess & Eilders, 2015). In the field of online discussions, studies have often employed deliberative norms to analyze the quality of user comments and thus whether they live up to the ideal of deliberation (e.g., Rowe, 2015; Ruiz et al., 2011; Stroud et al., 2015; Ziegele et al., 2020b).

Deliberation is a demanding mode of communication, and for it to take place, several criteria must be met. Like incivility, however, deliberation is not a uniform concept and is measured differently in various studies (Friess & Eilders, 2015, p. 320, pp. 328-331; Stromer-Galley, 2007, pp. 1-7). In their extensive review of online deliberation research, Friess and Eilders (2015, pp. 328-331) identified several criteria that are mostly applied to assess the deliberative quality of online discussions and can be subsumed under the dimensions inclusiveness, rationality/constructiveness, interactivity, and civility (see also Esau et al., 2017, p. 332; Friess et al., 2021, pp. 625-627; Ziegele et al., 2020b, pp. 863-866). Inclusiveness means that everyone should have equal chances to access and actively participate in deliberation (e.g., Habermas, 1996). Rationality and constructiveness refer to substantiating positions with arguments and empirical evidence, and to an orientation towards the common good and finding consensus (e.g., Gutmann & Thompson, 2004; Habermas, 1996). Interactivity asks participants to speak and listen, and thus to discuss reciprocally (e.g., Barber, 1984). Civility refers to the mutual recognition among discussion participants to be equal actors and is often equated with mutual respect (e.g., Gutman & Thompson, 2004; Habermas, 1990).

Against this backdrop, it is unsurprising that online incivility research has often approached incivility in public online discussions through the lens of deliberation and deliberative democracy. Incivility is then usually conceptualized as norm violating behavior undermining deliberation, or specifically referring to the civility dimension of deliberation, as the absence of civility and thus as disrespect (e.g., Anderson et al., 2014, p. 375; Coe et al., 2014, p. 660; Gervais, 2015, p. 169; Kenski et al., 2020, p. 797; Muddiman & Stroud, 2017, p. 588; Rösner et al., 2016, p. 462; Stroud et al., 2015, p. 190, 194; Ziegele & Jost, 2020, p. 893).

Among the most recognized concepts in this field is that of Coe and colleagues (2014) who developed one of the first typologies of incivility in online discussions, defining incivility as "features of discussion that convey an unnecessarily disrespectful tone toward the discussion forum, its participants, or its topics" (p. 660) (adapted by e.g., Kenski et al., 2020; Muddiman et al., 2017; Riedl et al., 2019; Rösner et al., 2016; Stroud et al., 2015; Ziegele et al., 2020b; Ziegele & Jost, 2020). Their typology consists of five types of online incivility, namely name-calling directed at other discussion participants, aspersion directed at a plan, policy, behavior or an idea, lying, vulgarity, and disparaging remarks about the communication style of other participants (Coe et al., 2014, p. 661). Three aspects of their concept are noteworthy, revealing some overlaps with other approaches to incivility. First, in line with scholars approaching incivility as impoliteness, Coe et al. (2014, p. 660) explicitly focus on the *tone* of communication in defining incivility and on uncivil behavior as made manifest in public discussions. Second, they also consider and specify the *target* of incivility in their definition. Finally, Coe et al. (2014) include another aspect of incivility in their concept, namely that uncivil behavior is "something unnecessary" (p. 660)*,* contributing *nothing of substance* to the discussion. This notion has already been expressed in earlier studies on incivility in offline context. Brooks and Geer (2007), for example, posited that incivility is "superfluous" because it "add[s] little in the way of substance to the discussion" (p. 5), and Mutz and Reeves (2005) considered uncivil behavior as "gratuitous" (p. 5).

Besides the aforementioned overlaps of deliberation approaches to incivility with politeness approaches in terms of definitional aspects (e.g., tone and target) and operationalizations of uncivil behavior (e.g., name-calling, vulgarity), however, the two are not equivalent. While sharp criticism and disagreements are already considered a face-threat and impolite in politeness theories, free and open discussions with disagreeing arguments are central in the deliberation concept. Thus, on a conceptual level, the violation of politeness norms and related understanding of incivility cannot be equated with violations of deliberative respect norms. Implicitly included in Coe et al.'s concept, this distinction becomes much clearer in the following studies.

Several other scholars who approached incivility as a violation of deliberative norms, have explicitly distinguished uncivil behavior from mere disagreement. They share the notion that harsh disagreement - as long as it is voiced respectfully – is an inevitable characteristic of and beneficial for deliberation because it functions as an indicator for diverse viewpoints (e.g., Herbst, 2010, pp. 10-26; Hwang et al., 2018, p. 217; Stryker et al., 2016, pp. 538-540; Ziegele et al., 2018, p. 529; Ziegele & Jost, 2020, p. 893). As Hwang et al. (2018) stated, mere disagreement is not uncivil but incivility rather means "expression of disagreement by denying and disrespecting the justice of the opposing views" (p. 217).

Despite some overlapping key categories of uncivil behavior pertaining to categories (1), (2), and (3) in the following, there are also varying operationalizations across studies in this field (e.g., Anderson et al., 2014, p. 375; Coe et al., 2014, p. 661; Hwang et al., 2018, p. 222; Gervais, 2015, p. 172; Muddiman & Stroud, 2017, pp. 588, 594-595; Riedl et al., 2019, p. 434; Rösner et al., 2016, p. 464; Stroud et al., 2015, p. 194; Ziegele & Jost, 2020, pp. 903-904): (1) disrespectful attacks that degrade participants as being not equal actors, including insults/name-calling, ad hominem attacks, and derogatory or condescending remarks, (2) threats to communications participants and her/his individual rights such as free speech, e.g., through verbal intimidation, (3) attacks against ideas, policies, or institutions, (4) foul language such as vulgarity and profanity, and (5) false or misleading information such as lies, conspiracy theories, defamation and character assassination, spin or misrepresentative exaggerations.

Although the operational definitions reveal partial overlap with the other incivility approaches, the three approaches are distinguishable on a conceptual level. They each define incivility as a violation of different norms – as a violation of interpersonal politeness norms, democratic norms, or deliberative norms. In contrast, contemporary theorizing has shifted to an integrative, multidimensional approach and has considered several norms in defining incivility. This approach will be outlined in the following chapter.

2.2.4 Incivility as a Violation of Multiple Norms

As already outlined, many studies have conceptualized incivility as a violation of a particular norm. These studies refer to different theoretical approaches that prescribe norms and norm violations in communication processes. Thus, it is prescribed which norm violations are uncivil in online discussions. This approach, however, is increasingly criticized as, for example, the statement by Chen and colleagues (2019, p. 3) points out: “When platforms and academics take it upon themselves to decide what is uncivil, they are imposing a particular definition of what counts and what doesn’t. And inevitably, these definitions may force a particular worldview.” Accordingly, more and more studies are taking a different approach and consider the widely shared notion that incivility lies in the eye of the beholder. In addition, the previous chapters have already shown that a large number of different types of norm violations have been identified, but have not yet been integrated into a unified framework. The following approaches take these aspects into account and argue for multidimensional concepts of incivility.

Muddiman (2017, 2019), for example, conducted a series of surveys and experiments to inquire what the public perceives as uncivil. Although Muddiman focused on uncivil behavior by politicians and not on incivility in public online discussions, the results provide general insights into perceptions of the construct of incivility. The author developed a two-dimensional model of incivility: “personal-level incivility” (Muddiman, 2017, p. 3183; Muddiman, 2019, p. 32) refers to violations of politeness norms and “public-level incivility” (Muddiman, 2017, p. 3184; Muddiman, 2019, p. 33) includes violations of democratic and deliberative norms. Personal-level incivility encompasses insults and personal attacks, obscenity, and emotional language such as extreme anger or yelling. Public-level incivility pertains to politicians showing a lack of comity and compromise by, for example, refusing to cooperate, demagogic or ideological extreme language, spreading misinformation, inciting riots, discriminating minorities, and executing non-public acts such as secretly taking donations or paying people for vote (Muddiman, 2017, p. 3187; Muddiman, 2019, p. 35). Survey and experimental data indicated that personal- and public-level incivility are distinct dimensions and that citizens perceive both as uncivil.

In a similar vein, Stryker and colleagues (2016, 2021) examined citizens’ perceptions of 23 different types of norm violations. In two surveys, they provided descriptions of the different norm violations and asked participants to rate how civil or uncivil they find such behaviors in offline or online political contexts. Based on the data, a three-dimensional concept of incivility was developed in the first study and confirmed in a replication study few years later: “Utterance Incivility, Discursive Incivility, and Deception” (Stryker et al., 2016, p. 547) were found as distinct dimensions of the underlying construct of perceived incivility (see also Stryker et al., 2021, pp. 6-8). The first dimension refers to, among others, personal attacks, disrespect, threats, demonizing an opponent, or using racial, sexist, ethnic, or religious slurs. The second dimension encompasses behaviors that intend to shut down or detract open and inclusive discussions by, for example, excluding participants with different political opinions from a discussion. Lastly, the third dimension deception refers to lying, failing to provide evidence for factual claims, and misrepresentative exaggerations (Stryker et al., 2016, p. 548; Stryker et al., 2021, p. 6).

The studies by Muddiman (2017, 2019) and Stryker et al. (2016, 2021) clearly indicate that perceived incivility includes violations of multiple norms. A number of other scholars have also pursued the multidimensional approach. Hopp (2019), for example, developed a two-dimensional model of incivility similar to that of Muddiman (2017, 2019), distinguishing between incivility as a “violation of speech-related norms” (p. 206) and as a “violation of inclusion-related norms” (p. 207). While the former includes types of incivility other scholars have considered violations of politeness norms such as insults, profane language, and threats, the latter pertains to violations of democratic, deliberative norms by denying other participants their individual rights, suppressing discussions on specific issues or undermining faith in democratic systems and institutions, for example (Hopp, 2019, pp. 206-208). In contrast to studies focusing on citizens’ perceptions of incivility, Hopp examined the senders of incivility in political online communication. Studying self-reported uncivil behavior, Hopp (2019, p. 205) consequently defines incivility as an intentional norm violation. Intentionality, however, is rarely linked to the definition of incivility in any other study.

In addition to the above approaches, Chen (2017, p. 6) has focused on the *degree of severity* in her concept of incivility. The author conceptualized incivility as a continuum based on the respective severity of the norm violations, ranging from impoliteness to the violation of democratic norms. Norm violations that fall under the category of impoliteness, such as “calling someone ‘stupid’” would be classified as more mildly, while “President Donald Trump’s assertion, in his 2015 campaign announcement speech, that Mexican immigrants were ‘rapists’ (…) is a sweeping pejorative statement that defames a group” (Chen, 2017, p. 6) and would be classified at the more harmful end of the continuum according to the author (for similar approaches, see also Su et al., 2018, p. 3681; Sydnor, 2018, p. 99).

In sum, various scholars have approached incivility as a multidimensional model considering several norms, perceptions of citizens, and the level of severity. First survey and experimental data reveals that the public perceives various forms of norm violations as uncivil, suggesting a multidimensional construct. In the following, I will provide a brief summary and discussion of this section before presenting our novel concept of incivility.

2.2.5 Summary and Discussion

Incivility is studied in a wide variety of academic disciplines, research contexts, and is highly subjective and elusive in nature. It is therefore unsurprising that research lacks a unified concept of incivility. In the field of online incivility research, however, one common denominator between various studies can be found, namely that incivility is usually conceived as a violation of norms. Scholars either approached incivility as a violation of (1) politeness norms based on face and politeness theories, (2) democratic norms based on constructionist theories of democracy, (3) deliberative norms based on deliberative theories of democracy, or (4) multiple norms.

Despite conceptual differences between the approaches rooted in the underlying theories, some overlap exists in terms of operational and definitional aspects. Regarding operationalization, politeness and deliberation approaches to incivility share, for example, the strong focus on rude and disrespectful behavior towards other communication participants. Key categories include insults, threats, and belittling. Likewise, some overlap is evident between approaches focusing on deliberative norms and those referring to democratic norms, with both operationalizing threats to individual liberty rights such as freedom of speech and discrimination or exclusion of certain social groups as types of incivility, for example. Regarding definitional aspects, different approaches distinguish between tone and substance of an uncivil message, identify several targets of incivility, and consider varying levels of severity. Probably the greatest similarity between different approaches, however, is that the vast majority conceives of incivility as a subjective construct. Yet, only few scholars have consistently implemented this aspect into the concept of incivility by approaching incivility as a *perceptual construct.*

Initial studies that examined perceptions of various types of (potentially) uncivil behavior indicated that incivility is a multidimensional construct. Violations of multiple norms are perceived as uncivil. These studies, however, also reveal some shortcomings. First, they focus primarily on incivility between political elites and thus develop a concept of incivility in a particular sub-context. Second, although they criticize that incivility concepts are largely based on prescriptive theoretical approaches, i.e., theories that set and prescribe norms in communication processes, as outlined in chapter 2.2.4, most of these studies nevertheless draw on prior research and thus on prescriptive theoretical approaches. Based on these theories, norms and norm violations are derived and the perceptions of these a priori defined (potential) types of incivility are examined. In order to obtain a comprehensive understanding of norms and norm violations from the perspective of those actively or passively involved in a discussion, however, the normative expectations of the communication participants should be considered.

The aim of the first research article of this dissertation was to address these shortcomings. In *Article I*, we developed a new theoretical approach that is based on analytical instead of prescriptive theories. Prescriptive approaches are mainly criticized for determining a particular normative view, i.e., in the context of incivility, defining what is civil and uncivil and what is good and bad according to a normative ideal (e.g., Benhabib, 1992; Chen et al., 2019, p. 3; N. Fraser, 1990; Papacharissi, 2004, pp. 265-266). However, it is questionable whether, in the case of incivility, these normatively determined categories of incivility actually reflect the perspective of those involved in online discussions and whether they are actually always bad. Another disadvantage is that by defining a priori what is uncivil, research limits itself to these categories aligned to a normative ideal. Analytical approaches instead seek to describe and explain rather than prescribe. By strictly following the perspective of communication participants and reconstructing their expectations toward other participants' communication, our analytical approach offers a more open and differentiated research perspective on the phenomenon of incivility, allows for a comprehensive view of what incivility can potentially encompass, and allows for describing and explaining differences between various participants without judging what is good and bad.

Therefore, in *Article I*, we theoretically reconstruct the normative expectations of communication participants against the backdrop of different analytical theories on communication, cooperation, and norms. These normative expectations are subsumed under five basic communication norms which serve as the basis for a multidimensional and perceptual concept of incivility. This new framework and its implications will be outlined in the following sections. Furthermore, the new concept of incivility will be contextualized within the literature and challenges and perspectives of incivility research will be discussed (*Article II*).

2.3 Incivility as a Violation of Communication Norms: Theoretical Approach

2.3.1 Article I: Incivility as a Violation of Communication Norms – A Typology Based on Normative Expectations toward Political Communication (Bormann, Tranow, Vowe, & Ziegele, 2022)

The starting point of *Article I* are three shortcomings of incivility research that have also been identified in the previous chapter: First, concepts of incivility diverge, making it difficult to measure its prevalence, causes, and effects in a valid and reliable manner, and to develop effective interventions. Second, studies refer to different norms when defining incivility and it has not yet been fully determined which specific norms are violated by uncivil behavior. Third, the majority of incivility research refers to theories that prescribe norms and norm violations and thus define uncivil behavior a priori. However, it is not yet clarified whether the postulates of these prescriptive theories reflect the normative expectations of the communication participants.

The article addressed these shortcomings in three steps: (1) Against the backdrop of different analytical approaches to cooperation, communication, and norms (e.g., Grice, 1975; Lindenberg, 2015; Tomasello, 2019) that are combined in a new framework of cooperative communication, a theoretical foundation for incivility research is developed that is not prescriptive. (2) Drawing on theoretical considerations of cooperative communication, the potential normative expectations of participants involved in (political) communication regarding the communicative behavior of other participants are reconstructed and five communication norms are derived from it. (3) The five communication norms serve as the basis for an integrative and multidimensional concept of incivility including a novel definition and comprehensive typology of incivility that can serve as a heuristic for future studies on incivility.

The theoretical framework combines arguments from Tomasello's evolutionary anthropology (2008, 2009, 2019), Grice's linguistic approach (1975, 1989), and Lindenberg's social rationality approach (2001, 2015). Specifically, it builds on the following premises that all three approaches share: (1) People can create shared intentions because they have the ability to adopt the perspective of others; (2) Shared intentions are the prerequisite for cooperation and cooperation is the elementary condition for social and political relationships and orders, from families to democratic societies; (3) For cooperation to succeed, it needs communication and communication is subject to normative expectations.

Drawing on these premises, we develop the concept of *cooperative communication*, which we understand as a communication that enables cooperation. Such communication is necessary because cooperation is a demanding mode of interaction: Through communication, participants can develop and align shared goals, develop a common ground, particularly with regard to shared knowledge, and coordinate their individual actions (e.g., Heath & Frey, 2004, pp. 182-183, 196-198; Kerr & Kaufman-Gilliland, 1994, pp. 513-514; Tomasello, 2009). We refer to this form of cooperative communication as *cooperation through communication*. However, in order for participants to align and achieve shared goals through communication, they must ensure that they understand each other (Schramm, 1954; Tomasello, 2008). In this sense, communication itself is a fundamental form of cooperation, which we define as *cooperation in communication*.

Cooperative communication, however, cannot be assumed as given but is endangered by three types of problems of the participants involved. *Problems of motivation* refer to the aspect that people tend to prioritize their individual benefits over the achievement of the shared goals. Thus, they need to be motivated to achieve the shared goals. *Cognition problems* can arise when participants do not have a common ground and thus do not share the same level of knowledge. Finally, *coordination problems* can occur when individual roles and responsibilities are unclear and thus cooperative actions cannot be coordinated.

To increase the likelihood of cooperative communication, five *communication norms* have emerged to address these problems. When participants adhere to these norms, the risk of failure of cooperative communication is minimized. Communication norms are defined as the normative expectations of participants regarding how to communicate in a given context (e.g., Homans, 1974, p. 96; Opp, 2015, p. 5). Based on our understanding of the elementary role of cooperation in societies, we assume that human communication is oriented towards cooperative communication in all kinds of communication contexts, thus also in public political (online) communication (see also Grice, 1975, p. 45; Jeffries & McIntyre, 2010, p. 106). Consequently, we assume that the basic normative expectation of participants in political communication is to "*communicate in a way that enables the understanding necessary for the respective cooperation"* (Bormann et al., 2022, p. 342). This central norm can be differentiated into five communication norms, which are linked to the central aspects of communication (e.g., Lasswell, 1948; Schaff, 1962) and address elementary challenges of cooperative communication.

The *information norm* refers to the substantial aspect of communication (content) and addresses the challenge of providing the information that are necessary for successful cooperation. As such, this norm subsumes participants' normative expectation to *"communicate what is informative with regard to the common cooperation goals"* (Bormann et al., 2022, p. 344). Based on the conversation maxims by Grice (1975, pp. 45-46), the information norm is subdivided into three dimensions referring to the quantity, quality, and relevance of the information provided in communication.

The *modality norm* pertains to the formal aspect of communication (mode) and addresses the challenge of ensuring mutual comprehension in cooperative communication. It expresses the normative expectation of participants to *"communicate comprehensibly with regard to the shared cooperation goals"* (Bormann et al., 2022, p. 344). Following Grice's (1975, p. 46) maxim of manner, we differentiate three dimensions of this norm, namely clarity, conciseness, and orderliness.

The *process norm* addresses the temporal aspect of communication (process) and is supposed to assure the connectivity of contributions in communication. Thus, the norm asks participants to "*communicate in such a way that the contributions in the respective cooperation context are connected"* (Bormann et al., 2022, p. 345). This abstract expectation can be specified by three dimensions, namely substantial connectivity which requires to stay on topic, reciprocal connectivity, which asks participants to refer to each other, and consistency of one's own contributions.

The *relation norm* refers to the social aspect of communication (actors) and ensures mutual trust between the participants. Respectful behavior is a clear indicator that the communication partner can be trusted (e.g., Lindenberg, 1998, pp. 85-89). Therefore, we presume that participants expect each other to *"communicate respectfully with the others involved in cooperative communication"* (Bormann et al., 2022, p. 346). The norm can be differentiated in three dimensions, namely politeness, appreciation, and deference.

Lastly, the *context norm* pertains to the spatial aspect of communication (context) and ensures that the specific context is considered. Depending on the social field, such as politics, religion, or science, and the degree of publicity, the specific normative expectations for communication can differ. Incivility, in this thesis, refers to the context of public political communication in liberal democracies. Following empirical democracy research (Coppedge et al. 2011; Diamond, 1999), we presume that participants of public political communication expect each other to *"communicate in such a way that cooperation in the political public sphere is rendered possible according to liberal-democratic principles"* (Bormann et al., 2022, p. 347) by considering liberal principles like individual and collective liberty rights, democratic principles like competitive elections, and democratic principles like rule of law.

Against the backdrop of this theoretical framework, we propose a new concept of incivility that is based on (1) an analytical approach, (2) the perceptions of the communication participants, and (3) several norms, thus addressing demands of incivility scholars and shortcomings of previous studies. Incivility is defined as the *"acts of communication in public political debates that participants disapprove of as severely violating communication norms of information, modality, process, relation, or context"* (Bormann et al., 2022, p. 348).

Three aspects of this definition are noteworthy. First, the concept draws on the *five communication norms* and thereby provides an integrative framework for incivility research. Several norm violations identified by previous research can be categorized within the framework. For example, violations of politeness norms (e.g., Su et al., 2018, pp. 3680-3681) can be classified as violating the relation norm and violations of democratic norms (e.g., Papacharissi, 2004, pp. 260-267) can be subsumed under the political context norm. Second, the concept explicitly refers to incivility in *public political contexts*. This includes debates in parliaments as well as online discussions that can be public or semi-public when they are publicly visibly but require a registration for active participation, such as social media platforms. Third, the concept strictly incorporates the *participants' perspective* and leaves the decision of what is civil and uncivil to the participants involved[5]. Moreover, the definition is based on the disapproval of participants. Disapproval is a two-step process: Participants must first perceive a violation of one or more of the communication norms and then evaluate it as a sanction-worthy violation. Thus, in *Article I* and the whole dissertation, it is distinguished between perception and evaluation. While both terms are understood as cognitive processes, *perception* means the pure organization and identification of a violation of the communication norms (Schacter et al., 2012, p. 123). Put differently, perception means that a communication participant organizes and identifies a communicative act as norm-violating when exposed to it. After perception, a norm violation is usually interpreted (Schacter et al., 2012, p. 123)[6] in terms of whether it is tolerable or sanction-worthy, among other aspects, which is defined as *evaluation*. Since norm violations can be tolerated in certain situations, only those that are evaluated as sanction-worthy are considered uncivil. Disapproval is a cognitive process and not necessarily linked to a visible behavioral reaction. Visible reactions that show disapproval, however, are defined as "explicit disapproval" (Bormann et al., 2022, p. 349).

Furthermore, a typology of incivility was developed (see Bormann et al., pp. 351-352). The typology is structured through the five communication norms. In the next step, types of incivility that previous empirical research identified were categorized as violations of the norms, and additional (possible) violations were supplemented. The typology is therefore

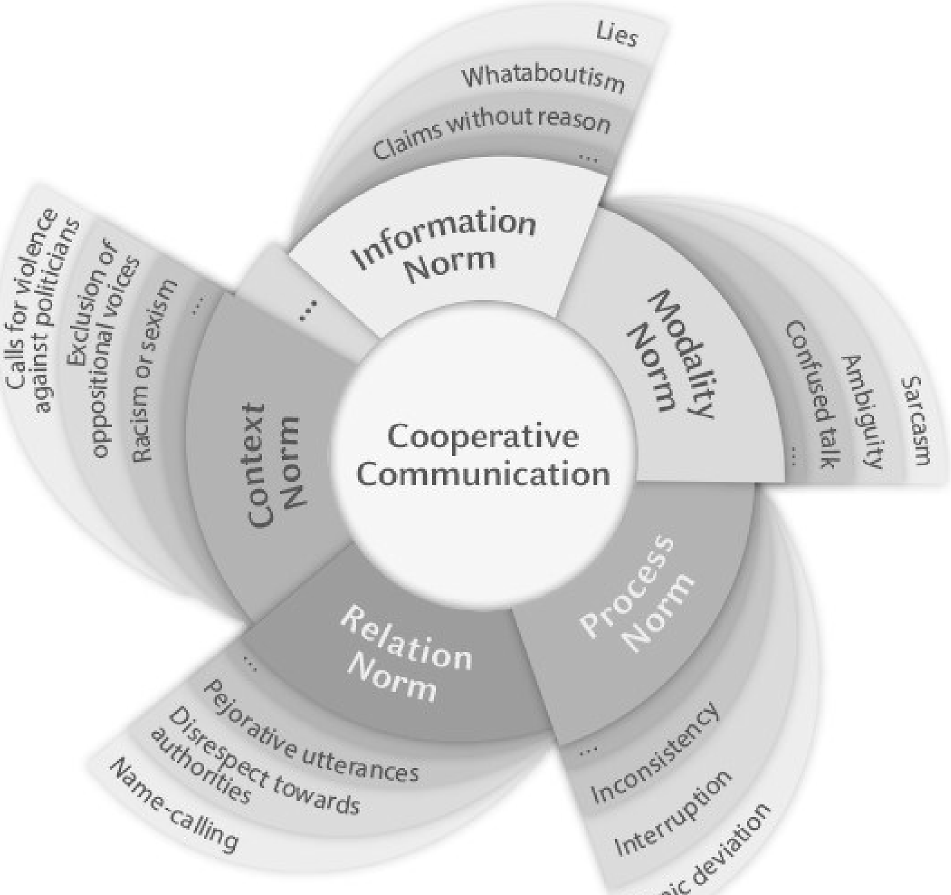
[5] Perception-oriented approaches are conceptualized and applied in other fields of communication research as well. Our perceptual approach can be linked to, for example, approaches of entertainment research, which examine entertainment from the recipient's perspective and let the recipients decide what they consider entertaining (for an overview, see e.g., Dohle & Vowe, 2014; Wünsch, 2006; Zillich, 2013), or to research on the "hostile media effect" addressing differing perceptions of media coverage based on preexisting attitudes (e.g., Gunther, 2017).

[6] Schacter et al. (2012) define the "organization, identification and interpretation of a sensation in order to form a mental representation" (p. 123) as perception. By sensation, the researchers mean the "simple awareness due to the stimulation of a sense organ" (p. 123), which is in this case the exposure to a communicative act that contains a violation of the communication norms. In this dissertation, however, the interpretation is distinguished from the organization and identification of the sensation and defined as "evaluation" because this cognitive process is of particular importance for our definition of incivility. For a similar but much more nuanced differentiation and approach, see e.g., Anderson and Carnagey (2004) who distinguished in the "General Aggression Model" several cognitive processes, among others, including "*perceptual schemata*, which are used to identify phenomena as (…) social events (e.g., personal insult)" (p. 174) and "several complex appraisal (…) processes" (p. 176).

developed deductively and inductively. Violations of the information norm include, for example, lies, factual claims without reasons, or misrepresentative exaggerations (e.g., Hopp, 2019, p. 210; Muddiman, 2017, p. 3187). Violations of the modality norm are, for example, sarcasm, irony, and ambiguity (e.g., Rowe, 2015, p. 128; Ziegele & Jost, pp. 896-897). Off-topic contributions and interruptions, among others, are violating the process norm (e.g., Sydnor, 2018, p. 99; Stryker et al., 2016, p. 542). Insults and belittling directed at other participants are examples for violations of the relation norm (e.g., Chen & Lu, 2017, p. 114; Coe et al., 2014, p. 661). Finally, violations of the political context norm include, for example, threats to democracy such as incitement to overthrow a democratically elected government by force, and threats to individual rights such as discriminating marginalized groups (e.g., Papacharissi, 2004, p. 274; Stryker et al., 2016, p. 542).

Figure 2 illustrates the approach to incivility based on cooperative communication and norms of cooperative communication.

Figure 2. Disc of Norms and Norm Violations (Bormann et al., 2022, p. 355).



2.3.2 Article II: Incivility (Bormann & Ziegele, in press)

Article II is part of an anthology that addresses different facets of challenges and perspectives of research on deviant online communication (for the article in full length, refer to the Appendix). The articles of the anthology were intended to be programmatic, reviewing and questioning the current state of research and giving impulses for future research. The aim of the article was to introduce the concept(s) of incivility, including one's own approach. Additionally, challenges of incivility research related to the concept were to be discussed and new directions identified. Thus, the article starts with an overview of the different approaches to incivility followed by an introduction of our own incivility concept (see chapter 2.2 and 2.3.1). Afterwards, three challenges are outlined that relate to (1) inconsistent operationalizations of incivility across studies, (2) the reliable measurement of incivility in content analyses, and (3) normative implications of incivility. Additionally, new perspectives on incivility in political communication are discussed.

The first problem of incivility research identified is that different studies can only be compared to a limited extend because of distinct operationalizations. Studies that conducted content analyses of online discussions came to very different conclusions about the prevalence of incivility (e.g., Coe et al., 2014; Rowe, 2015; Santana, 2014). Certainly, these studies focused on different platforms with diverging platform designs, topics, and communities, among others, but the varying number of uncivil comments can also be explained by distinct operationalizations. For example, while Coe et al. (2014, pp. 667-668) defined incivility as name-calling, vulgarity, aspersion, pejoratives, or lying and found a share of 22% incivility in user comments, Rowe (2015, p. 129) operationalized incivility as assignment of stereotypes, threats to democracy or individual rights, and reported a share of 3% incivility in comments on Facebook and 6% in website comments. Santana (2014, pp. 25-27) applied a broad concept of incivility, and unsurprisingly found a large share of incivility, reporting that up to 53% of user comments are uncivil. Likewise, experimental research applied different operationalizations and came to different conclusions regarding the democratic consequences of online incivility (e.g., Chen & Ng, 2017; Gervais, 2015; Kalch & Naab, 2017; Rösner et al., 2016). Moreover, distinct types of norm violations are often intermingled within these studies. In consequence, little is known about the effects of distinct types of incivility. The varying findings are particularly problematic because they suggest different normative and practical implications. For example, based on findings of low prevalence and harmfulness of incivility, policymakers may decide that incivility is not a pressing issue, while other studies may justify strong interventions. Future research should therefore develop and apply standardized, differentiated indicators to measure incivility.

The second challenge also refers to operationalization aspects of incivility, namely to develop instruments that reliably measure incivility in content analyses. Several manual content analyses struggle to achieve satisfactory inter-coder reliability (e.g., Coe et al., 2014, p. 661; Ross et al., 2018, pp. 3-4; Ziegele et al., 2018, pp. 539-540). Automated analyses applying dictionary-based approaches (e.g., Muddiman & Stroud, 2017) or machine learning (e.g., Su et al., 2018) also often display high rates of misclassification (e.g., Stoll et al., 2020). This applies particularly to subtle, culture- and context-specific forms of incivility, such as covert sexism, racism, irony or sarcasm. Such forms are more difficult to recognize and detect, and the coders' perception, knowledge, and experiences have a higher impact when it comes to deciding whether to code subtle forms as civil or uncivil. Based on our understanding of incivility as a perceptual construct, we propose an alternative, two-step procedure to analyze uncivil online comments that considers the disapproval of the communication participants (see also Bormann et al., 2022; *Article I*). First, comments are checked in terms of visible disapproval. Visible disapproval of participants serves as an indicator for incivility. Second, the disapproved comments are then analyzed by the coders for norm violations.

The third challenge relates to normative implications of incivility. The prevailing notion in communication science and practice is usually that incivility is undesirable, harmful, and needs to be eliminated (for an overview, see Chen et al., 2019). One reason for this is that several studies have demonstrated harmful effects of online incivility (e.g., Anderson et al., 2014; Gervais, 2015; Hsueh et al., 2015), and another one is that the majority of incivility studies are based on prescriptive theories that consider incivility, for example, as detrimental to deliberation or as a negative face-threat. In contrast, some studies have also reported positive effects of incivility (e.g., Borah, 2014; Brooks & Geer, 2007; Chen & Lu, 2017). Moreover, a large body of critical studies argues that calls for civility can serve as an instrument of a powerful elite to suppress minority voices (e.g., Baez & Ore, 2018, pp. 331-332; Lozano-Reich & Cloud, 2009, pp. 223-225; Stuckey & O’Rourke, 2014, p. 714, pp. 723-724). According to this research tradition, those in power decide over what is civil and who is thereby excluded from political discourse. Violations of civility norms are seen as a legitimate and effective instrument to differentiate an oppressor from the oppressed, to demonstrate belonging to a marginalized group, and to fight inequality and injustice. Against this background, scholars are well advised to not judge incivility as bad per se, but to consider the normative implications of the phenomenon from a more nuanced perspective, namely when and under what circumstances which type of incivility actually has a positive or negative effect.

In sum, we advocate a broad and nuanced perspective on the construct of incivility by approaching it as a perceptual concept based on five communication norms that are derived from elementary, non-prescriptive approaches to human cooperation and communication. The concept may provide three key benefits for future research: (1) The concept is very broad and integrates types of norm violations that previous studies have identified. However, the concept is also finely differentiated and enables nuanced analyses as it systematizes distinct types of violations along five communication norms. (2) The concept is not restricted to one particular context like, for example, that of Coe and colleagues (2014), but is applicable to communication between various types of actors in offline contexts as well as CMC. (3) The concept provides a non-prescriptive perspective and thus allows future research to gain a more fine-grain understanding about what the participants actually perceive as (un)civil in different countries, cultures, and contexts, and to better assess normative implications.

2.3.3 Summary and Next Steps

Incivility research has yielded considerable achievements, and a large and valuable body of studies on different facets of offline and online incivility among different actors exists. However, research lacks a unified concept of incivility. Thus, the aim of *Article I* was to develop a unifying concept of incivility that considers several aspects identified by previous research and addresses shortcomings of earlier concepts. The definition and typology of incivility developed within *Article I* build on a non-prescriptive, analytical approach and embed several types of incivility identified in previous research into an integrative framework by systematizing them based on five communication norms. In addition, the concept considers the demand for a *perceptual approach* by having the participant decide what is civil and uncivil. Furthermore, norm violations are covered that relate to the *tone* or *substance* of a communicative act, that are directed towards *different targets*, and the concept can be applied to *different political contexts* and includes *various potential communication participants*.

The concept, however, is based on theoretical assumptions about the normative expectations of communication participants. Therefore, the next step was to conduct empirical studies to examine what different participants in public political online discussions actually disapprove of as uncivil, which norm violations are perceived as how severe, and how different participants react to distinct types of violations of the communication norms. These empirical questions were addressed in three studies presented in the next section.

3. Investigation of Perceptions of and Reactions to Incivility

In the empirical part of the dissertation, qualitative and quantitative research was combined following a triangulation approach to obtain a more comprehensive yet nuanced answer to the dissertation's research question. In three studies, namely focus groups, an experimental study, and a quantitative content analysis, it was examined what various participants in public online discussions disapprove as uncivil and how they react to different types of incivility.

To examine incivility perceptions and thereby differentiate and test the concept of incivility developed in *Article I*, two empirical studies were conducted. The first study applied a *qualitative methodology* and explored in-depth what different participants of online discussions disapprove as uncivil, how they rate distinct types of incivility in terms of severity, and whether the perceptions of various communication participants differ. For this purpose, three different types of participants, namely lay participants (i.e., ordinary online users), semi-professional participants (i.e., members of online activist groups that combat incivility), and professional participants (i.e., community managers of different news media outlets), were brought together in five heterogeneous focus groups. The results are presented in *Article III* (chapter 3.2.1) and supported the assumption of a multidimensional concept of incivility, revealed quite large commonalities among the actors, and allowed to refine and differentiate the typology of incivility.

The second study aimed at empirically validating the concept of incivility, testing the assumptions derived from the focus groups, and examining reactions to incivility by lay communication participants in public online discussions. For this purpose, a *quantitative methodology* was applied. We conducted an online experiment in which participants were exposed to different types of norm violations in a simulated online discussion in a fully functional mock-up forum. The results as presented in *Article IV* (chapter 3.2.2) revealed that violations of all five communication norms are disapproved as uncivil, and that distinct types of norm violations vary in terms of perceived severity and elicit different responses.

Furthermore, in *Article V* (chapter 3.2.2) the results of a *quantitative manual content analysis* are presented that analyzed reactions to uncivil comments by professional communication participants, i.e., community managers. The content analysis in addition to the experiment was worthwhile because it allowed us to systematically examine an immense number of actual, real-life comments from community managers of different media outlets with regard to patterns and determinants in responses to incivility. The results suggested that distinct types of incivility elicit different responses by community managers. For example, violations that refer to the context norm were associated with more intervening responses, and community managers used varying moderation styles against different forms of violations.

Before presenting the studies and their findings, previous incivility research on perceptions of and reactions to incivility is reviewed. The first section 3.1.1 introduces studies on the perception of incivility in general and by different communication participants in specific. Afterwards, research on reactions to incivility in public online discussions by lay participants is discussed (chapter 3.1.2). The last review section 3.1.3 focuses on research dealing with reactions to online incivility by professional actors, i.e., community managers.

3.1 State of Research

3.1.1 Perceptions of Incivility

Studies on what different communication participants perceive as mildly and severely uncivil in public online discussions are scarce. Scholars have mostly studied perceptions of incivility in offline interactions between political elites by describing hypothetical scenarios of uncivil behavior instead of exposing study participants to actual uncivil messages. Moreover, most research is based on surveys, U.S.-centric and examined perceptions of the general public or specific sub-groups but not of the participants involved in a discussion (e.g., Kenski et al., 2020; Muddiman, 2017, 2019; Stryker et al., 2016, 2021). Some studies, however, focused on different participants in online discussions, namely either ordinary users (Kalch & Naab, 2017), semi-professional activists (Ziegele et al., 2020a) or professional community managers (Frischlich et al., 2019) and examined, among other aspects, their perceptions of specific types of incivility.

Muddiman (2017, 2019) studied public perceptions of her two-dimensional model of incivility consisting of personal-level and public-level incivility. As already mentioned earlier (see chapter 2.2.4), personal-level incivility refers to violations of politeness norms, namely insulting, obscene, and emotional language (Muddiman, 2017, p. 3187; Muddiman, 2019, pp. 32-33). Public-level incivility includes violations of deliberative and democratic norms, such as showing a lack of comity and compromise, using ideological extreme language, inciting riots, or discriminating minorities (Muddiman, 2017, p. 3187; Muddiman, 2019, pp. 33-34). While the former pertains to violations of the relation norm according to our concept of incivility, the latter refers to violations of the political context norm (Bormann et al., 2022; see chapter 2.4.1). In the first part of her studies, Muddiman (2017, 2019) exposed individuals to statements that described hypothetical uncivil behavior of politicians. In the second part, she asked study participants to provide one example of what they evaluate as uncivil behavior by politicians (Muddiman, 2019). All studies suggested that both sub-dimensions of incivility are perceived as uncivil. Interestingly, individuals perceived personal-level incivility as more uncivil than public-level incivility when rating scenarios of uncivil behavior (Muddiman, 2017, p. 3197; Muddiman, 2019, pp. 33-35). When asking to provide examples of incivility, however, respondents were more likely to cite violations of democratic norms and thus public-level incivility, or in other words violations of the political context norm (Muddiman, 2019, pp. 35-36).

Stryker et al. (2016, 2021) examined citizens‘ incivility perceptions by exposing study participants to 23 statements describing different types of offline or online incivility (see chapter 2.2.4). While all 23 types were perceived as at least slightly uncivil, the incivility ratings of the distinct types were quite differently in both studies. Using racial, sexist, religious, or ethnic slurs and encouraging or threatening harm in a political discussion were rated as most uncivil, followed by disrespect, insults and personal attacks against other discussion participants (Stryker et al., 2016, p. 543; Stryker et al., 2021, p. 5). Consequently, types of violations that can be categorized as violating the relation or political context norm were rated as severely uncivil. Types of violations pertaining to the information norm, such as failing to provide evidence or making statements that are exaggerated and distort the truth, and violations that can be defined as violating the process norm, like interrupting other communication participants, tended to be rated as more mildly uncivil (Stryker et al., 2016, p. 543; Stryker et al., 2021, p. 5).

Contrary to Muddiman (2017, 2019) and Stryker et al. (2016, 2021), Kenski and colleagues (2020) explicitly focused on public perceptions of incivility in online discussions among ordinary citizens and not politicians. More specifically, the authors asked study participants to rate real online comments produced by members of the lay public. Drawing on the concept by Coe et al. (2014, pp. 660-661; see chapter 2.2.3), they defined incivility as name-calling, vulgarity, lying, aspersion and pejoratives for speech, and found that name-calling and vulgarity were assessed as more uncivil than the other types (Kenski et al., 2020, p. 808). This result is in line with Muddiman's (2017) finding, both suggesting that violations that can be classified as violations of the relation norm are evaluated as most severe. However, survey data from Stryker et al. (2016, 2021) indicated that violations of the context norm are also assessed as quite severe, as do the results of Muddiman's second study (2019), in which study participants listed more violations pertaining to the context norm when they were asked to cite examples of incivility. These two norms in particular appear to elicit quite high severity evaluations.

In addition, some studies have investigated the perceptions of participants involved in online discussions, namely ordinary users (Kalch & Naab, 2017), online activists (Ziegele et al., 2020a) and journalists or community managers (Chen et al., 2020; Frischlich et al., 2019).

In their experiment on engagement against uncivil comments by ordinary users, Kalch and Naab (2017) exposed participants to a manipulated user comment on a mock-up discussion platform. The comments either contained explicit insulting, disrespectful attacks against a marginalized group or more implicit stereotypes and threats to individual rights (Kalch & Naab, 2017, pp. 404-405), both referring to violations of the context norm according to our concept (Bormann et al., 2022). Participants rated both comments as uncivil compared to the norm-compliant comment, and the direct attack against the marginalized group as more uncivil than the other violation of the context norm (Kalch & Naab, 2017, p. 406).

Ziegele and colleagues (2020a) conducted a survey with members of the German online activist group #Iamhere (English translation for #Ichbinhier). #Iamhere is the largest German activist group with around 45,000 members who collectively engage against incivility in public online discussions (for detailed information on #Iamhere, see Ley, 2018). Study participants were asked to rate how civil or uncivil they find user comments that contained either name-calling, threats (i.e., violations of the relation norm according to our concept), antagonistic stereotypes, rejections of democracy (i.e., violations of the context norm), or lies (i.e., violation of the information norm). More specifically, they were asked to rate the "harmfulness" as the term "incivility" is not common in German language (Ziegele et al., 2020a, p. 740). The online activists rated all five types of incivility and thus violations of the relation, context, and information norm as highly harmful, or put differently, as severely uncivil. Additionally, the results indicated relatively similar incivility ratings among the activists (Ziegele et al., 2020a, p. 742). In contrast, in-depth interviews with community managers from different German news media outlets suggested that their evaluations of incivility vary widely (Frischlich et al., 2019, pp. 2023-2027). Frischlich et al. (2019, p. 2021) asked the community managers how harmful they rate certain forms of incivility in public online discussions, namely trolling, spreading false information, and disrespectful attacks, which can be classified as violations of the information and relation norm applying our concept of incivility (Bormann et al., 2022). Ratings of these norm violations varied from relatively mildly to highly problematic and harmful among the different community managers (Frischlich et al., 2019, pp. 2023-2027).

In summary, prior studies indicate that violations of several norms constitute perceived incivility and that distinct types of norm violations are not equally severe. More specifically, types of incivility that can be classified as violating the relation norm and as violating the context norm tend to be rated as more uncivil than violations of the other communication norms. Moreover, violations of the process norm and modality norm, in particular, have rarely been examined in perceptual studies, and when studied, they were predominantly rated as only "somewhat uncivil" (Stryker et al., 2016, p. 543; Stryker et al., 2021, p. 5). Further, the question remains whether different participants in online discussions, such as ordinary users, activists, and community managers, differ in their perceptions of incivility. Initial findings reveal a mixed picture.

The next chapters will review empirical findings on reactions to incivility in public online discussions by different actors. The first section outlines reactions to incivility by lay participants, that is, ordinary users of Web 2.0 platforms (chapter 3.1.2). The second section focusses on professional participants, that is, community managers who professionally monitor and engage in online discussions (chapter 3.1.3).

3.1.2 Reactions to Incivility by Lay Participants

When confronted with incivility in public online discussions, participants usually have several options to react. These options differ between lay participants and professional participants. While professional participants are community managers of discussion forums who work for the platform provider and/or a specific medium, lay participants are ordinary users from the community (for a similar classification, see Friess et al., 2021, pp. 627-630). A key difference between these actors is that professional participants are entitled with administrative governance rights which expand their options to react to uncivil comments (e.g., Friess et al., 2021, pp. 627-628; Watson et al., 2019, pp. 1845-1846). Before elaborating on these extended rights in the next chapter, this section will first systematize responses by lay participants and outline empirical studies in this field.

In *Article I*, we drew on the framework by Hirschman (1970) to identify possible forms of responses to incivility. Hirschman (1970) provided a widely recognized concept on human responses to decline in firms, organizations, and nations. The basic argument is that members of an organization have different options when they perceive a decline in quality or individual utility: They can either "exit," that is, withdraw from the relationship or "voice," that is, communicating the disappointment, complaint, or suggestions for change in order to improve the situation and relationship (Hirschman, 1970, pp. 3-5). Moreover, they can remain "loyal," i.e., continue to support the organization (Hirschman, 1970, pp. 77-79). According to Hirschman (1970), however, loyalty usually comes with the expectation that "*someone* will act or *something* will happen to improve matters" (p. 78), and he assumes that if dissatisfaction increases, members will eventually react with exit or voice. Thereby, the degree of loyalty plays again an essential role. A high level of loyalty is associated with a lower likelihood of an exit. Loyal members are more likely to ignore initial deterioration signals and are more likely to voice their dissatisfaction in the next step before exiting (Hirschman, 1970, pp. 77-79).

The concept has been applied to various fields ranging from personal relationships to protest movements, political parties, and public policy (for an overview, see Dowding, 2016). In the field of online discussions, Hirschman's framework also offers a fruitful starting point to classify different types of responses. When participants are confronted with uncivil comments and their normative expectations are thus disappointed, they can either leave the discussion, or voice their disapproval. Moreover, they can ignore the norm violations and expect that someone else will intervene or something will happen to improve the discussion situation (Hirschman, 1970, p. 78). When participants voice their disapproval with a norm violation, we define it as "explicit disapproval" (Bormann et al., 2022, p. 349; see chapter 2.3.1) in *Article I.*

Research on lay participants' reactions to incivility in online discussions suggests that norm violations are usually not ignored but rather lead to (a willingness to) exiting the discussion or voicing the disapproval. Several studies showed that exposure to incivility makes participants more likely to leave the discussion or less willing to stay in the discussion and actively participate (e.g., Hwang et al., 2008; Kluck & Krämer, 2021; Lück & Nardi, 2019; Pang et al., 2016). When participants stay in the discussion after exposure to incivility, however, they are likely to voice their disapproval (e.g., Gervais, 2015; Kalch & Naab, 2017). Disapproval can usually be expressed in online discussions by writing a reprimanding comment or by pressing specific social buttons, such as dislike or flagging buttons (e.g., Crawford & Gillespie, 2016, p. 411; Gervais, 2015, pp. 169-170; Kalch & Naab, 2017, pp. 401-402; Naab et al., 2018, p. 779; Porten-Cheé et al., 2020, pp. 519-521; Wilhelm et al., 2020, p. 924). In comments, disapproval can be expressed by, for example, pointing out the norm violation, by reprimanding the uncivil commenter, or by demanding sanctions for the uncivil commenter (e.g., Gervais, 2015, pp. 169-170; Kalch & Naab, 2017, p. 401). Additionally, many platforms have implemented social buttons that allow participants to respond to a particular comment (e.g., Porten-Cheé et al., 2020, p. 520). The like/dislike button, for example, serves to express one's agreement or disagreement with a comment (e.g., Kalch & Naab, 2017, p. 402). Given that a dislike can also simply express one's disagreement with the political position or argument presented in a particular comment, it is not a clear indicator for disapproval of a norm violation. A much clearer indicator for a disapproved norm violation and thus also a clearer form of explicit disapproval is flagging.

By flagging comments, participants report them as norm violating to the platform providers or community managers of the respective online discussion (e.g., Crawford & Gillespie, 2016, p. 411; Naab et al., 2018, p. 779; Porten-Cheé et al., 2020, p. 220; Wilhelm et al., 2020, p. 924). These professional actors have more governance rights than lay participants, and therefore more response options to comments that violate their platform rules. Besides reprimanding the uncivil commenter, they can block her or him from the platform and thus from participating in further discussion, and they can change or delete the flagged comment (Friess et al., 2021, pp. 627-628; Watson et al., 2019, p. 1845).

Several studies revealed that lay participants in online discussions are likely to voice their disapproval by writing a reprimanding comment as a reaction to incivility, or by flagging uncivil comments (e.g., Gervais, 2015; Kalch & Naab, 2017; Naab et al., 2018; Wilhelm et al., 2020). In addition, recent survey data of German online users has suggested that, in comparison to previous years, more and more users report to engage against hate speech by writing reprimanding comments or using flagging buttons (LfM, 2021, p. 3). Flagging, in particular, seems to become the prevalent response to uncivil comments among lay participants, as indicated by survey data (LfM, 2021, p. 4), but also by an experiment conducted by Kalch and Naab (2017, p. 406), in which participants used more often flagging than reprimanding comments.

Although different forms of reactions, namely ignorance, exit, and voice, to various distinct types of norm violations have not yet been systematically examined, empirical findings from prior studies suggest that different types of incivility might elicit different responses (Kalch & Naab, 2017; Naab et al., 2018; Wilhelm et al., 2020). Kalch and Naab (2017, p. 406) found that participants were more likely to express explicit disapproval against a comment that explicitly attacked a vulnerable group than a comment that attacked the same vulnerable group more implicitly. Likewise, an experiment by Naab et al. (2018, p. 790) revealed that comments that directly attacked an individual person of a vulnerable group tended to elicit more flags than comments that attacked the whole group. Similarly, findings by Wilhelm et al. (2020, p. 934) who also studied flagging behavior, showed that participants were more likely to flag violations pertaining to the context norm (i.e., incitements for violence) than violations referring to the information norm (i.e., rumors and conspiracy theories).

In sum, lay actors in online discussions have several options to respond to communicative acts they disapprove as uncivil. Applying Hirschman's (1970) concept to online discussions, three types of behavior can be differentiated, namely ignorance, exit, and voice. Previous research showed that incivility is usually not ignored but that exposure to incivility increases the likelihood of (willingness to) exiting the discussion. When staying in the discussion, however, participants are likely to express their disapproval of incivility by writing a reprimanding reply comment to the uncivil commenter or by flagging the uncivil comment. Moreover, initial findings suggest that different forms of incivility might elicit different responses of lay participants in online discussions, however, a systematic analysis has not yet been conducted.

3.1.3 Reactions to Incivility by Professional Participants

The increase of uncivil comments in online discussions urged platform providers to undertake action against it. As a consequence, many news media restricted the comment functions on their websites or completely shut down their comment sections (e.g., Meedia, 2016; Stroud et al., 2015, p. 189; Wüllner, 2015). Yet, their sites on social media platforms continued to exist (e.g., Rowe, 2015, p. 122). Social media operators, however, also implemented technical intervention options against uncivil comments, such as flagging (see chapter 3.1.2) and features to block users from discussion or to modify and delete comments. To intervene against uncivil comments in online discussions and to promote a civil discussion culture, several media have employed community managers (i.e., professional participants) who monitor and moderate comment sections. Their role and function are usually seen as "governance mechanisms (…) to facilitate cooperation and prevent abuse" (Grimmelmann, 2015, p. 47) in an online discussion.

In contrast to lay participants who are independent members from a platform's community, professional participants work for a specific news outlet or platform provider, thus acting on behalf of them and representing their rules and values (e.g., Friess et al., 2021, pp. 627-628). Professional participants own administrative governance rights, usually enabling them to block users, delete or modify comments before and after publication, allow users to flag comments as uncivil, and to generally establish discussion rules and norms (Watson et al., 2019, pp. 1845-1846). Moreover, they can apply filtering methods to automatically detect and sort out comments that include predefined uncivil words (e.g., Diakopoulos & Naaman, 2011, p. 134; Friess et al., 2021, pp. 627-628; Ksiazek, 2018, p. 655). These forms of reaction to uncivil comments are defined as "content moderation" (Friess et al., 2021, p. 628; Wright, 2006,

p. 555; Ziegele et al., 2018, p. 532) in the extant literature. Content moderation in the sense of deleting comments and blocking users is especially necessary when comments violate the law. However, community managers can theoretically also delete or modify comments that do not violate the law. In addition, content moderation is usually non-discursive and non-transparent because it is not publicly visible and community managers do not have to justify their actions to the community. As such, content moderation has often been criticized as limiting participation and free speech (e.g., Janssen & Kies, 2005, pp. 321-322; Riedl et al., 2020, p. 440; Wright, 2006, pp. 553-556; Ziegele & Jost, 2020, p. 829).

Besides content moderation, community managers can also react to incivility by actively engaging in the discussion and writing own comments, which prior studies have termed "interactive moderation" (e.g., Friess et al., 2021, p. 628; Wright, 2006, p. 556; Ziegele & Jost, 2020, p. 829; Ziegele et al., 2018, p. 529). While the focus here is on professional participants writing reprimanding comments against incivility, interactive moderation also encompasses posting comments that answer questions of other participants, provide additional information, thank lay participants for constructive comments, and foster participation (e.g., Friess et al., 2021, p. 628; Stroud et al., 2015, pp. 190-192; Ziegele & Jost, 2020, p. 895).

In recent years, more and more news media have employed interactive moderation in the comment sections of their websites and social media sites (e.g., Ksiazek & Springer, 2020). From a normative perspective, scholars have emphasized the added value to content moderation in that interactive moderation could intervene against incivility and promote civil discussions without overly limiting free speech (Janssen & Kies, 2005, pp. 321-322; Friess et al., 2021, p. 628; Wright, 2006, pp. 553-556). Empirical studies have indicated that interactive moderation can indeed positively affect the discussion quality, and increase lay participants' willingness to actively participate (Stroud et al., 2015; Ziegele & Jost, 2020). In addition, studies revealed that most lay participants appreciate interactive moderation in general (e.g., Bergström & Wadbring, 2014; Diakopoulos & Naaman, 2011; Stroud et al., 2016). The results, however, also suggested that the positive effect and appreciation of interactive moderation refers to specific and not all forms of moderation. This implies that community managers have varying moderation styles. Put differently, professional participants seem to respond to uncivil comments in different ways.

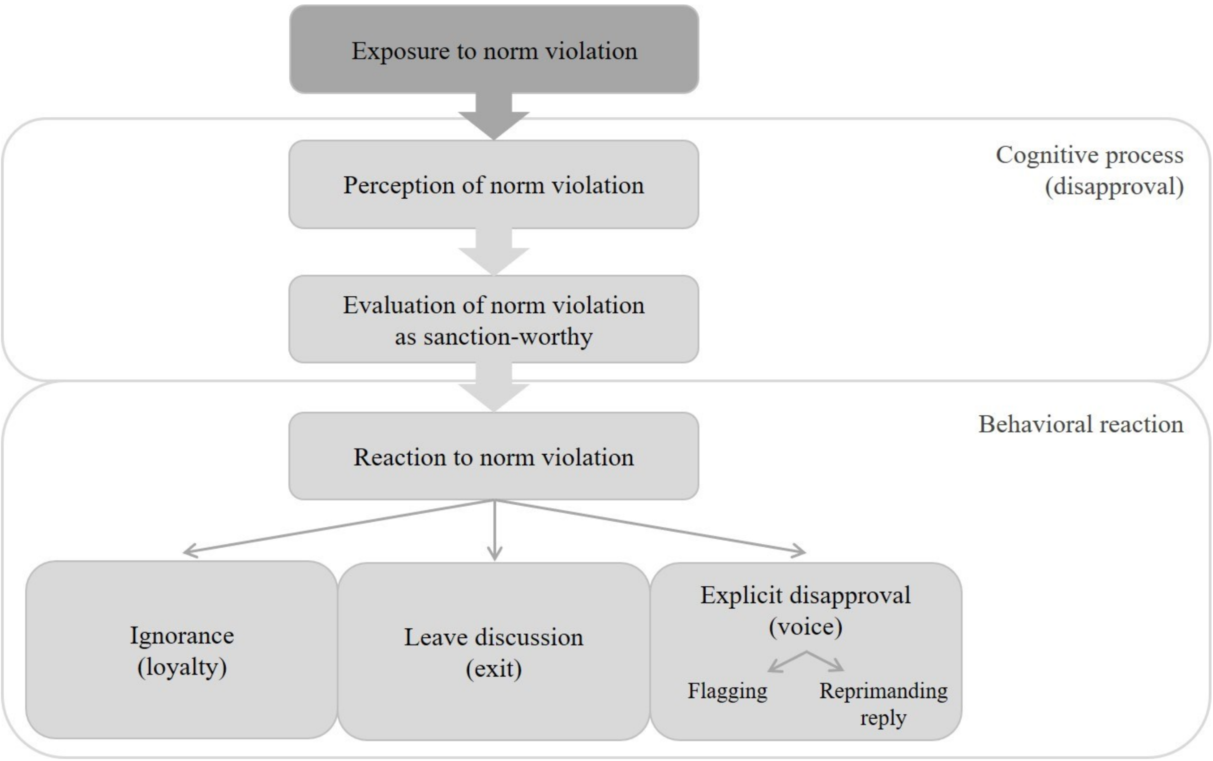
So far, different forms of interactive moderation against uncivil comments have hardly been systematized and systematically investigated in research. *Article V* therefore provides a typology of interactive moderation styles against uncivil comments and examines the patterns, determinants, and potential effects of different responses to distinct types of incivility by professional participants.

3.1.4 Summary and Next Steps

Only few studies have examined what various participants of public online discussions perceive as uncivil. What we can learn from prior research is that violations of several norms seem to constitute perceived incivility and that distinct types of incivility tend to be evaluated differently in terms of severity. What is largely unclear, however, is whether violations of all five identified communication norms (see *Article I* and chapter 2.3.1) are disapproved of as uncivil by communication participants and to what extent perceptions differ among various participants in online discussions, namely ordinary users defined as lay participants, online activists defined as semi-professional participants, and professional participants, that is, community managers. These questions are addressed within two empirical studies presented in *Article III* and *Article IV*, which are summarized in the next chapters 3.2.1 and 3.2.2.

Besides perceptions of incivility, the previous review focused on reactions to incivility by lay and professional participants in public online discussions. Drawing on Hirschman's (1970) influential framework, three potential types of reactions to incivility by lay participants were classified, namely ignorance (i.e., loyalty), leaving the discussion (i.e., exit), and explicit disapproval (i.e., voice). Figure 3 provides an overview of the different types of reactions to norm violations in online discussions, and the preceding cognitive process of disapproval.

Figure 3. Schematic Overview of the Cognitive Process after Exposure to Norm Violations and Different Types of Reactions to these Violations in Online Discussions.


Exposure to norm violation
Cognitive process (disapproval)
Perception of norm violation
Evaluation of norm violation as sanction-worthy
Behavioral reaction
Reaction to norm violation
Ignorance (loyalty)
Leave discussion (exit)
Explicit disapproval (voice)
Flagging
Reprimanding reply

Regarding lay participants' reactions to incivility, initial empirical studies suggested that uncivil behavior is usually not ignored, but participants are more likely to leave the discussion (i.e., exit) or to actively engage against incivility (i.e., voice). It remains open, however, whether violations of different communication norms elicit different reactions among lay participants. This question is answered in *Article IV,* outlined in chapter 3.2.2.

To combat online incivility, several platforms have employed community managers who perform interactive moderation against uncivil behavior. So far, there has been no systematic analysis regarding what types of incivility elicit responses from these professional participants, and whether they post different types of comments against distinct types of incivility. The content analysis presented in *Article V* and summarized in chapter 3.2.2 therefore aimed at providing insights into patterns, determinants, and potential consequences of interactive moderation against uncivil comments.

3.2 Incivility as a Violation of Communication Norms: Empirical Approach

3.2.1 Article III: Perceptions and Evaluations of Incivility in Public Online Discussions – Insights from Focus Groups with Different Online Actors (Bormann, 2022)

Approaching incivility as a perceptual construct, the logical next step was to ask different communication participants what they perceive as mildly and severely uncivil. Thus, the first empirical study addressed two research questions. *RQ1 asked what different actors in public online discussions perceive as norm violating, and where they agree and differ in their perceptions. RQ2 asked how the norm violations are evaluated in terms of severity and which evaluation criteria become apparent.*

To answer the research questions, five heterogeneous focus groups with three different actors of online discussions, namely, ordinary users, members of activist groups, and professional community managers, were conducted. The endeavor is highly relevant for particularly three reasons: (1) As already outlined in chapters 2.2 and 3.1.1, scholars largely agree that incivility is a perceptual concept, but only few empirical studies have examined what participants of online discussions perceive as (mildly and severely) uncivil. Prior research has focused on incivility in offline contexts among politicians, applied quantitative methods in which a priori defined types of incivility were tested (Kenski et al., 2020; Muddiman, 2017, 2019; Stryker et al., 2016, 2021), or conducted studies with one type of actors, such as activists (Ziegele et al., 2020a) or community managers (Frischlich et al., 2019). To gain a more precise and comprehensive understanding of perceived incivility in online discussions, in-depth studies with various online actors are necessary. (2) The typology of incivility developed in *Article I*, could be complemented and refined based on the study. The typology can function as a theoretical and empirical framework for future research. Furthermore, hypotheses can be generated from the results, for example on differences between the five communication norms or on differences and similarities of different online actors' perceptions. (3) Lastly, several practical implications can be derived from the findings, including development of intervention strategies and media companies learning what their users consider sanction-worthy.

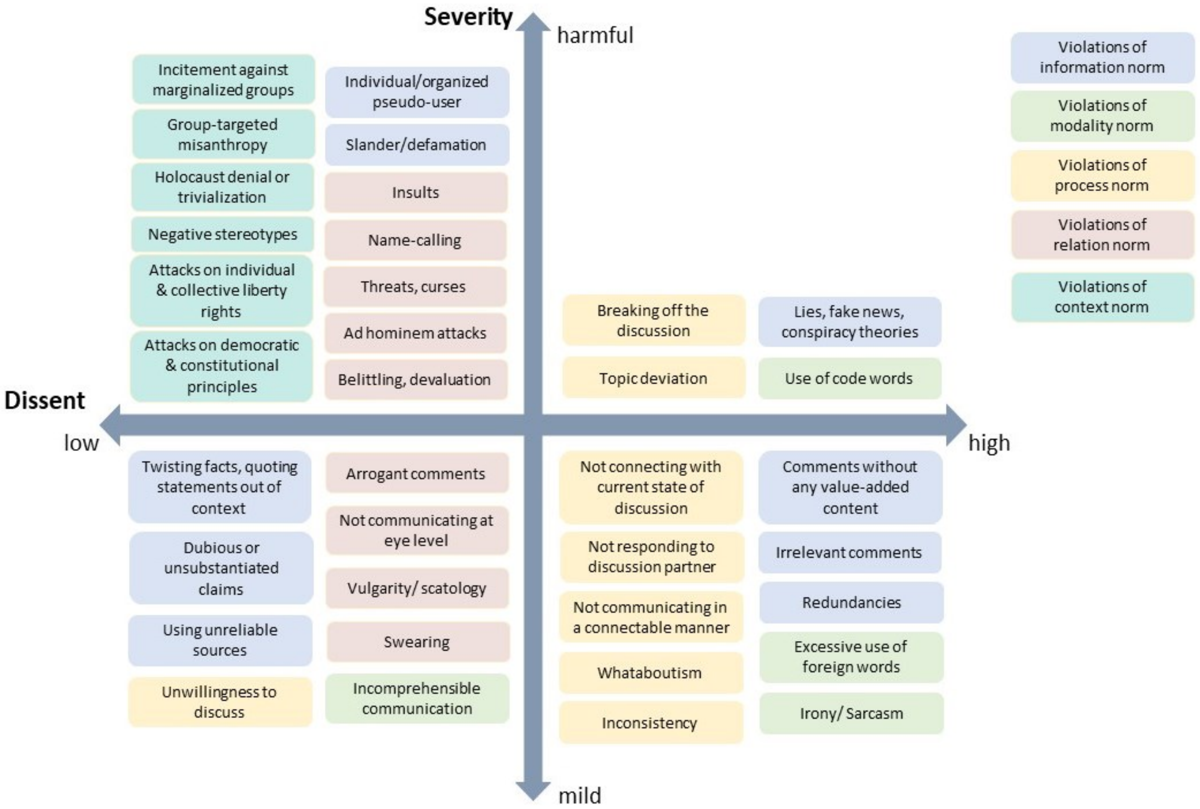
Against the backdrop of the theoretical framework developed in *Article I,* incivility is approached as a violation of one or several of the five communication norms that is disapproved of by communication participants. To explore in depth what different participants of public online discussions disapprove as uncivil, a qualitative semi-structured focus groups methodology was implemented. Five heterogeneous groups with a total of 25 participants were conducted across Germany. Each group included four to six participants and at least one representative of the three types of online actors: (1) ordinary users (*n* = 9), (2) members of the largest German online activist groups combating incivility, i.e., #Iamhere and No-Hate-Speech-Movement (*n* = 6), and (3) community managers from leading German news outlets (*n* = 10). The focus groups were held in November 2019 in five different German cities and were each moderated by two researchers. The semi-structured discussion guide was developed in accordance with common guidelines (e.g., Hennink et al., 2020, pp. 143-149; Krueger & Casey, 2015, pp. 35-62), pretested in an additional focus group, and included open questions and stimuli material to explore perceptions of norm violations in online discussions. The five focus groups were audio-recoded and transcribed. The transcripts were analysed using a "thematic qualitative content analysis" (Kuckartz, 2014, pp. 97-121) that consisted of a multilevel coding procedure including deductive and inductive elements, with the five communication norms forming the basis for the coding scheme.

The results suggest a multidimensional model of perceived incivility and support the broad concept developed in *Article I.* The different actors perceive various types of incivility in public online discussions. All types of violations reported by the online actors could be classified into our systematic of five communication norms, resulting in five categories of incivility: informational incivility including violations of the information norm; formal incivility, that is, violations of the modality norm; processual incivility, i.e., violations of the process norm; personal incivility consisting of violations of the relation norm; and anti-democratic incivility, that is, violations of the context norm.

The findings also revealed differences among the actors, and regarding severity evaluations of different types of norm violations. First, although the actors showed quite large similarities in what they perceive as mildly and severely uncivil, some differences became apparent. Particularly users and activists tended to be more sensitive to norm violations, reported more norm violations, and evaluated, for example, specific types of informational and processual incivility as more severe than community managers. This variance in incivility perceptions and evaluations could be caused by the different roles and associated functions of the actors in online discussions. Second, violations of different communication norms are not evaluated as equally severe. Overall, the actors tended to evaluate violations of the relation and context norms as more severe than violations of the other norms, with particularly violations of the context norm being consistently evaluated as highly harmful. Exceptions are specific violations of the information norm, such as lies and defamation, that were also evaluated as quite severe.

The results are subsumed and illustrated in Figure 4, which is a refined typology of perceived incivility in online discussions. It includes all reported types of violations and classifies them along the dimensions (1) dissent between the actors (low vs. high), and (2) severity of the type of norm violation (mild vs. harmful).

Figure 4. Typology of Perceived Incivility in Public Online Discussions (Bormann, 2022, p. 8).


Severity
harmful
mild
Dissent
low
high
Incitement against marginalized groups
Group-targeted misanthropy
Holocaust denial or trivialization
Negative stereotypes
Attacks on individual & collective liberty rights
Attacks on democratic & constitutional principles
Individual/organized pseudo-user
Slander/defamation
Insults
Name-calling
Threats, curses
Ad hominem attacks
Belittling, devaluation
Breaking off the discussion
Topic deviation
Lies, fake news, conspiracy theories
Use of code words
Twisting facts, quoting statements out of context
Dubious or unsubstantiated claims
Using unreliable sources
Unwillingness to discuss
Arrogant comments
Not communicating at eye level
Vulgarity/ scatology
Swearing
Incomprehensible communication
Not connecting with current state of discussion
Not responding to discussion partner
Not communicating in a connectable manner
Whataboutism
Inconsistency
Comments without any value-added content
Irrelevant comments
Redundancies
Excessive use of foreign words
Irony/ Sarcasm
Violations of information norm
Violations of modality norm
Violations of process norm
Violations of relation norm
Violations of context norm

Lastly, the results suggest that the process of evaluating the severity of norm violations is multi-layered. More specifically, several criteria were identified that shape the evaluation of norm violations. The criteria were categorized drawing on Lasswell's (1948) model of communication: Who (sender) says what (message) in which channel (here broadly defined as context[7]) to whom (recipient) with what effect (presumed consequences)? (Bormann, 2022, p. 11; see Table 1). When evaluating a norm violation, participants consider (1) the sender and evaluate norm violations by, for example, repeat offenders worse than one-time slips; (2) characteristics of the message and evaluate a norm violation as less severe when, for example, the message also contains constructive elements; (3) several contextual criteria and tend to evaluate norm violations less severely, for example, on platforms with a more casual conversational tone and a younger community; (4) potential consequences of the norm violation, evaluating those that potentially affect more people as more harmful. Lastly, individual characteristics of the recipient play a role when evaluating norm violations, such as gender and situational mood.

Table 1. Criteria Shaping Actors' Evaluation of the Severity of Norm Violations (Bormann, 2022, p. 11).

Category	Criteria	Actors
Sender	Intention of norm violation	All actors
	Discussion intentions	All actors
	Frequency of norm violations	All actors, esp. CM
	Real/pseudo-user	CM
	Political views	Users, activists
Message	Number of violations	All actors
	Target	All actors
	Constructive elements	All actors
Context	Discussion tone	All actors
	Community	All actors
	Discussion quality	All actors
	Medium genre	All actors
	Law	CM
Recipient	Socio-demographics	All actors
	Situational mood	All actors
	Thematic involvement	Users, activists
Presumed consequences	Fosters further violations	CM
	Number of people negatively affected	Users, activists

Notes. "All actors" means community managers, users, and activists. "CM" stands for community managers.

[7] The term "context" was used because it best summarized and describes the criteria that are subsumed under this category. Despite the proximity of the terms, there is no substantive and conceptual overlap with the context norm.

3.2.2 Article IV: Perceptions of and Reactions to Different Types of Incivility in Public Online Discussions – Results of an Online Experiment (Bormann, Heinbach, & Kluck, 2022)

Since the qualitative study presented in *Article III* suggests that incivility includes disapproved violations of all five communication norms, while indicating differences in severity, the next step was to empirically validate these assumptions applying a quantitative, experimental study. Further, given that different types of norm violations seem to be evaluated differently, one could assume that responses also vary among distinct forms of incivility. The second study therefore addressed three main research questions, namely *what participants in public online discussions perceive as uncivil (RQ1), how they evaluate distinct types of incivility in terms of severity (RQ2), and how they react to various types of incivility (RQ3).*

The study thus aimed at empirically validating the concept of incivility developed in *Article I* and gaining insights into severity evaluations and reactions to violations of different communication norms by ordinary communication participants, which has scientific and practical relevance. (1) This study can provide future research with an empirically validated, multidimensional concept of incivility that allows for differentiated measurement of various types of norm violations from a perceptual perspective. (2) The study also provides insights into what types of incivility are assessed as most harmful and how lay participants react to different types of incivility. Against this backdrop, platform providers and community management can develop more nuanced, graduated intervention strategies.

Drawing on our theoretical approach developed in *Article I* (see chapter 2.3.1) and the state of research on perceptions of and reactions to incivility in public online discussions as outlined in chapters 3.1.1 and 3.1.2, we derived five hypotheses and one research question. Hypotheses 1-2 referred to our concept of incivility. Following our definition of incivility as disapproved violations of communication norms, we conceptualized disapproval as a two-step process including (1) the perception of a norm violation, and (2) the classification of the violation as sanction-worthy. We therefore expected that violations of the communication norms are recognized (H1.1) and classified as sanction-worthy (H1.2) compared to norm-compliant behavior in public online discussions. Moreover, we assumed that the evaluation of a norm violation as sanction-worthy is mediated by the perception of the norm violation (H2). Regarding the severity of different types of violations of the communication norms, we expected that violations of the relation norm and context norm are rated as more severe than violations of the other norms (H3). This assumption can be derived from prior research (e.g., Kenski et al., 2020; Muddiman, 2017; Stryker et al., 2016, 2021; see chapter 3.1.1) and reflects the findings of the focus group study (*Article III*, see chapter 3.2.1). In terms of reactions to norm violations, we derived two additional hypotheses from previous empirical studies as reviewed in chapter 3.1.2, thereby distinguishing between three forms of possible reactions, namely ignorance (i.e., loyalty), leaving the discussion (i.e., exit), or explicit disapproval (i.e., voice; see chapter 3.1.2). The hypotheses posit that participants confronted with violations of the communication norms will show more explicit disapproval (H4) and will more often leave the discussion (H5). In addition, we asked whether different norm violations elicit different reactions, namely ignorance, leaving, or various forms of explicit disapproval (RQ1).

To investigate the hypotheses and research question, an online-experiment was conducted from 20 October to 3 November 2021, utilizing a six-condition, single-level, between-subjects design. During the study, participants (*N* = 433) were directed to a fully functional mock-up discussion forum named "Let's discuss." They were asked to participate in a simulated, but to them seemingly real public online discussion on one of two political topics. In the discussion, participants were exposed to a norm compliant comment, or a comment violating one of the five communication norms. For each communication norm, one representative type of violation was selected: Insults and vulgarity against another communication participant as violations of the relation norm, stereotypes and threats of violence against a social group as violations of the context norm, false information as violation of the information norm, irony/sarcasm as violation of the modality norm, and topic deviation as violation of the process norm. All other features of the comment were kept as constant as possible. The manipulated comments were pretested to ensure that the norm violations are perceived as intended.

To measure perceptions of norm violations and severity evaluations, scales were developed for perception of each of the communication norms (Cronbach's *α* > .90), sanction-worthiness (α = .86, *M* = 2.85, *SD* = 1.64), and deviance and harmfulness (*α* = .97, *M* = 3.16, *SD* = 1.62). Regarding reactions, we followed our differentiation of responses to incivility as outlined in chapter 3.1.2, and programmed these options within the forum: The participants could leave the discussion, explicitly disapprove of the norm violation by flagging or writing a reprimanding comment, or ignore the norm violation, which was operationalized as the absence of leaving and explicit disapproval. All comments written by the participants (*N* = 373) were content analyzed by two researchers to examine whether they contain an explicit disapproval (Krippendorff's *α* = .87).

The results provide empirical evidence for the multidimensional concept of incivility developed in *Article I*. More specifically, participants perceived violations of the communication norms, and evaluated the distinct types as more sanction-worthy than the norm-compliant comment (H1). In addition, the results mostly support the assumption of the two-stage process of disapproval as the effect of the norm violation on sanction-worthiness was at least partially mediated by the perceived violation of the intended communication norm except for the information norm (H2). Yet, in most cases, a direct effect of the norm violation on the evaluation of sanction-worthiness remains, suggesting that even if a norm violation is not correctly recognized, recipients intuitively evaluate it as deviant and sanction-worthy. Moreover, one additional finding is noteworthy: Although each violation was mainly perceived as violating the intended communication norm, in some cases the violation also negatively affected the perception of other norms. Particularly insults and vulgarity were not only perceived as violating the relation norm, but the comment was also perceived as a little less informative (information norm), comprehensibly (modality norm), connective (process norm), and respectful to liberal democratic principles (context norm). Thus, the conceptual differences of the five norms are reflected in the perceptions, but they blur to some extent.

The findings also showed that violations of the relation and context norms are evaluated as more severe than violations of the other norms, and thus support H3. Moreover, insults and vulgarity were rated as the most severe norm violation, and were most likely to elicit explicit disapproval by flagging or writing a reprimanding comment. However, participants exposed to the violations of the other communication norms were also more likely to show explicit disapproval compared to those exposed to norm-compliant comments, which confirms H5. The assumption that participants rather leave the discussion when confronted with norm violations was rejected (H4). Participants were more likely to stay in the discussion. Moreover, the findings suggest that participants tend to use different forms of explicit disapproval depending on the type of norm violation. While participants tended to write reprimanding replies against comments containing violations of the information, process, relation, and context norm, only violations of the relation and context norms led to flagging, with insults and vulgarity being more often flagged as stereotypes and threats of violence against a social group.

3.2.3 Article V: Journalistic Counter-Voices in Comment Sections: Patterns, Determinants, and Potential Consequences of Interactive Moderation of Uncivil User Comments (Ziegele, Jost, Bormann, & Heinbach, 2018)

Article V focuses on responses to incivility by professional participants in public online discussions. More specifically, it addressed three research questions, namely *whether different types of incivility are related to increased interactive moderation (RQ1), how community* *managers respond to different types of incivility (RQ2), and how these responses relate to the level of incivility in the subsequent discussion (RQ3).*

Since the concept of incivility provided in *Article I* had not been finalized at the time of the study, a two-dimensional concept of incivility was applied relating to Muddiman's (2017) model of "personal- and public-level incivility" (p. 3184). Types of incivility that were defined as personal-level incivility in *Article V* can be classified as violations of the relation norm (e.g., insults, name-calling, vulgarity) and types considered public-level incivility can be defined as violations of the context norm (e.g., antagonistic stereotypes, threats of violence against social groups) according to our concept. Hence, I will refer to these two norms in the following.

As already outlined in chapter 3.1.3, to the date of the study presented in *Article V,* research has lacked a systematic overview of different interactive responses of community managers to incivility. Therefore, we provided a typology of interactive moderation of uncivil comments (see Table 2). The typology draws on deliberation (e.g., Friess & Eilders, 2015), behavioral psychology (e.g., Cheng et al., 2014), and prior research on interactive moderation (Grimmelman, 2015; Stroud et al., 2015; Ziegele & Jost, 2020). In addition to the theoretical considerations, a qualitative content analysis of 100 moderation comments of the study's sample was conducted. The result of this deductive and inductive reasoning is the typology in Table 2. It classifies professional participants' reactions to incivility based on their *deliberativeness*, that is, adhering (i.e., deliberative) or not adhering (i.e., non-deliberative) to norms of deliberative discussions such as rationality, constructiveness, and mutual respect, and the *kind of sanction*, that is, either positive (i.e., reward) or negative (i.e., punishment).

Table 2. Typology of Interactive Moderation of Incivility (Ziegele et al., 2018, p. 534).

		Kind of sanction	
		Reward	**Punishment**
Deliberativeness	**Deliberative**	**Discursive moderation** Factually engaging with comments; providing additional information; clarifying questions; adding arguments.	**Regulative moderation** Factually complaining about comments; asking users to behave more civilly; pointing to violations of predefined rules.
	Non-deliberative	**Sociable moderation** Informally complimenting comments; creating an informal and pleasant discussion atmosphere.	**Confrontational moderation** Offensively attacking comments; using irony/sarcasm to expose comments to ridicule.

To answer the research questions, a quantitative manual content analysis of 9,763 comments by lay and professional participants on the Facebook sites of 15 German news outlets was conducted. The sample included private and public service media, national and regional formats, and liberal as well as conservative outlets. In February 2016 and October 2016 all posts and subsequent comments published on the 15 Facebook sites were crawled, and a stratified random sample was selected consisting of (1) lay participants' comments that received a moderation comment, (2) the respective moderation comments, (3) the subsequent reply comments, and (4) comments of the same thread that had not received moderation comments. The resulting $N = 9{,}762$ user and moderation comments were coded by 52 trained undergraduate students regarding violations of the relation and context norms, and the different types of interactive moderation.

Overall, the results revealed that 25% of the lay participants' comments included one or several violations of the relation norm, and 14% contained one or several violations of the context norm. The prevalence of violations of the relation and context norms varied significantly between different media outlets, ranging from 7% (Berliner Morgenpost) to 35% (Deutschlandfunk). The prevalence of moderation comments was quite low in the whole sample and also varied across the media outlets. While the lowest share of moderation comments was 0.08% on the Facebook sites of ZDF and Deutschlandfunk, the radio channels hr-info (4.05%), BR24 (2.55%), and the newspaper Die Welt (1.23%) moderated the highest share of comments.

Regarding RQ1, the results further suggested that violations of the context norm relate to increased responses by professional participants. Violations of the relation norm in comments neither increased nor decreased the likelihood of interactive moderation. Regarding RQ2, the analysis revealed that when responding to uncivil comments, community managers mostly apply discursive moderation (61%), followed by sociable moderation (31%), and less frequently regulative moderation (18%) and confrontational moderation (16%). Compared to responses to civil comments, however, the share of discursive, regulative, and confrontational moderation increased, and the share of sociable moderation declined, suggesting that community managers tend to react to uncivil comments more often with negative sanctions and less often in a sociably style. However, responses to distinct types of norm violations did not differ significantly. Lastly, it was examined how different forms of moderation relate to the level of incivility in the subsequent discussion (RQ3). The results indicated that different moderation styles are associated with different levels of incivility in the reply comments: While a sociable moderation decreased the prevalence of incivility, a regulative moderation increased the level of incivility in the subsequent discussion.

4. Discussion

4.1 Summary of the Results

The overarching research question of this dissertation was: *How can incivility in public online discussions be systematized, how do communication participants perceive it, and how do they react to it?* This dissertation thus aimed at (1) providing a theoretically well-founded systematization of incivility in public online discussions, (2) empirically examining incivility perceptions of participants involved in public online discussions, thereby refining and validating the systematization of incivility, and (3) investigating participants' reactions to incivility in public online discussions. The aims were achieved. The research question was answered within a research program consisting of a methodological triangulation. Qualitative and quantitative research methods were applied within this dissertation, addressing several shortcomings of incivility research and resulting in five research articles:

(1) *Systematization*: Incivility research has been lacking a uniform systematization of the construct that considers the perspective of the communication participants, is theoretically well-founded and empirically applicable. Therefore, in *Article I,* we provided a new theoretical approach to incivility drawing on analytical theories on cooperation, communication, and norms. We developed a new definition and comprehensive typology of incivility in public online discussions based on five communication norms and the disapproval of communication participants. In *Article II,* our new concept of incivility was contextualized within the extant literature, and it was discussed why the concept could be beneficial for future incivility research.

(2) *Perceptions:* Given that only few previous studies have focused on incivility perceptions of different participants in public online discussions, and those that examined perceptions considered either one group of participants such as lay participants (e.g., Kalch & Naab, 2017) or professional participants (e.g., Frischlich et al., 2019), or focused on particular a priori defined types of incivility (e.g., Kenski et al., 2020), the next step was to conduct a qualitative study with different online actors, thereby refining the developed typology. *Article III* presents the results of a series of focus groups with different participants in online discussions, namely lay participants (i.e., ordinary users), semi-professional participants (i.e., online activists collectively combating incivility), and professional participants (i.e., community managers). The results suggest that (a) incivility encompasses violations of all five communication norms, (b) different participants share a quite large common ground as to what they perceive as uncivil, (c) different types of

norm violations are not assessed as equally severe, and (d) several criteria shape the processing of norm violations. To empirically test and validate the concept of incivility and thus the assumptions (a) and (c) derived from the focus groups, an online experiment was conducted within *Article IV*. Participants were confronted with violations of the five communication norms in a mock-up online discussion forum. The results revealed that violations of all five norms are indeed disapproved as uncivil, and that violations of the relation and context norm are evaluated as more severe than violations of the other norms. Since previous studies have not yet systematically investigated the perception of different types of norm violations by participants in online discussions in an experimental setting, another research gap was filled with this study. Furthermore, the study empirically validated the concept of incivility and thus provided future research with a theoretically well-founded and empirically validated systematization of incivility.

(3) *Reactions:* Applying Hirschman's (1970) framework of responses to decline in firms, organizations, and states to lay participants' reactions to incivility, three types of responses were distinguished: ignorance, leaving the discussion, and explicit disapproval. The online experiment presented in *Article IV* suggested that participants are likely to stay in the discussion and show explicit disapproval when exposed to incivility. However, reactions were not uniform: Participants used different forms of explicit disapproval (i.e., writing a reprimanding comment or flagging) in response to distinct types of norm violations. In terms of flagging, for example, violations of the relation and context norm were more likely to receive a flag than the other norm violations. This study also addressed a shortcoming of previous incivility research, as there has not yet been a systematic study of various forms of lay participants' responses to five distinct types of norm violations. Finally, *Article V* addressed reactions to different forms of incivility by professional participants, that is, community managers from different news media outlets. This study was one of the first attempts to systematize and analyze different types of intervening reactions by professional participants to distinct types of incivility in online discussions, thus also addressing a shortcoming of incivility research. Drawing on a typology of different forms of interactive moderation as a response to incivility, results of a content analysis of 15 Facebook sites showed that professional participants react more often to violations of the context norm than to violations pertaining to the relation norm, but that their moderation styles do not differ considerably in response to violations of these two norms. Compared to reactions to norm-compliant comments, however, professional

participants more often apply discursive, regulative, and confrontational moderation in response to uncivil comments, suggesting that they tend to react with more reprimands.

The results of the individual research articles will be discussed in more detail and across studies in the next chapter 4.2. Notably, this dissertation has several limitations, which are outlined in chapter 4.3. Finally, theoretical, empirical, and practical implications are discussed in chapter 4.4.

4.2 Discussion of the Results

4.2.1 Incivility as a Violation of Communication Norms

Overall, the studies provide support for the multidimensional concept of incivility as a disapproved violation of one or several of five communication norms developed in *Article I*.

The focus groups (*Article III*) revealed that different actors of online discussions perceive various types of norm violations in public online discussions (see Figure 4, chapter 3.2.1). Moreover, all of the examples reported by the actors could be classified as violating one of the communication norms. Put differently, the systematic of the five communication norms covered all perceived types of incivility, there was no type that could not be categorized within the systematic. This finding further supports the systematic of five communication norms, although we explicitly emphasized in *Article I* and indicated through the dots in Figure 2 (see chapter 2.3.1) that the systematic is not necessarily exhaustive and that further communication norms might exist.

To empirically validate the incivility concept and assumptions from the qualitative focus groups, the online experiment was conducted (*Article IV*). The results indicated that compared to comments in online discussions that were compliant with the five communication norms, comments that included a violation of the information norm, modality norm, process norm, relation norm, or context norm were disapproved as uncivil by the participants. Thus, the findings of the two empirical studies strengthen the broad concept of incivility encompassing violations of five norms. These results reflect the findings of prior studies on incivility perceptions that also suggested that violations of several norms are perceived as uncivil (e.g., Kalch & Naab, 2017; Kenski et al., 2020; Muddiman, 2017, 2019; Stryker et al., 2016, 2021; Ziegele et al., 2020a). However, previous studies and their underlying incivility concepts are extended: While earlier studies have primarily focused on violations that pertain to the relation norm or context norm (e.g., Kalch & Naab, 2017, pp. 404-405; Kenski et al., 2020, p. 802; Muddiman, 2017, p. 3187, 3194; Muddiman, 2019, pp. 33-34; Ziegele et al., 2020a, p. 740), the results of the focus groups and the online experiment suggest that violations of the information norm, modality norm, and process norm such as lying, sarcasm, and topic deviation are also disapproved as uncivil by participants in online discussions.

The studies also strengthen the theoretical consideration of linking the definition of incivility to the *disapproval* of communication participants. We defined disapproval as (1) perceiving a violation of the communication norms, and (2) evaluating it as sanction-worthy (Bormann et al., 2022, p. 348; see chapter 2.3.1 and Figure 3). From a theoretical perspective, we argued that violations of the communication norms are not always uncivil, but are sometimes unavoidable and tolerated (see also e.g., Grice, 1975, pp. 49-56). Therefore, norm violations are only defined as uncivil when the participants classify them as sanction-worthy. The empirical data support this argument: The focus groups revealed that processing norm violations is multi-layered with several evaluation criteria playing a role. Under certain circumstances participants evaluate norm violations as acceptable, which pertains primarily to violations of the information, modality, and process norms; the tolerance level tends to be lower for violations of the relation and context norms. The participants mentioned, for example, that they find sarcasm sometimes appropriate, such as to defuse a situation that is coming to a head. Or that a white lie was acceptable if it circumvented an insult, suggesting that norm violations are tolerated if they prevent worse norm violations. However, insults are also tolerated under certain circumstances, for example if the participants think that the person “deserves” the insult - examples cited in the focus groups were insults directed at neo-Nazis or pedophiles.

Moreover, the experimental data largely confirmed the two-stage process of disapproval. The effect of the norm violation on the evaluation of sanction-worthiness was in all cases - except for the violation of the information norm - at least partly mediated by the perceived violation of the respective communication norm. Interestingly, a certain direct effect from the norm violation on the sanction-worthiness remained in the case of the modality norm and relation norm, and the effect of the violation of the information norm on the perceived sanction-worthiness was not mediated by the perception of the norm violation. Thus, it can be assumed that even when participants do not recognize the norm violation as intended, or in other words, assign the norm violation to the correct communication norm on the perceptual level (step 1 of disapproval), they intuitively evaluate norm violations as inappropriate and sanction-worthy (step 2 of disapproval). Consequently, it can be reasoned that the second step of disapproval, that is, the negative evaluation of a norm violation, is particularly relevant for the definition of incivility.

Another interesting finding of the experimental study was that participants tend to perceive individual types of norm violations not as distinctly as we conceptually distinguish them. More specifically, the individual types were not only perceived as violating the intended communication norm, but also as violating the other communication norms. In particular, insults and vulgarity were not only perceived as violations of the relation norm, but were also classified as violations of all other norms compared to the norm-compliant comment. In other words, insults and vulgarity are not only perceived as disrespectful towards the communication partner, but also as less informative, less comprehensible, less connective, and as disrespecting liberal democratic principles. The violation of the context norm, i.e., stereotype and threat of violence against a social group, was also rated as less informative, less connective, and as disrespectful to the communication participants. The violation of the modality norm, that is, sarcasm, was also perceived as disrespectful and not considering the context. The violation of the process norm, i.e., topic deviation, was also perceived as less informative. Although all types were perceived as most strongly violating the intended norm, the communication norms are not completely distinguishable on a perceptual level. Rather, individual norm violations also have a negative impact on the perception of the other norms. In terms of our concept of cooperation (*Article I*, see chapter 2.3.1), it can be inferred that even a single type of violation can already hinder cooperative communication and cooperation can fail.

Although the studies provide support for the broad concept of incivility as a violation of one or several of five communication norms, violations of the different norms are not to be considered equally in terms of dissent and severity. The focus groups revealed that particularly violations of the information, modality, and process norms were controversial among different actors. Moreover, both empirical studies suggested that violations of the relation norm and context norm are overall evaluated as more severe than violations of the information, modality, and process norm. This finding could be explained by the target of the norm violations: Contrary to the other norms, violations of the relation and context norms often directly target people – either present communication participants or non-present social groups. Previous research has already suggested to distinguish between different targets of incivility (e.g., Coe et al., 2014, p. 660; Papacharissi, 2004, p. 274; Rowe, 2015, p. 128; Su et al., 2018, p. 3681) and it seems to indeed have an impact on severity evaluations whether humans or non-human objects such as topic-related aspects or organizations are targeted.

Further, in the focus groups, violations of the context norm tended to be evaluated as most severe. In the experiment, however, violations of the relation norm were evaluated as most severe, which reflects the findings of the vast majority of previous studies (e.g., Kenski et al., 2020; Muddiman, 2017; Stryker et al., 2016, 2021). Interestingly, the studies by Muddiman (2017, 2019) revealed quite similar results: When her study participants were asked to provide one example of incivility, they primarily cited violations pertaining to the context norm (Muddiman, 2019, pp. 35-36). In her experimental study, however, violations referring to the relation norm were rated worse (Muddiman, 2017, p. 3197). Accordingly, it can be assumed that participants tend to react more strongly to violations of the relation norm when they *see* it, but violations of the context norm and thus uncivil behavior that endangers broader democratic values and the common good, are *top-of-mind* when they are asked to provide examples. Yet, this result also hints to what Papacharissi (2004) claimed: “impeccable incivility, that should frighten us” (p. 279). Impeccable incivility refers to covert stereotypes and discriminations or attacks against other liberal democratic principles, which we classified as violations of the context norm. According to several scholars, these types of incivility are more difficult to identify than explicit insults because they are not recognizable on the level of words or need certain background knowledge (e.g., Kalch & Naab, 2017, pp. 409-410; Papacharissi, 2004, p. 279; Ross et al., 2018, pp. 3-4; Stoll et al., 2020, pp. 126-139).

In summary, it can be concluded that while violations of all five communication norms in online public discussions are disapproved of as uncivil by different participants, the five norms cannot be considered equal. While violations of the relation and context norms are consistently rated as more severe, violations of the information, modality, and process norms are more controversial among different participants and tend to be rated as less severe. Consequently, and as displayed in Figure 5, the concept of incivility can be assumed as consisting of a core and a somewhat more contested edge. The evaluation of norm violations as (mildly and severely) uncivil, however, depends on various criteria discussed in the following.

4.2.2 Evaluation Criteria

In *Article I*, we already argued that, from a theoretical perspective, the evaluation process of norm violations is relatively complex because of four reasons (Bormann et al., 2022, pp. 349-350): (1) Communicative acts can be disrespectful and at the same time informative and comprehensible. Thus, participants always have to weigh and interpret the violations of or compliance with the different *communication norms* and arrive at an overall evaluation whether the act is civil or uncivil. (2) Given that norms are subject to zeitgeist, evaluations of norm violations can also change over *time*. Finally, evaluations can differ between (3) specific *situations*, and (4) *participants* because of various individual characteristics. These theoretical considerations can be complemented and differentiated by the focus group data (*Article III*, see chapter 3.2.1). The results suggested that several criteria seem to shape the evaluation of norm violations. These criteria were classified drawing on Lasswell's (1948) model of communication:

(1) *Sender (Who?):* The participants attribute various characteristics to the sender of a norm violating comment, which seem to have an impact on the evaluation of the norm violation. They consider, for example, whether the sender principally shows an intention to discuss, whether she or he is a repeated offender, and whether the norm violations was intended or not intended. Previous research has already indicated that attribution processes play an essential role in the processing of discussion comments (Kluck & Krämer, 2021; Kluck et al., 2021). Applying our concept of incivility and drawing on assumptions on cooperative communication and attribution processes, Kluck et al. (2021) found that participants' attributions about the (a) discussion intentions, (b) tolerable political views, and (c) discussion skills of the sender were often lower when exposed to incivility. Moreover, attributions differed among violations of the five communication norms, with especially violations of the relation norm leading to negative attributions on all three levels. In an additional study, Kluck and Krämer (2021) showed that negative attributions had an impact on the evaluation of the sender and on participants' response intentions. More specifically, their results revealed that participants attribute relatively high levels of aggressive discussion motives to uncivil senders, which results in generally negative evaluations of the sender, and in participants being less willing to participate. These studies provide valuable insights into the role of attribution processes in the processing of different types of incivility, further supporting to approach incivility in a multidimensional way and from a perceptual perspective (for an overview, see Kluck, 2021).

(2) *Message (says what?):* Participants evaluate several message characteristics and specifics of the norm violation when evaluating whether it is civil or uncivil. Participants consider, for example, the number of violations, whether it contains constructive elements besides violations, and the target of the norm violation, which has already been discussed to be important in prior research (e.g., Coe et al., 2014, p. 660; Papacharissi, 2004, p. 274; Rowe, 2015, p. 128; Su et al., 2018, p. 3681).

(3) *Context (in which channel?):* Norm violations are evaluated differently depending on the context. This includes, for example, broad contextual aspects such as the cultural context and national law, but also more specific aspects such as platform design, genre of the medium, characteristics of the community, and the discussion tone and topic. Several studies have already indicated that prevalence, perceptions, and effects of incivility vary

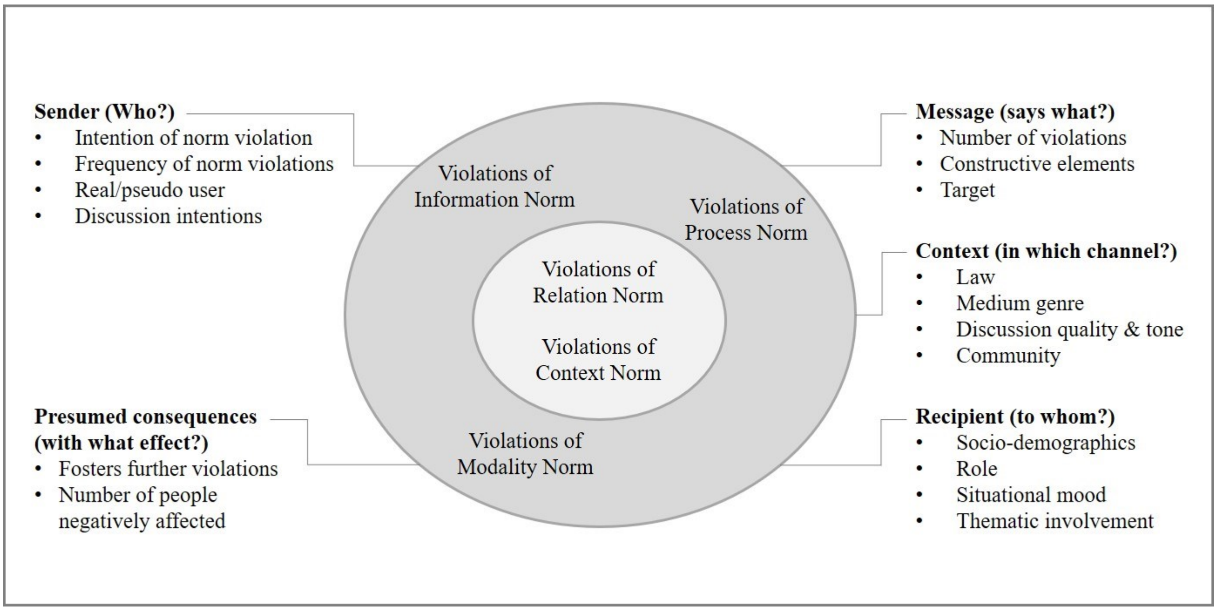
between contexts, such as between different topics (e.g., Coe et al., 2014; Stroud et al., 2015), between different platforms and media (e.g., Sobieraj & Berry, 2011; Sydnor, 2018), and between specific platform designs or elements (e.g., Esau et al., 2017; Santana, 2014; Ziegele et al., 2018).

(4) *Recipient (to whom?):* Individual characteristics of the recipient also seem to have an impact on the evaluation process, such as socio-demographics, the situational mood, and the thematic involvement. Initial studies have already suggested that, for example, different personality traits and gender (e.g., Kenski et al., 2020, pp. 807-809) as well as the political views and party identification (e.g., Gervais, 2015, p. 181; Muddiman, 2017, p. 3197) affect the processing of incivility.

(5) *Presumed consequences (with what effect?):* Participants consider potential effects of the norm violation when they evaluate it. For example, they estimate whether the norm violations could foster further violations and thus lead to an incivility spiral. Moreover, it is assessed how many people are potentially affected by the norm violation. This aspect has received scant study in incivility research so far.

In summary, it can be concluded that the processing of norm violations is multi-layered with several criteria consciously or subconsciously playing a role. From the theoretical considerations and empirical findings, a multi-dimensional model of incivility can be drawn that includes a core and an edge of norm violations as well as various criteria that potentially affect the evaluation of these norm violations (see Figure 5). Since violations of the relation and context norms tended to be evaluated as the most severe in the focus groups and experimental study, elicited the strongest reactions, and the least disagreement among different participants in online discussions, it can be concluded that they form the core of incivility. Violations of the information, modality, and process norms tended to be rated as less severe, elicited less explicit disapproval, and were more contentious among different participants, and can therefore be considered as the edge of incivility. However, the various criteria depicted in Figure 5 seem to play a role in the evaluation of norm violations and presumably in the reactions to different norm violations. For example, under certain circumstances, a violation of the information norm may be rated worse than a violation of the relation norm. Tendentially, however, the results from the studies suggest that the impact of the criteria tends to diminish from the edge to the core of incivility. In other words, violations of the relation and context norms tend to be more consistently rated as severe than violations of the other norms, also under varying circumstances. As the criteria are inductively derived from qualitative data, they require further research. Future studies should investigate which criteria have what impact on the evaluation of and response to violations of the five norms and also whether certain criteria, such as characteristics of the recipient, have a greater impact than others, such as contextual criteria.

Figure 5. Core and Edge of Incivility with Criteria Shaping the Evaluation Process of Norm Violations.


Sender (Who?)
Intention of norm violation
Frequency of norm violations
Real/pseudo user
Discussion intentions
Presumed consequences (with what effect?)
Fosters further violations
Number of people negatively affected
Violations of Information Norm
Violations of Process Norm
Violations of Relation Norm
Violations of Context Norm
Violations of Modality Norm
Message (says what?)
Number of violations
Constructive elements
Target
Context (in which channel?)
Law
Medium genre
Discussion quality & tone
Community
Recipient (to whom?)
Socio-demographics
Role
Situational mood
Thematic involvement

4.2.3 Reactions to Incivility

Just as distinct types of norm violations are evaluated differently, they also elicit partly different responses from the participants in public online discussions.

First of all, the experiment presented in *Article IV* revealed that lay participants were more likely to stay in the discussion and express their disapproval when exposed to a norm violating comment. From a democratic and deliberative perspective, this is a desirable result because it suggests that political participation is not compromised by incivility (e.g., Gastil, 2008; Rowe, 2015; Ruiz et al., 2011). However, several other studies indicated that incivility, particularly when combined with political disagreement, impedes participation or willingness to participate (e.g., Hwang et al., 2008; Kluck & Krämer, 2021; Lück & Nardi, 2019; Pang et al., 2016). Two explanations could account for the conflicting result. First, the programmed discussion forum *Let's discuss* featured an innovative platform design with specific rules that explicitly asked participants, among others, to interactively discuss, which might have motivated the participants to stay in the discussion and write a comment. As learned from the focus group data and prior research (e.g., Esau et al., 2017), the design of the platform matters when processing a norm violation and regarding the general quality of the discussion (see chapter 4.1.2). Thus, the platform specifics might have also had an impact on how lay participants respond to norm violations. Second, the two comments displayed before the norm violating comment were norm-compliant. Therefore, the majority of comments was norm-compliant, and all participants responded to each other so far. According to "social learning theory" (Bandura, 1986)*,* it can be argued that participants learn and adopt discussion behavior by observing those of other participants. Moreover, based on the "focus theory of normative conduct" (Cialdini et el., 1991), it could be argued that the descriptive, salient norms might have been to actively participate and behave norm-compliant. This could also explain why lay participants were likely to write reprimanding comments against almost all types of norm violations (expect for violations of the modality norm) and why these reprimanding comments were mostly compliant with the communication norms.

In terms of ignorance and flagging, differences became evident. Only violations of the relation and context norms were likely to be flagged by lay participants, and these two types of violations were also least likely to be ignored. Moreover, violations of the relation norm received more flags than those of the context norm. Lay participants seem to prescribe different functions to different forms of explicit disapproval. Flagging might be seen as a harsh sanction as it usually has immediate and serious consequences for the uncivil commenter, and therefore might only be executed when the norm violations are evaluated as quite severe. Other studies also suggested that participants are more likely to flag severe, explicit norm violations directed at humans (e.g., Kalch & Naab, 2017, pp. 409-410; Wilhelm et al., 2020, p. 934, 936).

The content analysis examining professional participants' reactions also suggested that they do not consistently engage with two different forms of incivility, that is, violations of the relation norm and context norm. Only violations pertaining to the context norm were associated with an increased level of responses. Violations of the relation norm neither increased nor decreased the likelihood of professional participants' reactions. One explanation could be that community managers, in their professional role as representatives of a medium, might focus above all on accordance with liberal democratic principles and the democratic merit of an online discussion, thereby performing their role as producers and preservers of a democratic public discourse (e.g., Chen & Pain, 2017, p. 888; Frischlich et al., 2018, pp. 2017-2018; Ziegele et al., 2018, p. 546), which became also apparent in the focus groups. The style of responses, however, did not significantly differ between violations of the two norms.

4.3 Limitations

Overall, the research program with methodological triangulation conducted within this dissertation offered several advantages and provided multi-faceted insights into the research subject. However, it is not without limitations. Several relevant aspects raised in the different studies could not be further examined within the dissertation. For example, the criteria that presumably shape evaluation processes inductively developed from the focus group data, could not be further investigated in experimental studies. Moreover, instruments measuring perceptions of violations of the communication norms (scales) or the occurrence of violations (codebook) could not be validated or developed, which would have been relevant to provide future research with not only a theoretical concept of incivility but also empirically validated instruments measuring the concept. It would have been also important, for example, to conduct further content analyses to examine the real-life reactions of semi-professional and lay participants to various uncivil comments. In addition, experiments with semi-professional and professional actors would have provided valid findings on the disapproval of and responses to different types of norm violations. Conducting a field experiment instead of the online experiment would have increased the external validity of the results. And a quantitative, representative survey could have been conducted to examine perceptions of the vast amount of norm violations identified in the focus groups and also to be able to test the incivility model using confirmatory factor analyses (following Stryker et al., 2016, 2021). However, all of these aspects could not be implemented in the research program due to time and cost constraints.

Furthermore, the individual research articles have several shortcomings, which also lead to overall limitations of this dissertation and the incivility model illustrated in Figure 5.

Particularly noteworthy is the restriction of all studies to the German-speaking population or to comment sections of German media outlets. Hence, the perception of norm violations and reactions to it were investigated in a specific cultural context. Since norms are highly context-sensitive and the cultural background presumably also impacts the perception of norm violations, the results cannot be generalized to other countries and cultures. Especially liberal democratic principles are interpreted differently in various countries since every nation has developed its own specific constitutional culture (Müller, 2007, pp. 56-58) based on historical experiences and cultural developments. For example, liberty rights are formalized in different ways in various democratic states: Several statements protected by freedom of speech in the USA are prosecuted in Canada or Germany (Michelman, 1999, pp. 1025-1026). Accordingly, it can be concluded that violations of the context norm in particular are perceived differently in different countries. Future studies should examine perceptions of violations of the five communication norms in other countries, and conduct cross-national studies.

With regard to the focus groups (*Article III*), it should also be mentioned that the sample was not only limited to the German-speaking population, but participants were also predominantly highly educated and came from a similar social and political milieu. Future studies should investigate whether the results also count for participants with lower education levels from heterogeneous milieus. Since the results of the focus groups in terms of incivility perceptions and evaluation criteria are based on qualitative data, they cannot be generalized. The online experiment partly validated some assumptions derived from the focus groups. However, the criteria that presumably shape the evaluation process of norm violations were not experimentally tested and should thus be systematically examined in future experimental studies. Furthermore, the focus group findings on differences and similarities between the incivility perceptions of different actors of online discussions (i.e., ordinary users, online activists, and community managers) have not been studied experimentally. Future studies should experimentally examine different actors of online discussions, their perceptions, and their reactions to incivility.

The experiment (*Article IV*) has also several limitations. Above all, the discussion forum *Let's discuss* was not completely comparable to a real discussion forum, because we had to implement specific discussion rules to ensure that the participants react to the manipulated comment. Moreover, we could only test one representative type of violation for each communication norm. Thus, the results are limited to one specific type of norm violation per norm and could be different if we had selected other types of violations for the five norms. Future studies should therefore examine and compare perceptions of multiple types of violations of the five norms. In addition, comparisons of several violations of one communication norm would be interesting. For example, is discrimination of a social group evaluated as more severe than calls to the overthrow a democratically elected government by force? In addition, only one severity level and one norm violation within a comment were tested; the severity level and number of norm violations should also be varied in future studies. Finally, a scale had to be developed to measure perceptions of norm violations. Scale reliability was very satisfying, but the scale should be validated for future studies.

The content analysis (*Article V*) also exhibits multiple limitations, however, one of the most important ones with regard to this dissertation is that the study was conducted before the concept of incivility (*Article I*) was fully developed. Hence, it applied a two-dimensional approach to incivility that included some violations pertaining to the relation norm and violations that can be classified as violating the context norm, and did not examine violations of all five communication norms. Similarly, the typology of responses to incivility by professional participants in online discussions was aligned neither with the systematics of communication norms and norm violations nor with the differentiation of responses to norm violations developed and applied within *Article IV*. For more precise insights into reactions to different norm violations by professional participants, further studies would be necessary.

Despite its limitations, the dissertation has some relevant implications for research and practice, which will be outlined in the following chapter.

4.4 Implications for Research and Practice

Demands for a unified concept of incivility that is both theoretically well-founded and empirically applicable have strongly increased in recent years (e.g., Boatright, 2019, p. 4; Chen et al., 2019, pp. 1-2; Jamieson et al., 2018, p. 206, p. 213; Masullo, in press; Muddiman, 2017, p. 3183). A lack of conceptual clarity leads to inconsistent assumptions about the prevalence, causes, and consequences of incivility in public online discussions. As a consequence, it is difficult to determine normative implications of incivility and to develop and legitimize intervention measures. The dissertation project therefore aimed at developing a concept of incivility that meets the theoretical and empirical requirements of incivility research. The results have several implications for research and practice.

The *theoretical contribution* of this dissertation is the theoretically well-founded and empirically refined and validated concept of incivility consisting of a definition and comprehensive typology that build on a newly developed theoretical framework of cooperative communication. The dissertation thus provides conceptual clarity on the phenomenon of incivility and strengthens the theoretical foundation of incivility research. By approaching incivility as a violation of communication norms and providing a fine-grained systematization of communication norms and violations, the concept incorporates previous concepts into an integrative framework. Although the concept is inclusive and broad, it does not inherently conflate or intermingle distinct types of incivility, which is a common critique of prior incivility concepts (e.g., Chen et al., 2019, pp. 1-2; Hopp, 2019, pp. 203-206; Jamieson et al., 2018, pp. 207-208; Kluck, 2021, p. 5; Masullo, in press; Rossini, 2020, p. 2). Instead, different types of incivility are treated as such by systematizing them along five conceptually distinct communication norms. Furthermore, the concept follows a consistent perceptual approach. It is anchored in the definition that the communication participants decide what is civil and what is uncivil. This implies a theoretical shift in incivility research: Although scholars have largely agreed that incivility is in the eye of the beholder (e.g., Chen, 2017, p. 5; Chen et al., 2019, p. 2; Coe et al., 2014, p. 660; Jamieson et al., 2018, p. 206; Kalch & Naab, 2017, p. 400; Kluck & Krämer, 2021, p. 3; Sydnor, 2018, p. 97), the perception aspect has not yet been consistently anchored in incivility models.

Finally, the theoretical foundation developed in *Article I* provides a theoretical contribution that goes beyond incivility research and can be beneficial for communication science in general. Drawing on theoretical considerations regarding cooperation, communication, and norms (e.g., Grice, 1975; Tomasello, 2019), we developed an empirical-analytical approach as an alternative to prescriptive approaches that prescribe norms in communication processes against the backdrop of normative ideals, such as politeness theories or normative theories of democracy (e.g., Brown & Levinson, 1987; Habermas, 1996). Yet, we derive and justify the five communication norms based on their function for cooperative communication, which is a communication that enables cooperation. We argue that cooperation and thus cooperative communication are the key elements in almost all human relationships, from families to societies and political processes and policy making (Tomasello, 2008, 2009, 2019). In doing so, our framework proposes a methodological shift by approaching communication norms and norm violations empirical-analytically instead of prescribing them. Moreover, approaching incivility as a perceptual construct, requires to consequently consider the perceptions of the communication participants when conducting empirical studies, also in textual studies such as quantitative manual content analyses.

The dissertation has also several *empirical contributions*. First, the dissertation provides an empirically refined and validated concept of incivility that can be used as a research model in future empirical studies. The concept enables incivility research to measure different facets of incivility in various political offline and online contexts. More specifically, the typology allows a much more differentiated and precise determination of the prevalence, causes, perceptions, reactions, and consequences of distinct types of norm violations in public online discussions, but also in other contexts. This can result in a profound and nuanced understanding of incivility, its normative implications and consequences for democratic societies. The typology can serve, for example, as the basis for a coding scheme in content analyses, for training algorithms that automatically detect incivility in online discussions, for operationalizing stimuli in experimental studies, and for developing standardized indicators in surveys.

Second, the focus groups and online experiment (*Article III*, *Article IV*) shed light on the perceptions of norm violations of different participants in public online discussions. The findings not only showed that violations of some norms are perceived as more harmful than violations of other norms, but also indicated how complex the processing of norm violations is. Future studies can build on the results and, for example, examine in more detail the differences in the perceptions of norm violations by different participants, or systematically investigate the influence of the evaluation criteria on the evaluation of norms and norm violations. Several hypotheses can be derived from the results of the focus groups for these purposes, which can be tested in quantitative studies. Furthermore, the quite innovative method of heterogeneous focus groups proved to be very fruitful and provided valuable insights into perceptions and evaluations of incivility by various online actors. As such, the conception and implementation of the study can serve as an orientation for future studies applying qualitative methodologies. Finally, the scale developed to measure perceptions of conformity with or violation of the five communication norms could be empirically validated in future studies and then used as a standardized measurement for the systematics of norms and norm violations in surveys and experiments.

Third, the experiment and content analysis data (*Article IV*, *Article V*) provided valuable insights into lay and professional participants' reactions to different types of incivility. Both studies suggested that reactions to different types of incivility are not uniform. When eliciting different responses, one could also assume that different types of incivility might have different effects in terms of, for example, participants' emotions and attitudes or discussion dynamics. The results can serve as starting points for systematically studying reactions to different norm violations by various participants including lay, semi-professional and professional online actors, and other effects of different violations on these participants.

Overall, the triangulation employed in this dissertation has proven to be very fruitful: first developing a theoretical approach and then enriching and testing it with different methods has yielded multi-layered and valid findings on the systematization and perception of incivility as well as reactions to it. Without conducting the qualitative study, for example, it would not have become apparent how broad and multifaceted the phenomenon of incivility is. The inductive development of various criteria that seem to have an impact on the evaluation of norm violations and that could provide explanations for differences in incivility evaluations was also a relevant output of the qualitative study. However, qualitative studies cannot identify causal effects and test hypotheses, which is why the experiment was a valuable addition. The experiment confirmed several assumptions of the focus group study and validated the concept of incivility. A vast amount of actual responses to incivility in real online discussions, on the other hand, could be examined descriptively and systematically with the content analysis, which yielded additional insights. The dissertation could thus serve as an orientation for future research programs employing triangulation.

After all, the theoretical and empirical findings of the research articles can serve as a starting point for a *research program on communication norms and norm violations in online and offline political communication*. The program could investigate how the communication norms are manifested under which conditions, how they are enforced in different contexts, and under which circumstances norm violations occur. In addition, it could be systematically investigated under which circumstances norm violations are tolerated and when they are disapproved of as uncivil by whom. Such a research program would provide valuable insights into political communication processes in general.

The results of this dissertation also have some *implications for practice,* particularly for platform providers and community managements in terms of intervention measures.

First, the typology could be used in moderation practice to identify and systematize different types of incivility. Based on the typology, training sessions for community managers could also be designed and conducted, for example, on what all is considered uncivil and what comments could be responded to in what form. In the focus groups, community managers expressed that they would find such formats helpful. In addition, the systematics of communication norms and typology of norm violations could serve as a basis for developing discussion rules on online platforms, i.e., netiquettes.

The typology can not only support with the manual identification of norm violations, but can also be used to develop algorithms that would automatically detect incivility in online discussions. Previous approaches to machine learning either focused primarily on types of incivility that can be detected at the word level such as name-calling, or do not work well for more subtle forms of incivility (e.g., Davidson et al., 2017; Risch & Krestel, 2020; Stoll et al., 2020). However, if a comprehensive codebook is developed from the typology and manually coded data is available, algorithms could be trained using the data and possibly work better for detecting multiple forms of incivility in the future.

In addition, based on the typology and on the results from the empirical studies, tailored forms of intervention can be developed and tested for different types of incivility. For example, the results from the focus groups and the experiment showed that violations of the relation and context norms are evaluated as most problematic and sanction-worthy. In these cases, lay participants (i.e., ordinary user) are likely to expect stricter interventions than for violations of the other three norms, and consequently, are likely to perceive stronger sanctions as legitimate. Thus, community managers should always intervene against violations of the relation and context norms and could respond with reprimanding comments, threatening sanctions, or directly carrying out sanctions such as deleting or modifying comments or blocking the uncivil commenter. Violations of the information, modality, and process norms, on the other hand, could be met with milder interventions, such as clarifying the misinformation, writing a reminder to please stay on topic, or a friendly inquiry as to whether a sarcastic comment could be communicated again in a clearly comprehensible manner. Such tailored and graded interventions are also relevant regarding the criticism on general restrictions on freedom of speech or attempts to silence or exclude certain social groups (e.g., Baez & Ore, 2018, pp. 331-332; Chen et al., 2019, pp. 2-3; Papacharissi, 2004, p. 266; Lozano-Reich & Cloud, 2009, pp. 223-225). The results suggest that stronger interventions should only be exercised, and thus might only be legitimized for specific violations.

Lastly, the results from the experiment suggest that lay participants also intervene against incivility, i.e., respond with explicit disapproval. From a democratic theory perspective, this is a very desirable outcome, because it would indicate that incivility does not hinder participation and that a community could in principle regulate itself. To encourage engagement against norm violations and promote communication norms, platform operators could, for example, implement innovative platform designs and clear requests for assistance in the event of perceived norm violations. Thereby, they could involve their community in the moderation process and take joint action against incivility.

References

Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2014). The "nasty effect:" Online incivility and risk perceptions of emerging technologies. *Journal of Computer-Mediated Communication, 19*(3), 373–387. https://doi.org/10.1111/jcc4.12009

Anderson, C. A., & Carnagey, N. L. (2004). Violent evil and the general aggression model. In A. Miller (Ed.), *The social psychology of good and evil* (pp. 168–192). Guilford Publications.

Baez, K. L., & Ore, E. (2018). The moral imperative of race for rhetorical studies: on civility and walking-in-white in academe. *Communication and Critical/Cultural Studies, 15*(4), 331–336. https://doi.org/10.1080/14791420.2018.1533989

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall, Inc.

Barber, B. R. (1984). *Strong democracy: Participatory politics for a new age*. University of California Press.

Bejan, T. (2017). *Mere civility: Disagreement and the limits of toleration.* Harvard University Press. https://doi.org/10.4159/9780674972728

Benhabib, S. (1992). Models of the public space. In C. Calhoun (Ed.), *Habermas and the public sphere* (pp. 73–98). MIT Press.

Ben-Porath, E. N. (2010). Interview effects: Theory and evidence for the impact of televised political interviews on viewer attitudes. *Communication Theory*, *20*(3), 323–347. https://doi.org/10.1111/j.1468-2885.2010.01365.x

Benson, T. W. (2011). The rhetoric of civility: Power, authenticity, and democracy. *Journal of Contemporary Rhetoric, 1*(1), 22–30.

Bergström, A., & Wadbring, I. (2014). Beneficial yet crappy: Journalists and audiences on obstacles and opportunities in reader comments. *European Journal of Communication, 30*(2), 137–151. https://doi.org/10.1177/0267323114559378

Bjorklund, W. L., & Rehling, D. L. (2009). Student perceptions of classroom incivility. *College Teaching, 58*(1), 15–18. https://doi.org/10.1080/87567550903252801

Blanchard, A. L., & Markus, M. L. (2004). The experienced sense of a virtual community: Characteristics and processes. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems, 35*(1), 64–79. https://doi.org/10.1145/968464.968470

Boatright, R. G. (2019). A crisis of civility? In R. G. Boatright, T. Shaffer, S. Sobieraj, & D. Goldthwaite Young (Eds.), *A crisis of civility? Political discourse and its discontents* (pp. 1–6). Routledge.

Borah, P. (2013). Interactions of news frames and incivility in the political blogosphere: Examining perceptual outcomes. *Political Communication, 30*(3), 456–473. https://doi.org/10.1080/10584609.2012.737426

Borah, P. (2014). Does it matter where you read the news story? Interaction of incivility and news frames in the political blogosphere. *Communication Research*, *41*(6), 809–827. https://doi.org/10.1177/0093650212449353

Bormann, M. (2022). Perceptions and evaluations of incivility in public online discussions – Insights from focus groups with different online actors. *Frontiers in Political Science*, *4*(812145). https://doi.org/10.3389/fpos.2022.812145

Bormann, M., Heinbach, D., & Kluck, J. P. (2022). *Perceptions of and reactions to different types of incivility in public online discussions – Results of an online experiment.* Manuscript submitted for publication.

Bormann, M., Tranow, U., Vowe, G., & Ziegele, M. (2022). Incivility as a violation of communication norms – A typology based on normative expectations toward political communication. *Communication Theory,* *32*(3), 332–362. https://doi.org/10.1093/ct/qtab018

Bormann, M., & Ziegele, M. (in press). Incivility. In C. Strippel, S. Paasch-Colberg, M. Emmer, & J. Trebbe (Eds.), *Challenges and perspectives of hate speech analysis. An interdisciplinary anthology*. Digital Communication Research.

Brooks, D. J., & Geer, J. G. (2007). Beyond negativity: The effects of incivility on the electorate. *American Journal of Political Science, 51*(1), 1–16. https://doi.org/10.1111/j.1540-5907.2007.00233.x

Brown, P., & Levinson, S. C. (1987). *Politeness. Some universals in language usage.* Cambridge University Press.

Buckels, E. E., Trapnell, P. D., & Paulhus, D. L. (2014). Trolls just want to have fun. *Personality and Individual Differences,* *67*, 97–102. https://doi.org/10.1016/j.paid.2014.01.016

Bullock, S. C. (2019). The patron saint of civility? Benjamin Franklin and the problems of civil discourse. In R. G. Boatright, T. Shaffer, S. Sobieraj, & D. Goldthwaite Young (Eds.), *A crisis of civility? Political discourse and its discontents* (pp. 176–187). Routledge.

Cahoon, L. (2000). Civic meetings, cultural meetings. In L. S. Rouner (Ed.), *Civility* (pp. 40–64). Notre Dame University Press.

Chen, G. M. (2015). Losing face on social media: Threats to positive face lead to an indirect effect on retaliatory aggression through negative affect. *Communication Research, 42*(6), 819–838. https://doi.org/10.1177%2F0093650213510937

Chen, G. M. (2017). *Online incivility and public debate: Nasty talk*. Palgrave Macmillan.

Chen, G. M., & Lu, S. (2017). Online political discourse: Exploring differences in effects of civil and uncivil disagreement in news website comments. *Journal of Broadcasting & Electronic Media*, *61*(1), 108–125. https://doi.org/10.1080/08838151.2016.1273922

Chen, G. M., Muddiman, A., Wilner, T., Pariser, E., & Stroud, N. J. (2019). We should not get rid of incivility online. *Social Media and Society*, *5*(3), 1–5. https://doi.org/10.1177/2056305119862641

Chen, G. M., & Ng, Y. M. M. (2017). Nasty online comments anger you more than me, but nice ones make me as happy as you. *Computers in Human Behavior*, *71*, 181–188. https://doi.org/10.1016/j.chb.2017.02.010

Chen, G. M., & Pain, P. (2017). Normalizing online comments. *Journalism Practice, 7*(11), 876–892. http://dx.doi.org/10.1080/17512786.2016.120595

Chen, G. M., Pain, P., Chen, V. Y., Mekelburg, M., Springer, N., & Troger, F. (2020). 'You really have to have a thick skin': A cross-cultural perspective on how online harassment influences female journalists. *Journalism, 21*(7), 877–895. https://doi.org/10.1177/1464884918768500

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 24, pp. 201–234). Academic Press.

Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, *64*(4), 658–679. https://doi.org/10.1111/jcom.12104

Coppedge, M., Gerring, J., Altman, D., Bernhard, M., Fish, S., Hicken, A., Kroenig, M., Lindberg, S. I., McMann, K., Paxton, P., Semetko, H. A., Skaaning, S.-E., Staton, J., & Teorell, J. (2011). Conceptualizing and measuring democracy: A new approach. *Perspectives on Politics, 9*(2), 247–267. https://doi.org/10.1017/S1537592711000880

Costello, M., Hawdon, J., & Ratliff, T. N. (2017). Confronting online extremism: The effect of self-help, collective efficacy, and guardianship on being a target for hate speech. *Social Science Computer Review, 35*(5), 587–605. https://doi.org/10.1177/0894439316666272

Crawford, K., & Gillespie, T. (2016) What is a flag for? Social media reporting tools and the vocabulary of complaint. *New Media & Society, 18*(3), 410–428. https://doi.org/10.1177/1461444814543163

Dahlberg, L. (2001). The Internet and democratic discourse. Exploring the prospects of online deliberative forums extending the public sphere. *Information, Communication & Society, 4*(4), 615–633. https://doi.org/10.1080/13691180110097030

Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). *Automated hate speech detection and the problem of offensive language*. https://arxiv.org/abs/1703.04009

Diakopoulos, N. A., & Naaman, M. (2011). Towards quality discourse in online news comments. In *CSCW '11 Proceedings of the ACM 2011 conference on Computer supported cooperative work* (pp. 133–142). ACM. https://doi.org/10.1145/1958824.1958844

Diamond, L. (1999). *Developing democracy: Toward consolidation.* Johns Hopkins University Press.

Dohle, M., & Vowe, G. (Eds.) (2014). *Politische Unterhaltung – unterhaltende Politik. Forschung zu Medieninhalten, Medienrezeption und Medienwirkungen* [Political entertainment - entertaining politics. Research on media content, media reception and media effects]. Herbert von Halem.

Dowding, K. (2016). Albert O. Hirschman, exit, voice and loyalty: Responses to decline in firms, organizations, and states. In M. Lodge, E. C. Page, & S. J. Ball (Eds.), *The Oxford handbook of classics in public policy and administration* (pp. 256–271). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199646135.013.30

Dryzek, J. S. (2000). *Deliberative democracy and beyond: Liberals, critics, contestations*. Oxford University Press.

Dryzek, J. S., Bächtiger, A., Chambers, S., Cohen, J., Druckman, J. N., Felicetti, A., Fishkin, J. S., Farrell, D. M., Fung, A., Gutmann, A., Landemore, H., Mansbridge, J., Marien, S., Neblo, M. A., Niemeyer, S., Setälä, M., Slothuus, R., Suiter, J., Thompson, D., & Warren, M. E. (2019). The crisis of democracy and the science of deliberation. *Science, 363*(6432), 1144–1146. https://doi.org/10.1126/science.aaw2694

Esau, K., Friess, D., & Eilders, C. (2017). Design matters! An empirical analysis of online deliberation on different news platforms. *Policy & Internet*, *9*(3), 321–342. https://doi.org/10.1002/poi3.154

Ferree, M. M., Gamson, W. A., Gerhards, J., & Rucht, D. (2002). Four models of the public sphere in modern democracies. *Theory and Society, 31*(3), 289–324. https://www.jstor.org/stable/658129

Fraser, B. (1990). Perspectives on politeness. *Journal of Pragmatics, 14*(2), 219–236. https://doi.org/10.1016/0378-2166(90)90081-N

Fraser, N. (1990). Rethinking the public sphere: A contribution to the critique of actually existing democracy. *Social Text*, *25/26*, 56–80. https://doi.org/10.2307/466240

Friess, D., & Eilders, C. (2015). A systematic review of online deliberation research. *Policy & Internet*, *7*(3), 319–339. https://doi.org/10.1002/poi3.95

Friess, D., Ziegele, M., & Heinbach, D. (2021). Collective civic moderation for deliberation? Exploring the links between citizens' organized engagement in comment sections and the deliberative quality of online discussions. *Political Communication*, *38*(5), 624–646. https://doi.org/10.1080/10584609.2020.1830322

Frischlich, L., Boberg, S., & Quandt, T. (2019). Comment sections as targets of dark participation? Journalists' evaluation and moderation of deviant user comments. *Journalism Studies, 20*(14), 2014–2033. https://doi.org/10.1080/1461670X.2018.1556320

Garton Ash, T. (2016). Redefreiheit: Prinzipien für eine vernetzte Welt [Free speech: Ten principles for a connected world]. Hanser.

Gastil, J. (2008). *Political communication and deliberation.* Sage Publications.

Gervais, B. T. (2015). Incivility online: Affective and behavioral reactions to uncivil political posts in a web-based experiment. *Journal of Information Technology & Politics, 12*(2), 167–185. https://doi.org/10.1080/19331681.2014.997416

Gervais, B. T. (2017). More than mimicry? The role of anger in uncivil reactions to elite political incivility. *International Journal of Public Opinion Research, 29*(3), 384–405. https://doi.org/10.1093/ijpor/edw010

Geschke, D., Klaßen, A., Quent, M., & Richter, C. (2019). *#Hass im Netz: Der schleichende Angriff auf unsere Demokratie* [#Online hate: The creeping attack on our democracy]. https://www.idz-jena.de/fileadmin/user_upload/_Hass_im_Netz_-_Der_schleichende_Angriff.pdf

Goffman, E. (1955). On face-work. *Psychiatry, 18*(3). 213–231. https://doi.org/10.1080/00332747.1955.11023008

Goffman, E. (1967). *Interaction ritual: Essays on face-to-face behavior*. Doubleday.

Graham, S. L. (2007). Disagreeing to agree: Conflict (im)politeness and identity in a computer-mediated community. *Journal of Pragmatics*, *39*(4), 742–759. https://doi.org/10.1016/j.pragma.2006.11.017

Graham, T., & Witschge, T. (2003). In search of online deliberation: Towards a new method for examining the quality of online discussions. *Communications*, *28*(2), 173–204. https://doi.org/10.1515/comm.2003.012

Grice, P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics* (pp. 41–58). Academic Press. https://doi.org/10.1163/9789004368811_003

Grice, P. (1989). *Studies in the way of words.* Harvard University Press.

Grimmelmann, J. (2015). The virtues of moderation. *Yale Journal of Law and Technology, 17*(1), 42–109.

Gunther, A. C. (2017). Hostile media effect. In P. Rössler, C. A. Hoffner, & L. van Zoonen (Eds.), *The international encyclopedia of media effects* (pp. 1–10). Wiley-Blackwell.

Gutmann, A., & Thompson, D. F. (2004). *Why deliberative democracy?* Princeton University Press. https://doi.org/10.1515/9781400826339

Habermas, J. (1962/1989). *The structural transformation of the public sphere: An inquiry into a category of a bourgeois society* (trans. by T. Burger and F. Lawrence). MIT Press.

Habermas, J. (1990). *Moral consciousness and communicative action*. MIT Press.

Habermas, J. (1991). The public sphere. In C. Mukerji, & M. Schudson (Eds.), *Rethinking popular culture: Contemporary perspectives in cultural studies* (pp. 398–404). University of California Press.

Habermas, J. (1996). *Between facts and norms: Contributions to a discourse theory of law and democracy.* MIT Press. https://doi.org/10.7551/mitpress/1564.001.0001

Heath, R. G., & Frey, L. R. (2004). Ideal collaboration: A conceptual framework of community collaboration. *Annals of the International Communication Association, 28*(1), 189–231. https://doi.org/10.1080/23808985.2004.11679036

Hennink, M., Hutter, I., & Bailey, A. (2020). *Qualitative research methods* (2nd ed.). Sage.

Herbst, S. (2010). *Rude democracy: Civility and incivility in American politics*. Temple University Press.

Herwartz, C. (2022). *Wie die EU eine neue Social-Media-Welt schaffen will* [How the EU wants to create a new social media world]. https://www.handelsblatt.com/technik/it-internet/regulierung-von-algorithmen-wie-die-eu-eine-neue-social-media-welt-schaffen-will/27947728.html

Hirschman, A. O. (1970). *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states.* Harvard University Press.

Homans, G. C. (1974). *Social behavior. Its elementary forms*. Harcourt, Brace & World.

Hopp, T. (2019). A network analysis of political incivility dimensions. *Communication and the Public, 4*(3), 204–223. https://doi.org/10.1177%2F2057047319877278

Hsueh, M., Yogeeswaran, K., & Malinen, S. (2015). "Leave your comment below": Can biased online comments influence our own prejudicial attitudes and behaviors? *Human Communication Research, 41*(4), 557–576. https://doi.org/10.1111/hcre.12059

Hwang, H., Borah, P., Kang, N., & Veenstra, A. (2008, May). *Does civility matter in the blogoshpere? Examining the interaction effects of incivility and disagreement on citizen attitudes.* Paper presented at the 58th annual conference of the International Communication Association (ICA), Montreal, Canada.

Hwang, H., Kim, Y., & Kim, Y. (2018). Influence of discussion incivility on deliberation: An examination of the mediating role of moral indignation. *Communication Research*, *45*(2), 213–240. https://doi.org/10.1177/0093650215616861

Jamieson, K. H. (2000). *Incivility and its discontents: Lessons learned from studying civility in the US House of Representatives*. Allyn and Bacon.

Jamieson, K. H., & Falk, E. (2000). Continuity and change in civility in the House. In J. Bond and R. Fleisher (Eds.), *Polarized politics: Congress and the President in a partisan era* (pp. 96–108). Congressional Quarterly Press.

Jamieson, K. H., Volinsky, A., Weitz, I., & Kenski, K. (2018). The political uses and abuses of civility and incivility. In K. Kenski & K. H. Jamieson (Eds.), *The Oxford handbook of political communication* (pp. 205–218). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199793471.013.79_update_001

Jansen, F. (2021). *Höchststand bei Kriminalität von Extremisten* [Peak in crime by extremists]. https://www.tagesspiegel.de/politik/hoechststand-bei-kriminalitaet-von-extremisten-straftaten-ohne-ende-im-jahr-der-pandemie/27155866.html

Janssen, D., & Kies, R. (2005). Online forums and deliberative democracy. *Acta Politica, 40*(3), 317–335. https://doi.org/10.1057/palgrave.ap.5500115

Jeffries, L., & McIntyre, D. (2010). *Stylistics*. Cambridge University Press.

Kalch, A., & Naab, T. K. (2017). Replying, disliking, flagging: How users engage with uncivil and impolite comments on news sites. *SCM Studies in Communication and Media*, *6*(4), 395–419. https://doi.org/10.5771/2192-4007-2017-4-395

Kenski, K., Coe, K. & Rains, S. A. (2020). Perceptions of uncivil discourse online. An examination of types and predictors. *Communication Research, 47*(6), 795-814. https://doi.org/10.1177/0093650217699933

Kerr, N. L., & Kaufman-Gilliland, C. M. (1994). Communication, commitment, and cooperation in social dilemma. *Journal of Personality and Social Psychology, 66*(3), 513–529. https://doi.org/10.1037/0022-3514.66.3.513

Khedkar S., Karsi P., Ahuja D., & Bahrani A. (2021). Hateful memes, offensive or non-offensive! In A. Khanna, D. Gupta, S. Bhattacharyya, A. E. Hassanien., S. Anand, & A. Jaiswal (Eds.), *Proceedings of International conference on innovative computing and communications. Advances in intelligent systems and computing* (Vol. 1388). Springer. https://doi.org/10.1007/978-981-16-2597-8_52

Kim, J., Wyatt, R. O., & Katz, E. (1999). News, talk, opinion, participation: The part played by conversation in deliberative democracy. *Political Communication*, *16*(4), 361–385. https://doi.org/10.1080/105846099198541

Kluck, J. P. (2021). *It's not the message, it's the sender! An integrative approach to investigate incivility in online political discussions from the perspective of social perception* [Doctoral dissertation, University of Duisburg-Essen]. https://doi.org/10.17185/duepublico/75896

Kluck, J. P., Bormann, M., Rieß, A. & Krämer, N. C. (2021). *Unwilling, intolerable views, or incapable? – How different types of incivility can be distinguished based on attributions about cooperative communication*. Manuscript submitted for publication.

Kluck, J. P., & Krämer, N. C. (2021). “What an idiot!” – How the appraisal of the writer of an uncivil comment impacts discussion behavior. *New Media & Society*. Advance online publication. https://doi.org/10.1177/14614448211000666

Krueger, R. A., & Casey, M. A. (2015). *Focus groups: A practical guide for applied research* (5th ed.). Sage.

Ksiazek, T. B. (2018). Commenting on the news: Explaining the degree and quality of user comments on news websites. *Journalism Studies*, *19*(5), 650–673. https://doi.org/10.1080/1461670X.2016.1209977

Ksiazek, T. B., & Springer, N. (2020). *User comments and moderation in digital journalism: Disruptive engagement.* Routledge.

Kuckartz, U. (2014). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung* (3rd ed.) [Qualitative text analysis. A guide to methods, practice and using software]. Beltz Juventa.

Kümpel, A. S., & Rieger, D. (2019). *Wandel der Sprach- und Debattenkultur in sozialen Online-Medien* [Change of speech and debate culture in social media]. Konrad-Adenauer-Stiftung.

Lasswell, H. D. (1948). The structure and function of communication in society. In L. Bryson (Ed.), *The Communication of ideas* (pp. 37–51). Harper and Brothers.

LfM – Landesanstalt für Medien NRW (2021). *Ergebnisbericht der forsa-Befragung zu Hate Speech 2021* [Results report of the forsa survey on hate speech 2021]. https://www.medienanstalt-nrw.de/fileadmin/user_upload/NeueWebsite_0120/Themen/Hass/forsa_LFMNRW_Hassrede2021_Ergebnisbericht.pdf

Lindenberg, S. (1998). Solidarity, its microfoundation and macrodependence: A framing approach. In P. Doreian, & T. J. Fararo (Eds.), *The problem of solidarity* (pp. 61–112). Gordon and Breach Publishers.

Lindenberg, S. (2001). Social rationality versus rational egoism. In J. H. Turner (Ed.), *Handbook of sociological theory* (pp. 635–668). Kluwer Academic/Plenum Publisher. https://doi.org/10.1007/0-387-36274-6_29

Lindenberg, S. (2015). Social Rationality and Weak Solidarity: A Coevolutionary Approach to Social Order. In E. Lawler, S. R. Thye, & J. Yoon (Eds.), *Order on the edge of chaos: Social psychology and the problem of social order* (pp. 43–62). Cambridge University Press. https://doi.org/10.1017/CBO9781139924627.004

Lobinger, K., Krämer, B., Venema, R., & Benecchi, E. (2020). Pepe – Just a funny frog? A visual meme caught between innocent humor, far-right ideology, and fandom. In B. Krämer, & C. Holtz-Bacha (Eds.), *Perspectives on populism and the media: Avenues for research* (pp. 333–352). Nomos.

Lozano-Reich, N. M., & Cloud, D. L. (2009). The uncivil tongue: Invitational rhetoric and the problem of inequality. *Western Journal of Communication, 73*(2), 220–226. https://doi.org/10.1080/10570310902856105

Lück, J. & Nardi, C. (2019). Incivility in user comments on online news articles: Investigating the role of opinion dissonance for the effects of incivility on attitudes, emotions and the willingness to participate. *SCM – Studies in Communication & Media, 8*(3), 311–337. https://doi.org/10.5771/2192-4007-2019-3-311

Lyotard, J. F. (1984). *The postmodern condition*. University of Minnesota Press.

Masullo, G. (in press). Future directions for online incivility research. In C. Strippel, S. Paasch-Colberg, M. Emmer, & J. Trebbe (Eds.), *Challenges and perspectives of hate speech analysis. An interdisciplinary anthology*. Digital Communication Research.

McSwiney, J., Vaughan, M., Heft, A., & Hoffmann, M. (2021). Sharing the hate? Memes and transnationality in the far right's digital visual culture. *Information, Communication & Society*, *24*(16), 2502–2521. https://org/10.1080/1369118X.2021.1961006

Meedia (2016). *Überfordert vom Leser-Hass: Zeitungsredaktionen schränken Kommentarfunktion ein* [Overwhelmed by readers' hatred: Newspaper editors restrict the comment function]. https://meedia.de/2016/03/01/ueberfordert-vom-leser-hass-zeitungsredaktionen-schraenken-kommentarfunktion-ein/

Meyer, H. K., & Carey, M. C. (2014). In moderation: Examining how journalists' attitudes toward online comments affect the creation of community. *Journalism Practice, 8,* 213–228. https://doi.org/10.1080/17512786.2013.859838

Michelman, F. I. (1999). Morality, identity and constitutional patriotism. *Denver Law Review, 76(*4), 1009–1028.

Mower, D. S. (2019). Conclusion: The real morality of public discourse: Civility as an orienting attitude. In R. G. Boatright, T. Shaffer, S. Sobieraj, & D. Goldthwaite Young (Eds.), *A crisis of civility? Political discourse and its discontents* (pp. 210–232). Routledge.

Muddiman, A. (2017). Personal and public levels of political incivility. *International Journal of Communication, 11*, 3182–3202.

Muddiman, A. (2019). How people perceive political incivility. In R. G. Boatright, T. Shaffer, S. Sobieraj, & D. Goldthwaite Young (Eds.), *A crisis of civility? Political discourse and its discontents* (pp. 31–44). Routledge.

Muddiman, A., & Stroud, N. J. (2017). News values, cognitive biases, and partisan incivility in comment sections. *Journal of Communication*, *67*(4), 586–609. https://doi.org/10.1111/jcom.12312

Müller, J-W. (2007). *Constitutional Patriotism*. Princeton University Press.

Mutz, D. C. (2007). Effects of "In-your-face" television discourse on perceptions of a legitimate opposition. *American Political Science Review*, *101*, 621–635. https://doi.org/10.1017/S000305540707044X

Mutz, D. C. (2015). *In-your-face politics: The consequences of uncivil media*. Princeton University Press.

Mutz, D. C., & Reeves, B. (2005). The new videomalaise: Effects of televised incivility on political trust. *American Political Science Review*, *99*(1), 1–15. https://doi.org/10.1017/S0003055405051452

Naab, T. K., Kalch, A., & Meitz, T. G. (2018). Flagging uncivil user comments: Effects of intervention information, type of victim, and response comments on bystander behavior. *New Media & Society*, *20*(2), 777–795. https://doi.org/10.1177/1461444816670923

Naab, T. K., Naab T., & Brandmeier, J. (2021). Uncivil user comments increase users' intention to engage in corrective actions and their support for authoritative restrictive actions. *Journalism & Mass Communication Quarterly*, *98*(2), 566–588. https://doi.org/10.1177/1077699019886586

Neuberger, C. (2017). Soziale Medien und Journalismus [social media and journalism]. In J.-H. Schmidt, & M. Taddicken (Eds.), *Handbuch Soziale Medien* [handbook social media] (pp. 101–127). Springer.

Neuerer, D. (2022). *Straftaten gegen Politiker nehmen deutlich zu – Innenministerin kündigt Konsequenzen an* [Crimes against politicians increase significantly - Interior Minister announces consequences]. https://www.handelsblatt.com/politik/deutschland/kriminalitaet-straftaten-gegen-politiker-nehmen-deutlich-zu-innenministerin-kuendigt-konsequenzen-an/28040606.html

Newman, N., Fletcher, R., Levy, D., & Nielsen, R. K. (2016). *Reuters Institute digital news report 2016*. Reuters Institute for the Study of Journalism, University of Oxford.

Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D., & Nielsen, R. K. (2017). *Reuters Institute digital news report 2017*. Reuters Institute for the Study of Journalism, University of Oxford.

O'Sullivan, P. B., & Flanagin, A. J. (2003). Reconceptualizing 'flaming' and other problematic messages. *New Media & Society,* *5*(1), 69–94. https://doi.org/10.1177/1461444803005001908

Opp, K.-D. (2015). Norms. In J. D. Wright (Ed.), *International encyclopedia of social and behavioral sciences* (Vol. 17, 2nd ed., pp. 5–10). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.32103-1

Otto, L. P., Lecheler, S., & Schuck, A. R. T. (2020). Is context the key? The (non-)differential effects of mediated incivility in three European countries. *Political Communication, 37*(1), 88–107. https://doi.org/10.1080/10584609.2019.1663324

Oz, M., Zheng, P., & Chen, G. M. (2018). Twitter versus Facebook: Comparing incivility, impoliteness, and deliberative attributes. *New Media & Society, 20*(9), 3400–3419. https://doi.org/10.1177/1461444817749516

Pang, N., Ho, S. S., Zhang, A. M.R., Ko, J. S.W., Low, W. X., & Tan, K. S.Y. (2016). Can spiral of silence and civility predict click speech on Facebook? *Computers in Human Behavior, 64,* 898–905. https://doi.org/10.1016/j.chb.2016.07.066

Papacharissi, Z. (2004). Democracy online: civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society, 6*(2), 259–283. https://doi.org/10.1177/1461444804041444

Porten-Cheé, P., Kunst, M., & Emmer, M. (2020). Online civic intervention: A new form of political participation under conditions of a disruptive online discourse. *International Journal of Communication, 14*, 514–534.

Quandt, T. (2018). Dark participation. *Media & Communication, 6*(4), 36–48. https://doi.org/10.17645/mac.v6i4.1519

Reeves, B., & Nass, C. (1996). *The media equation*. Cambridge University Press.

Riedl, M. J., Naab, T. K., Masullo, G., Jost, P., & Ziegele, M. (2021). Who is responsible for interventions against problematic comments? Comparing user attitudes in Germany and the United States. *Policy & Internet*, *13*(3), 433–451. https://doi.org/10.1002/poi3.257

Rieger, D., Dippold, J., & Appel, M. (2020). Trolle gibt es nicht nur im Märchen – Das Phänomen Trolling im Internet [Trolls don't only exist in fairy tales - The phenomenon of trolling on the Internet]. In M. Appel (Ed.), *Die Psychologie des Postfaktischen: Über Fake News, „Lügenpresse", Clickbait & Co* [The psychology of the postfactual: On fake news, "Lügenpresse," clickbait & co.] (pp. 45–58). Springer.

Risch, J., Krebs, E., Löser, A., Riese, A., & Krestel, R. (2018). Fine-grained classification of offensive language. *Proceedings of GermEval*, 38–44.

Risch, J., & Krestel, R. (2020). Toxic comment detection in online discussions. In B. Agarwal, R. Nayak, N. Mittal, & S. Patnaik (Eds.), *Deep learning-based approaches for sentiment analysis. Algorithms for intelligent systems*. Springer. https://doi.org/10.1007/978-981-15-1216-2_4

Rösner, L., Winter, S., & Krämer, N. C. (2016). Dangerous minds? Effects of uncivil online comments on aggressive cognitions, emotions, and behavior. *Computers in Human Behavior*, *58*, 461–470. https://doi.org/10.1016/j.chb.2016.01.022

Ross, B., Rist, M., Carbonell, G., Cabrera, B., Kurowsky, N., & Wojatzki, M. (2018). *Measuring the reliability of hate speech annotations: The case of the European refugee crisis*. https://arxiv.org/pdf/1701.08118.pdf

Rossini, P. (2020). Beyond incivility: Understanding patterns of uncivil and intolerant discourse in online political talk. *Communication Research*. Advance online publication. https://doi.org/10.1177/0093650220921314

Rowe, I. (2015). Civility 2.0: A comparative analysis of incivility in online political discussion. *Information, Communication & Society*, *18*(2), 121–138. https://doi.org/10.1080/1369118X.2014.940365

Ruiz, C., Domingo, D., Micó, J. L., Díaz-Noci, J., Meso, C., & Masip, P. (2011). Public sphere 2.0? The democratic qualities of citizen debates in online newspapers. *The International Journal of Press/Politics*, *16*(4), 463–487. https://doi.org/10.1177/1940161211415849

Santana, A. D. (2014). Virtuous or vitriolic: The effect of anonymity on civility in online newspaper reader comment boards. *Journalism Practice, 8*(1), 18–33. https://doi.org/10.1080/17512786.2013.813194

Schacter, D. L., Gilbert, D. T., & Wegner, D. M. (2012). *Psychology*. Palgrave Macmillan.

Schaff, A. (1962). *Introduction to semantics.* Pergamon Press.

Schelling, T. (1960). *Strategy of conflict*. Oxford University Press.

Schilpzand, P., Pater, I. E., & Erez, A. (2016). Workplace incivility: A review of the literature and agenda for future research. *Journal of Organizational Behavior, 37*(S1), 57–88. https://doi.org/10.1002/job.1976

Schleif, L. & Kettemann, M. C. (2021). *Komplementär oder konkurrierend: NetzDG und DSA* [Complementary or competing: NetzDG and DSA]. https://www.hans-bredow-institut.de/de/blog/komplementaer-oder-konkurrierend-netzdg-und-dsa

Schramm, W. (1954). How communication works. In W. Schramm (Ed.), *The process and effects of communicatio*n (pp. 3–26). University of Illinois Press.

Schudson, M. (1997). Why conversation is not the soul of democracy. *Critical Studies in Mass Communication, 14*(4), 1–13.

Schwerhoff, G. (2020). Invektivität und Geschichtswissenschaft. Konstellationen der Herabsetzung in historischer Perspektive – ein Forschungskonzept [Invectivity and historical research. Constellations of defamation in historical perspective – A new research concept]. *Historische Zeitschrift*, *311*(1), 1-36. https://doi.org/10.1515/hzhz-2020-0024

Silk, J., & Connor, R. (2021). *Capitol Hill riots prompt Germany to revisit online hate speech law.* https://www.dw.com/en/capitol-hill-riots-prompt-germany-to-revisit-online-hate-speech-law/a-56171516

Simpson, D. (1960). *Cassell's new Latin dictionary*. Funk and Wagnalls.

Sobieraj, S., & Berry, J. M. (2011). From incivility to outrage: Political discourse in blogs, talk radio, and cable news. *Political Communication*, *28*(1), 19–41. https://doi.org/10.1080/10584609.2010.542360

Springer, N., Engelmann, I., & Pfaffinger, C. (2015). User comments: Motives and inhibitors to write and read. *Information, Communication & Society, 18*(7), 798–815. https://doi.org/10.1080/1369118X.2014.997268

Stoll, A., Ziegele, M., & Quiring, O. (2020). Detecting incivility and impoliteness in online discussions. *Computational Communication Research, 2*(1), 109–134. https://doi.org/10.5117/CCR2020.1.005.KATH

Stromer-Galley, J. (2007). Measuring deliberation's content: A coding scheme. *Journal of Public Deliberation, 3*(1), 1–35. https://doi.org/10.16997/jdd.50

Stroud, N. J., Scacco, J.M., Muddiman, A., & Curry, A. L. (2015). Changing deliberative norms on news organizations' facebook sites. *Journal of Computer-Mediated Communication, 20*(2), 188–203. https://doi.org/10.1111/jcc4.12104

Stroud, N. J., van Duyn, E., & Peacock, C. (2016). *News commenters and news comment readers*. https://mediaengagement.org/wp-content/uploads/2016/03/ENP-News-Commenters-and-Comment-Readers1.pdf

Stryker, R., Conway, B. A., Bauldry, S., & Kaul, V. (2021). Replication note: What is political incivility? *Human Communication Research*. https://doi.org/10.1093/hcr/hqab017

Stryker, R., Conway, B. A., & Danielson, J. T. (2016). What is political incivility? *Communication Monographs, 83*(4), 535–556. https://doi.org/10.1080/03637751.2016.1201207

Stuckey, M. E., & O'Rourke, S. P. (2014). Civility, democracy, and national politics. *Rhetoric and Public Affairs, 17*(4), 711–736.

Su, L. Y. F., Xenos, M. A., Rose, K. M., Wirz, C., Scheufele, D. A., & Brossard, D. (2018). Uncivil and personal? Comparing patterns of incivility in comments on the Facebook pages of news outlets. *New Media & Society, 20*(10), 3678–3699. https://doi.org/10.1177/1461444818757205

Sydnor, E. (2018). Platforms for incivility: examining perceptions across different media formats. *Political Communication*, *35*(1), 97–116. https://doi.org/10.1080/10584609.2017.1355857

Theocharis, Y., Barberá, P., Fazekas, Z., & Popa, S. A. (2020). The Dynamics of Political Incivility on Twitter. *SAGE Open*. https://doi.org/10.1177/2158244020919447

Tomasello, M. (2008). *Origins of human communication.* MIT Press.

Tomasello, M. (2009). *Why we cooperate.* MIT Press.

Tomasello, M. (2019). *Becoming human: A theory of ontogeny.* Harvard University Press. https://doi.org/10.4159/9780674988651

Tuckerman, N., & Dunnan, N. (1995). *The Amy Vanderbilt complete book of etiquette.* Doubleday.

Vanderbilt, A., & Baldridge, L. (1978). *The Amy Vanderbilt complete book of etiquette: A guide to contemporary living*. Doubleday.

Van 't Riet, J., & Van Stekelenburg, A. (2021). The effects of political incivility on political trust and political participation: A meta-analysis of experimental research. *Human Communication Research*. Advance online publication. https://doi.org/10.1093/hcr/hqab022

Wachs, S., Koch-Priewe, B., & Zick, A. (Eds.) (2021). *Hate Speech - Multidisziplinäre Analysen und Handlungsoptionen. Theoretische und empirische Annäherungen an ein interdisziplinäres Phänomen* [Hate speech - multidisciplinary analyses and options for action. Theoretical and empirical approaches to an interdisciplinary phenomenon]. Springer VS.

Wang, M. Y., & Silva, D. E. (2018). A slap or a jab: An experiment on viewing uncivil political discussions on Facebook. *Computers in Human Behavior*, *81*, 73–83. https://doi.org/10.1016/j.chb.2017.11.041

Watson, B. R., Peng, Z., & Lewis, S. C. (2019). Who will intervene to save news comments? Deviance and social control in communities of news commenters. *New Media & Society*, *21*(8), 1840–1858. https://doi.org/10.1177/1461444819828328

Weber Shandwick (2019). Civility in America 2019: Solutions for tomorrow. https://www.webershandwick.com/wp-content/uploads/2019/06/CivilityInAmerica2019SolutionsforTomorrow.pdf

Wilhelm, C., Joeckel, S., & Ziegler, I. (2020). Reporting hate comments: Investigating the effects of deviance characteristics, neutralization strategies, and users' moral orientation. *Communication Research, 47*(6), 921-944. https://doi.org/10.1177/0093650219855330

Wright, S. (2006). Government-run online discussion fora: Moderation, censorship and the shadow of control. *The British Journal of Politics and International Relations*, *8*(4), 550–568. https://doi.org/10.1111/j.1467-856x.2006.00247.x

Wright, S., & Street, J. (2007). Democracy, deliberation and design: The case of online discussion forums. *New Media & Society,* *9*(5), 849–69. https://doi.org/10.1177/1461444807081230

Wüllner, D. (2015). *Lassen Sie uns diskutieren* [Let's discuss]. http://www.sueddeutsche.de/kolumne/ihre-sz-lassen-sie-uns-diskutieren-1.2095271

Wünsch, C. (2006). *Unterhaltungserleben. Ein hierarchisches Zwei-Ebenen-Modell affektiv-kognitiver Informationsverarbeitung* [Entertainment experience. A hierarchical two-level model of affective-cognitive information processing]. Herbert von Halem.

Ybarra, M. L., Mitchell, K. J., Wolak, J., & Finkelhor, D. (2006). Examining characteristics and associated distress related to Internet harassment: Findings from the Second Youth Internet Safety Survey. *Pediatrics, 118*(4), 1169–1177.

Young, M. (1996). Communication and the other: Beyond deliberative democracy. In S. Benhabib (Ed.), *Democracy and difference* (pp. 120–135). Princeton University Press. https://doi.org/10.1515/9780691234168-007

Yun, G. W., Allgayer, S., & Park, S.-Y. (2020). Mind your social media manners: Pseudonymity, imaginary audience, and incivility on Facebook vs. YouTube. *International Journal of Communication*, *14*, 3418–3438.

Ziegele, M. (2016). *Nutzerkommentare als Anschlusskommunikation. Theorie und qualitative Analyse des Diskussionswerts von Online-Nachrichten* [User comments as follow-up communication. Theory and qualitative analysis of the discussion value of online news]. Springer VS.

Ziegele, M. (2019). Reader commenting. In T. P. Voss, F. Hanusch, D. Dimitrakopoulou, M. Geertsema-Sligh, & A. Sehl (Eds.), *The international encyclopedia of journalism studies* (pp. 1-8). Wiley. https://doi.org/10.1002/9781118841570.iejs0059

Ziegele, M., & Jost, P. (2020). Not funny? The effects of factual versus sarcastic journalistic responses to uncivil user comments. *Communication Research*, *47*(6), 891–920. https://doi.org/10.1177/0093650216671854

Ziegele, M., Jost, P., Bormann, M., & Heinbach, D. (2018). Journalistic counter-voices in comment sections: Patterns, determinants, and potential consequences of interactive moderation of uncivil user comments. *SCM Studies in Communication and Media, 7*(4), 525–554. https://doi.org/10.5771/2192-4007-2018-4-525

Ziegele, M., Köhler, C., & Weber, M. (2017, May). *Socially destructive! Effects of hateful user comments on recipients' prosocial behavior.* Paper presented at the 67th annual conference of the International Communication Association (ICA), San Diego, USA.

Ziegele, M., Naab, T. K., & Jost, P. (2020a). Lonely together? Identifying the determinants of collective corrective action against uncivil comments. *New Media & Society*, *22*(5), 731–751. https://doi.org/10.1177/1461444819870130

Ziegele, M., Quiring, O., Esau, K. & Friess, D. (2020b). Linking news value theory with online deliberation: How news factors and illustration factors in news articles affect the deliberative quality of user discussions in SNS' comment sections. *Communication Research, 47*(6), 860–890. https://doi.org/10.1177/0093650218797884

Zillich, A. F. (2013). *Fernsehen als Event. Unterhaltungserleben bei der Fernsehrezeption in der Gruppe* [Television as an event. Entertainment experience during television reception in the group]. Herbert von Halem.

Zimmermann, F., & Kohring, M. (2018). „Fake News" als aktuelle Desinformation. Systematisch Bestimmung eines heterogenen Begriffs ["Fake news" as current disinformation. Systematic determination of a heterogeneous concept]. *M&K Medien & Kommunikationswissenschaft*, *66*(4), 526–541. https://doi.org/10.5771/1615-634X-2018-4-526

Zompetti, J. P. (2019). Rhetorical incivility in the Twittersphere: A comparative thematic analysis of Clinton and Trump's tweets during and after the 2016 presidential election. *Journal of Contemporary Rhetoric*, *9*(1/2), 29–54.

Appendix

The Appendix consists of two research articles of the cumulative dissertation, which are currently under review or in press. The appendices are presented in the following order:

Article II in full length

Article IV in full length

Article II

Incivility

Status of the Article:

The article is currently in press as a chapter of the anthology "Challenges and perspectives of hate speech analysis. An interdisciplinary anthology" edited by C. Strippel, S. Paasch-Colberg, M. Emmer, and J. Trebbe. The book will be published as part of the open access book series "Digital Communication Research."

Authors and Affiliation:

First author: Marike Bormann is a research assistant at the Department of Social Sciences at the Heinrich Heine University Düsseldorf, Germany.

Second author: Marc Ziegele is an assistant professor at the Department of Social Sciences at the Heinrich Heine University Düsseldorf, Germany.

Contact Corresponding Author:

Marike Bormann, Department of Social Sciences, Heinrich Heine University Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany. E-Mail: marike.bormann@hhu.de

Funding:

This research was supported by the Digital Society research program funded by the Ministry of Culture and Science of the German State of North Rhine-Westphalia.

Declaration:

The authors have no potential conflict of interests regarding this article. All authors have agreed to the submission.

Incivility

Abstract

Incivility is considered a significant challenge for democratic discourse and has been the subject of many studies in a variety of contexts. Although political incivility has a long research tradition, and scholarly attention toward the phenomenon has increased with the advance of social media, there is academic controversy regarding the concept and normative implications of incivility in political contexts. This chapter provides an overview of different incivility approaches in the extant literature, discusses key challenges in incivility research, and outlines normative implications. Further, we suggest future directions for incivility research and argue why an integrative, multidimensional concept of incivility offers great potential for incivility research in the field of political (online) communication.

Keywords

Incivility, norm violations, online discussions, political communication

Incivility in Political Communication—An Established Yet Elusive Concept

Incivility has been studied in a variety of contexts, ranging from workplace environments (e.g., Schilpzand et al., 2014) to political contexts (e.g., Jamieson, 2000; Papacharissi, 2004). For this chapter, we focus on incivility in political communication. Incivility in public political discourse is a recurring subject of concern across different countries. Recently, various speakers have feared a decline or even a "crisis of civility" (Boatright et al., 2019). Polls have shown that 68% of Americans think that incivility in political communication is a major social issue. Moreover, most Americans have reported personal encounters with incivility (Weber Shandwick, 2020). Surveys among German online users reveal a similar picture, with 73% of users reporting that they have already been exposed to uncivil or hateful comments (LfM, 2020). Even the German federal president urgently called for more "reason and civility" (Steinmeier, 2019) in online discussions.

Political incivility, similar to the general phenomenon of incivility, has been the subject of many studies in a variety of contexts. These include, for example, incivility in political news articles, political campaigns, and advertising, and in political debates in Congress, television, and radio talk shows or interviews. Studies in this field usually analyze uncivil portrayals of politicians or incivility in the interactions between political elites, such as politicians, journalists, and experts (e.g., Ben-Porath, 2010; Jamieson, 2000; Mutz & Reeves, 2005). Besides incivility among political elites, scholars have become increasingly interested in studying incivility in online discussions among ordinary citizens on social media platforms or on the websites of traditional news media. Online incivility research has yielded significant output, including findings on the causes, determinants, and patterns of incivility (e.g., Coe et al., 2014; Rossini, 2020), the perceptions of incivility (e.g., Stryker et al., 2016), the effects of incivility (e.g., Rösner et al., 2016), and interventions against incivility (e.g., Kalch & Naab, 2016; Ziegele et al., 2018a).

Although political incivility has a long research tradition and academic attention to the phenomenon has increased with the advance of the Internet, there is academic controversy regarding the concept, theory, operationalization, and normative implications of incivility in political contexts. In the following section, we first provide an overview of different approaches to the phenomenon of political incivility in the extant literature and argue for an integrative, multidimensional concept. We then discuss the challenges of different approaches and outline the normative implications of incivility. Lastly, we argue why an integrative approach offers great potential for incivility research in the field of political (online) communication.

Concepts of Political Incivility

Incivility is a broad phenomenon that encompasses a wide spectrum of communication in offline and online contexts. Owing to its Latin word stem *civis* (citizen) and *civitas* (citizenship), which historically refer to the civic role and the order of the polity (Simpson, 1960), the concept of incivility and much research on incivility explicitly focus on the political sphere and public political communication.

Incivility has a long tradition of research, but scholars are still having trouble finding an agreed-upon conceptual definition and operationalization. Herbst (2010) noted that the decision of where to draw the line between civility and incivility lies “very much in the eye of the beholder” (p. 3). Similarly, Coe and colleagues (2014) stated that “incivility is a notoriously difficult term to define, because what strikes one person as uncivil might strike another person as perfectly appropriate” (p. 660). Benson (2011) pointed out that civility and incivility “are always situational and contestable” (p. 22). Hence, defining incivility is challenging, and a variety of approaches to the phenomenon can be found. Nevertheless, most definitions—at least implicitly—share the notion that *incivility is a violation of norms*. The majority of scholars approach incivility as a violation of *respect norms*, *democratic norms*, or *politeness norms*. These studies usually refer to normative theories of democracy or politeness theories. Additionally, recent studies have conceptualized incivility as a violation of *multiple norms*. Although these different perspectives are not always entirely clear-cut, it is helpful to briefly outline them in the following sections before proposing a new approach that integrates the different perspectives (for an overview of the different approaches, see also Bormann et al., 2021).

Incivility as a Violation of Respect Norms

Studies analyzing incivility as a *violation of (deliberative) respect norms* usually refer to normative theories of democracy, mostly deliberation theory. Deliberation theory sketches a public sphere accessible to everyone in which citizens debate matters of public interest in a reciprocal, rational, and respectful manner (Gastil, 2008; Gutmann & Thompson, 2004; Habermas, 1996). Within this framework, civility is understood as mutual respect between discussants. Thus, studies have often defined incivility as *disrespectful behavior in public discussions* toward other participants, the forum, or specific topics (e.g., Anderson et al., 2014; Coe et al., 2014; Gervais, 2014, 2015; Sobieraj & Berry, 2011). It is important to note that such disrespectful behavior differs from mere disagreement. Disagreement, if voiced respectfully, is an inevitable characteristic of discussions with political opponents and is beneficial for deliberation (Herbst, 2010; Stromer-Galley, 2007). From this perspective, only disagreement (or negativity) combined with disrespect constitutes incivility (e.g., Hwang et al., 2018). Despite partly overlapping definitions, studies analyzing incivility as a violation of respect norms vary regarding their operationalizations of incivility. These operationalizations range from *name-calling*, *emotional displays*, and *ideologically extremize language* (Sobieraj & Berry, 2011) to *lying* (Coe et al., 2014) and the *use of conspiracy theories* (Gervais, 2014).

Incivility as a Violation of Liberal Democratic Norms

Many scholars have also approached incivility as a *violation of liberal democratic norms* (e.g., Kalch & Naab, 2017; Oz et al., 2017; Papacharissi, 2004; Rowe, 2015). These studies often refer to Papacharissi's (2004) distinction between impoliteness and incivility.

According to Papacharissi (2004), many earlier concepts of incivility have, in fact, measured impoliteness, which is “etiquette-related” (p. 260) and something that is not undesirable per se, as “adherence to etiquette (…) frequently restricts conversation” (p. 260), especially in political discussions. The author argued that incivility goes further than impoliteness, threatens democratic norms, and has negative implications for democracy. Consequently, impoliteness and incivility are operationalized differently, with the latter focusing on *threats to democracy*, *threats to individual rights*, and *antagonistic stereotypes*, such as *racism* or *sexism* (Papacharissi, 2004). This approach has since been used by various researchers. Rossini (2020), for example, similarly argued that violations of politeness norms cannot be equated with violations of democratic norms, and that only violations of the latter would be detrimental to democracy. Violations of democratic norms in Rossini’s operationalization include discriminatory expressions and threats to individual liberty rights or denial of political participation. Contrary to Papacharissi (2004), however, Rossini defined violations of interpersonal politeness or respect norms as *incivility*, and norm violations that pose a threat to democracy as *intolerance*. Here, we clearly observe some inconsistencies in contemporary concepts of incivility. The resulting challenges will be discussed in more detail below.

Incivility as a Violation of Interpersonal Politeness Norms

Similar to Rossini (2020), various studies have analyzed incivility as a violation of *interpersonal politeness norms* (e.g., Ben-Porath, 2010; Chen & Lu, 2017; Chen & Ng, 2017; Mutz, 2007, 2015; Mutz & Reeves, 2005). These studies draw on politeness theories that deal with the rules of interpersonal interaction in public spaces, such as *social norm approaches* (Fraser, 1990) or *face theory* (Brown & Levinson, 1987; Goffman, 1959). Social norm approaches often follow a Western understanding of etiquette; within this understanding, incivility is usually defined as a violation of the social norms of politeness for a given culture (e.g., Ben-Porath, 2010; Mutz, 2007; Mutz & Reeves, 2005). Against the backdrop of face theory, researchers have also conceptualized incivility as a threat to people’s positive face, which is the socially desired and constructed public identity that people act out during a communication process (e.g., Chen & Lu, 2017; Chen & Ng, 2017). According to these approaches, incivility manifests, among others, in *insults*, *name-calling*, *yelling* (or using capital letters to indicate yelling in online communication), *interruption*, *profanity*, and *vulgarity* (Ben-Porath, 2010; Chen & Lu, 2017; Chen & Ng, 2017; Mutz, 2007; Mutz & Reeves, 2005).

Incivility as a Violation of Multiple Norms

Contemporary theorizing about incivility has shifted to a constructionist perspective, suggesting that incivility is "multifaceted, individual, and context specific" (Wang & Silva, 2018, p. 73). Consequently, current research often approaches incivility as *perceived violations of multiple norms*. Muddiman (2017), for example, derived from the perceptions of participants in two experiments a two-dimensional model of perceived incivility. In this model, "personal-level incivility" (Muddiman, 2017, p. 3183) includes violations of interpersonal politeness norms, and "public-level incivility" (Muddiman, 2017, p. 3184) includes violations of deliberative norms, such as *ideological extremity* and *lack of comity*. Chen (2017) also approached incivility as a perceptual continuum, with impoliteness being on the mild end and hate speech being on the harmful end of the continuum. In their extensive survey, Stryker et al. (2016) found that besides violations of politeness and democratic norms, participants perceived *deception* as a third dimension of incivility. This dimension includes *lies* as well as *misleading* and *exaggerating claims*, which can be considered violations of honesty norms.

Toward an Integrative Concept of Political Incivility

In our own research, we propose a new concept of political incivility that incorporates previous concepts into an integrative framework, while following a bottom-up approach from the perspective of communication participants (Bormann et al., 2021). Based on theories on cooperation, communication, and norms (e.g., Grice, 1975; Lindenberg, 2015; Tomasello, 2008, 2009), we suggest five communication norms that individuals can disapprove of violating. The five communication norms build on the central aspects of communication, namely, the substantial aspect (content; information), the formal aspect (mode), the temporal aspect (process), the social aspect (actors; relation), and the spatial aspect (context; Bormann et al., 2021; Lasswell, 1948; Schaff, 1962). Violations of the five norms potentially constitute incivility. The *information norm* refers to the substance of the information provided in a discussion. It can be violated when, for example, participants lie, spread conspiracy theories, or communicate misleading, irrelevant information. The *modality norm* concerns the formal aspect of communication and can be violated when participants communicate ambiguously, for example, by using sarcasm. The *process norm* refers to the interconnectedness of contributions and can be violated when, for example, participants deviate from the topic of the discussion or refuse to be responsive. The *relation norm* expresses the expectation of participants to be respectful and polite; it can be violated when, for example, participants use name-calling, insults, or vulgarity. Lastly, the *political context norm* encompasses the normative expectations of participants in political discussions to consider essential liberal democratic principles in their contributions. This norm can be violated when, for example, participants threaten the rights of other individuals, question the democratic constitution, or incite violence against democratic governments or minority groups. In our concept, incivility occurs when participants disapprove of an act of communication as severely violating one or several of these five communication norms (Bormann et al., 2021).

In summary, it becomes clear that political incivility is a multi-faceted and complex phenomenon. A common denominator of the existing concepts that we can identify is that incivility refers to violations of norms. Depending on the research tradition, these norms include deliberative norms of mutual respect, liberal-democratic norms, or norms derived from politeness research. We also proposed an attempt toward an integrative concept of incivility in political communication. This concept describes incivility as a perceived violation of one or several of five basic communication norms, namely, the information norm, the modality norm, the process norm, the relation norm, and the political context norm. In the following sections, we discuss the challenges and perspectives related to these different approaches to political incivility.

Challenges of Research on Political Incivility

Challenges Related to Inconsistent Definitions and Measures

A major challenge in research on political incivility is related to the difficulty of comparing the findings of different studies. Content analyses of online discussions, for example, have reported varying shares of incivility in user comments, ranging from 3% to more than 50% (e.g., Rowe, 2015; Santana, 2014). Some of these variations are clearly due to the fact that studies have analyzed different platforms and topics, among others. Yet, the *operationalizations of incivility* also vary significantly from study to study; thus, different phenomena are studied under the same term. Coe et al. (2014), for example, found that 22% of the user comments posted on a newspaper's website contained incivility, which the authors operationalized as name-calling, vulgarity, aspersion, pejoratives, or lying accusations. Rowe (2015) operationalized these norm violations as impoliteness and found that 32% of the comments posted on a newspaper's Facebook site and 35% of the comments posted on the newspaper's website were impolite. Incivility in terms of the assignment of stereotypes and threats to democracy or individual's rights was only visible in 3% of the Facebook comments and in 6% of the website comments (Rowe, 2015). Similarly, Santana (2014) compared incivility in anonymous and non-anonymous news website comments. Applying a broad operationalization of incivility as personal attacks, threats, vulgarities, abusive, foul, or hateful language, assignment of stereotypes, epithets, ethnic slurs, and racist or bigoted speech, Santana found that up to 53% of the comments were uncivil.

What renders these diverging findings particularly problematic is that they suggest different normative and practical implications for governing online discussion spaces. While policymakers or journalists may conclude that incivility is not a pressing issue based on studies that report low shares of incivility, research that has reported otherwise may justify calls for strong interventions. Future research should thus invest in reaching agreed-upon standardized operationalizations of incivility to increase the comparability of findings and to provide more reliable assessments of the development of incivility over time.

Diverging operationalizations of uncivil behavior are also problematic in experimental research (e.g., Chen & Ng, 2017; Gervais, 2015; Kalch & Naab, 2017; Rösner et al., 2016). Some studies on the effects of incivility, for example, have operationalized incivility as a unidimensional construct or as a "monolith" (see Masullo in this collection). These studies mingle different types of uncivil behavior, such as name-calling, vulgarity, histrionics, and lies. Consequently, the distinct effects of the different types of incivility cannot be assessed (e.g., Gervais, 2015; Rösner et al., 2016). Yet, the few studies that have investigated people's perceptions of different types of incivility suggest that participants evaluate each type differently in terms of severity (e.g., Muddiman, 2017; Stryker et al., 2016), and that different types of incivility have varying effects on people's behavioral intentions (e.g., Kalch & Naab, 2017). Distinct forms of uncivil behavior should therefore not be viewed and investigated as unidimensional in future studies (see also Masullo in this collection for a similar appeal).

Challenges Related to the Reliable Measurement of Incivility in Content Analyses

As previously mentioned, many studies on political incivility have applied content analyses to investigate the patterns, determinants, and potential consequences of uncivil communication (e.g., Coe et al., 2014; Rowe, 2015; Ziegele et al., 2018a, 2020). For these analyses, it is often challenging to achieve satisfactory levels of reliability and external validity for the measures that are used. Some manifestations of incivility, such as name-calling, can easily be recognized by all coders. However, when it comes to more subtle, culture-specific, or context-specific norm violations, such as implicit stereotypes, coders regularly struggle to detect these forms of incivility reliably. Similarly, it is difficult to detect norm violations in online discussions that perpetrators intentionally camouflage to circumvent algorithms and word filters, for example.

Ross et al. (2018) demonstrated that even among researchers who are familiar with incivility-related concepts, there is sometimes low agreement on what should be classified as civil and uncivil. Particularly for subtle norm violations, the coders' individual perceptions, knowledge, and experiences impact whether they classify a speech act as uncivil. Human speech is a rich and complex phenomenon, and so are the potential manifestations of political incivility. Although many studies provide clear coding instructions for various types of incivility, it is challenging or even impossible to consider all or even the most possible manifestations of these types in a coding scheme. Some researchers tackle this problem by coding only incivility that is measurable on the level of words. This, however, reduces the validity of incivility measures. The problem is no less urgent in automated analyses of political incivility. Previous studies have already applied dictionary-based approaches (e.g., Muddiman & Stroud, 2017) and machine learning (e.g., Su et al., 2018) to study online incivility. Similar to manual content analyses, these methods work best for explicit forms of incivility that are clearly expressed through the use of specific words, such as offensive language or extreme forms of hate speech (e.g., Davidson et al., 2017). Automatically detecting subtle or ambiguous forms of incivility, such as covert racism or sarcasm, is far more challenging, and many automated measures suffer from high rates of misclassification (Stoll et al., 2020).

In understanding incivility as a perceptual construct and accepting that even the work of professional coders in content analyses will be, to some extent, affected by individual biases, we can think about alternative or complementary ways to classify incivility in content analyses. For example, each contribution in online discussions could be checked to determine whether it was visibly disapproved of by other participants. Disapproval here can be expressed, among others, through a sanctioning reply comment. If a comment has been visibly disapproved, coders can analyze it regarding the specific type(s) of norm violations (Bormann et al., 2021). Although this procedure will certainly work only for a small fraction of uncivil contributions, it would account for the fact that incivility is often a matter of the perceptions of the people involved in the respective communication.

Challenges Related to the Normative Implications of Incivility

Normative implications of incivility are controversial among scholars. This can be partly explained by the fact that studies have reported different consequences of incivility. Experimental research, for example, has found various negative effects of being exposed to uncivil content: incivility in political talk shows can reduce viewers' trust in politics and politicians (Mutz & Reeves, 2005). Uncivil online discussions have been found to increase readers' opinion polarization (Anderson et al., 2014), stimulate negative emotions and aggressive cognitions (Gervais, 2015; Rösner et al., 2016), and promote further incivility (Gervais, 2015; Ziegele et al., 2018c). Moreover, uncivil comments can adversely affect the perceived quality of news articles (Prochazka et al., 2018) and increase prejudice against social minorities (Hsueh et al., 2015). Beyond that, specific types of incivility, also known as *hate speech* (e.g., Ziegele et al., 2018b; see also Frischlich and Sponholz in this collection), have raised strong concerns among researchers, since these types are often used to further marginalize certain groups. Uncivil attacks against women in online discussions, for example, often aim to silence and exclude them from political discourse (e.g., Chen et al., 2020). However, various studies have also reported beneficial outcomes of incivility; exposure to uncivil content can, for example, increase people's interest in politics (Brooks & Geer, 2007) and their intentions to participate politically (Borah, 2014; Chen, 2017; Chen & Lu, 2017).

Taken together, empirical studies analyzing the consequences of incivility arrive at different conclusions regarding whether incivility is a good or bad thing. Overall, however, the prevailing claim in public discourse is that incivility is undesirable and needs to be eliminated (Chen et al., 2019). This claim is not only based on empirical findings but also on prescriptive theories. From a deliberation perspective, for example, incivility is mainly considered as undermining deliberative discourse, and from a politeness perspective, it is predominantly assessed as a negative threat to the constructed public self-image of individuals. These prescriptive theories, however, neglect an important argument: just as incivility itself can serve as a tool to silence minorities, calls for civility can also be used as silencing mechanisms (see also Litvinenko in this collection). As of today, various researchers have argued that democracy can endure heated discussions and that high demands for civil discourse can exclude certain social groups, such as educationally disadvantaged milieus (e.g., Bejan, 2017; Estlund, 2008; Garton Ash, 2016). Therefore, calls for "robust civility" (Garton Ash, 2016, p. 316) or "mere civility" (Bejan, 2017) are being voiced—a civility that is robust and broad, tolerates disagreement, various language styles, and heated discussions.

In a similar vein, a large body of *critical studies* conceive of civility as a set of norms that a powerful elite establishes to suppress marginalized groups. From this perspective, calls for civility mainly serve as an instrument of the powerful to suppress the powerless and reinforce existing power relations and social inequality (e.g., Baez & Ore, 2018; Lozano-Reich & Cloud, 2009; Stuckey & O'Rourke, 2014). According to these studies, the powerful can decide what is considered (un)civil, perform social control, and thus exclude minority voices from political discourse.

When conceptualizing calls for civility as a strategy to exclude and suppress certain groups, the positive implications of incivility emerge. For example, critical studies have acknowledged incivility as an instrument of the powerless to express their identity. From this perspective, incivility is a powerful means of differentiating an oppressor from an oppressed, and thus an out-group from an in-group (e.g., Jamieson et al., 2017). Violations of civility norms can then demonstrate self-assertion and belonging to a marginalized group (e.g., Lozano-Reich & Cloud, 2009; Stuckey & O'Rourke, 2014). Further, marginalized groups can use incivility to draw attention to their problems and fight for their rights. In fact, incivility has been described as the weapon of the powerless (Scott, 1985) and as a strategic instrument of marginalized groups to denounce injustice and seek change. Incivility is then seen as an act of dissent and democratic activism and has important mobilizing functions (Edyvane, 2020; Jamieson et al., 2017). Thus, protest, threats, insults, and several other uncivil expressions against social injustice can sometimes be considered legitimate, and some scholars even plead for an "uncivil tongue" (Lozano-Reich & Cloud, 2009, p. 221). Other scholars, however, explicitly call for "responsible incivility" (Edyvane, 2020, p. 105). From this perspective, incivility is legitimate only when its positive democratic consequences outweigh the negative ones.

Overall, the normative implications of incivility depend on various factors. An across-the-board evaluation of incivility as something bad seems inappropriate because such an evaluation neglects the sometimes positive effects of incivility and the sometimes legitimate use of an "uncivil tongue" (Lozano-Reich & Cloud, 2009, p. 221) to fight inequality and injustice. Researchers should, therefore, withstand the temptation to justify the relevance of their own research solely by referring to the destructive effects of incivility. Thereby, they can help to promote a more differentiated perspective on the phenomenon.

Towards New Perspectives on Incivility in Political Communication

Incivility is a multi-faceted, dynamic, and, partly, elusive phenomenon. What we can say with some confidence is that incivility is mostly situated in the fields of politics and political communication. Additionally, studies are relatively consistent in conceptualizing incivility as a violation of norms, although the specific norms that incivility violate cover a broad range and include interpersonal politeness norms, deliberative respect norms, liberal democratic norms, and communication norms. Further, an increasing number of studies agree that incivility is a matter of perceptions and, as such, often a violation of multiple norms.

In this chapter, we have outlined various conceptual, methodological, and normative challenges that arise from a multitude of approaches toward incivility. From these challenges, we have derived some potential directions for future research on incivility. More specifically, we recommend developing more consistent operationalizations of incivility, rethinking the ways in which perceived incivility can be measured in content analyses, and broadening the view on when and why incivility is a "good" or "bad" thing.

Despite the challenges related to the concept of incivility, we should not disregard its benefits. Most importantly, by broadly focusing on *norm violations*, incivility resonates with other concepts that investigate specific deviant communicative behaviors, such as *flaming*, *offensive speech, and hate speech* (see Sponholz and Frischlich in this collection). Compared to other concepts of deviant communication, such as *toxicity* (see Risch in this collection), incivility is a strongly theory-based construct that has a long research tradition. Research has provided far-reaching insights into the causes, patterns, and consequences of incivility in offline and online contexts, and future studies can build on established experiences and measurements. Incivility is also tailored to the analysis of political communication among elites and citizens. At the same time, the concept is flexible enough to be applied to non-political contexts, such as the analysis of social interactions in the workplace.

Nevertheless, to exploit the full potential of the incivility concept, we advocate a broad view of the phenomenon that integrates different previous approaches. More specifically, we sketched a perceptual and multidimensional model of incivility (Bormann et al., 2021). This model is built on fundamental concepts of human cooperation and communication, and includes five communication norms (information, process, modality, relation, and political context) that are largely compatible with the multitude of the norm concepts suggested in previous incivility research. Within our integrative approach, we conceive of incivility as disapproved violations of one or several of these communication norms. This concept offers various benefits for future research. First, although our concept is broad enough to cover most norm violations that previous research has identified, it does not conceive of incivility as a monolith. Rather, the model specifies different types of norm violations in a distinctive way by systematizing them along the five communication norms. Second, owing to its roots in the fundamental processes of communication and cooperation, the concept can be applied to a variety of contexts, ranging from offline political interactions between politicians to online discussions among citizens. Lastly, the concept is based on perceptions or, more specifically, on the disapproval of those involved in the respective communication. Consequently, our concept allows for a less prescriptive and more differentiated perspective regarding which potential norm violations can actually be considered uncivil in specific contexts.

Social norms have always been in flux and are constantly being renegotiated among citizens and elites. The Internet and the social web have accelerated this development, as currently demonstrated by debates around *political correctness* or *canceling culture*, to name only a few. In these debates, we observe that the perceptions of civility and incivility clash among different camps and that the perceived civil behavior of one's own camp is disapproved of as uncivil by members of the other camp. Further, various communication and behavior that societies have evaluated as civil back in history may be considered uncivil today. For example, denying women the right to publicly raise their voice on political issues and to participate politically was not considered uncivil a few decades ago but certainly would be today. Similarly, in many societies, the use of racial stereotypes was widely perceived as appropriate for a long time but would today be evaluated as an act of incivility. Since incivility is—and will likely always be—subject to individual perceptions and zeitgeists, future research would benefit from paying more attention to the contexts of uncivil communication, such as time, culture, situation, social groups, or issues, for example. With these arguments in mind, we argue that future incivility research should investigate more comprehensively the circumstances under which different individuals and social groups perceive specific norm violations as civil or uncivil and evaluate them as (democratically) legitimate or harmful. Our multidimensional concept offers a fruitful starting point for such research in that it distinguishes between distinct norm violations, considers individual perceptions and evaluations of communication participants, and is applicable to a wide variety of contexts.

References

Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2014). The "nasty effect:" Online incivility and risk perceptions of emerging technologies. *Journal of Computer-Mediated Communication, 19*(3), 373–387. https://doi.org/10.1111/jcc4.12009

Baez, K. L., & Ore, E. (2018). The moral imperative of race for rhetorical studies: on civility and walking-in-white in academe. *Communication and Critical/Cultural Studies, 15*(4), 331–336. https://doi.org/10.1080/14791420.2018.1533989

Bejan, T. (2017). *Mere civility: Disagreement and the limits of toleration*. Harvard University Press.

Ben-Porath, E. N. (2010). Interview effects: Theory and evidence for the impact of televised political interviews on viewer attitudes. *Communication Theory*, *20*(3), 323–347. https://doi.org/10.1111/j.1468-2885.2010.01365.x

Benson, T. W. (2011). The rhetoric of civility: Power, authenticity, and democracy. *Journal of Contemporary Rhetoric, 1*(1), 22–30.

Boatright, R., Shaffer, T., Sobieraj, S., & Young, D. G. (Eds.) (2019). *A crisis of civility? Political discourse and its discontents*. Routledge. https://doi.org/10.4324/9781351051989

Borah, P. (2014). Does It matter where you read the news story? Interaction of incivility and news frames in the political blogosphere. *Communication Research*, *41*(6), 809–827. https://doi.org/10.1177/0093650212449353

Bormann, M., Tranow, U., Vowe, G., & Ziegele, M. (2021). Incivility as a violation of communication norms: A typology based on normative expectations toward political communication. *Communication Theory.* https://doi.org/10.1093/ct/qtab018

Brooks, D.J., & Geer, J.G. (2007). Beyond negativity: The effects of incivility on the electorate. *American Journal of Political Science, 51*(1), 1–16. https://doi.org/10.1111/j.1540-5907.2007.00233.x

Brown, P., & Levinson, S. C. (1987). *Politeness. Some universals in language usage.* Cambridge University Press.

Chen, G. M. (2017). *Online incivility and public debate: Nasty talk*. Palgrave Macmillan.

Chen, G. M., & Lu, S. (2017). Online political discourse: Exploring differences in effects of civil and uncivil disagreement in news website comments. *Journal of Broadcasting & Electronic Media*, *61*(1), 108–125. https://doi.org/10.1080/08838151.2016.1273922

Chen, G. M., Muddiman, A., Wilner, T., Pariser, E., & Stroud, N. J. (2019). We should not get rid of incivility online. *Social Media and Society*, *5*(3), 1–5. https://doi.org/10.1177/2056305119862641

Chen, G. M., & Ng, Y. M. M. (2017). Nasty online comments anger you more than me, but nice ones make me as happy as you. *Computers in Human Behavior*, *71*, 181–188. https://doi.org/10.1016/j.chb.2017.02.010

Chen, G. M., Pain, P., Chen, V. Y., Mekelburg, M., Springer, N., & Troger, F. (2020). 'You really have to have a thick skin': A cross-cultural perspective on how online harassment influences female journalists. *Journalism*, *21*(7), 877–895. https://doi.org/10.1177/1464884918768500

Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, *64*(4), 658–679. https://doi.org/10.1111/jcom.12104

Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. https://arxiv.org/abs/1703.04009

Edyvane, D. (2020). Incivility as dissent. *Political Studies, 68*(1), 93–109. https://doi.org/10.1177/0032321719831983

Estlund, D. M. (2008). *Democratic authority: A philosophical framework*. Princeton University Press.

Fraser, B. (1990). Perspectives on politeness. *Journal of Pragmatics, 14*(2), 219–236. https://doi.org/10.1016/0378-2166(90)90081-N

Garton Ash, T. (2016). *Redefreiheit: Prinzipien für eine vernetzte Welt* [Free speech: Ten principles for a connected world]. Hanser.

Gastil, J. (2008). *Political communication and deliberation.* Sage Publications.

Gervais, B. T. (2014). Following the news? Reception of uncivil partisan media and the use of incivility in political expression. *Political Communication*, *31*(4), 564–583. https://doi.org/10.1080/10584609.2013.852640

Gervais, B. T. (2015). Incivility online: Affective and behavioral reactions to uncivil political posts in a web-based experiment. *Journal of Information Technology & Politics, 12*(2), 167–185. https://doi.org/10.1080/19331681.2014.997416

Goffman, E. (1959). *The presentation of self in everyday life.* Doubleday.

Grice, P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics* (pp. 41–58). Academic Press.

Gutmann, A., & Thompson, D.F. (2004). *Why deliberative democracy?* Princeton University Press.

Habermas, J. (1996). *Between facts and norms: Contributions to a discourse theory of law and democracy*. MIT Press.

Herbst, S. (2010). *Rude democracy: Civility and incivility in American politics*. Temple University Press.

Hsueh, M., Yogeeswaran, K., & Malinen, S. (2015). “Leave your comment below”: Can biased online comments influence our own prejudicial attitudes and behaviors? *Human Communication Research, 41*(4), 557–576. https://doi.org/10.1111/hcre.12059

Hwang, H., Kim, Y., & Kim, Y. (2018). Influence of discussion incivility on deliberation: An examination of the mediating role of moral indignation. *Communication Research*, *45*(2), 213–240. https://doi.org/10.1177/0093650215616861

Jamieson, K. H. (2000). *Incivility and its discontents: Lessons learned from studying civility in the US House of Representatives.* Allyn and Bacon.

Jamieson, K. H., Volinsky, A., Weitz, I., & Kenski, K. (2017). The political uses and abuses of civility and incivility. In K. Kenski & K. H. Jamieson (Eds.), *The Oxford handbook of political communication.* Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199793471.013.79_update_001

Kalch, A., & Naab, T. K. (2017). Replying, disliking, flagging: How users engage with uncivil and impolite comments on news sites. *SCM Studies in Communication and Media*, *6*(4), 395–419. https://doi.org/10.5771/2192-4007-2017-4-395

Lasswell, H. D. (1948). The structure and function of communication in society. In L. Bryson (Ed.), *The communication of ideas* (pp. 37–51). Harper and Brothers.

LfM – Landesanstalt für Medien NRW (2020). *Ergebnisbericht der forsa-Befragung zu Hate Speech 2020* [Results report of the forsa survey on hate speech 2020]. https://www.medienanstalt-nrw.de/fileadmin/user_upload/NeueWebsite_0120/Themen/Hass/forsa_LFMNRW_Hassrede2020_Ergebnisbericht.pdf

Lindenberg, S. (2015). Solidarity: Unpacking the social brain. In A. Laitinen & A.B. Pessi (Eds.), *Solidarity. Theory and practice* (pp. 30–54). Lexington Books.

Lozano-Reich, N. M., & Cloud, D. L. (2009). The uncivil tongue: Invitational rhetoric and the problem of inequality. *Western Journal of Communication, 73*(2), 220–226. https://doi.org/10.1080/10570310902856105

Muddiman, A. (2017). Personal and public levels of political incivility. *International Journal of Communication, 11*, 3182–3202.

Muddiman, A., & Stroud, N. J. (2017). News values, cognitive biases, and partisan incivility in comment sections. *Journal of Communication, 67*(4), 586–609. https://doi.org/10.1111/jcom.12312

Mutz, D. C. (2007). Effects of "In-your-face" television discourse on perceptions of a legitimate opposition. *American Political Science Review*, *101*, 621–635. https://doi.org/10.1017/S000305540707044X

Mutz, D. C. (2015). *In-your-face politics: The consequences of uncivil media*. Princeton University Press.

Mutz, D. C., & Reeves, B. (2005). The new videomalaise: Effects of televised incivility on political trust. *American Political Science Review*, *99*(1), 1–15. https://doi.org/10.1017/S0003055405051452

Oz, M., Zheng, P., & Chen, G. M. (2018). Twitter versus Facebook: Comparing incivility, impoliteness, and deliberative attributes. *New Media & Society*, *20*(9), 3400–3419. https://doi.org/10.1177/1461444817749516

Papacharissi, Z. (2004). Democracy online: civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society, 6*(2), 259–283. https://doi.org/10.1177/1461444804041444

Prochazka, F., Weber, P., & Schweiger, W. (2018). Effects of civility and reasoning in user comments on perceived journalistic quality. *Journalism Studies, 19*(1), 62–78. https://doi.org/10.1080/1461670X.2016.1161497

Rösner, L., Winter, S., & Krämer, N. C. (2016). Dangerous minds? Effects of uncivil online comments on aggressive cognitions, emotions, and behavior. *Computers in Human Behavior*, *58*, 461–470. https://doi.org/10.1016/j.chb.2016.01.022

Ross, B., Rist, M., Carbonell, G., Cabrera, B., Kurowsky, N., & Wojatzki, M. (2018). Measuring the reliability of hate speech annotations: The case of the European refugee crisis. https://arxiv.org/pdf/1701.08118.pdf

Rossini, P. (2020). Beyond incivility: Understanding patterns of uncivil and intolerant discourse in online political talk. *Communication Research*. https://doi.org/10.1177/0093650220921314

Rowe, I. (2015). Civility 2.0: A comparative analysis of incivility in online political discussion. *Information, Communication & Society*, *18*(2), 121–138. https://doi.org/10.1080/1369118X.2014.940365

Santana, A. D. (2014). Virtuous or vitriolic: The effect of anonymity on civility in online newspaper reader comment boards. *Journalism Practice, 8*(1), 18–33. https://doi.org/10.1080/17512786.2013.813194

Schaff, A. (1962). *Introduction to semantics.* Pergamon Press.

Schilpzand, P., Pater, I. E., & Erez, A. (2016). Workplace incivility: A review of the literature and agenda for future research. *Journal of Organizational Behavior*, *37*(S1), 57–88. https://doi.org/10.1002/job.1976

Scott, J. C. (1985). *Weapons of the weak: Everyday forms of peasant resistance.* Yale University Press.

Simpson, D. (1960). *Cassell's new Latin dictionary*. Funk and Wagnalls.

Sobieraj, S., & Berry, J. M. (2011). From incivility to outrage: Political discourse in blogs, talk radio, and cable news. *Political Communication*, *28*(1), 19–41. https://doi.org/10.1080/10584609.2010.542360

Steinmeier, F.-W. (2019, May 6-8). *Eröffnung der re:publica 2019* [keynote address]. 13th re:publica, Berlin, Germany. https://www.bundespraesident.de/SharedDocs/Reden/DE/Frank-Walter-Steinmeier/Reden/2019/05/190506-Eroeffnung-Republica.html

Stoll, A., Ziegele, M., & Quiring, O. (2020). Detecting incivility and impoliteness in online discussions. *Computational Communication Research, 2*(1), 109–134.

Stromer-Galley, J. (2007). Measuring deliberation's content: A coding scheme. *Journal of Public Deliberation, 3*(1), Article 12. https://doi.org/10.16997/jdd.50

Stryker, R., Conway, B. A., & Danielson, J. T. (2016). What is political incivility? *Communication Monographs, 83*(4), 535–556. https://doi.org/10.1080/03637751.2016.1201207

Stuckey, M. E., & O'Rourke, S. P. (2014). Civility, democracy, and national politics. *Rhetoric and Public Affairs, 17*(4), 711–736.

Su, L. Y. F., Xenos, M. A., Rose, K. M., Wirz, C., Scheufele, D. A., & Brossard, D. (2018). Uncivil and personal? Comparing patterns of incivility in comments on the Facebook pages of news outlets. *New Media & Society, 20*(10), 3678–3699. https://doi.org/10.1177/1461444818757205

Tomasello, M. (2008). *Origins of human communication*. MIT Press.

Tomasello, M. (2009). *Why we cooperate.* MIT Press.

Wang, M. Y., & Silva, D. E. (2018). A slap or a jab: An experiment on viewing uncivil political discussions on Facebook. *Computers in Human Behavior*, *81*, 73–83. https://doi.org/10.1016/j.chb.2017.11.041

Weber Shandwick (2020). Civility in America 2019: Solutions for tomorrow. Retrieved from https://www.webershandwick.com/wp-content/uploads/2019/06/CivilityInAmerica2019SolutionsforTomorrow.pdf

Ziegele, M., Jost, P. B., Bormann, M., & Heinbach, D. (2018a). Journalistic counter-voices in comment sections: Patterns, determinants, and potential consequences of interactive moderation of uncivil user comments. *SCM Studies in Communication and Media*, *7*(4), 525–554. https://doi.org/10.5771/2192-4007-2018-4-525

Ziegele, M., Koehler, C., & Weber, M. (2018b). Socially destructive? Effects of negative and hateful user comments on readers' donation behavior toward refugees and homeless persons. *Journal of Broadcasting & Electronic Media, 62*(4), 636–653. https://doi.org/10.1080/08838151.2018.1532430

Ziegele, M., Quiring, O., Esau, K., & Friess, D. (2020). Linking news value theory with online deliberation: How news factors and illustration factors in news articles affect the deliberative quality of user discussions in SNS' comment sections. *Communication Research, 47*(6), 860–890. https://doi.org/10.1177/0093650218797884

Ziegele, M., Weber, M., Quiring, O., & Breiner, T. (2018c). The dynamics of online news discussions: Effects of news articles and reader comments on users' involvement, willingness to participate, and the civility of their contributions. *Information, Communication & Society*, *21*(10), 1419–1435. https://doi.org/10.1080/1369118X.2017.1324505

Article IV

Perceptions of and Reactions to Different Types of Incivility in Public Online Discussions – Results of an Online Experiment

Status of the Article:

The article has been submitted to a journal and is currently under review.

Authors and Affiliation:

First author: Marike Bormann is a research assistant at the Department of Social Sciences at the Heinrich Heine University Düsseldorf, Germany.

Second author: Dominique Heinbach is a research assistant at the Department of Social Sciences at the Heinrich Heine University Düsseldorf, Germany.

Third author: Jan Philipp Kluck is a postdoctoral researcher at the Department of Social Psychology at the University of Duisburg-Essen, Germany.

Contact Corresponding Author:

Marike Bormann, Department of Social Sciences, Heinrich Heine University Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany. E-Mail: marike.bormann@hhu.de

Funding:

This research was supported by the Digital Society research program funded by the Ministry of Culture and Science of the German State of North Rhine-Westphalia.

Declaration:

The authors have no potential conflict of interests regarding this article. All authors have agreed to the submission.

Perceptions of and Reactions to Different Types of Incivility in Public Online Discussions – Results of an Online Experiment

Abstract

While incivility in public online discussions is considered a pressing concern, there are few empirical findings on what participants in an online discussion perceive and evaluate as (mildly and severely) uncivil and on how they react to different types of incivility. Based on a novel approach to incivility as a disapproved violation of communication norms, the present study examines perceptions of and reactions to norm violations. In a fully functional mock-up discussion forum, participants were confronted with comments that contained violations of five different communication norms. The results suggest that violations of all five communication norms are disapproved as uncivil, and that distinct types of incivility vary in perceived severity and elicit different responses. In particular, participants evaluated insults and vulgarity directed at another person as most severe and were more likely to respond to these with sanctioning replies or flags compared to other norm violations.

Keywords

Incivility perceptions, incivility evaluations, communication norms, norm violations, online discussions

Introduction

Incivility in public online discussions has been highlighted as a serious challenge by scholars, journalists, politicians, and the general public (e.g., Boatright, 2019; Chen et al., 2019), with studies revealing that between 20% and 50% of user comments on various news websites or social media platforms contained uncivil elements (Coe et al., 2014; Rowe, 2015; Santana, 2014). This is particularly problematic given that incivility can have negative effects on participants and on discussion dynamics. For instance, research has found that uncivil comments can increase aggressive cognitions and emotions (Gervais, 2015; Rösner et al., 2016), promote prejudicial attitudes and behavior (Hsueh et al., 2015), lead to attitude polarization (Anderson et al., 2014), and provoke more incivility in the subsequent comments (Chen & Lu, 2017; Gervais, 2015; Ziegele et al., 2018).

Although there is nowadays a valuable body of scholarship addressing incivility in online discussions, a major challenge of incivility research is that definitions and operationalizations of the construct are often inconsistent. Thus, different studies are only comparable to a limited extent. Moreover, the majority of incivility research has approached the phenomenon as a "monolith" (Masullo, in press, p. 1) instead of treating distinct types of incivility as such. Experimental studies often only consider one type of incivility or use distinct types of incivility interchangeably, for example when examining cognitive, affective, and behavioral reactions to incivility (e.g., Anderson et al., 2014; Chen & Lu, 2017; Gervais, 2015; Rösner et al., 2016). Consequently, the effects of distinct types of incivility have not yet been sufficiently examined. Since a growing amount of research suggests that incivility is a multidimensional rather than a monolithic construct, and that the perception of different types of incivility varies (e.g., Kenski et al., 2020; Muddiman, 2017; Stryker et al., 2016, 2021), it is necessary to rethink the concept. Taking this into account, Bormann et al. (2021) recently developed a new and multidimensional approach to incivility as a disapproved violation of communication norms. The approach integrates several existing incivility concepts into a comprehensive typology and follows a perceptual perspective by considering the disapproval of the participants involved in the respective discussion. However, the typology is based on theoretical assumptions and has not yet been empirically validated.

The present study thus aims to validate this typology of incivility and to shed light on perceptions and evaluations of as well as reactions to distinct types of incivility. We therefore build on recent studies that considered the question of how people perceive incivility in its various forms and how participants react to distinct types of uncivil comments in (online) discussions. However, the present study moves beyond the realm of incivility between political elites (Muddiman, 2017), surveys of public perceptions of incivility (Kenski et al., 2020; Stryker et al., 2016, 2021), and user engagement against one or two specific types of incivility (Kalch & Naab, 2017) by testing, in an experimental setting, *(1) what participants who are actually engaged in an online discussion perceive as uncivil, (2) how they evaluate distinct types of incivility in terms of severity, and (3) how they react to various types of incivility*. For this purpose, participants discussed political topics and were confronted with different forms of uncivil comments in a fully functional mock-up discussion forum.

The contributions of this study are twofold. First, to facilitate future research, we contribute to providing an integrative concept of incivility that meets the demands of a perceptual, multidimensional approach by validating it empirically. The approach examines incivility from a perceptual perspective and measures differential effects of distinct types of incivility. Second, we offer new insights into how distinct types of incivility are evaluated, which types of incivility participants perceive as the most problematic, and how participants react to different forms of incivility. This is highly relevant in light of the frequently discussed normative goal of reducing incivility and deploying various interventions. Such insights allow for an improved targeting of the types of incivility that are perceived as most harmful, and can inform optimally tailored intervention strategies.

What is Incivility?

Defining incivility is a difficult undertaking, as the phenomenon is "multifaceted (…) and context-specific" (Wang & Silva, 2017, p. 73) and "[o]ne person's incivility is another's civility" (Chen et al., 2019, p. 3). It is therefore unsurprising that incivility research has yielded several different definitions and operationalizations. One common denominator of the various conceptualizations is that most scholars approach the phenomenon as a violation of norms (e.g., Jamieson et al., 2018; Muddiman, 2017; Papacharissi, 2004; Coe et al., 2014; for an overview, see Bormann et al., 2021). There has been great debate in the literature regarding whether incivility is a violation of politeness norms such as insults, vulgarity, and name-calling (e.g., Chen, 2017; Mutz, 2007; Rossini, 2020) or a violation of liberal democratic norms such as threats to democracy or individual liberty rights (e.g., Papacharissi, 2004). Additionally, recent research has suggested that incivility is a multidimensional construct, since empirical findings indicate that violations of both politeness and democratic norms are perceived as uncivil (e.g., Muddiman, 2017; Stryker et al., 2016, 2021). Moreover, Stryker and colleagues (2016, 2021) found that besides the types of incivility that fit into these two norm categories, there are further forms of deviant communication that are perceived as uncivil, such as lying, exaggerating, or interrupting.

Taking the findings by Stryker et al. (2016, 2021) into account, Bormann and colleagues (2021) recently proposed an integrative and multidimensional approach to incivility as a disapproved violation of communication norms. Drawing on different approaches to communication and cooperation, namely action theory, evolutionary anthropology, and linguistics (Grice, 1975; Lindenberg, 2015a, 2015b; Tomasello, 2008, 2009, 2019), the authors argued that human cooperation is the essential element of social and political relations and systems, and that communication plays a key role in enabling cooperation. Such communication is defined as "cooperative communication" (Bormann et al., 2021, p. 6). Yet, cooperative communication is demanding and faces several challenges, i.e., assuring mutual trust through respect, considering the specific communication context, providing information that is necessary for cooperation, assuring mutual comprehension among the communication partners, and ensuring that communication participants' contributions are connected (Bormann et al., 2021). These challenges are addressed by five communication norms, namely, a *relation norm*, a *context norm*, an *information norm*, a *modality norm*, and a *process norm* (Bormann et al., 2021), which subsume several normative expectations held by participants involved in a communication towards the behavior of other participants. The five communication norms build the basis for their concept of incivility, which Bormann et al. (2021, p. 16) define as *communicative acts in public political discussions that participants disapprove of as severely violating one or several of the five communication norms*.

The concept integrates different approaches from previous research. The *relation norm* subsumes normative expectations of participants to communicate respectfully with other participants, which overlaps with politeness norms (e.g., Brown & Levinson, 1987; Fraser, 1990). Violations of this norm include, for example, insults, name-calling, and vulgarity, which are the types of incivility on which the majority of previous research has focused (e.g., Chen & Lu, 2017; Mutz, 2007; Rossini, 2020). The *context norm* expresses participants' expectations to consider the respective context of the discussion. In public political (online) discussions, participants expect each other to align their contributions with liberal democratic principles, or in other words, to consider democratic norms (Bormann et al., 2021; Papacharissi, 2004). Violations of this norm include, for example, threats to democracy or individual rights, stereotyping, or discriminating against social groups (e.g., Hopp, 2019; Kalch & Naab, 2017; Oz et al., 2018; Papacharissi, 2004). The other three norms have been less studied in incivility research, yet individual types of violations have been identified and should be studied in more detail. The *information norm* expresses participants' expectations to only communicate what is informative, that is, true and important, in the respective discussion

(Bormann et al., 2021). Violations of this norm include, for example, lies and misleading exaggerations (e.g., Stryker et al., 2016; Wang & Silva, 2018). The *modality norm* involves the expectation to communicate comprehensibly (Bormann et al., 2021), and can be violated, for example, by the use of sarcasm and irony or by ambiguous communication (e.g., Rowe, 2015; Ziegele & Jost, 2020). Finally, the *process norm* expresses normative expectations of participants to link their contributions, i.e., to discuss reciprocally, consistently, and to stay on topic (Bormann et al., 2021). Violations of this norm include interruptions, not responding to one's discussion partner, and off-topic contributions (e.g., Hopp, 2019; Ruiz et al., 2011; Sydnor, 2018).

Besides these five norms, there is a further crucial aspect inherent in defining incivility: In line with previous research (e.g., Chen et al., 2019; Herbst, 2010; Muddiman, 2017; Stryker et al., 2016, 2021), incivility is approached as a perceptual construct, according to which each communication participant decides what she/he disapproves of as uncivil. *Disapproval* is described as a two-step process (Bormann et al., 2021). First, the participants involved in a discussion recognize a violation of one or several of the five communication norms. Second, the participants evaluate the violation as worthy of sanction. As it is almost inevitable that communication norms will be violated from time to time, for instance in order to avoid worse violations (e.g., using a white lie to avoid an insult), it can be assumed that norm violations are sometimes tolerated. Therefore, this second step is of particular importance. A perceived violation of the communication norms is only considered uncivil if it is evaluated negatively, that is, as worthy of sanction.

Against the backdrop of the outlined incivility concept, and since violations of all five communication norms appear in incivility research and perceptual studies suggest a multidimensional model, we assume that violations of all five norms are disapproved as uncivil in online political discussions. This implies that participants in online political discussions (1) recognize violations of the communication norms and (2) classify them as worthy of sanction. Thus, we expect that

H1.1: Compared to norm-compliant behavior, violations of the relation norm, context norm, information norm, modality norm, and process norm are recognized as such by participants in online political discussions.

H1.2: Compared to norm-compliant behavior, violations of the relation norm, context norm, information norm, modality norm, and process norm are evaluated as worthy of sanction by participants in online political discussions.

As mentioned above, the concept by Bormann et al. (2021) assumes that incivility disapproval is a two-step process consisting of recognizing a norm violation first and then evaluating it as worthy of sanction. Therefore, we expect that

H2: The effect of the norm violation on the evaluation of sanction-worthiness is mediated by the recognition of the norm violation.

Severity of Incivility

Studies on perceptions of incivility suggest that distinct types of incivility vary considerably in perceived severity. For example, Muddiman (2017) examined perceptions of different forms of uncivil behavior in interactions among politicians, and defined the concepts of “personal-level incivility” (p. 3183), which can be classified as violations of the relation norm, and “public-level incivility” (p. 3183), which pertains to violations of the political context norm. The findings revealed that while both types were perceived as uncivil, personal-level incivility is rated as more uncivil than public-level incivility. However, the author only used a single-item scale to measure perceptions of incivility. Likewise, Kenski and colleagues (2020) surveyed perceptions of five different types of incivility in online discussions and found that uncivil behavior that can be classified as a violation of the relation norm (name-calling and vulgarity) was rated as most uncivil. Furthermore, Stryker and colleagues (2016) surveyed 23 forms of norm-violating behavior and found that the distinct types were rated quite differently. Specifically, inciting or threatening harm in a political discussion and "racial, sexist, ethnic, or religious slurs", which can be classified as violations of the political context norm, were rated as most uncivil (Stryker et al., 2016, p. 542). Types of incivility that can be classified as violations of the relation norm (e.g., insults and name-calling directed at other discussion participants) were also predominantly rated as very uncivil, while types of incivility referring to violations of the information norm (making misleading or exaggerated statements, failing to provide evidence) or the process norm (e.g., interrupting other participants) tended to be evaluated as more mildly uncivil (Stryker et al., 2016). In their replication study with a broader sample, Stryker et al. (2021) found very similar response patterns. However, in both studies, the authors used a single-item measurement scale asking participants to rate descriptions of norm-violating interactions from "not at all uncivil" to "very uncivil" (Stryker et al., 2016, p. 542; Stryker et al., 2021, p. 5). Moreover, the authors merely described the different types of incivility and did not study them experimentally. Hence, the question remains whether participants who are engaged in an online discussion also assess the severity of distinct types in the same way when they are actually exposed to them.

In sum, previous studies indicate that distinct types of norm violations are evaluated differently in terms of their severity. Although these studies were primarily based on survey data, applied different operationalizations of incivility, and did not undertake a differentiated measurement of the severity of distinct norm violations, some assumptions can be derived from their results. We expect violations of the relation norm and political context norm, in particular, to be evaluated as more severe than violations of the other three norms. Previous research suggests that norm violations are rated as especially severe when they are directed against people – either participants who are present in the respective discussion or absent persons or groups (e.g., Kenski et al., 2020; Stryker et al., 2016, 2021). Violations of the information norm have rarely been examined in incivility research, but initial findings suggest a milder degree of severity. Violations of the modality and process norm have also scarcely been researched, although the few available findings likewise suggest them to be milder types of incivility (Stryker, 2016, 2021). In light of the aforementioned findings, we propose one additional hypothesis:

H3: Violations of the relation norm and political context norm will be evaluated as more severe than violations of the information norm, modality norm, and process norm.

Reactions to Incivility

Similar to perceptions and evaluations, little is known about the effects that distinct types of incivility have on participants' behavior. In discussions on social media, on the websites of news media, or in political forums, lay participants usually have different options to react to uncivil comments by other participants (e.g., Kalch & Naab, 2017; Naab et al., 2021; Porten Cheé et al., 2020; Ruiz et al., 2011; Wilhelm et al., 2020). Bormann et al. (2021) identified three different types of reactions to uncivil comments, namely, leaving the discussion, ignoring the uncivil comment, and protesting against the norm violation.

Protesting against an uncivil act is defined as "explicit disapproval" (Bormann et al., 2021, p. 18). In online discussions, lay participants can usually express their explicit disapproval in *interactive* and *non-interactive* ways. Participants can verbalize explicit disapproval interactively, that is, in a discursive and verbal form by *writing a reply* to the uncivil commenter pointing out the norm violation, reprimanding or criticizing the perpetrator, or demanding sanctions (e.g., Gervais, 2015; Kalch & Naab, 2017; Porte-Cheé et al., 2020). They can also express their explicit disapproval non-interactively by using technical tools on discussion platforms (e.g., Porten-Cheé et al., 2020). The most common tool that discussion platforms provide to lay participants for non-interactive explicit disapproval is *flagging* (Kalch & Naab, 2017; Porten-Cheé et al., 2020; Wilhelm et al., 2020). By flagging comments, participants report perceived norm violations to platform providers or community managers (e.g., Crawford & Gillespie, 2016; Naab et al., 2018; Porten-Cheé et al., 2020; Wilhelm et al., 2020). Providers and community managers have several governance rights and options to sanction inappropriate behavior that lay participants have not. For example, they can delete or change the inappropriate comment and block the uncivil commenter, whereupon the commenter can no longer participate in the discussion (e.g., Friess et al., 2020; Stroud et al., 2015; Ziegele et al., 2018). Due to the tremendous number of user comments on websites and social media, almost all platforms have integrated flagging options. For platform providers, flagging is a useful mechanism to identify content that violates their guidelines and to involve the discussion participants in this process (e.g., Crawford & Gillespie, 2016; Porten-Cheé et al., 2020; Wilhelm et al., 2020). For participants in online discussions, flagging can be more effective than writing a reply, since sanctions by providers or community managers have immediate consequences for uncivil commenters due to their special governance rights (Friess et al., 2020; Naab et al., 2018).

Several studies have shown that inappropriate comments in online discussions are not usually ignored by participants but rather lead to a decreased willingness to stay in the discussion or to participants leaving the discussion (e.g., Hwang et al., 2008; Kluck & Krämer, 2021; Lück & Nardi, 2019; Pang et al., 2016). When participants stay in the discussion, they are likely to sanction norm-violating comments, i.e., show explicit disapproval by writing a sanctioning reply or by flagging (e.g., Gervais, 2015; Kalch & Naab, 2017; Naab et al., 2018; Ruiz et al., 2011; Singer, 2009; Wilhelm et al., 2020). A recent survey among German online users even revealed that compared to previous years, more and more users are explicitly counteracting hateful comments by flagging them or writing sanctioning replies (Media Authority of North Rhine-Westphalia, 2020). In particular, the study found that flagging as a reaction to hateful comments has steadily increased. This could be due to the fact that a growing number of platform providers have integrated flagging buttons and they are meanwhile a well-known, established intervention option (e.g., Crawford & Gillepse, 2016; Porten-Cheé et al., 2020).

Against this backdrop, we expect that

H4: Participants exposed to violations of the communication norms in online political discussions will show more explicit disapproval than participants who are not exposed to norm violations.

H5: Participants exposed to violations of the communication norms will more often exit the discussion than participants who are not exposed to norm violations.

Given the conceptual differences among the five communication norms, we further assume that distinct types of norm violations are likely to elicit different reactions from participants. However, scientific knowledge on the effects of distinct types of incivility is lacking. In a study by Naab and colleagues (2018), participants were more likely to flag comments that directly attacked individuals who identified as part of the LQBTQI+-community rather than comments that attacked a whole group. However, operationalizations of violations of the relation and context norm were not distinguished. Likewise, an experiment by Kalch and Naab (2017) revealed that participants were more likely to flag or write a sanctioning reply to a comment that directly attacked Muslims by using insults, vulgarity, and abusive language as compared to a comment that indirectly stereotyped Muslims and demanded the death penalty without referring to the right to a fair trial. Again, both examples violated democratic principles and thus, by our definition, violate the context norm. Therefore, it can only be concluded that different types of incivility seem to elicit different responses, but it remains unclear which type evokes which response.

Another interesting finding from Kalch and Naab’s (2017) study was that participants in online discussions usually limited themselves to one particular reaction. Instead of combining flagging and writing a sanctioning reply, for example, participants chose one option to engage against uncivil comments. Overall, participants used flagging to a greater extent than sanctioning replies, potentially because writing replies requires more effort than clicking a flagging button (Porten-Cheé et al., 2020). Another reason might be that participants consider flagging to be a more effective sanctioning measure. However, one can also interpret from these findings that participants attribute different functions to different forms of explicit disapproval and therefore use them depending on the type of norm violation (Kalch & Naab, 2017). This would suggest that distinct types of norm violations evoke different forms of explicit disapproval.

Following a similar approach, Wilhelm and colleagues (2020) tested whether five different types of norm violations, namely calls for violence, agitation, defamation, rumor, and conspiracy theories, predict flagging behavior. According to our definition, these types of norm violations can be classified as violations of the context and information norm. Their findings suggested that flagging behavior differed significantly among the distinct types of norm violations, with incitements of violence (i.e., violation of the context norm) being more likely to be flagged than rumors or conspiracy theories (i.e., violation of the information norm). However, a clear-cut hierarchy in flagging behavior across the five types of norm violations was not found.

In sum, no clear assumptions can be derived from previous studies about how participants respond to distinct violations of the communication norms. Therefore, we ask

RQ1: Do distinct types of norm violations lead to different reactions, namely ignoring the norm violation, exiting the discussion, or different forms of explicit disapproval?

Method

To investigate the hypotheses and research questions, we conducted an online experiment that utilized a six-condition, single-level, between-subjects design. Subjects were randomly assigned to exposure to a norm-compliant comment or to a comment violating one of the five communication norms. The comments were posted in a fully functional mock-up online discussion forum. During the study, participants were asked to take part in a simulated, but to them seemingly real, discussion in this forum. The experiment was conducted from 20 October to 3 November 2021 in Germany.

Sample

In total, 447 participants who were recruited from a commercial online panel completed the survey. We excluded 14 participants who indicated that they had technical problems with the forum or who showed suspicious response behavior, which we identified through manual data checks. The analysis thus refers to 433 participants, and the sample is representative of the German population with regard to gender and age. Overall, 216 (49.9%) of the participants identified as female, 214 (49.4%) as male, and three (0.7%) as gender-diverse. The participants ranged in age from 18 to 74 years (*M* = 46.3, *SD* = 15.9). The majority had a high educational level, with either a university degree (31.4%), PhD (1.8%) or the general qualification for university entrance (*Abitur*) (30%). Furthermore, most of the participants used social media (55.2%) or the websites of news media (72.3%) frequently, i.e., daily or one to several times a week, to get political news. Moreover, 53% of the participants reported reading user comments on social media or news websites/forums daily or one to several times a week. Writing user comments was less common, with 80.2% of the participants indicating writing user comments once a month or less.

Procedure

The experiment was embedded in an online survey. Participants were introduced to a cover story stating that the purpose of the study was to evaluate a newly developed public online discussion forum named *Let's discuss*[8] (please refer to Appendix A for screenshots of the forum). We used a fully functional platform and deployed the cover story to simulate a realistic discussion environment and to actually engage participants in an online discussion. In line with our conceptual definition of incivility, we thus wanted to examine the perceptions of and reactions to distinct types of norm violations by participants who are actually involved in

[8] The forum was originally developed for a previous, unpublished study and modified for the purpose of the present study. Detailed information about the previous study can be found here: https://osf.io/p8z6u?view_only=ad81ffd6d58742cf919b69e1ac128f9c

an online discussion. Moreover, by implementing such a realistic environment and simulating public interaction, external validity can be increased.

The survey began with questions about socio-demographic characteristics and the use of social media and other discussion platforms. Subsequently, participants were introduced to *Let's discuss* and to two specific procedures of the platform: First, discussions are round-based, with all participants writing a comment one after the other in the first round. It was explained that this procedure is to ensure that everyone has the chance to post a comment voicing her/his opinion. Second, participants are only allowed to react to the previous comment. Participants were told that they had several options how to react: They could either reply to the comment, or leave the current discussion and would be automatically redirected to another one. In actual fact, however, these procedures were only introduced to control that the participants reacted to the manipulated comment. Finally, before directing the participants to the forum, we informed them about the *flagging* option. It was stated that by flagging a comment, they would anonymously report it to the platform providers, and that the providers would decide whether to delete the comment and/or block the commenter.

In the forum, participants were randomly assigned to one of two different political topics, namely universal basic income or direct democracy. Each topic was introduced by a short and unbiased information text to ensure that all participants shared the same level of knowledge, and participants were asked to discuss the pros and cons of introducing a universal basic income or more forms of direct democracy in Germany. The topics were not treated as an experimental factor but were merged in the analysis in order to increase the generalizability of the results and to eliminate effects of the individual topics. The introductory texts were each followed by two norm-compliant comments that represented an ambivalent and a contra position. The third comment represented a pro-position and was manipulated: It was either norm-conforming or violated one of the five communication norms. To create a realistic impression of a real-time discussion, the comments appeared one after another with a time delay (please refer to Table A1 in the Appendix for the time intervals between the comments). After the three comments appeared, participants were asked to react to the manipulated comment, with the option of replying or leaving the discussion. If the participant chose to reply, she/he could write a comment and was then redirected to the survey. If the participant chose to leave, she/he was returned directly to the survey.

Back in the survey, participants answered questions referring to their perception of violations of the five communication norms and to their evaluation of the comment as deviant, harmful, or worthy of sanction. Lastly, they were asked to evaluate the forum, which served to check for technical problems, comprehension and credibility of the introductory texts, perceived authenticity of the comments, and honesty of the participants' behavior. After the experiment, the participants were fully debriefed.

Stimuli

Due to the large number of different forms of violations of the five communication norms, we could not examine all potential types of violations. Therefore, we selected and operationalized one representative type of violation per communication norm (see Table 1). All of these types have been defined as uncivil in the existing incivility literature, albeit with varying frequency. As a baseline, we first constructed the two comments, i.e. one for each topic, that were compliant with the communication norms. Next, we developed the comments that each violated one of the five communication norms[9]. All of the manipulated comments consisted of the same pro-argument regarding the topics and addressed the previous commenter, except for the comment that included the violation of the process norm, namely topic deviation. Moreover, the comments were comparable in terms of length and language usage between and within the two topics, and we included, for example, minor spelling errors

[9] The stimulus materials were used in the present study and for a different purpose in another study. Detailed information about the unpublished study can be found here: https://osf.io/p8z6u?view_only=ad81ffd6d58742cf919b69e1ac128f9c

to make them appear more realistic. The severity of the different types of norm violations was also kept as constant as possible and the individual types of norm violations were operationalized in a very distinctive manner to avoid conceptual overlap between different types of violations.

[Table 1 here]

The comment that included a violation of the relation norm was operationalized by an insult against the previous commenter, accompanied by vulgar language. The comment that included a violation of the context norm was operationalized by an antagonistic stereotype and a threat of violence against a certain group. In both topics, the stereotype and threat of violence were directed against an elite group, namely politicians and employers. The comment including a violation of the information norm was operationalized by obviously false information about the respective topic that could be identified based on the information provided in the introductory texts. The ironic/sarcastic comment included an overemphasis of counterintuitive arguments, serving as a marker of irony. Finally, the comment that included a violation of the process norm did not address the previous commenter and deviated from the topic, stating that another topic is more relevant and should be discussed instead. We pretested all manipulated comments to ensure that the different types of norm violations were perceived as intended, namely as insulting, ironic/sarcastic, deviating from the topic, or as including lies or stereotypes and threats of violence. Please refer to Appendix B for detailed information on the procedure and the results of the pretest, and to Appendix C for the comments used in the study.

Measures

Perceived Violation of Communication Norms

We developed scales to measure whether participants perceived violations of the communication norms. For this purpose, each of the five communication norms was operationalized using four items (or five in the case of the more complex context norm). The operationalization was based on the definition of the norms and their individual dimensions (Bormann et al., 2021), and the items were formulated positively, i.e., in conformity with the norms. Participants rated the items on a 7-point Likert scale from 1 = "strongly disagree" to 7 = "strongly agree", with high values thus indicating that the manipulated comment was perceived as compliant with the respective communication norm. Regarding the *relation norm*, participants were asked to rate statements like "The comment is respectful to the other participants in the discussion" or "The comment is polite to the other participants in the discussion", and the four items were aggregated to an index with high reliability (Cronbach's α = .98, *M* = 5.1, *SD* = 1.9). The *context norm* was operationalized using items such as "The comment respects the fundamental values of our democracy", or "The comment respects the rights of all people, regardless of which group they belong to" (α = .95, *M* = 5.3, *SD* = 1.7). The *information norm* was operationalized by items such as "The comment contains information on the discussion topic that is clearly true" or "The comment contains important information for the discussion" (α = .92, *M* = 4.4, *SD* = 1.6). Regarding perceived compliance with or violation of the *modality norm*, participants were asked to rate statements like "The comment is clearly understandable" or "The comment is worded concisely" (α = .90, *M* = 5.3, *SD* = 1.4). Lastly, the *process norm* was operationalized using items such as "The comment ties in with previous comments" or "The comment relates to the discussion topic" (α = .91, *M* = 5.3, *SD* = 1.4).

Evaluation of Norm Violations

Sanction-worthiness of the norm violations was measured with five items such as "The comment should be intervened against" or "The author of the comment should be warned" on a scale from 1 = "strongly disagree" to 7 = "strongly agree" (α = .86, *M* = 2.9, *SD* = 1.6). *Deviance and harmfulness* of the norm violations were measured using semantic differential items adapted from Naab (2016), Naab et al. (2018), and Ziegele et al. (2020), supplemented with additional items developed by us. The eleven items were rated on 7-point scales with anchors such as “appropriate/inappropriate,” “tolerable/intolerable,” “unproblematic/problematic,” “harmless/harmful,” “non-dangerous/dangerous” (α = .97, *M* = 3.2, *SD* = 1.6).

Reactions to Norm Violations

Leaving the discussion: The forum captured whether the participants did (1) or did not choose (0) to leave the discussion. *Ignoring* the norm violation was operationalized as staying in the discussion and as the absence of explicit disapproval. *Explicit disapproval* was operationalized as flagging and/or writing a sanctioning reply. The forum captured whether the participants flagged (1) or did not flag (0) the manipulated comment. To examine whether the participants wrote sanctioning replies, all comments (*N* = 373) were content analyzed. We followed Gervais (2014) and extended his operationalization to ascertain whether a *disapproving criticism* (Krippendorff’s α = .87) was evident regarding the content, expression, behavior, or personal characteristics of the manipulated comment or commenter. Direct indications of a violation of the communication norms or reprimands were also coded within this category (e.g., “this is an insult,” “this is a lie,” “this comment should be sanctioned”). The comments were coded by two of the authors, who completed coding training and the reliability test using 75 comments (20% of the total comments).

Controls for the Setting

Comprehensibility of the platform rules (*M* = 6.4, *SD* = .9), believability of the information text (*M* = 6.2, *SD* = 1.1), realism of the comments and discussion (*M* = 5.6, *SD* = 1.6; *M* = 5.6, *SD* = 1.7), honesty of one’s own reaction (*M* = 6.5, *SD* = 1.1), and whether the participants had no technical issues on the platform (*M* = 6.6, *SD* = .9) were measured as controls for the setting (from 1 = “strongly disagree” to 7 = “strongly agree”).

Results

Effects of Norm Violations on Perceived Norm Conformity and Evaluation of Sanction-worthiness

To test H1.1, we conducted six factorial ANOVAs with the type of norm violation as independent variable. To control for the topic, it was added as a second independent variable. First, we added perceptions of norm conformity in general as dependent variable. Next, to investigate whether the norm violations were perceived as intended, we conducted five ANOVA models (one for each norm) and added perceived norm conformity in the intended dimension as dependent variable (e.g. perceived conformity with the relation norm, context norm etc.). To test H1.2, we conducted an ANOVA model with participants' evaluation of sanction-worthiness as dependent variable. The main effect of norm violation on perceived general norm conformity was statistically significant, $F(5, 421) = 19.07, p < .001, \eta_p{}^2 = .19$. Furthermore, there was a significant main effect of norm violation on the evaluation of sanction-worthiness, $F(5, 421) = 35.03, p < .001, \eta_p{}^2 = .29$. To investigate whether norm violations were recognized and evaluated as worthy of sanction compared to norm-compliant behavior, we conducted planned contrasts with norm-compliant behavior as reference category. To investigate whether the five norm violations were perceived as uncivil according to the theoretical considerations by Bormann et al. (2021), we ran mediation analyses using the PROCESS macro (Hayes, 2018), with evaluation of sanction-worthiness as dependent variable and perceived norm conformity in the intended dimension as mediator (H2). In the following, we report the remaining results regarding H1 and H2 sorted by norm violations.

Violation of the Relation Norm

The contrast analysis revealed that a comment that included a violation of the relation norm was perceived as significantly less norm-compliant in terms of general norm conformity ($M = 4.1$, $SD = 1.3$) compared to norm-compliant behavior ($M = 5.8$, $SD = 1.0$, $p < .001$). The main effect of norm violation on perceived conformity with the relation norm was also significant, *F*(5, 421) = 48.89, *p* < .001, $\eta_{p}{}^{2}$ = .37. Planned contrasts revealed that a comment that violated the relation norm was perceived not only as less norm-compliant in general, but also in the intended dimension (*M* = 2.8, *SD* = 2.1) compared to the control group (*M* = 6.4, *SD* = 1.0, *p* < .001). Furthermore, such comments were perceived as significantly more worthy of sanction (*M* = 4.6, *SD* = 1.9) than norm-compliant behavior (*M* = 1.9, *SD* = 0.9, *p* < .001). The mediation analysis (see Figure D1 in the Appendix D) confirmed these findings. Furthermore, there was a significant indirect effect on the evaluation of sanction-worthiness through perceived norm conformity, *b* = 1.62, BCa CI [1.20, 2.05]. However, the direct effect was still strong and significant, *b* = 1.11, *p* < .001. In conclusion, the effect on sanction-worthiness was partly mediated by perceived norm conformity.

Violation of the Context Norm

A comment that violated the context norm was also perceived as significantly less norm-compliant in general (*M* = 4.7, *SD* = 1.4) compared to the control group (*M* = 5.8, *SD* = 1.0, *p* < .001). Again, there was a significant main effect of norm violation on perceived norm conformity in the intended dimension, *F*(5, 421) = 31.07, *p* < .001, $\eta_{p}{}^{2}$ = .27. Planned contrasts revealed that a violation of the context norm was perceived as significantly less compliant with the context norm (*M* = 4.2, *SD* = 2.0) than a comment that contained no norm violations (*M* = 6.2, *SD* = 1.1, *p* < .001). Furthermore, participants perceived the comment as significantly more worthy of sanction (*M* = 3.3, *SD* = 1.7) compared to the control group (*M* = 1.9, *SD* = 0.9, *p* < .001). The mediation analysis (see Figure D2 in Appendix D) confirmed these findings, and also revealed a significant indirect effect on the evaluation of sanction-worthiness through perceived norm conformity, b = 1.00, BCa CI [0.68, 1.36]. The direct effect was substantially smaller and only marginally significant, *b* = 0.38, *p* = .074. Thus, we conclude that the effect on sanction-worthiness was mostly mediated by perceived norm conformity.

Violation of the Information Norm

Regarding general norm conformity, perceptions of participants that were exposed to a comment that violated the information norm (*M* = 5.6, *SD* = 0.9) did not differ significantly from perceptions in the control group (*M* = 5.8, *SD* = 1.0, *p* = .094). Nevertheless, a comment that contained a violation of the information norm was perceived as significantly less compliant with that particular norm (*M* = 4.6, *SD* = 1.6) compared to norm-compliant behavior (*M* = 5.1, *SD* = 1.3, *p* = .043), *F*(5, 421) = 5.98, *p* < .001, $\eta_p{}^2$ = .07. Additionally, a violation of the information norm was perceived as significantly more worthy of sanction (*M* = 2.5, *SD* = 1.2) than a comment that did not contain any norm violations (*M* = 1.9, *SD* = 0.9, *p* = .011). The mediation analysis (see Figure D3 in Appendix D) mostly confirmed these results. However, the negative effect of the violation of the information norm on perceived norm conformity was only marginally significant (*b* = -0.47, *p* = .077). Consequently, there was no significant indirect effect on evaluation of sanction-worthiness through perceived norm conformity, *b* = 0.15, BCa CI [0.00, 0.32]. The effect on sanction-worthiness was not mediated by perceived norm conformity.

Violation of the Modality Norm

Planned contrasts revealed that a comment that violated the modality norm was perceived as significantly less norm-compliant with respect to general norm conformity (*M* = 5.0, *SD* = 1.2) compared to the control group (*M* = 5.8, *SD* = 1.0, *p* <.001). Furthermore, participants perceived the comment as less compliant with the modality norm (*M* = 5.0, *SD* = 1.4) compared to norm-compliant behavior (*M* = 5.6, *SD* = 1.4, *p* = .005), *F*(5, 421) = 2.45, *p* = .03, $\eta_p{}^2$ = .03. A violation of the modality norm was also perceived as significantly more worthy of sanction (*M* = 2.7, *SD* = 1.3) compared to a comment that contained no norm violations (*M* = 1.9, *SD* = 0.9, *p* < .001). The mediation analysis (see Figure D4 in Appendix D) confirmed these findings. There was a significant indirect effect on the evaluation of sanction-worthiness through perceived norm conformity, *b* = 0.20, BCa CI [0.06, 0.38]. The direct effect of the treatment on the evaluation of sanction-worthiness remained significant, *b* = 0.60, *p* = .007. Hence, the effect on sanction-worthiness was partly mediated by perceived norm conformity.

Violation of the Process Norm

Finally, participants who were exposed to a violation of the process norm perceived the comment as significantly less norm-compliant in general (*M* = 5.2, *SD* = 1.2) compared to norm-compliant behavior (*M* = 5.8, *SD* = 1.0, *p* < .001). A comment that violated the process norm was also perceived as less compliant with that norm (*M* = 4.7, *SD* = 1.9) compared to the control group (*M* = 5.9, *SD* = 1.2, *p* < .001), *F*(5, 421) = 8.00, *p* < .001, η_p^2 = .09. In contrast, the comment was not perceived as more worthy of sanction (*M* = 2.2, *SD* = 0.9) compared to a comment that contained no norm violations (*M* = 1.9, *SD* = 0.9, *p* = .190). The mediation analysis (see Figure D5 in Appendix D) confirmed these effects. However, there was a significant indirect effect on the evaluation of sanction-worthiness through perceived norm conformity, *b* = 0.35, BCa CI [0.18, 0.55]. The direct effect was minimal and not significant, *b* = -0.07, *p* = .75. Thus, we conclude that the effect on sanction-worthiness was fully mediated by perceived norm conformity.

In summary, H1 and H2 can be largely confirmed. For the most part, participants recognized norm violations and assessed them as worthy of sanction, and the effects of most norm violations on perceived sanction-worthiness were at least partially mediated by perceived norm conformity. However, it should also be noted that in most cases, the norm violations were not only perceived as less compliant with the intended communication norm but also as violations of other communication norms. For example, insults and vulgarity against another participant were not only perceived as less compliant with the relation norm (*M* = 2.8, *SD* = 2.1, *p* < .001), but also as violating the context norm (*M* = 4.0, *SD* = 1.8, *p* < .001), the information norm (*M* = 3.8, *SD* = 1.6, *p* < .001), the modality norm (*M* = 5.1, *SD* =

1.5, *p* = .008), and the process norm (*M* = 5.0, *SD* = 1.3, *p* < .001) compared to norm-compliant behavior.

Perceived Severity of Norm Violations

H3 postulated that violations of the relation norm and the political context norm would be evaluated as more severe than other norm violations. Severity was measured according to participants' evaluations of (1) the sanction-worthiness of a comment and (2) its deviance and harmfulness. To test H3 regarding sanction-worthiness, we used the same ANOVA model that was used to investigate H1.2. As reported above, there was a significant effect of norm violation on perceived sanction-worthiness, and all norm violations except for the process norm were perceived as more worthy of sanction compared to norm-compliant behavior. Furthermore, there was a significant effect of norm violation on perceived deviance and harmfulness, *F*(5, 421) = 26.65, *p* < .001, $\eta_p{}^2$ = 0.24. Bonferroni post hoc tests revealed that a violation of the relation norm was perceived as significantly more worthy of sanction (*M* = 4.6, *SD* = 1.9) and as more deviant and harmful (*M* = 4.5, *SD* = 1.7) than violations of the information norm (sanction-worthiness, *M* = 2.5, *SD* = 1.2, *p* < .001; deviance and harmfulness, M = 2.8, SD = 1.3, p < .001), the modality norm (sanction-worthiness, *M* = 2.7, *SD* = 1.3, *p* < .001; deviance and harmfulness, *M* = 3.1, *SD* = 1.4, p < .001), and the process norm (sanction-worthiness, *M* = 2.2, *SD* = 0.9, *p* < .001; deviance and harmfulness, *M* = 2.7, *SD* = 1.1, *p* < .001). Likewise, a violation of the context norm was perceived as more worthy of sanction (*M* = 3.3, *SD* = 1.7) and as more deviant and harmful (*M* = 3.8, *SD* = 1.8) than violations of the information norm (sanction-worthiness, *p* = .009; deviance and harmfulness, *p* < .001) and the process norm (sanction-worthiness, *p* < .001; deviance and harmfulness, *p* < .001). However, compared to a violation of the modality norm, a comment that violated the context norm was not perceived as significantly more worthy of sanction (*p* = .259) or more deviant and harmful (*p* = .085), although the difference regarding deviance and harmfulness was marginally significant. Therefore, H3 was mostly confirmed. Additionally, a violation of the relation norm was perceived as more worthy of sanction ($p < .001$) and as more deviant and harmful ($p = .023$) than a violation of the context norm.

Reactions to Norm Violations

H4 assumed that participants who were exposed to norm violations would show more explicit disapproval compared to the control group. To test this hypothesis, we first computed the variable "explicit disapproval" and coded whether a participant either flagged a comment or wrote a reply that contained disapproving criticism. Next, we conducted a logistic regression with the norm violations as independent variable and explicit disapproval as dependent variable. First, it should be stated that not a single participant in the control group reacted with explicit disapproval, leading to a quasi-complete separation caused by the control group, which could consequently not be used as a reference category. Therefore, to form the reference category, we combined the control group with the group with the lowest occurrence of explicit disapproval, which was the group that was exposed to a violation of the modality norm (Allison, 2008). The logistic regression revealed significant positive effects of violations of the information norm, process norm, relation norm, and context norm on explicit disapproval (see Table 2). Therefore, H4 can be confirmed. The effect of a violation of the relation norm was by far the strongest.

[Table 2 here]

To investigate whether participants exposed to norm violations would more often exit the discussion compared to the control group (H5), we conducted another logistic regression with participants' decisions to leave the discussion as dependent variable. As can be seen in Table 3, there were no significant effects of any norm violation. Therefore, H5 could not be confirmed.

[Table 3 here]

To investigate whether distinct types of norm violations lead to different reactions (RQ1), we again conducted logistic regressions with flagging, disapproving criticism, and exiting the discussion as dependent variables. Furthermore, we computed the variable "ignoring the norm violation", that is, if a participant neither showed explicit disapproval nor left the discussion, and added this as another dependent variable. As mentioned above (results for H4), a violation of the relation norm was by far the most likely to lead to explicit disapproval (see Table 2). That was also true for disapproving criticism alone (see Table 4). With regard to flagging, violations of the relation norm and the context norm were the only types of violation that participants flagged at all[10], meaning that four of the six categorical variables led to quasi-complete separation. Therefore, we conducted two regression models, first with a violation of the relation norm and then with a violation of the context norm as independent variable, with all other groups as reference category. Tables 5 and 6 show a significant effect of a violation of the relation norm on flagging, but no significant effect of a violation of the context norm. As mentioned with respect to H5, no norm violation led to leaving the discussion (see Table 3). Additionally, violations of the relation and context norm were less likely to be ignored compared to the control group (see Table 7). A violation of the context norm was more likely to be ignored than a violation of the relation norm. There were no significant effects of other norm violations.

[Tables 4, 5, 6, and 7 here]

Discussion

The main purpose of this study was to examine what participants who are actually engaged in an online discussion perceive as uncivil, how they evaluate different types of norm violations in terms of severity, and how they react to them. Based on the approach of Bormann et al. (2021), incivility was defined as a disapproved violation of one or several of five communication norms. In an experimental setting with a fully functional discussion

[10] n = 18 in the condition "violation of the relation norm" and n = 6 in the condition "violation of the context norm"

forum, participants were confronted with comments that contained violations of the communication norms.

Incivility as a Multidimensional Concept

Following Bormann et al. (2021), we defined disapproval of norm violations as a two-step process including (1) perception of a violation of the communication norms and (2) evaluation of the violation as worthy of sanction. Thus, our first assumption (H1.1) was that violations of the five communication norms would be recognized by participants in online political discussions. This hypothesis was confirmed: Insults and vulgarity against another discussion participant were perceived as a violation of the relation norm; stereotypes and threats of violence against a specific group were recognized as violations of the context norm; irony/sarcasm was perceived as violating the modality norm; topic deviation was perceived as violating the process norm; and false information was perceived as violating the information norm. The violation of the information norm was the least well recognized, although interestingly, it was nevertheless evaluated as being worthy of sanction. Thus, while participants apparently perceived a norm violation as worthy of sanction, they may have been unable to clearly identify it as a violation of the information norm. We know from previous research that processing false information is a complex endeavor (e.g., Lewandowsky et al., 2012), and combining this with the results of our study, it can be assumed that there is a difference between the perception of a norm violation and the correct classification of this violation, especially for more implicit norm violations.

Furthermore, a very interesting finding regarding H1.1 was that most norm violations were not only perceived as violations of the intended communication norm but also as violations of other communication norms. For example, insults and vulgarity against a discussion participant were not only perceived as disrespectful to the discussion partner (relation norm) but also as less respectful of liberal democratic principles (context norm), less informative (information norm), less comprehensible (modality norm), and less connective (process norm). Thus, while the five communication norms do seem to exist, they seem to be intermingled on a perceptual level. Violations of individual norms are not perceived in a distinctive manner, but rather also have negative effects on the perception of the other communication norms. Nevertheless, all violations were most strongly perceived as violating the intended communication norm. It would be interesting for future research to systematically examine the perceptual overlaps between the norms.

Our next assumptions were that violations of the communication norms would be evaluated as worthy of sanction (H1.2) and that this effect would be mediated by the perception of violations of the communication norms (H2). These hypotheses were largely confirmed and the results thus support the broad incivility concept of Bormann et al. (2021) and reflect the findings of previous perception-oriented studies (e.g., Muddiman, 2017; Stryker et al., 2016, 2021). Incivility does not merely encompass insults and stereotypes. Rather, participants in online political discussions also disapprove of violations of information, modality, and process norms as uncivil. With the exception of the information norm, which was not well recognized as such, the two-stage process of disapproval was also confirmed: The effects of the norm violations on sanction-worthiness were mediated at least in part by the perceived violation of the intended communication norm, except for the information norm.

Severity of Incivility

We further expected the severity of distinct types of norm violations to vary insofar as violations of the relation and context norms would be rated as more severe than violations of the other norms (H3). This assumption was largely confirmed and the results are thus in line with previous studies (e.g., Kenski et al., 2020; Muddiman, 2017; Stryker et al., 2016, 2021). Notably, the violation of the relation norm was rated as the most severe violation by far. Participants perceive explicit attacks against others as the most harmful, deviant, and sanction-worthy type of incivility. This is likely because such attacks can be easily recognized, in contrast to *impeccable incivility* (Papacharissi, 2004), which at first glance appears to be polite due to the avoidance of name-calling or obscene language, but contains, for example, implicit stereotypes and discrimination. Moreover, the violation of the relation norm targeted (1) a person rather than an object, and (2) a participant involved in the discussion, unlike the violation of the context norm, which targeted absent third parties. Other researchers have already suggested distinguishing between the targets of incivility, such as *interpersonal* vs. *other-directed* (Papacharissi, 2004; Rowe, 2015) or *personal* vs. *impersonal* (Su et al., 2018), and such distinctions indeed seem to have an impact on evaluations of incivility. Additionally, it appears to be important to distinguish whether the target is people or objects such as institutions or topic-related aspects.

When whole groups are attacked, the perceived severity can also vary depending on the group. The present findings indicate that the violation of the context norm is perceived as worse when it refers to the topic of direct democracy compared to the topic of a universal basic income. In other words, the attack against politicians was rated as worse than the attack against employers, even though two reasonably comparable elite groups were being attacked. Presumably, the differences would have been even greater if other social groups had been chosen as the stimulus, for example marginalized, vulnerable groups.

Reactions to Incivility

We further expected that participants exposed to violations of the communication norms would show more explicit disapproval (H4) and would more often exit the discussion (H5) than participants who were not exposed to norm violations. Rather than leaving, the participants tended to stay in the discussion and explicitly disapproved the norm-violating comment, which is a favorable outcome from a democratic and deliberative perspective (e.g., Ellis, 2012; Gastil, 2008). The lack of support for H5 is an interesting finding, since several previous studies reported that incivility (particularly when not like-minded) hinders participation (e.g., Hwang et al., 2008; Lück & Nardi, 2019; Pang et al., 2016). However, the finding might also be due to our study design, the rules in the forum, and the cover story that asked the participants to test the discussion forum. Moreover, before the norm-violating comment, two norm-compliant comments were shown, and thus the majority of comments were norm-compliant. According to the *focus theory of normative conduct* (e.g., Cialdini et al., 1991), it might be argued that the descriptive and salient norms in the forum were to actively participate and behave in a norm-compliant manner, which may have positively impacted participants' motivation to stay in the discussion and write a comment reprimanding the uncivil commenter.

Furthermore, the participants showed explicit disapproval of almost all types of norm violations by writing a sanctioning comment. More specifically, violations of the relation norm, context norm, information norm, and process norm led to *disapproving criticism* in participants' reply comments, but violations of the relation norm were by far the most likely to elicit disapproval. Compared to writing a sanctioning reply, *flagging* was used less often as a form of explicit disapproval. Only violations of the relation and context norm had an effect on flagging, with the former being flagged more often than the latter.

The findings suggest that lay participants in online discussions are likely to engage against most types of incivility, but use different forms of explicit disapproval depending on the type of norm violation and its level of severity. Compared to writing a reply with disapproving criticism, flagging might be seen as a stricter form of disapproval, which is better tailored to more severe types of incivility as flagging has more serious and direct consequences for the uncivil commenter. Flagging was only used in the case of violations of relation or context norms that were evaluated as quite severe, and participants were also less likely to ignore violations of these two norms. Findings by Wilhelm et al. (2020) and Kalch and Naab (2017) point in a similar direction, as the authors found that distinct norm violations elicited different responses and that participants were more likely to flag severe, explicit norm violations directed against an individual or a group.

Theoretical and Practical Implications

Our study yields several theoretical and practical implications. We extended existing research on perceptions of and reactions to incivility (e.g., Kalch & Naab, 2017; Kenski et al., 2020; Stryker et al., 2016, 2021) by exposing participants to multiple potential types of incivility in an experimental setting. Moreover, we measured perceptions and evaluations of distinct types of norm violations, as well as reactions to these violations, in a nuanced manner. The results suggest that incivility is a (1) perceptual and (2) multidimensional construct, and that distinct types of incivility vary in their level of severity and lead to different reactions. If distinct types elicit different responses, other effects of incivility are also likely to vary by type and perceived severity. This might explain the heterogeneous findings of several previous studies regarding the consequences of incivility, as these studies often used different types or mixed distinct types of incivility within one study (e.g., Borah, 2012; Chen & Lu, 2017; Hwang et al., 2008). Future studies should approach incivility from a multidimensional and perceptual perspective, and examine what effects distinct types of incivility have on different participants, on third parties, and on discussion dynamics.

Our results additionally provide some relevant insights for platform operators and moderation practice. User engagement against uncivil comments was very high, which would also be desirable on real platforms such as comment sections on social media or on the websites of news media. A self-regulating community could reduce the demands on professional moderators and avoid the frequent accusation of restricting freedom of speech (e.g., Meyer & Carey, 2014; Wright, 2006). The innovative techniques and rules of the mock-up discussion forum *Let's discuss* which was used in the present study, such as replying to the previous comment and the precise information about rules and possibilities of intervention, might have encouraged the participants to engage against uncivil comments. The specific study situation certainly encouraged engagement as well, although studies with comparable settings tended to show less engagement, for example in writing comments (e.g., Kalch &

Naab, 2017). Therefore, it might be worthwhile to develop and apply innovative techniques and rules on discussion platforms. Moreover, it could be beneficial for platform providers and community managers to precisely inform their users about the rules, communication norms, and possibilities of intervention on their platform, and to explicitly involve the users in moderation practice. Professional moderators should additionally pay particular attention to impeccable incivility, which is harmful but difficult to detect by discussion participants and by algorithms that are frequently employed in moderation practice (e.g., Stoll et al., 2020).

Limitations

This study has several limitations. First, external validity was decreased and increased at the same time by (specifics of) the discussion forum. On the one hand, we had to ensure that the participants responded to the manipulated comment, and therefore developed specific discussion rules. Consequently, the discussion on *Let's discuss* is not comparable to a discussion on social media or on news websites, and the participants did not necessarily behave as they would in a real discussion on familiar platforms. On the other hand, we designed the forum to be as realistic as possible and the vast majority of the participants felt that the discussion was real and authentic. A major benefit of the forum was that it allowed us to examine the perceptions of participants involved in an online discussion and to measure their actual behavior in response to norm violations. Second, we only examined one representative type of violation for each communication norm, meaning that conclusions for the entire norm are limited. Third, we did not manipulate the level of severity. Future studies should analyze and compare several types of violations of each communication norm and varying levels of severity. Fourth, as there are no standardized measures that would have been suitable for our constructs, we had to develop several new scales. Although the scales are theoretically well-founded, their validity can only be determined to a limited extent. Finally, we considered few factors that might have an additional impact on processing and reacting to different types of incivility. Future studies should examine, for example, whether personality traits (e.g., Kenski et al., 2020) or political partisanship (e.g., Gervais, 2015) have an impact on perceptions and reactions to different types of norm violations.

Conclusion

Overall, our study provides relevant insights into how participants of online political discussions process distinct types of norm violations and how they react to them. The results support a multidimensional model of incivility as a disapproved violation of one or several of five communication norms. While violations of all five norms can be classified as uncivil, we identified considerable differences among the distinct types of incivility in terms of their severity and the responses to them. Future studies can draw on this differentiated incivility concept in order to examine patterns, determinants, causes and consequences of distinct types of norm violations, and to better tailor intervention strategies against different forms of uncivil behavior.

References

Allison, P. D. (2008). *Covergence failures in logistic regression.* In SAS Global Forum 2008, San Antonio, Texas, USA. https://support.sas.com/resources/papers/proceedings/pdfs/sgf2008/360-2008.pdf

Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2014). The "nasty effect:" Online incivility and risk perceptions of emerging technologies. *Journal of Computer-Mediated Communication, 19*(3), 373–387. https://doi.org/10.1111/jcc4.12009

Boatright, R. G. (2019). A crisis of civility? In R. G. Boatright, T. Shaffer, S. Sobieraj, & D. Goldthwaite Young (Eds.), *A crisis of civility? Political discourse and its discontents* (pp. 1–6). Routledge.

Borah, P. (2012). Does it matter where you read the news story? Interaction of incivility and news frames in the political blogosphere. *Communication Research, 41*(6), 809–827. https://doi.org/10.1177/0093650212449353

Bormann, M., Tranow, U., Vowe, G., & Ziegele, M. (2021). Incivility as a violation of communication norms: A typology based on normative expectations toward political communication. *Communication Theory.* https://doi.org/10.1093/ct/qtab018

Brown, P., & Levinson, S. C. (1987). *Politeness. Some universals in language usage.* Cambridge University Press.

Chen, G.M. (2017). *Online incivility and public debate: Nasty talk*. Palgrave Macmillan.

Chen, G. M., & Lu, S. (2017). Online political discourse: Exploring differences in effects of civil and uncivil disagreement in news website comments. *Journal of Broadcasting & Electronic Media*, *61*(1), 108–125. https://doi.org/10.1080/08838151.2016.1273922

Chen, G. M., & Ng, Y. M. M. (2017). Nasty online comments anger you more than me, but nice ones make me as happy as you. *Computers in Human Behavior*, *71*, 181–188. https://doi.org/10.1016/j.chb.2017.02.010

Chen, G. M., Muddiman, A., Wilner, T., Pariser, E., & Stroud, N. J. (2019). We should not get rid of incivility online. *Social Media and Society*, *5*(3), 1–5. https://doi.org/10.1177/2056305119862641

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 24, pp. 201–234). Academic Press.

Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, *64*(4), 658–679. https://doi.org/10.1111/jcom.12104

Crawford, K., & Gillespie, T. (2016) What is a flag for? Social media reporting tools and the vocabulary of complaint. *New Media & Society, 18*(3), 410–428. https://doi.org/10.1177/1461444814543163

Ellis, D. G. (2012). *Deliberative communication and ethnopolitical conflict. Language as social action: Vol. 13*. Peter Lang.

Fraser, B. (1990). Perspectives on politeness. *Journal of Pragmatics, 14*(2), 219–236. https://doi.org/10.1016/0378-2166(90)90081-N

Friess, D., Ziegele, M. & Heinbach, D. (2020). Collective Civic Moderation for Deliberation? Exploring the Links between Citizens' Organized Engagement in Comment Sections and the Deliberative Quality of Online Discussions. *Political Communication*, 1–23. https://doi.org/10.1080/10584609.2020.1830322

Gastil, J. (2008). *Political communication and deliberation*. Los Angeles: Sage.

Gervais, B. T. (2015). Incivility online: Affective and behavioral reactions to uncivil political posts in a web-based experiment. *Journal of Information Technology & Politics, 12*(2), 167–185. https://doi.org/10.1080/19331681.2014.997416

Gervais, B. T. (2017). More than mimicry? The role of anger in uncivil reactions to elite political incivility. *International Journal of Public Opinion Research, 29*(3), 384-405. https://doi.org/10.1093/ijpor/edw010

Grice, P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics* (pp. 41–58). Academic Press. https://doi.org/10.1163/9789004368811_003

Han, S. H., & Brazeal, L. M. (2015). Playing nice: modeling civility in online political discussions. *Communication Research Reports, 32*(1), 20-28. https://doi.org/10.1080/08824096.2014.989971

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach (Second edition). Methodology in the social sciences.* The Guilford Press.

Herbst, S. (2010). *Rude democracy: Civility and incivility in American politics*. Temple University Press.

Hopp, T. (2019). A network analysis of political incivility dimensions. *Communication and the Public, 4*(3), 204–223. https://doi.org/10.1177%2F2057047319877278

Hsueh, M., Yogeeswaran, K., & Malinen, S. (2015). "Leave your comment below": Can biased online comments influence our own prejudicial attitudes and behaviors? *Human Communication Research, 41*(4), 557–576. https://doi.org/10.1111/hcre.12059

Hwang, H., Borah, P., Kang, N., Veenstra, A. (2008, May). Does civility matter in the blogoshpere? Examining the interaction effects of incivility and disagreement on citizen attitudes. *Paper presented at the annual conference of the International Communication Association*, Montreal, Canada.

Jamieson, K. H., Volinsky, A., Weitz, I., & Kenski, K. (2018). The political uses and abuses of civility and incivility. In K. Kenski & K. H. Jamieson (Eds.), *The Oxford handbook of political communication*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199793471.013.79_update_001

Kalch, A., & Naab, T. K. (2017). Replying, disliking, flagging: How users engage with uncivil and impolite comments on news sites. *SCM Studies in Communication and Media*, *6*(4), 395–419. https://doi.org/10.5771/2192-4007-2017-4-395

Kenski, K., Coe, K. & Rains, S. A. (2020): Perceptions of uncivil discourse online. An examination of types and predictors. *Communication Research, 47*(6), 795-814.

Kluck, J. P., & Krämer, N. C. (2021). "What an idiot!" – How the appraisal of the writer of an uncivil comment impacts discussion behavior. *New Media & Society*. https://doi.org/10.1177/14614448211000666

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction. *Psychological Science in the Public Interest, 13*(3), 106–131. https://doi.org/10.1177/1529100612451018

Lindenberg, S. (2015a). Social Rationality and Weak Solidarity: A Coevolutionary Approach to Social Order. In E. Lawler, S. R. Thye, & J. Yoon (Eds.), *Order on the edge of chaos: Social psychology and the problem of social order* (pp. 43–62). Cambridge University Press. https://doi.org/10.1017/CBO9781139924627.004

Lindenberg, S. (2015b). Solidarity: Unpacking the social brain. In A. Laitinen & A. B. Pessi (Eds.), *Solidarity. Theory and practice* (pp. 30–54). Lexington Books.

Lück, J. & Nardi, C. (2019). Incivility in user comments on online news articles: Investigating the role of opinion dissonance for the effects of incivility on attitudes, emotions and the willingness to participate. *SCM – Studies in Communication & Media, 8*(3), 311–337. https://doi.org/10.5771/2192-4007-2019-3-311

Masullo, G. (in press). Future directions for online incivility research. In M. Emmer, J. Trebbe, S. Paasch-Colberg, & C. Strippel (Eds.), *Challenges and perspectives of hate speech analysis.*

Media Authority of North Rhine-Westphalia (2020). *Ergebnisbericht der forsa-Befragung zu Hate Speech 2020* [Results report of the forsa survey on Hate Speech 2020). https://www.medienanstalt-nrw.de/fileadmin/user_upload/NeueWebsite_0120/Themen/Hass/forsa_LFMNRW_Hassrede2020_Ergebnisbericht.pdf

Meyer, H. K., & Carey, M. C. (2014). In Moderation: Examining how journalists' attitudes toward online comments affect the creation of community. *Journalism Practice, 8*(2), 213–228. https://doi.org/10.1080/17512786.2013.859838

Muddiman, A. (2017). Personal and public levels of political incivility. *International Journal of Communication, 11*, 3182–3202.

Mutz, D. C. (2007). Effects of “In-your-face” television discourse on perceptions of a legitimate opposition. *American Political Science Review*, *101*, 621–635. https://doi.org/10.1017/S000305540707044X

Naab., T. K. (2016). Der Sanktionsbedarf von Facebook-Inhalten aus Sicht von NutzerInnen und seine Determinanten [The need for sanctioning Facebook content from the perspective of users and its determinants]. *M&K Medien & Kommunikationswissenschaft, 64*(1), 56-73. https://doi.org/10.5771/1615-634X-2016-1-56

Naab, T. K., Kalch, A., & Meitz, T. G. (2018). Flagging uncivil user comments: Effects of intervention information, type of victim, and response comments on bystander behavior. *New Media & Society*, *20*(2), 777–795. https://doi.or/10.1177/1461444816670923

Naab, T. K., Naab T., & Brandmeier J. (2021). Uncivil user comments increase users’ intention to engage in corrective actions and their support for authoritative restrictive actions. *Journalism & Mass Communication Quarterly*, *98*(2), 566-588. https://doi.org/10.1177/1077699019886586

Oz, M., Zheng, P., & Masullo Chen, G. (2018). Twitter versus Facebook: Comparing incivility, impoliteness, and deliberative attributes. *New Media & Society, 20*(9), 3400-3419. https://doi.org/10.1177/1461444817749516

Pang, N., Ho, S. S., Zhang, A. M.R., Ko, J. S.W., Low, W. X., & Tan, K. S.Y. (2016). Can spiral of silence and civility predict click speech on Facebook? *Computers in Human Behavior, 64,* 898–905. https://doi.org/10.1016/j.chb.2016.07.066

Papacharissi, Z. (2004). Democracy online: civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society, 6*(2), 259–283. https://doi.org/10.1177/1461444804041444

Porten-Cheé, P., Kunst, M., & Emmer, M. (2020). Online civic intervention: A new form of political participation under conditions of a disruptive online discourse. *International Journal of Communication, 14*, 514–534.

Rösner, L., Winter, S., & Krämer, N. C. (2016). Dangerous minds? Effects of uncivil online comments on aggressive cognitions, emotions, and behavior. *Computers in Human Behavior*, *58*, 461–470. https://doi.org/10.1016/j.chb.2016.01.022

Rossini, P. (2020). Beyond incivility: Understanding patterns of uncivil and intolerant discourse in online political talk. *Communication Research*. https://doi.org/10.1177/0093650220921314

Rowe, I. (2015). Civility 2.0: A comparative analysis of incivility in online political discussion. *Information, Communication & Society*, *18*(2), 121–138. https://doi.org/10.1080/1369118X.2014.940365

Santana, A. D. (2014). Virtuous or vitriolic: The effect of anonymity on civility in online newspaper reader comment boards. *Journalism Practice, 8*(1), 18–33. https://doi.org/10.1080/17512786.2013.813194

Singer, J. B. (2009). Separate spaces: Discourse about the 2007 Scottish elections on a national newspaper web site. *The International Journal of Press/Politics*, *14*(4), 477–496. https://doi.org/10.1177/1940161209336659

Stoll, A., Ziegele, M., & Quiring, O. (2020). Detecting incivility and impoliteness in online discussions. *Computational Communication Research, 2*(1), 109–134.

Stroud, N. J., Scacco, J. M., Muddiman, A., & Curry, A. L. (2015). Changing deliberative norms on news organizations' facebook sites. *Journal of Computer-Mediated Communication, 20*(2), 188–203. https://doi.org/10.1111/jcc4.12104

Stryker, R., Conway, B. A., Bauldry, S., & Kaul, V. (2021). Replication note: What is political incivility? *Human Communication Research*. https://doi.org/10.1093/hcr/hqab017

Stryker, R., Conway, B. A., & Danielson, J. T. (2016). What is political incivility? *Communication Monographs, 83*(4), 535–556. https://doi.org/10.1080/03637751.2016.1201207

Su, L. Y.-F., Xenos, M. A., Rose, K. M., Wirz, C., Scheufele, D. A., & Brossard, D. (2018). Uncivil and personal? Comparing patterns of incivility in comments on the Facebook pages of news outlets. *New Media & Society, 20*(10), 3678-3699. https://doi.org/10.1177/1461444818757205

Sydnor, E. (2018). Platforms for incivility: examining perceptions across different media formats. *Political Communication*, *35*(1), 97–116. https://doi.org/10.1080/10584609.2017.1355857

Tomasello, M. (2008). *Origins of human communication.* MIT Press.

Tomasello, M. (2009). *Why we cooperate.* MIT Press.

Tomasello, M. (2019). *Becoming human: A theory of ontogeny.* Harvard University Press. https://doi.org/10.4159/9780674988651

Wang, M. Y., & Silva, D. E. (2018). A slap or a jab: An experiment on viewing uncivil political discussions on Facebook. *Computers in Human Behavior*, *81*, 73–83. https://doi.org/10.1016/j.chb.2017.11.041

Watson, C., Clark, L. A., & Tellegen, A. (1988). Development and measurement of brief measures of positive and negative affect: The PANAS scale. *Journal of Personality and Social Psychology, 54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Wilhelm, C., Joeckel, S., & Ziegler, I. (2020). Reporting hate comments: Investigating the effects of deviance characteristics, neutralization strategies, and users' moral orientation. *Communication Research, 47*(6), 921-944. https://doi.org/10.1177/0093650219855330

Wright, S. (2006). Government-run Online Discussion Fora: Moderation, Censorship and the Shadow of Control. *The British Journal of Politics and International Relations, 8*(4), 550–568. https://doi.org/10.1111/j.1467-856X.2006.00247.x

Ziegele, M., & Jost, P. B. (2020). Not funny? The effects of factual versus sarcastic journalistic responses to uncivil user comments. *Communication Research*, *47*(6), 891-920. https://doi.org/10.1177%2F0093650216671854

Ziegele, M., Naab, T. K. & Jost, P. (2020). Lonely together? Identifying the determinants of collective corrective action against uncivil comments. New Media & Society, 22(5), 731-751. https://doi.org/10.1177/1461444819870130

Ziegele, M., Weber, M., Quiring, O., & Breiner, T. (2018). The dynamics of online news discussions: Effects of news articles and reader comments on users' involvement, willingness to participate, and the civility of their contributions. *Information, Communication & Society*, *21*(10), 1419–1435. https://doi.org/10.1080/1369118X.2017.1324505

Tables and Figures

Table 1. Violations of the communication norms used in this study.

Communication norm	Type of norm violation used in this study	Incivility studies considering this type
Relation norm *Communicate respectfully with other participants*	Insult and vulgarity against another participant	e.g., Chen & Lu, 2017; Coe et al., 2014

Context norm
Consider liberal democratic principles
	Negative stereotype of and threat of violence against a social/political group	e.g., Papacharissi, 2004;
		Kalch & Naab, 2017;
		Oz et al., 2018

Information norm *Communicate only what is informative*	False information	e.g., Muddiman, 2017; Stryker et al., 2016
Modality norm *Communicate comprehensibly*	Irony, sarcasm	e.g., Anderson & Huntington, 2017; Rowe, 2015
Process norm *Connect your contributions*	Topic deviation	e.g., Hopp, 2019

Table 2. Logistic regression of norm violations on explicit disapproval.

Dependent variable: Explicit disapproval	b	95% CI for Odds Ratio		
		Lower	Odds	Upper
Constant	-3.17 *** [-4.74, -2.33]			
Norm violation: information norm (ref = none, modality norm)	1.89 ** [0.72, 3.75]	2.01	6.65	21.94
Norm violation: process norm (ref = none, modality norm)	2.12 *** [1.06, 4.03]	2.57	8.31	26.89
Norm violation: relation norm (ref = none, modality norm)	3.80 *** [2.88, 5.67]	14.48	44.81	138.75
Norm violation: context norm (ref = none, modality norm)	2.52 *** [1.44, 4.28]	3.97	12.41	38.86
Topic (1 = universal basic income)	-0.61 * [-1.21, -0.04]	0.31	0.54	0.95
n		433		

Note. R^2 *= .19 (Cox & Snell) .29 (Nagelkerke). Model* χ^2 *= 78.33,* $p < .001$*. *** $p < 0.05$*; *** $p < 0.01$*; **** $p < 0.001$*.*

Table 3. Logistic regression of norm violations on leaving the discussion.

Dependent variable: Leaving the discussion	b	95% CI for Odds Ratio		
		Lower	Odds	Upper
Constant	1.44 *** [0.89, 2.33]			
Norm violation: information norm (ref= none)	0.54 [-0.37, 1.66]	0.66	1.71	4.42
Norm violation: modality norm (ref= none)	0.45 [-0.56, 1.49]	0.62	1.56	3.97
Norm violation: process norm (ref= none)	0.32 [-0.68, 1.30]	0.56	1.38	3.39
Norm violation: relation norm (ref= none)	0.39 [-0.54, 1.56]	0.59	1.48	3.72
Norm violation: context norm (ref= none)	0.66 [-0.35, 1.74]	0.72	1.94	5.19
Topic (1 = universal basic income)	0.19 [-0.46, 0.79]	0.68	1.21	2.14
n		433		

Note. $R^2 = .006$ *(Cox & Snell) .01 (Nagelkerke). Model* $\chi^2 = 2.50$, $p = .87$. $* p < 0.05$; $** p < 0.01$; $*** p < 0.001$.

Table 4. Logistic regression of norm violations on disapproving criticism.

Dependent variable: Disapproving criticism	b	95% CI for Odds Ratio		
		Lower	Odds	Upper
Constant	-3.22 *** [-4.75, -2.42]			
Norm violation: information norm (ref= none, modality norm)	1.89 ** [0.76, 3.51]	2.00	6.60	21.74
Norm violation: process norm (ref= none, modality norm)	2.10 *** [1.05, 3.81]	2.53	8.18	26.41
Norm violation: relation norm (ref= none, modality norm)	3.29 *** [2.32, 5.07]	8.63	26.77	83.03
Norm violation: context norm (ref= none, modality norm)	2.09 *** [0.96, 3.80]	2.51	8.12	26.20
Topic (1 = universal basic income)	-0.48 [-1.14, 0.10]	0.35	0.62	1.11
n		433		

Note. $R^2 = .13$ *(Cox & Snell) .21 (Nagelkerke). Model* $\chi^2 = 51.56$, $p < .001$. $* p < 0.05$; $** p < 0.01$; $*** p < 0.001$.

Table 5. Logistic regression of a violation of the relation norm on flagging.

Dependent variable: Flagging	b	95% CI for Odds Ratio		
		Lower	Odds	Upper
Constant	-3.77 ***			
	[-5.17, -3.07]			
Norm violation: relation norm (1 = yes)	3.13 ***	8.48	22.95	62.07
	[2.24, 4.43]			
Topic (1 = universal basic income)	-0.83	0.17	0.44	1.10
	[-1.78, 0.05]			
n		433		

Note. $R^2 = .10$ (Cox & Snell) .29 (Nagelkerke). Model $\chi^2 = 46.63$, $p < .001$. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table 6. Logistic regression of a violation of the context norm on flagging.

Dependent variable: Flagging	b	95% CI for Odds Ratio		
		Lower	Odds	Upper
Constant	-2.79 ***			
	[-3.47, -2.30]			
Norm violation: context norm (1 = yes)	0.59	0.69	1.81	4.74
	[-0.71, 1.61]			
Topic (1 = universal basic income)	-0.36	0.30	0.70	1.61
	[-1.35, 0.51]			
n		433		

Note. $R^2 = .005$ (Cox & Snell) .01 (Nagelkerke). Model $\chi^2 = 2.02$, $p = .36$. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table 7. Logistic regression of norm violations on ignoring the norm violation.

Dependent variable: Ignoring the norm violation	b	95% CI for Odds Ratio		
		Lower	Odds	Upper
Constant	1.35 ***			
	[0.76, 2.17]			
Norm violation: information norm (ref = none)	-0.53	0.27	0.59	1.32
	[-1.42, 0.26]			
Norm violation: modality norm (ref = none)	0.06	0.45	1.06	2.49
	[-0.86, 1.01]			
Norm violation: process norm (ref = none)	-0.75	0.22	0.47	1.03
	[-1.61, 0.01]			
Norm violation: relation norm (ref = none)	-1.98 ***	0.06	0.14	0.30
	[-2.97, -1.31]			
Norm violation: context norm (ref = none)	-0.88 *	0.19	0.41	0.90
	[-1.77, -0.09]			
Topic (1 = universal basic income)	0.36	0.92	1.44	2.23
	[-0.10, 0.84]			
n		433		

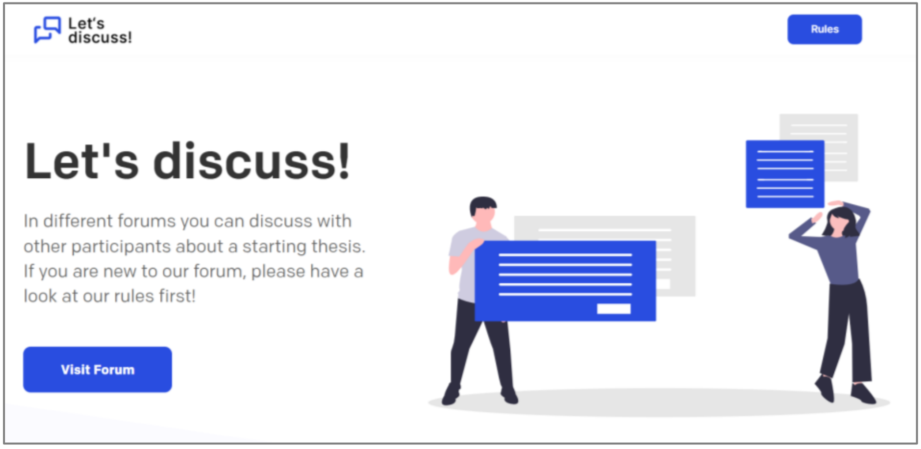
Note. $R^2 = .09$ (Cox & Snell) .13 (Nagelkerke). Model $\chi^2 = 41.59$, $p < .001$. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Appendix A: Forum

Table A1. Time intervals of the notices "user is writing" and replies in the forum.

Element	Writing 1	Answer 1	Writing 2	Answer 2	Writing 3	Answer 3
Time in Seconds	13	20	13	22	12	21,5

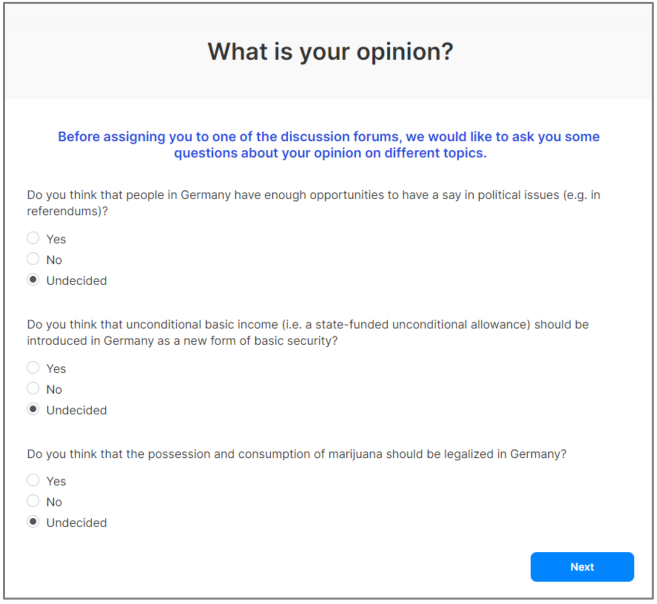
Figure A1. Welcome page of the discussion forum



Let's discuss!

In different forums you can discuss with other participants about a starting thesis. If you are new to our forum, please have a look at our rules first!

Figure A2. 2nd page of the forum with questions about participants' opinions.



What is your opinion?

Before assigning you to one of the discussion forums, we would like to ask you some questions about your opinion on different topics.

Do you think that people in Germany have enough opportunities to have a say in political issues (e.g. in referendums)?

○ Yes
○ No
● Undecided

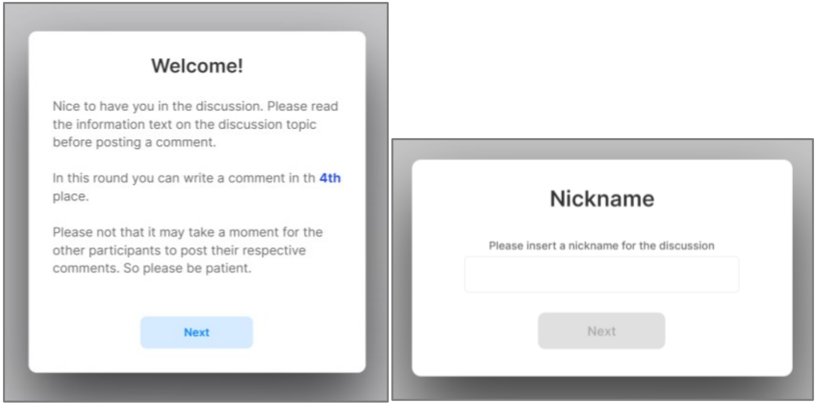
Do you think that unconditional basic income (i.e. a state-funded unconditional allowance) should be introduced in Germany as a new form of basic security?

○ Yes
○ No
● Undecided

Do you think that the possession and consumption of marijuana should be legalized in Germany?

○ Yes
○ No
● Undecided

Figure A3. 3rd and 4th page, information before discussion and request of a nickname.



Welcome!

Nice to have you in the discussion. Please read the information text on the discussion topic before posting a comment.

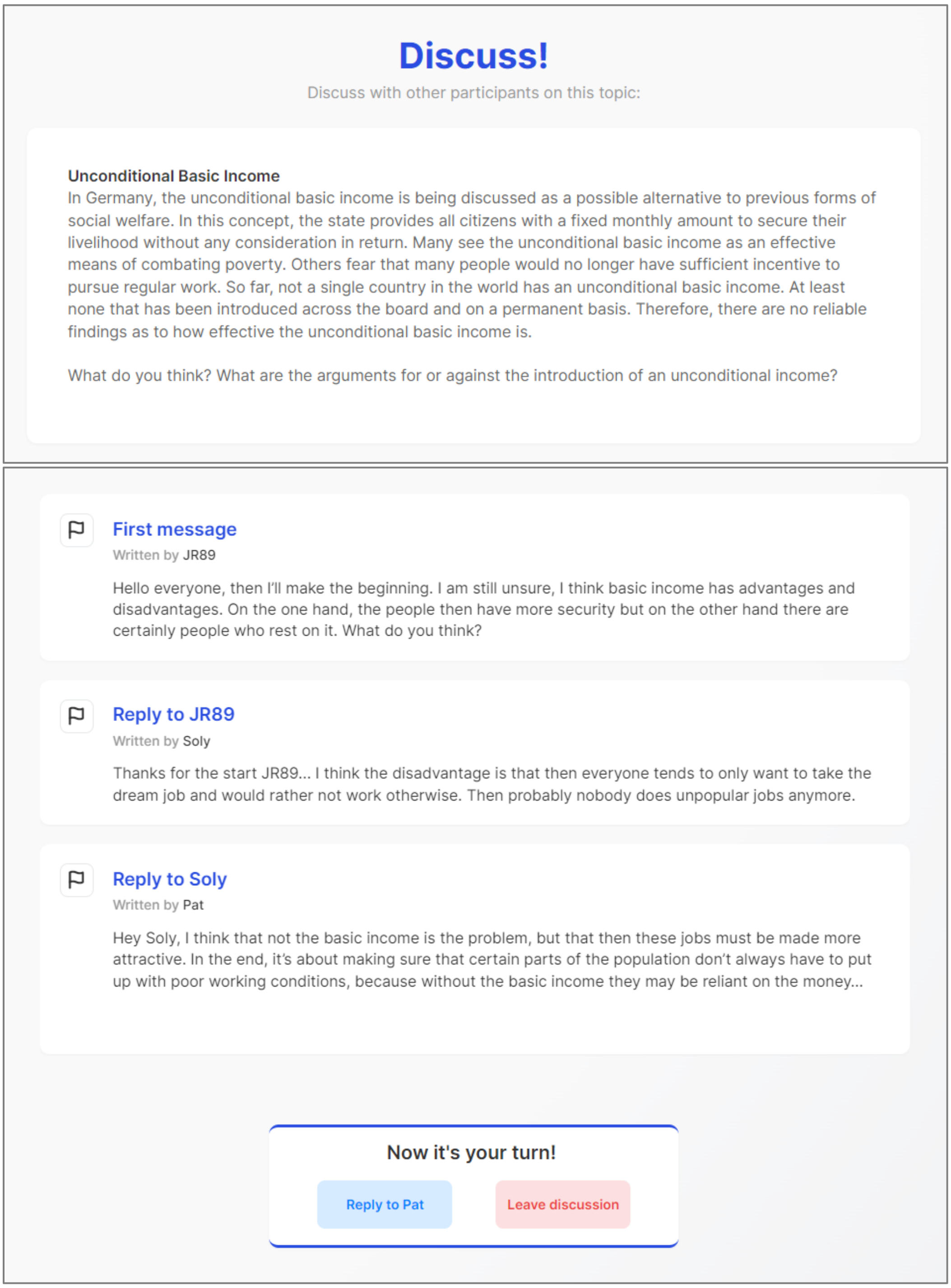
In this round you can write a comment in th **4th** place.

Please not that it may take a moment for the other participants to post their respective comments. So please be patient.

Nickname

Please insert a nickname for the discussion

Figure A4. Screenshot of the discussion forum at participants' turn.



Discuss!

Discuss with other participants on this topic:

Unconditional Basic Income

In Germany, the unconditional basic income is being discussed as a possible alternative to previous forms of social welfare. In this concept, the state provides all citizens with a fixed monthly amount to secure their livelihood without any consideration in return. Many see the unconditional basic income as an effective means of combating poverty. Others fear that many people would no longer have sufficient incentive to pursue regular work. So far, not a single country in the world has an unconditional basic income. At least none that has been introduced across the board and on a permanent basis. Therefore, there are no reliable findings as to how effective the unconditional basic income is.

What do you think? What are the arguments for or against the introduction of an unconditional income?

First message

Written by JR89

Hello everyone, then I'll make the beginning. I am still unsure, I think basic income has advantages and disadvantages. On the one hand, the people then have more security but on the other hand there are certainly people who rest on it. What do you think?

Reply to JR89

Written by Soly

Thanks for the start JR89... I think the disadvantage is that then everyone tends to only want to take the dream job and would rather not work otherwise. Then probably nobody does unpopular jobs anymore.

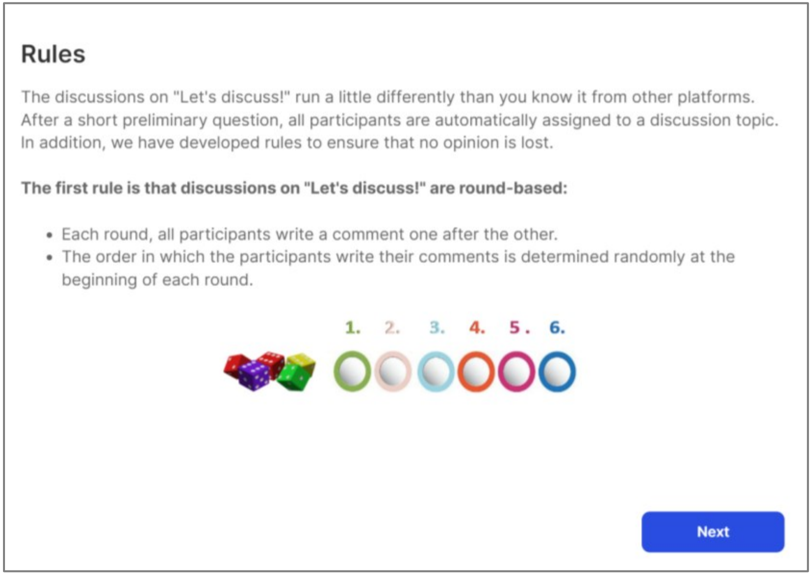
Reply to Soly

Written by Pat

Hey Soly, I think that not the basic income is the problem, but that then these jobs must be made more attractive. In the end, it's about making sure that certain parts of the population don't always have to put up with poor working conditions, because without the basic income they may be reliant on the money...

Now it's your turn!

Figure A5. First rule of “Let’s discuss.”



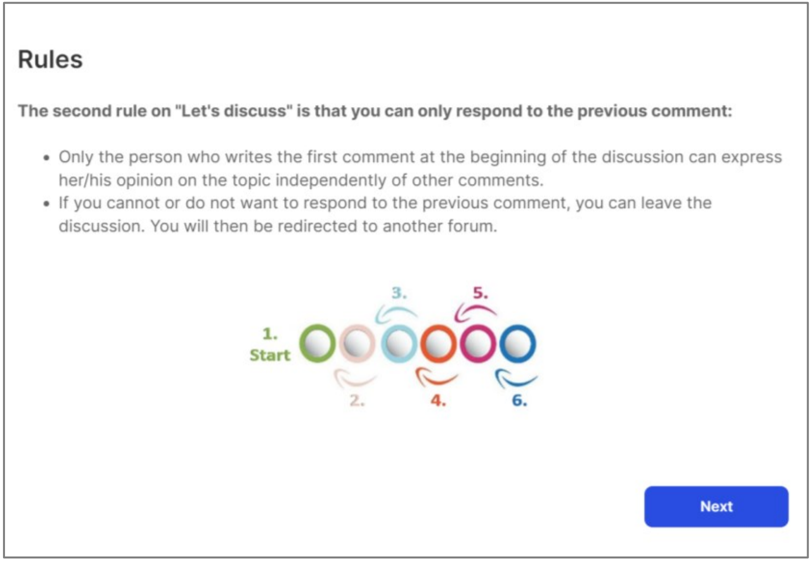
Rules

The discussions on "Let's discuss!" run a little differently than you know it from other platforms. After a short preliminary question, all participants are automatically assigned to a discussion topic. In addition, we have developed rules to ensure that no opinion is lost.

The first rule is that discussions on "Let's discuss!" are round-based:

- Each round, all participants write a comment one after the other.
- The order in which the participants write their comments is determined randomly at the beginning of each round.

Figure A6. Second rule of “Let’s discuss.”

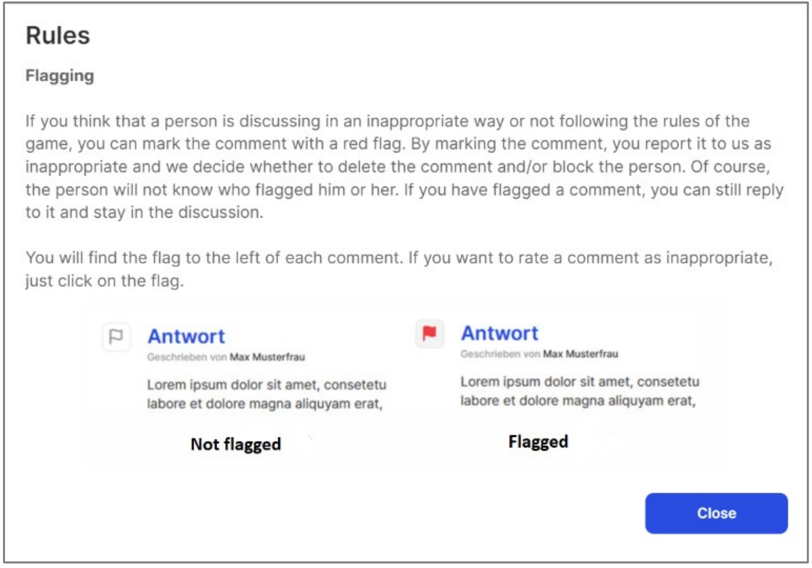


Rules

The second rule on "Let's discuss" is that you can only respond to the previous comment:

- Only the person who writes the first comment at the beginning of the discussion can express her/his opinion on the topic independently of other comments.
- If you cannot or do not want to respond to the previous comment, you can leave the discussion. You will then be redirected to another forum.

Figure A7. Information on the flagging option.



Rules

Flagging

If you think that a person is discussing in an inappropriate way or not following the rules of the game, you can mark the comment with a red flag. By marking the comment, you report it to us as inappropriate and we decide whether to delete the comment and/or block the person. Of course, the person will not know who flagged him or her. If you have flagged a comment, you can still reply to it and stay in the discussion.

You will find the flag to the left of each comment. If you want to rate a comment as inappropriate, just click on the flag.

Antwort

Geschrieben von **Max Musterfrau**

Lorem ipsum dolor sit amet, consetetu labore et dolore magna aliquyam erat,

Not flagged

Antwort

Geschrieben von **Max Musterfrau**

Lorem ipsum dolor sit amet, consetetu labore et dolore magna aliquyam erat,

Flagged

Appendix B: Pretest

First, we created a comment compliant with the five communication norms for each topic, namely the introduction of a universal basic income in Germany and the expansion of direct democracy in Germany. Then, within an iterative process, the comments violating the communication norms were constructed. In developing the comments, we considered several aspects: (1) The core argument of the comment was kept consistent among the different conditions, except for the condition of topic violation; (2) Each comment violated only one of the five communication norms; (3) The severity of the distinct types of norm violations was kept as equal as possible. A total of twelve comments were developed and pretested before the main study to check whether they were perceived as intended and to improve their quality.

The pretest ($N = 47$, $M_{\text{age}} = 42.60$, $SD = 13.70$, 68.1% female)[11] was conducted using a within-subjects design. The comments were presented in a random order and without preceding comments. All 47 participants rated eight items for each comment on a 7-point Likert scale from 1 = *strongly disagree* to 7 = *strongly agree*. The items were worded in a straightforward and clearly understandable manner, and asked about the presence of characteristics that correspond to the distinct types of norm violations. The results revealed that most types of norm violations were recognized as intended (see Table B1 and Table B2). However, some norm violations were also rated high on items that did not address the intended norm violation. After in-depth discussions with several participants, we were able to identify problematic wordings and refine the comments. For example, the violation of the context norm was rated high on the item *doubtful fact without evidence* (violation of the information norm). The discussions revealed that participants recognized the stereotype as intended, but were unsure whether the comment could also be classified as a false fact. As a result, we framed the respective comment more strongly as opinion than as a fact.

[11] The pretest was conducted for the present study and for another study using the same stimulus material for a different purpose. Information about the previous study can be found here: https://osf.io/p8z6u?view_only=ad81ffd6d58742cf919b69e1ac128f9c

Table B1. Mean values of the ratings of eight items for comments addressing the topic of basic income.

	Topic: Basic income											
	Norm-compliant comment		Lies/false information		Irony, sarcasm		Topic deviation		Insults, vulgarity		Stereotypes, threats of violence	
The comment…	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
...presents doubtful facts as being proven without giving concrete references to sources	2.79	1.94	6.43	1.14	3.98	2.15	3.40	1.93	4.34	2.03	4.89	2.08
...contains ironic/sarcastic elements	1.32	0.81	1.81	1.44	6.34	1.55	1.38	1.03	2.70	2.03	2.06	1.50
...deviates from the actual discussion topic	2.72	1.70	2.13	1.28	3.21	1.52	5.36	1.63	3.55	1.70	4.26	1.57
... responds directly to another person	5.51	1.80	5.02	1.90	5.00	1.87	2.53	1.89	6.02	1.64	5.06	1.90
...contains vulgar expressions	1.23	0.76	1.28	0.77	1.72	1.43	1.21	0.75	4.77	2.21	3.26	2.14
...personally attacks a person from the discussion	1.72	1.26	1.94	1.28	3.64	2.37	1.53	1.37	6.51	1.00	2.74	2.06
...generally ascribes negative characteristics to people of a certain group	1.98	1.47	1.79	1.25	3.09	2.07	1.62	1.21	3.23	2.07	6.60	0.83
...threatens the well-being of others because of a group affiliation	1.53	1.08	1.66	1.13	1.94	1.50	1.51	1.20	2.13	1.68	4.57	2.23

Table B2. Mean values of the ratings of eight items for comments addressing the topic of direct democracy.

	Topic: Direct democracy											
	Norm-compliant comment		Lies/false information		Irony, sarcasm		Topic deviation		Insult, vulgarity		Stereotype, threats of violence	
The comment…	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
...presents doubtful facts as being proven without giving concrete references to sources	2.83	2.07	6.19	1.33	3.13	2.23	2.02	1.42	3.23	2.06	4.26	2.31
...contains ironic/sarcastic elements	1.47	1.20	1.87	1.50	6.26	1.55	1.17	0.43	2.77	2.21	2.57	1.80
...deviates from the actual discussion topic	1.94	1.41	2.19	1.36	2.51	1.49	5.00	2.04	2.98	1.76	3.66	1.83
... responds directly to another person	5.32	1.96	5.11	1.98	5.40	1.70	2.85	2.21	6.23	1.09	4.96	2.00
...contains vulgar expressions	1.32	1.05	1.34	1.05	1.38	0.97	1.26	0.85	6.13	1.33	3.77	2.34
...personally attacks a person from the discussion	1.38	1.03	2.53	1.79	3.70	2.02	1.81	1.45	6.85	0.42	2.72	2.12
...generally ascribes negative characteristics to people of a certain group	1.47	1.23	2.04	1.62	2.55	1.84	1.43	0.93	2.91	2.07	6.40	1.36
...threatens the well-being of others because of a group affiliation	1.47	1.14	1.66	1.22	1.89	1.56	1.43	1.18	1.96	1.41	4.83	2.22

Appendix C: Stimuli

Table C1. Introduction texts and standard comments in the discussion topic of basic income.

Introduction text	In Germany, the universal basic income is being discussed as a possible alternative to previous forms of social welfare. In this concept, the state provides all citizens with a fixed monthly amount to secure their livelihood without any consideration in return. Many see the universal basic income as an effective means of combating poverty. Others fear that many people would no longer have sufficient incentive to pursue regular work. So far, not a single country in the world has a universal basic income. At least none that has been introduced across the board and on a permanent basis. Therefore, there are no reliable findings as to how effective the universal basic income is. What do you think? What are the arguments for or against the introduction of an unconditional income?
Comment 1 by "JR89"	Hello everyone, then I'll make the beginning. I am still unsure, I think basic income has advantages and disadvantages. On the one hand, the people then have more security but on the other hand there are certainly people who rest on it. What do you think?
Comment 2 by "Soly"	Thanks for the start JR89... I think the disadvantage is that then everyone tends to only want to take the dream job and would rather not work otherwise. Then probably nobody does unpopular jobs anymore.

Note. The study was conducted in Germany. The texts and comments were translated for publication. In the original German version, comments included minor misspelling to increase the authenticity of the comments.

Table C2. Manipulated comments by "Pat" in the discussion topic of basic income.

Type of norm violation	Comment
Norm-compliant	Hey Soly, I think that not the basic income is the problem, but that then these jobs must be made more attractive. In the end, it's about making sure that certain parts of the population don't always have to put up with poor working conditions, because without the basic income they may be reliant on the money…
Insults, vulgarity to other participants	@Soly, I think this is a really asocial attitude from you. Have you idiot ever thought about the aspect that not the basic income is the problem, but that these jobs must be made more attractive? In the end, it's about making sure that certain parts of the population don't always have to put up with shitty working conditions, because without the basic income they may be reliant on the money…
Stereotypes, threats of violence against social/political groups	Hey Soly, I think that not the basic income is the problem, but that then these jobs must be made more attractive. In my opinion, it is due to the money-grubbing capitalists that certain parts of the population have to put up with the poor working conditions, because without the basic income they may be reliant on the money of these exploiters. I

	think that these exploiters themselves should be sent to forced labor camps and do the work…
Lies/false information	Hey Soly, the basic income has already worked great in 1000 other countries. So it is already proven that the basic income works great. It has been proven that unpopular jobs are made more attractive because certain parts of the population don't have to put up with poor working conditions anymore, because without the basic income they would be reliant on the money...
Irony/sarcasm	Hey Soly, the basic income is of course the absolutely only reason that no one wants to do these unattractive jobs, there can be noooothing else. The poor working conditions that certain parts of the population often have to put up with, because without the basic income they might be reliant on the money, of course have noooothing at aaaall to do with it...
Topic deviation	I think there are more urgent social issues at the moment. In my opinion, the last year and a half has shown that it is much more important to talk about the state of our education and healthcare system than about the universal basic income...

Note. The study was conducted in Germany. The comments were translated for publication. In the original German version, comments included minor misspelling to increase the authenticity of the comments.

Table C3. Introduction texts and standard comments in the discussion topic of direct democracy.

Introduction text	There is always a debate about whether citizens should have more of a say in political decisions. In Germany, there have so far only been forms of direct democracy at the municipal and state levels, i.e. elections in which people vote on whether they are for or against a particular issue. Many would like to see these votes expanded, also at the federal level. Others believe that complex political issues cannot be broken down into yes and no questions. From a scientific perspective, it has not yet been determined whether direct democracy achieves better democratic results than representative democracy. Studies have not provided reliable statements in this regard. What do you think? What are the arguments for or against expanding forms of direct democracy?
Comment 1 by "JR89"	Hello everyone, then I'll make the beginning. I am still unsure, I think direct democracy has advantages and disadvantages. On the one hand, people then have more influence on certain political decisions but on the other hand, the various parties can manipulate opinions in their interest. What do you think?
Comment 2 by "Soly"	Thanks for the start JR89... I think the problem is that citizens don't always have the expertise to properly weigh political decisions. Therefore, politicians may have the better oversight of what is best for everyone.

Note. The study was conducted in Germany. The texts and comments were translated for publication. In the original German version, comments included minor misspelling to increase the authenticity of the comments.

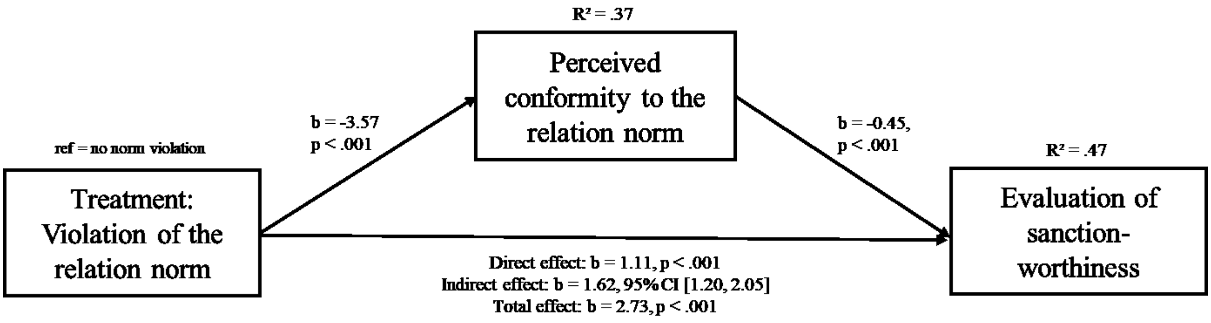
Table C4. Manipulated comments by "Pat" in the discussion topic of direct democracy.

Type of norm violation	Comment
Norm-compliant	Hey Soly, I think that you can also give the citizens the necessary information via information campaigns. In the end, from my point of view, there is no better representation of interests than when everyone gives a vote directly for themselves...
Insults, vulgarity to other participants	@Soly, I think this is a really stupid argument from you, apparently your mind is somewhat limited. Still, you can give the citizens the necessary information via information campaigns. I think it sucks anyway if not everyone with his voice directly can represent their own interests...
Stereotypes, threats of violence against social/political groups	Hey Soly, you can also give the citizens the necessary information via information campaigns. I personally believe that our mendacious politicians do not really represent the interests of the citizens anyway. Therefore, these political actors should be beaten out of the parliaments by force if we can represent our interests directly with our vote...
Lies/false information	Hey Soly, there are a million studies that clearly show that direct democracy is the only true democracy. In fact, information campaigns can give the citizens the necessary information without any problems. In the end, this is the only true representation of interests for each individual...
Irony/sarcasm	Hey Soly, as well, it is aaaaaabsolutely not conceivable to give the necessary information to the citizens by information campaigns. It would be completely crazy if everyone could directly represent their own interests - that would be muuuuuch too democratic...
Topic deviation	I think there are currently more important issues that should concern us than the extent of direct democracy in Germany. It is much more important to discuss who will win the next election and how well Germany can then position itself economically after Corona...

Note. The study was conducted in Germany. The comments were translated for publication. In the original German version, comments included minor misspelling to increase the authenticity of the comments.

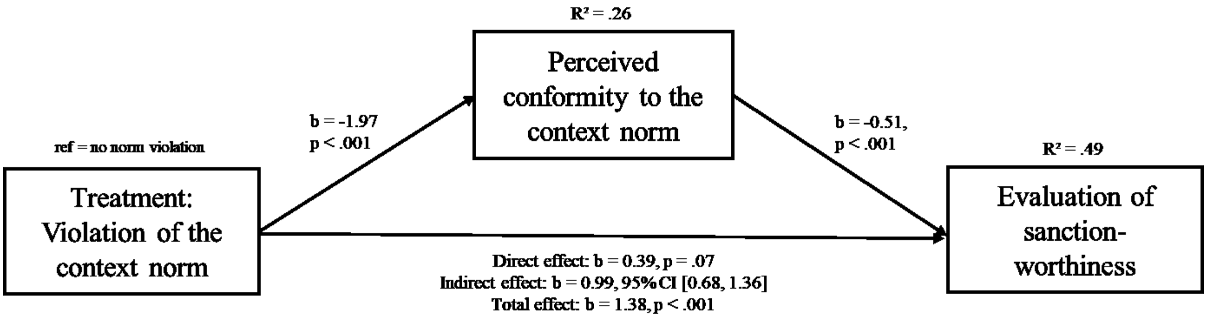
Appendix D: Mediation analyses

Figure D1. Mediation analysis of the effect of violation of the relation norm on evaluation of sanction-worthiness via perceived norm conformity.


R² = .37
Perceived conformity to the relation norm
b = -3.57
p < .001
b = -0.45,
p < .001
ref = no norm violation
R² = .47
Treatment: Violation of the relation norm
Evaluation of sanction-worthiness
Direct effect: b = 1.11, p < .001
Indirect effect: b = 1.62, 95% CI [1.20, 2.05]
Total effect: b = 2.73, p < .001

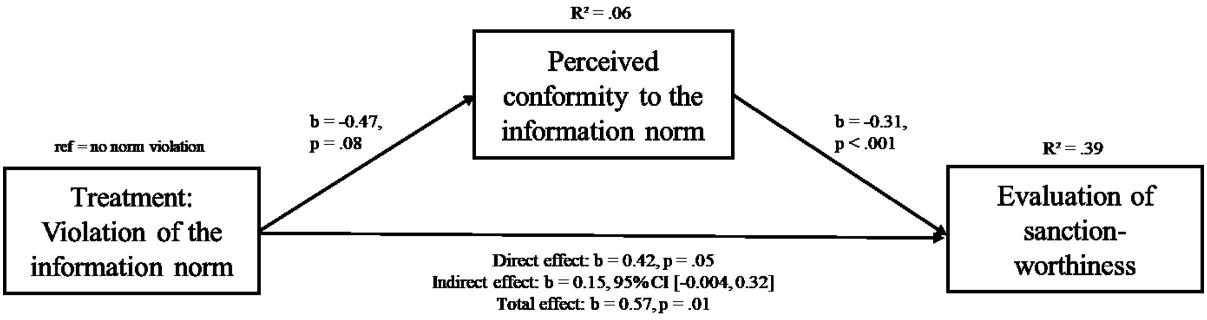
Note. Estimation with PROCESS, unstandardized coefficients. Covariates: Topic; violation of the information norm; violation of the modality norm; violation of the process norm; violation of the context norm. *N* = 433.

Figure D2. Mediation analysis of the effect of violation of the context norm on evaluation of sanction-worthiness via perceived norm conformity.


R² = .26
Perceived conformity to the context norm
b = -1.97
p < .001
b = -0.51,
p < .001
ref = no norm violation
R² = .49
Treatment: Violation of the context norm
Evaluation of sanction-worthiness
Direct effect: b = 0.39, p = .07
Indirect effect: b = 0.99, 95% CI [0.68, 1.36]
Total effect: b = 1.38, p < .001

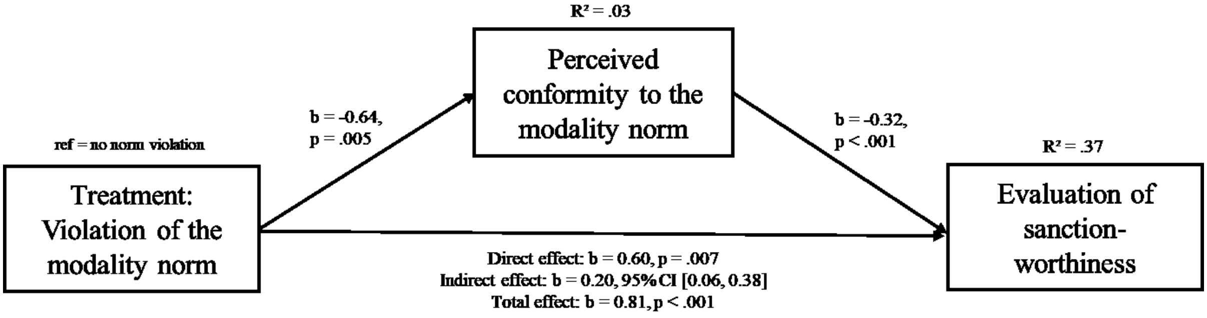
Note. Estimation with PROCESS, unstandardized coefficients. Covariates: Topic; violation of the information norm; violation of the modality norm; violation of the process norm; violation of the relation norm. *N* = 433.

Figure D3. Mediation analysis of the effect of violation of the information norm on evaluation of sanction-worthiness via perceived norm conformity.


R² = .06
Perceived conformity to the information norm
ref = no norm violation
b = -0.47, p = .08
b = -0.31, p < .001
R² = .39
Treatment: Violation of the information norm
Evaluation of sanction-worthiness
Direct effect: b = 0.42, p = .05
Indirect effect: b = 0.15, 95% CI [-0.004, 0.32]
Total effect: b = 0.57, p = .01

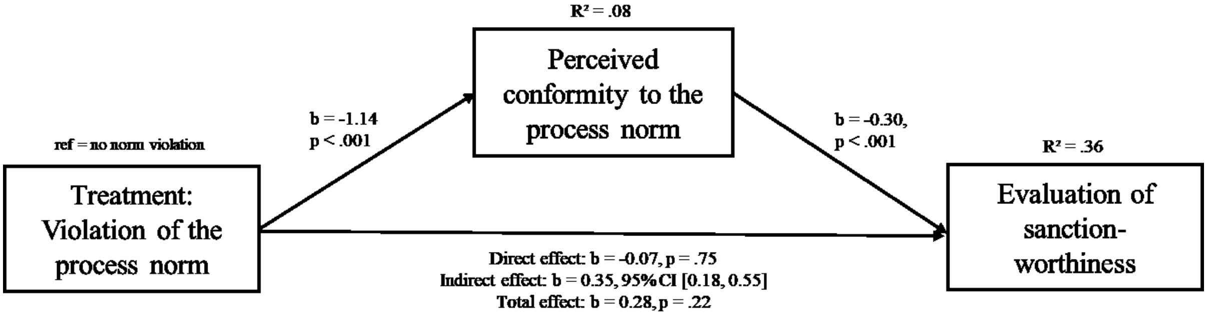
Note. Estimation with PROCESS, unstandardized coefficients. Covariates: Topic; violation of the modality norm; violation of the process norm; violation of the relation norm; violation of the context norm. *N* = 433.

Figure D4. Mediation analysis of the effect of violation of the modality norm on evaluation of sanction-worthiness via perceived norm conformity.


R² = .03
Perceived conformity to the modality norm
ref = no norm violation
b = -0.64, p = .005
b = -0.32, p < .001
R² = .37
Treatment: Violation of the modality norm
Evaluation of sanction-worthiness
Direct effect: b = 0.60, p = .007
Indirect effect: b = 0.20, 95% CI [0.06, 0.38]
Total effect: b = 0.81, p < .001

Note. Estimation with PROCESS, unstandardized coefficients. Covariates: Topic; violation of the information norm; violation of the process norm; violation of the relation norm; violation of the context norm. *N* = 433.

Figure D5. Mediation analysis of the effect of violation of the process norm on evaluation of sanction-worthiness via perceived norm conformity.


R² = .08
Perceived conformity to the process norm
b = -1.14
p < .001
ref = no norm violation
Treatment: Violation of the process norm
b = -0.30,
p < .001
R² = .36
Evaluation of sanction-worthiness
Direct effect: b = -0.07, p = .75
Indirect effect: b = 0.35, 95% CI [0.18, 0.55]
Total effect: b = 0.28, p = .22

Note. Estimation with PROCESS, unstandardized coefficients. Covariates: Topic; violation of the information norm; violation of the modality norm; violation of the relation norm; violation of the context norm. *N* = 433.